Integrative taxonomy of *Polygonia* Hübner 1819 (Lepidoptera: Nymphalidae) in Alberta

by

Christianne Margaret McDonald

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Systematics and Evolution

Department of Biological Sciences

University of Alberta

© Christianne Margaret McDonald, 2015

ABSTRACT

Speciation can be an elaborate process. Delimiting species and reconstructing evolutionary relationships may be similarly complex, revealing gene tree discordance, cryptic species, geographic structuring or hybridization. In order to solve such systematic problems, a careful balance should be struck between evidence from morphology and molecules. Relationships among *Polygonia* species have been explored using mitochondrial genes (ND1, COI), nuclear genes (*wgl*, EF-1α, GAPDH, RpS5) and morphology (wing patterns, venation, genitalia), but their topology remains inconclusive, due at least in part to phylogenetic discordance. Here, I used mitochondrial COI gene sequence in tandem with genomic single nucleotide polymorphisms (SNPs) genotyped using genotyping-by-sequencing (GBS) methods to assess species and subspecies boundaries. I also reconstructed phylogenetic relationships in the genus to further investigate phylogenetic discordance. Distinct genetic clusters resulted from discriminant analysis on principal components (DAPC) of SNPs, while COI sequencing revealed a new mitochondrial lineage, making *P. gracilis* paraphyletic. Genetic clusters were carried forward into the morphological analysis to serve as prior categories for the specimens. Ten visually scored diagnostic characters selected based on personal observations and appearance in the taxonomic literature clustered the specimens into the same groups as genetic characters, while digital colour analysis of wing areas gave less congruent groupings. I used discriminant correspondence analysis (DCA) of the visually scored characters to compare their diagnostic utility and construct a new species-level dichotomous key. This integrative approach to constructing diagnostic keys supports species identifications that are designed to correspond more closely to genetic clusters.

ACKNOWLEDGEMENTS

I couldn't have done this without the support of so many people. First, John Acorn and Felix Sperling – your different styles of mentorship and motivation helped me grow as a person and as a researcher. I am extremely fortunate to have had the opportunity to have you both as supervisors, and I look forward to working with you in the future. Graduate school was a lot like adolescence for me – I didn't realize that my parents (or supervisors, in this case) were right all along until I was almost done. Thank you for your continued patience while I learned this.

Thank you to Dr. Sally Leys, who took the time to serve on my examining committee in the middle of the summer. Field seasons are short, and I really appreciate your time. Likewise, thank you Dr. Alison Murray for chairing my exam.

The Sperling lab, past and present, notably: Bryan Brunet, Heather Leibel, Giovanny Fagua Gonzalez, Erin Campbell, Stephen Trevoy, Jasmine Janes, Phil Batista, Gwylim Blackburn, but most of all, I need to thank Julian Dupuis – you have helped me with so many aspects of my project when you really did not have the time, and I appreciate it very much. I owe you a few manuscript proofreads and co-authorship, at the very least. Thank you all for your patience with me, especially when I wasn't sure how to run STRUCTURE, PAUP*, R, TASSEL and a few other things too. Thank you for taking the time to explain things that were simple to you, but that I was only just learning. Your expertise saved me from drowning.

My loving husband, Eric Nylund. I'm not sure I can express how much your friendship means to me. I could not have done this without you by my side. I can't wait to see what adventures come next for us. An enormous thank you to my parents, Don and Sue McDonald. Things weren't always easy, but I could always count on you for never-ending support and encouragement. Cat McDonald, my sister – thank you for reminding me to have fun when I could. You helped keep me sane in little ways that went a long way.

Last, but certainly not the least, the Leys lab – Rachel Brown, Danielle Pendlebury, Amanda Kahn, Nathan Farrar, Pam Reid, and Jasmine Mah. Your friendship and understanding helped me get through the tough times. Thank you for teaching me about sponges and joining me for coffee breaks, mini donuts, Slurpee runs and all the good times in between. To my colleagues in CW312 – notably Carla Starchuk, Mark Wolansky and Stephanie Blais – thank you for the advice, the coffee breaks, and for helping me climb a steep learning curve while juggling a thesis.

Field and lab work for this thesis was funded by the ACA Grants in Biodiversity (supported by the Alberta Conservation Association) and an NSERC Discovery Grant to Felix Sperling.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

List of Abbreviations

bp: base pair

COI: cytochrome c oxidase I

DAPC: discriminant analysis on principal components

DCA: discriminant correspondence analysis

GBS: genotyping-by-sequencing

GTR: general time-reversible

LDA: linear discriminant analysis

MCA: multiple correspondence analysis

ML: maximum likelihood

MP: maximum parsimony

mtDNA: mitochondrial DNA

NGS: next-generation sequencing

nDNA: nuclear DNA

PCA: principal components analysis

RAD: restriction-site-associated

RGB: red green blue (luminance values)

SNP: single nucleotide polymorphism

CHAPTER 1

1. GENERAL INTRODUCTION

1.1 THE SPECIES PROBLEM

The importance of the species rank cannot be overstated, given that species are a fundamental unit in biology, regardless of field (Mayr 1982a). Despite the utility of species as a unit and the ferocity with which this topic has been discussed for over a century, the scientific community has not agreed upon a single definition.

In addition to the obvious importance of species to our understanding of evolution and their contribution to basic science, accurate species identification is essential to the conservation of biodiversity (Wiens 2007). Species and subspecies are the primary units of conservation; as such, it is crucial to understand the boundaries of these units to effectively determine their conservation status (Vogler and Desalle 1994; Wiens 2007). The definition of species is also critical in optimizing pest control in forestry and agriculture, or diagnosing vectors of disease and host-parasite interactions (Hausdorf 2011). Bad taxonomy – that is, identification or classification of species without following the appropriate rules of taxonomy (Winston 1999) – can have catastrophic effects that may seriously impact our understanding of the natural world we study, or in some cases can affect nature itself (Bortolus 2008). All of these issues hinge upon species delimitation, the act of assessing the number of species and determining the boundaries between them, which is intrinsically tied to the definition of a species, or species concept (De Queiroz 2007).

Philosophers and biologists have attempted to explain species since Aristotelian times, defining them by "essences" and "differentia" long before evolutionary theory came about (Gerson 1999). Analytic philosophers argued that natural groupings - as opposed to artificial or man-made groupings - existed as "kinds" that were bound by natural law, and would exist as a group regardless of whether mankind was present to group them (Quine 1970). Natural kinds are recognized as such only if the group can be extrapolated beyond its members; membership of a new individual can be assessed by the intrinsic criteria the other

members share (Quine 1970). Species may be natural kinds that exist independently of our minds, a philosophical position termed realism and espoused by Mayr (1992) among others. Alternatively, species might only exist as arbitrary categories that were created by humans for convenience, much like genera, a view held by Darwin (1859) and known as nominalism. Regardless of whether these groups are natural or artificial, their boundaries are not absolute; morphological intermediates, hybrids, subspecies and polymorphs exist in great number, clouding the categorical criteria for assigning membership (Mayr 1992).

Species delimitation traditionally relied on type specimens, the similarity of a group of specimens to this prototype and clear discontinuity with other such groups, resulting in limits being imposed on variation permitted within a species (Mayr 1992). This typological species concept gives no criteria for species membership other than similarity, and hinges on the expertise of taxonomists alone. Multivariate analysis methods like Principal Component Analysis and Multiple Correspondence Analysis can lend credibility through mechanical objectivity when it comes to assessing species boundaries by morphological similarity; however, phenotypic clusters do not necessarily correspond to biological species (Hebert et al. 2003; Ward et al. 2005; Dupuis et al. 2012).

In defining what a species is, many taxonomists also attempt to define how a species comes into being (Sperling 2003; De Queiroz 2007; Hausdorf 2011). The biological act of speciation can be extremely intricate and difficult to fit to any particular set of rules, so it stands to reason that it is not easily defined (Hausdorf 2011). In early stages of speciation, gene flow occurs readily between species and hybrids are generally reproductively viable; in later stages of linear separation, even when intrinsic reproductive isolation occurs, species barriers are still permeable to gene flow (Mallet 2005). While it may be simple to define a species based on two fully diverged taxa, it becomes increasingly difficult to do so for species undergoing lineage separation (De Queiroz 2007). As a result, numerous authors have coined their own species concepts, many of which are incompatible with those of their rivals (Orr 2001; Sperling 2003; De Queiroz 2007; Hausdorf 2011).

Many species concepts emphasize reproductive isolation as the primary criterion for species status, whether this isolation was developed intrinsically by biological incompatibility or extrinsically through geographic isolation (Mayr 1942; van Valen 1976). The classic

example is taught in high school and undergraduate biology classrooms around the world: the biological species concept, which is based explicitly on the production of fertile offspring within species and reproductive incompatibility between them (Mayr 1942). This definition is challenged by many examples of viable hybrids in nature, uniparental species and those with horizontal gene transfer (Harrison 1989; Mallet 2005; Hausdorf 2011). One concept that addresses this issue is the genomic integrity species concept (Sperling 2003), where genomic integrity refers to whether populations maintain a cohesive genomic core where their ranges overlap, despite some gene flow. Since allopatric and parthenogenetic populations cannot be assessed by their reproductive continuity, this definition proposes calibrating overall genetic divergences between similar taxa with sympatric populations.

The phylogenetic species concept requires a population to exhibit a pattern of phylogenetic ancestry and descent to be considered a species (Eldredge and Cracraft 1980); this definition is applicable to uniparental species, however, it neglects introgressed lineages and parallel speciation, both of which contribute to incomplete "parental patterns of ancestry and descent" (Nagel and Schluter 1998; Hausdorf 2011). In addition, assessing phylogenetic relationships between groups is complicated by the difference between gene trees and species trees. When using multiple genes to reconstruct evolutionary relationships, strong discordance is often found between markers (Maddison 1997; Nylin et al. 2001; Wahlberg et al. 2009; Dupuis et al. 2012). Another factor to consider when using the phylogenetic species concept to delineate species is the arbitrary nature of the decision about which groups are "deemed worthy" of recognition, since the worth of a clade can vary greatly between taxonomists.

Other concepts focus on the act of lineage separation and the recognition of species as independent lineages. The most inclusive of these definitions is the unified species concept, or general lineage concept, which defines a species as a metapopulation evolving independently from other units (De Queiroz 2007). This concept is amenable to defining species at the onset of speciation, before substantial differentiation between populations is detectable; however, it is not useful for the act of drawing species boundaries, as there are no measurable criteria given (Hausdorf 2011). Even the ecological species concept, which emphasizes observable traits such as niche occupation and adaptive zones, relies on arbitrary boundaries based on a minimal difference between lineages (van Valen 1976).

The species concepts named above – as well as a long list of others – vary in their inclusiveness, diagnostic application and utility in varying stages of speciation. The debate to define a species continues, yet no singular definition seems appropriate for all biological cases to which the term species has been applied, as evidenced by the series of rebuttals each publication on species definitions faces (De Queiroz 2007; Hausdorf 2011; Orr 2001, and references therein). Several authors have come to the conclusion that there is no "correct definition" and that each has merits of its own; consequently the use of a species concept is ultimately up to the taxonomist (Orr 2001; Hausdorf 2011).

1.1.1 SPECIES DELIMITATION METHODS

Defining species can be equivalent to the act of defining speciation, but the task of drawing species boundaries has a distinct set of challenges (Hausdorf 2011). Before molecular techniques were developed, morphological and ecological traits were used successfully for many years (Mayr 1963), and continues today (Will and Rubinoff 2004; Leaché et al. 2009). In Lepidoptera, the study of genitalic characters, wing venation, wing patterning (Mutanen 2005), and life history traits are particularly useful for delimiting species and have long been effective tools for drawing species boundaries (Sperling 2003; Brunet et al. 2013). Host plant use, niche occupation, phenology and development are all still used to help diagnose even complex groups of species (Lumley and Sperling 2010; Brunet et al. 2013).

Quantitative morphological study, or morphometrics, is useful in delimiting species boundaries in cryptic species complexes such as the spruce budworm (*Choristoneura fumiferana*) species group (Lumley and Sperling 2010). Using 47 digital wing colour, length or area characters, Lumley and Sperling (2010) were able to diagnose species boundaries between nearly all species in the complex, despite strong morphological similarity across all specimens. Another example of the use of morphometrics to delimit species is in the hymenopteran genus *Ophion* (Schwarzfeld and Sperling 2014) where a combination of geometric morphometrics, traditional morphology and qualitative morphology were used to discover six new species. Similarly, landmark geometric morphometrics has been used to describe inter- and intraspecific variation in a genus of syrphid flies, *Merodon*, with a high rate of success compared to molecular clusters from mitochondrial DNA (mtDNA) and allozymes (Francuski et al. 2009).

With the development of molecular techniques, systematists added a new set of tools to their repertoires. Such methods brought an abundance of new, highly accessible characters, making it easier to detect cryptic species, gain new insights into mimics and models, appropriately assign juveniles to conspecific adults in cases of metamorphosis, and find correspondence of males to females in cases of sexual dimorphism (Janzen et al. 2009; van Nieukerken et al. 2012). The introduction of next generation sequencing (NGS) of DNA is advancing taxonomy further still, not only by increasing the yield of sequence data but also decreasing the cost and time required to obtain the data (McCormack et al. 2013). The quantity of data produced by these methods, up to 400 times as many nucleotides per species compared to traditional Sanger sequencing (nucleotides per species) (Cruaud et al. 2014), has allowed biologists to overcome many of the challenges of traditional molecular methods. Next generation sequencing samples a greater number of genes to better approximate species phylogenies, acquires many rapidly mutating genetic markers to allow large-scale population genetics studies, and results in sufficient sequence data to resolve relationships even among recently diverged taxa (Lerner and Fleischer 2010; Lemmon and Lemmon 2013). With these obstacles overcome, NGS provides the sampling needed to more accurately assess gene flow and genomic differentiation of populations. One example, genotyping by sequencing (GBS) (Elshire et al. 2011), is a technique that allows partial sequencing across the genome at short DNA regions associated with restriction sites to provide complexity reduction. When implemented using the Illumina sequencing platform (Bennett 2004), thousands of single nucleotide polymorphisms (SNPs) may be discovered and consistently genotyped within species. Due to the ubiquity of SNPs across the genome, GBS allows systematists to closely approximate the species tree by sampling many genes simultaneously (Cariou et al. 2013; Cruaud et al. 2014).

Next generation sequencing (NGS) methods are extremely powerful for delimiting species and reconstructing phylogenies of recently diverged taxa; however, they are not without their drawbacks. Whole genome sequencing requires high-quality DNA, which can be challenging if samples are historical, improperly preserved or contain chemical impurities (McCormack et al. 2012; Lemmon and Lemmon 2013). Many NGS techniques have a high initial cost for lab equipment, reagents and sequencing equipment, creating a cost barrier to reaping the benefits of the low cost-per-base of these methods (Lemmon and Lemmon 2013). Sequencing accuracy is lower than for traditional sequencing methods, especially in

repetitive regions, so stringent read depth filters are necessary to prevent mis-called bases, which can result in spurious identification of polymorphic loci (El-Metwally et al. 2014). Another hazard of this protocol is the presence of paralogous reads that align due to sequence similarity, producing spurious calls that are difficult to differentiate from biological polymorphisms (Etter et al. 2011).

Combining methods across different disciplines of biology can contribute different sources of character information and increase accuracy and confidence in taxonomic studies (Dayrat 2005; Will et al. 2005; Schlick-Steiner et al. 2010). More robust than any one method alone, integrative taxonomy uses a broad range of characters to elucidate species boundaries in a multisource approach – morphology, behavioral characters, ecological traits, genetic data and geographic ranges (Will et al. 2005; Lumley and Sperling 2010; Schlick-Steiner et al. 2010). The aim of this approach is to move taxonomy toward understanding the processes involved in speciation by combining several compatible, yet independent disciplines (Fujita et al. 2012). The use of many character types can drastically reduce error rates in species delimitation studies (from 0.30 to 0.027, based on failure rates among character types) (Schlick-Steiner et al. 2010) and can also reveal discordance, or disagreement among disciplines, which can help expand current hypotheses on speciation of a group.

In a similar approach to integrative taxonomy, iterative taxonomy advocates the use of multiple lines of evidence to delimit and identify species; however, rather than promoting one total-evidence round of analysis using many sources of data, as its name implies, iterative taxonomy employs successive testing of species boundaries (Yeates et al. 2011). Where there is discordance between datasets, a new null hypothesis for species boundaries is formed and tested using a new line of evidence. Many taxonomic papers use integrative or iterative taxonomy to confidently describe species boundaries in complex groups (Padial and De La Riva 2009; Lumley and Sperling 2010; Schwarzfeld and Sperling 2014; Gebiola et al. 2015; Hansson et al. 2015).

1.2 THE GENUS *POLYGONIA* HÜBNER (1819)

Polygonia is a genus of nymphalid butterflies in the subfamily Nymphalinae, tribe Nymphalini (Wahlberg et al. 2003, 2005), along with thirteen other genera, including

Nymphalis (Kluk, 1780), *Kaniska* (Moore, 1899) and *Aglais* (Dalman, 1816) (Wahlberg 2006). *Polygonia* butterflies are commonly referred to as commas because of the silver punctuation on their ventral hind wings, or as anglewings from their angular wing margins (Acorn 1993). These charismatic butterflies are leaf mimics, emulating dead foliage on their ventral side, with a starkly contrasting orange and brown dorsal side (Pyle 2002). Habitat preferences of *Polygonia* vary between species globally; however, in Western North America they tend to inhabit mixed forests near water sources like rivers, ponds, lakes and streams (Bird et al. 1995; Shepard and Guppy 2011). They can often be found imbibing water and minerals from mud puddles, gravel walkways, animal dung, and sap flows on wounded trees (Bird et al. 1995).

1.2.1 LIFE HISTORY

Polygonia butterflies overwinter as adults, and are among the first butterflies seen after winter – or even during winter, if the weather is mild (Acorn 1993). They are among the longest-lived butterflies because of the long winters they endure, living 10 months to a year, provided temperatures are mild and moisture is available (Bird et al. 1995). *Polygonia* are generally univoltine in temperate climates; however, depending on the location, species and the environmental conditions, *Polygonia* can be bivoltine or even trivoltine (Nylin 1991), where the offspring of the mid-summer emergence will mature rapidly and mate immediately rather than enter into diapause.

Females of some species in *Polygonia* exhibit dimorphism, either lacking striations and contrast on the ventral side of the wings (referred to as smeared morph in this study) or are heavily marked on the ventral side (contrasted morph), like their male counterparts (Shepard and Guppy 2011) (Figure A.6). Additionally, several species within the genus exhibit two seasonal colour morphs, depending on the time of their emergence: a bright summer morph and a dark overwintering morph called *umbrosa* (meaning "shadowed"). These forms caused some confusion in the early years of *Polygonia* taxonomy, with the *umbrosa* form frequently identified either as a subspecies or a variant (Pelham, 2008), although dimorphic females were recognized as being conspecific based on their dorsal markings. The differences between seasonal morphs go beyond physical appearance and phenology; the autumn morph does not mate immediately like its summer form, but rather it enters into reproductive diapause before overwintering (Nylin 1991). Triggers for development into the summer

versus the autumn morph have been studied extensively and are a combination of photoperiod and temperature during larval development (Endo 1969; Endo et al. 1988; Fukuda and Endo 2006; Soderlind and Nylin 2011).

In Alberta, the overwintered flight of anglewings occurs in late April to late June (Bird et al. 1995). These first adults of the season are often quite ragged, as they have endured a harsh winter in addition to months of flight the previous summer. The males immediately search for a mate, patrolling an area and approaching any resting butterfly to investigate opportunities for copulation (Endo 1973).

Female *Polygonia* butterflies oviposit on the underside of host plant leaves, laying eggs singly or in groups of up to 3 or 4 (James and Nunnallee 2011). Eggs generally hatch 3-5 days after oviposition and emergent larvae begin feeding on the host plant immediately; some *Polygonia* larvae have been observed curling leaves into nests and securing them with silk for shelter until they pupate (James and Nunnallee 2011). Host plant use varies widely in the genus, although the most common hosts used are *Urtica* (nettle), *Ulmus* (Elm), *Humulus* (hop), *Ribes* (gooseberry), *Salix* (willow) and *Betula* (birch) (Weingartner et al. 2006). Some species utilize several genera of host plants, like *P. faunus* (*Alnus*, *Betula*, *Rhododendron*, *Vaccinium*, *Ribes*, *Populus*, *Salix* and *Urtica*), while others are more specific in their host use, like *P. c-aureum* (*Humulus* and *Urtica*) (Weingartner et al. 2006). Researchers have hypothesized that the ancestral host plant for *Polygonia* was an urticalean rosid, based on the plants used by extant species in the genus (Weingartner et al. 2006).

Within 30 days of eggs hatching in the Pacific Northwest, larvae will mature into adult butterflies (James and Nunnallee 2011) and emerge in late July or early August, depending on environmental conditions. Adults will continue to fly while weather is mild (above 10^{o}C or so) then will find a sheltered place to overwinter (Bird et al. 1995).

1.2.1 TAXONOMIC HISTORY

The genus *Polygonia* has a Holarctic distribution, with lineages in North America, Europe and Asia (Wahlberg et al. 2009). There are 15 currently accepted species of *Polygonia* worldwide; 6 Palearctic (*Polygonia c-album* (Linneus, 1758), *P. c-aureum* (Linneus, 1758), *P. giganteum* (Leech, 1883), *P. undina* (Grum-Grschimailo, 1890), *P. egea* (Cramer, 1775), *P. interposita* (Staudinger, 1881)) and 9 Nearctic (*P. faunus* (W.H. Edwards, 1869), *P. comma*

(Harris, 1842), *P. progne* (Cramer, 1776), *P. gracilis* (Grote and Robinson, 1867), *P. oreas* (W.H. Edwards, 1869), *P. interrogationis* (Fabricius, 1798), *P. haroldii* (Dewitz, 1877), *P. satyrus* (W.H. Edwards, 1869), *P. g-argenteum* (Doubleday and Hewitson, 1846) (Wahlberg et al. 2005, 2009; Pelham 2008).

In North America, the range of *P. faunus*, *P. gracilis* and *P. satyrus* can reach as far north as the Arctic Circle, while other species such as *P. g-argenteum*, *P. interrogationis* and *P. haroldii* inhabit areas as far south as central Mexico (Scott 1992; Bird et al. 1995; Layberry et al. 1998; Brock et al. 2003). Based on current understanding of the evolutionary history of the group, Wahlberg et al. (2009) suggest that there have been multiple colonization events into the Nearctic from the Palearctic followed by diversification events. The geographic ranges of North American species have been strongly affected by glaciation and deglaciation of the continent over the last 2 million years. Historic populations may have followed the movement of glaciers and inhabited ecosystems near glacial lakes, producing modern distributions across the Canadian boreal forest, through the cordillera, and throughout the northwest forested mountains.

The type species for the genus, *Polygonia c-aureum*, was described in 1758 by Linnaeus in *Systema Naturae* as *Papilio c-aureum* (Pelham 2008). Other members of the genus were described following the colonization and exploration of North America, beginning in 1779 with the description of *P. progne* by Pieter Cramer, also described under *Papilio*. In 1816, Hübner proposed the genera *Polygonia* and *Eugonia* in the same publication. *Papilio angelica* was designated as the type species for *Eugonia* 1873 by Grote (Grote 1873), although most authors followed Scudder's invalid designation in 1875 of *Papilio polychloros* as the type species (Hemming 1967). The same type was therefore designated for *Polygonia* and *Eugonia*, although under different names (*Polygonia c-aureum* and *Eugonia angelica*, respectively). In its 1934 meeting, the International Commission for Zoological Nomenclature recognized this error and awarded precedence to *Polygonia* (Hemming 1943), as it was named in a preceding line on the same page in the same publication as *Eugonia* (Hemming 1967). The species of *Polygonia* have variously been included in *Vanessa* (Fabricius, 1807), *Grapta* (Kirby, 1837), *Nymphalis* (Kluk, 1781) and *Tachyptera* (Berge, 1842) before the recognition of *Polygonia* as a distinct group in 1934 (Hemming 1943, 1967).

Due to morphological and ecological similarity between its members, *Polygonia* species have undergone several changes in rank within the genus, and many species contain several recognized subspecies (Scott 1984; Shapiro 1990; Wahlberg and Nylin 2003; Pelham 2008; Wahlberg et al. 2009). Species boundaries are further complicated by the presence of seasonal and sexual dimorphism (Scott 1992; Brock et al. 2003), increasing the number of named species – an overwintered female and a summer male of *P. comma* or *P. interrogationis* would look like different species even to a trained taxonomist without experience in rearing these species.

Despite phenotypic variation within species and similarity between species, identification resources generally agree on diagnostic characters for identifying species in the genus. The presence of a row of green spots and jagged wing edges indicates that *P. faunus* is the specimen in hand; golden tones dorsally and brown striations ventrally indicate *P. satyrus* (Scott 1992; Bird et al. 1995; Pyle 2002). *Polygonia progne* is discernible from *P. gracilis* by the uniform grey-brown of the ventral hind wing, while *P. gracilis* displays stark contrast between basal and marginal areas of the ventral hind wing (Hooper and Long 1973; Pyle 2002; Brock et al. 2003). Some characters, however, that are treated as highly diagnostic in some resources are not listed in others, such as the triangular dorsal hind wing spot in *P. satyrus* and *P. faunus*, which is lacking in *P. gracilis* (Hooper and Long 1973; Bird et al. 1995; Glassberg 2001), or the shape of the silver comma on the ventral side. This incongruence in the literature suggests a need to reassess the diagnostic characters in identification resources.

Phenotypic variation within species can be very high, yielding potentially more numerous subspecies and species diagnoses than the biological groups present. This variation confused understanding of species boundaries in the 1800s and continued in modern examples of species and subspecies diagnoses in the last century. A rarely collected Asian butterfly, *Polygonia interposita*, was considered a subspecies of *P. c-album* due to morphological and genetic similarities in mtDNA (Wahlberg et al., 2009). But after analysis of nuclear DNA (nDNA), it became clear that *P. interposita* was distinct from *P. c-album* and is now recognized as a separate species (Wahlberg et al. 2009).

The Nearctic species *P. progne* currently has no named subspecies; however, in 1984, *P. oreas* was considered a subspecies of *P. progne*, along with two others, *P. p. nigrozephyrus* and *P. p. silenus*, based on genitalia, ventral wing morphology, host plant use and larval characters

(Scott 1984). Only six years later, *P. oreas* was elevated to species rank and all but the nominate subspecies of *P. progne* were placed under *P. oreas* (Shapiro 1990; Pelham 2008).

The placement of *P. g. zephyrus* has been a matter of debate, although the most recent publications have treated it as a subspecies of *P. gracilis* (Weingartner et al. 2006; Pelham 2008; Wahlberg et al. 2009). In 1977, the two were treated as different species, *P. gracilis* and *P. zephyrus* (Shapiro 1990). In 1984, James Scott argued that the two were clearly subspecies based on genitalia, host plant use, overall morphological similarity and apparent introgression throughout western Canada (Scott 1984). Despite these compelling arguments, the two species have been inconsistently treated either as subspecies (Layberry et al. 1998; Opler 1999; Glassberg 2001) or distinct species (Bird et al 1995, Guppy and Sheppard 2011) in several publications since. In this study I treat *P. g. zephyrus* and *P. g. gracilis* as subspecies based on their observed morphological similarity and insubstantial genetic differentiation.

There is some discussion about the usefulness of subspecies as evolutionary units (Zink 2004; Patten 2010; Kodandaramaiah et al. 2012a; Proshek et al. 2015), which would have a significant impact on *Polygonia*, since as many as 5 or 6 subspecies are in use for some species (Pelham 2008). Kodandaramaiah et al. (2012) found that there was no significant population structure corresponding to five of the named subspecies of *P. faunus*, only to geographic distance between populations, despite morphological differences between geographically distinct populations. The variation seen within *P. faunus* is striking – from brown, mossy tones to brilliant emeralds and bright chartreuse – and to some, the morphological differences between regional populations seem pronounced enough to justify recognition of multiple subspecies (Scott 1992; Bird et al. 1995). Given the extent of the species range of *P. faunus*, its phenotypic variation should be unsurprising; however, designation as separate units should depend on the presence of distinct geographic boundaries as opposed to the magnitude of observable difference between specimens (Amadon 1949; Wilson and Brown 1953). To further investigate the findings of Kodandaramaiah et al. (2012), a study measuring phenotypic variation in *P. faunus* across North America would be helpful for determining whether the variation observed is clinal, indicating that subspecific designations are not merited, or whether there are discrete boundaries to these phenotypes that correspond to geographic location.

1.2.2 CHALLENGES IN PHYLOGENETIC RECONSTRUCTION

In *Polygonia*, there are several examples of discordance between nuclear (nDNA) and mitochondrial (mtDNA) genetic markers (Figure 1.1) (Nylin et al. 2001; Wahlberg and Nylin 2003; Weingartner et al. 2006; Wahlberg et al. 2009). Based on nDNA and morphology, *P. satyrus* is sister to *P. comma* and *P. g-argenteum*; however, mtDNA suggests a closer relationship to *P. gracilis* (Nylin et al. 2001; Wahlberg and Nylin 2003; Weingartner et al. 2006; Wahlberg et al. 2009). Nuclear DNA places *P. haroldii* squarely between *P. g. gracilis* and *P. g. zephyrus*, while mitochondrial markers suggest that it is sister to a clade composed of *P. satyrus*, *P. gracilis*, and *P. oreas*. Similarly, *P. interposita* in Asia is nearly identical to *P. c-album* in mtDNA, yet nDNA indicates that it is quite distinct (Wahlberg et al. 2009).

Phylogenetic discordance can be caused by a number of biological processes that can be difficult to reconstruct – ancient or recent introgression, retention of ancestral polymorphisms, cytoplasmic parasites such as *Wolbachia*, or incomplete lineage sorting (Beltran et al. 2002; Degnan and Rosenberg 2009; Kodandaramaiah et al. 2013). It can be difficult to tease apart evolutionary histories when there is uncertainty about which of these events have occurred in the past or are occurring now.

Mitochondrial DNA, specifically the first 600 bp of Cytochrome Oxidase I (COI), is a commonly used molecular marker for diagnosing species (Hebert et al. 2003; Ward et al. 2005; van Nieukerken et al. 2012) and reconstructing phylogenetic relationships between recently diverged animal taxa (Wahlberg et al. 2009; Dupuis et al. 2012). The relatively fast mutation rate of mtDNA compared to nDNA and its maternal inheritance give mtDNA a higher resolving power for recently diverged groups, making its utility for phylogenetic analysis appealing. (Avise et al. 1987; Hebert et al. 2004; Dupuis et al. 2012). These characteristics, along with ease of amplification, make mtDNA an ideal candidate for reconstructing phylogenetic histories of recently diverged taxa (Barraclough and Vogler 2000).

Unfortunately, discordance between gene trees is common; therefore relying on a single gene to delimit species or infer phylogenetic relationships is unwise and using multiple genes is suggested for greater confidence in the topology (Maddison 1997; Roe and Sperling 2007; Wahlberg et al. 2009; Dupuis et al. 2012). The question that remains is, when faced

with strong discordance between genes, how does one estimate the evolutionary history of the group?

Polygonia presents an ideal opportunity to elucidate the problem of genetic discordance; several examples of incongruence combined with morphological and ecological similarities make this group especially puzzling (Wahlberg et al. 2009). This genus exhibits complex interspecific relationships and shows evidence of recent, rapid speciation (Nylin et al. 2001; Wahlberg and Nylin 2003; Weingartner et al. 2006). This has created a complex conceptual topology for delimiting species using morphological methods and traditional Sanger sequencing of COI, and for testing the effectiveness of GBS for phylogenetic inference.

1.3 THESIS OVERVIEW

This study tests the species boundaries of members of the genus *Polygonia* within Alberta and reconstructs their phylogenetic history using a combination of SNPs and mtDNA. It also assesses the effectiveness of diagnostic field markings based on the genetic clusters using multivariate analysis.

The pathways of speciation are complex and difficult to reconstruct; however, a thorough understanding of phylogenetic relationships can provide insights into many aspects of a species' biology. To reconstruct phylogenetic relationships between species within *Polygonia,* I used GBS to discover SNPs throughout the genome, concatenated them and performed maximum parsimony and maximum likelihood analyses in Chapter 2. I compared that topology to one constructed with a 1450-base-pair region of the mitochondrial gene COI, in order to relate this work to previous studies and confirm the discordance between mtDNA and nDNA (Nylin et al. 2001; Wahlberg and Nylin 2003; Wahlberg et al. 2009). Species boundaries within the genus were assessed using Structure analysis, discriminant analysis on principal components (DAPC) and parsimony networks.

Correct identification of study species is crucial in all fields of biology; if the characters used to diagnose species are ineffective or misleading, the implications of these errors can be perpetuated in the literature for years, affecting our understanding and interpretations of the environment (Bortolus 2008). In Chapter 3, I selected characters based on diagnostic field markings in the literature, ten of which were scored visually and six were

scored digitally; I then used multivariate analyses to assess their diagnostic utility on the groups defined by genetic markers obtained in Chapter 2. These analyses serve as a method to test the efficacy of diagnostic field markings, which may in turn be useful in assessing the effectiveness of field guides and constructing new identification resources.

Species delimitation is an important process for all biological fields, and may be one of the most complex (Nadeau et al. 2013; Martin et al. 2015). An accurate, cost-effective method to delimit closely related species would be helpful in conservation, not to mention the advancement of our knowledge of the biological world. The intent of the work in this thesis is to provide a conceptual framework for a multisource approach to species delimitation using next generation sequencing techniques, traditional visually scored morphological characters and digital morphometric analyses. In addition to supporting the existing framework for multisource species delineation methods (Roe and Sperling 2007; Schlick-Steiner et al. 2010; Yeates et al. 2011), this work highlights the importance of field characters for accurate diagnoses. The work thereby proposes a method for assessing the accuracy of widely used characters in the literature and a new, evidence-based method to support diagnostic key construction.

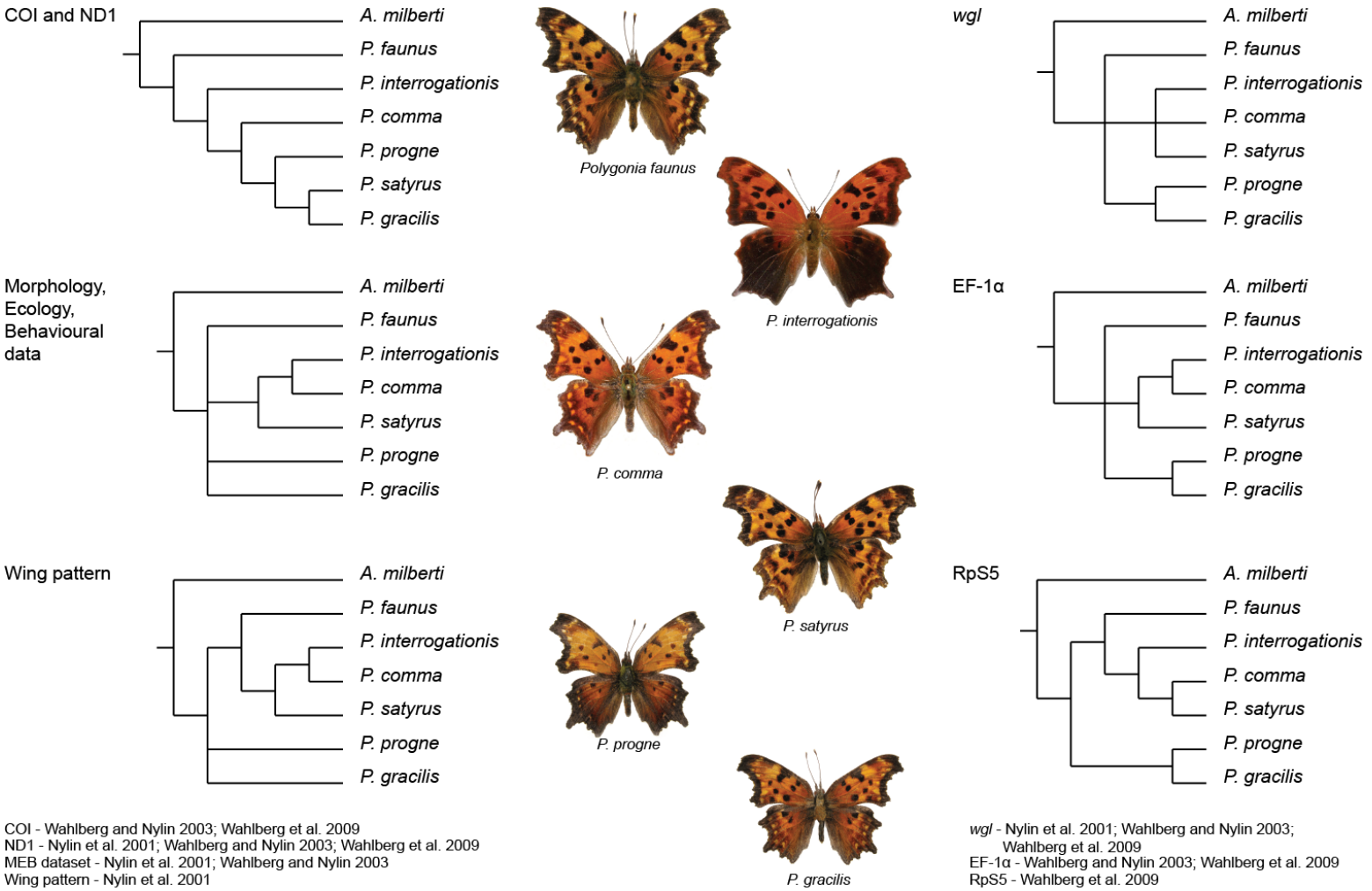


Figure 1.1: Previous published topologies for *Polygonia* species phylogenies, constructed using morphological data, mitochondrial and nuclear genes. Images of *Polygonia interrogationis* and *P. comma* were obtained from Butterflies of America (www.butterfliesofamerica.com) with permission from the photographers.

CHAPTER 2

2. SPECIES DELIMITATION AND PHYLOGENETIC DISCORDANCE IN THE GENUS *POLYGONIA* HÜBNER (1819) (LEPIDOPTERA: NYMPHALIDAE)

2.1 INTRODUCTION

Phylogenetic discordance occurs where two or more gene trees provide conflicting topologies for the same organisms, which can cause challenges for phylogenetic inference and species delimitation (Maddison 1997; Beltran et al. 2002; Degnan and Rosenberg 2009). Discordance is not uncommon and can have numerous causes, such as ancient or recent introgression, incomplete lineage sorting, retention of ancestral haplotypes, or parasites such as *Wolbachia* that alter sex ratios and gene flow (Hoelzer 1997; Narita et al. 2006; Wahlberg et al. 2009; Kodandaramaiah et al. 2013). Investigation into the causes of discordance can provide us with valuable insights into the evolutionary history of a group.

Mitochondrial DNA (mtDNA) analysis has been preferred to that of nuclear genes (nDNA) for investigations of recently diverged taxa, because of mtDNA's rapid mutation rate, uniparental inheritance and a generally shorter coalescence time (Avise et al. 1983; Harrison 1989; Dupuis et al. 2012). On the other hand, nDNA tends to be more useful for studying deep divergences in taxa, due to the relatively slow rate of mutation of many nuclear genes (Slade et al. 1994; Brower and DeSalle 1998; Lin and Danforth 2004). To mitigate errors that may arise from discordance, it is recommended that phylogenetic studies use multiple independent genetic markers to infer phylogenetic relationships, rather than relying on a single gene (Edwards and Beerli 2000; Degnan and Rosenberg 2009; Dupuis et al. 2012; Kodandaramaiah et al. 2013). However, it can be difficult and costly to obtain sequence data for multiple gene regions in the same set of individuals. Given a balance between budget and quantity of data, the development of next generation sequencing technologies has provided new opportunities in many fields of biology.

Next-generation sequencing (NGS) techniques have increased the yield of DNA data and decreased the cost and time to acquire it (McCormack et al. 2013). The large quantity of data produced by these methods, over 400 times that of Sanger sequencing (nucleotides per species) (Cruaud et al. 2014), has created new analytical opportunities for phylogenetics, phylogeography, and population genetics (Lerner and Fleischer 2010). NGS allows biologists to overcome challenges such as insufficient sequence length for resolving relationships, inadequate number of genes to approximate species phylogenies, and the low frequency of rapidly mutating genetic markers for large-scale population genetics studies (McCormack et al. 2013). NGS methods such as genotyping-by-sequencing (GBS) can also reveal single nucleotide polymorphisms (SNPs), which are particularly helpful for resolving relationships of recently diverged taxa, because of their frequency and ubiquity in the genome (Lerner and Fleischer 2010; McCormack et al. 2013), and may provide strong support for clades that were unresolved until the use of genomic SNPs (Wagner et al. 2013; Leaché et al. 2014; Pante et al. 2015)

Next-generation sequencing methods are not without their drawbacks. Sequencing accuracy is generally lower than that of Sanger sequencing, particularly in repetitive regions, requiring stringent read depth filters to prevent spurious SNP genotypes caused by mis-called bases (El-Metwally et al. 2014). In addition, many NGS methods are prone to allele bias, where one allele is amplified more efficiently than another, or only one is amplified due to mutations in the associated restriction site that begins sequence reads (Arnold et al. 2013).

Due to the utility of genomic SNPs in resolving complex phylogenetic relationships at the species level (Nadeau et al. 2013; Cruaud et al. 2014), the broad sampling of these markers across the genome (Davey et al. 2011), and their independence from a reference genome (Baird et al. 2008; Elshire et al. 2011), restriction-site-associated (RAD) sequencing was selected for use in this study. As well, the mitochondrial gene COI (1450 base pairs) was sequenced for comparison to genomic SNPs, which are largely contained in nuclear DNA, and to relate the present study to previous research on the genus *Polygonia* that used sequence from COI (Nylin et al. 2001; Weingartner et al. 2006; Wahlberg et al. 2009).

Polygonia is a genus of nymphalid butterflies that inhabits Palearctic and Nearctic regions (Wahlberg et al. 2009). Of the 15 species of *Polygonia* described worldwide, five live in Alberta, making this region the center of diversity for the genus and an ideal place to

study species interactions (Layberry et al. 1998; Pohl et al. 2010). Similarities in morphology and ecology make identification of *Polygonia* challenging in the field (Bird et al. 1995; Layberry et al. 1998; Acorn and Sheldon 2006). These similarities also make it difficult to assess taxonomic ranks, and phylogenetic discordance between mitochondrial and nuclear markers makes their evolutionary relationships problematic, although together they form a well-supported, monophyletic clade (Wahlberg and Nylin 2003; Wahlberg et al. 2009). *Polygonia* has been the focus of several studies over the last 20 years in an attempt to better understand the process of speciation (Nylin et al. 2001; Wahlberg 2006; Weingartner et al. 2006; Wahlberg et al. 2009). By sampling SNPs throughout the genome, my goal is to test species boundaries, particularly those that have been disputed (i.e., *Polygonia gracilis gracilis* and *P. g. zephyrus*) for *Polygonia* species that occur in Alberta, and to identify critical evolutionary events that may have led to phylogenetic discordance.

2.2 MATERIALS AND METHODS

2.2.1 COLLECTIONS AND VOUCHERS

From May 2012 to September 2013, I accumulated 268 specimens of *Polygonia* (97 *Polygonia faunus*, 54 *P. satyrus*, 69 *P. progne* 8 *P. gracilis gracilis* and 11 *P. g. zephyrus*) and several outgroup species (Figure 2.1) from Alberta and British Columbia, Canada, with the help of many colleagues, especially Julian Dupuis. Specimens were identified using field guides (Bird et al. 1995; Brock et al. 2003; Acorn and Sheldon 2006) with contributions from Felix Sperling and John Acorn. Due to species diversity and proximity to the research lab, sampling was focused in Alberta. Specimens were collected mostly by aerial netting, although nearly a dozen of them were collected in Leptraps brand butterfly traps baited with banana. Most specimens were kept alive in a refrigerator or cooler until they were frozen at -60^{o}C, where they remained until DNA extraction.

To relate my results to previous research, the COI data set used in this study was supplemented with GenBank DNA sequences submitted by Wahlberg et al (2009). A total of 44 sequences of 15 species were added to the data set for the Bayesian analysis, and 83 sequences were added to the haplotype network data set (Table A.1).

Specimens used for DNA analysis will be deposited as pinned vouchers in the E. H. Strickland Entomological Museum at the University of Alberta, and all unique COI sequences will be been uploaded to Genbank. Thoracic tissue was fortified with Lineco Archival Quality Neutral pH Adhesive during the pinning process, and specimens lacking thoraces were stored in glassine envelopes.

2.2.2 LABORATORY METHODS

Genomic DNA was extracted from legs and a portion of thoracic tissue using the DNeasy® Blood and Tissue kit (Qiagen, Venlo, Limburg, Netherlands). The entire thorax was used if the initial DNA yield was too low (less than 20 ng/μL) and additional tissue extraction was needed for further processing. Genomic DNA quality was assessed using a Nanodrop (ND-1000, Thermo Scientific, Waltham, MA, USA) and then a Qubit™ 2.0 fluorometry assay kit (dsDNA BR, Invitrogen by Life Technologies, Burlington, Ontario, Canada).

Samples that contained sufficient quantities of pure DNA (200ng of DNA suspended in 10μL mqH_2O, 260/280 ratio of 1.7-1.9 and 260/230 ratio of 0.9-2.0 were considered pure) were shipped to IBIS (Institut de Biologie Intégrative et des Systèmes) at Université Laval, Québec where RAD-seq libraries were prepared (restriction enzyme pair PstI and MspI) and sequenced by Illumina® for GBS (Elshire et al., 2011). The GBS process requires digestion of genomic DNA with a pair of restriction enzymes - in this case PstI and MspI - and then ligation of barcode forward adapters and reverse adapters (Y-adapters) to the "sticky ends" of the DNA fragments (Elshire et al. 2011; Poland et al. 2012). Then up to 96 samples at a time are multiplexed and amplified by PCR (Elshire et al. 2011). Single end sequencing was completed using the Illumina® high throughput sequencing platform (Bennett 2004). During the demultiplexing stage of the library preparation, we were informed by the IBIS manager (Brian Boyle) that mis-priming may have occurred at the 3' adapter, potentially leading to false SNPs being called. Steps were taken during the filtering process to eliminate these loci, as explained under Alignment and Filtering.

The entire mitochondrial COI gene was amplified, in two fragments, by PCR. The first 650 bp, commonly referred to as the barcode region (Hebert et al. 2003), used the primers LepF1 and HCO (Table 2.1). The final 800 bp were obtained using the primers

JerryI and PatII (Table 2.1). Some samples were not successfully amplified using the latter pair, so JerryI and MilaIV were used instead, producing a shorter COI fragment (1160 bp instead of 1450 in 9 individuals). PCR cycles consisted of an initial denaturation phase at 94°C for 5 minutes, followed by 35 cycles of 94°C for 30 s, 45°C for 30 s and 72°C for 2 min. The final elongation step took 5 minutes at 74°C, and then the samples were held at 4°C until their removal from the cycler. Samples were stored at -4°C until sequencing.

PCR products were cleaned using Exonuclease I and shrimp alkaline phosphatase (ExoSap - New England Biolabs) before dye-termination using BigDye® sequencing pre-mix v2.1 (Life Technologies). Fragments were sequenced using the ABI 3730 capillary system in the Molecular Biology Service Unit (MBSU) at the University of Alberta.

2.2.3 Alignment and filtering

Raw SNP GBS output was filtered to remove loci with a read depth lower than 5 and a minor allele frequency less than 0.05 using the UNEAK de novo pipeline (Universal Network Enabled Analysis Kit) (Lu et al. 2013) in TASSEL 3.0 (Bradbury et al. 2007) on the Jasper cluster on the WestGrid server (https://westgrid.ca). Sequences were not filtered for heterozygosity, as phylogenetic analyses of RAD-seq genomic data are not effected by high rates of heterozygosity, even when artificially imposed on the data (Cariou et al. 2013). Once the SNPs were filtered through the pipeline, 5 base pairs were manually trimmed from the 3' end of all reads, where mis-priming may have occurred during library preparation, and then all loci were removed that were duplicates of the query sequence, reducing the number of loci from 41,622 to 37,147. Sequences were then imported into a spreadsheet and sorted by percent coverage - the proportion of individuals that had been genotyped at a particular locus.

A minimum coverage threshold of 20%, a maximum of 80% missing data per locus, (6627 loci) was used for phylogenetic analyses, since phylogenetic analyses are robust to missing data, and may even be improved by the inclusion of loci with missing data (Wiens 2006; Huang and Knowles 2014). A threshold of 80% (827 loci) was used for population structure (STRUCTURE) (Pritchard et al. 2000) and discriminant analysis on principal components (DAPC) (Jombart 2008) analyses, as their accuracy is sensitive to large amounts of missing data (Pritchard et al. 2000). Programs treat IUPAC (International Union of Pure

and Applied Chemistry) ambiguity codes as uncertainties instead of genotypes (*e.g.* A or G instead of A and G), and some will impute the most likely character state, resulting in a loss of data. To retain phylogenetic signal from genotypic data, concatenated SNP loci were analyzed in a two-column format (one column per allele) as opposed to IUPAC ambiguity codes.

To process the COI data, forward and reverse sequence files from the ABI 3730 sequencer were inspected and aligned in Geneious v7.0.6 (Kearse et al. 2012) before performing multiple sequence alignments in Mesquite v2.75 (Maddison and Maddison 2008) using the ClustalW algorithm (Thompson et al. 2002). The subsequent alignment was examined in Mesquite and modified manually based on codon positions. The two regions of COI were prepared and aligned separately before being concatenated in Mesquite, after which the primer sequences were trimmed from both ends.

2.2.4 DATA ANALYSIS

Maximum parsimony analysis was performed using PAUP* (Swofford 2003) on the COI sequence data, and TNT (Goloboff et al. 2008), due to computational time constraints, on 6627 concatenated SNP loci (20% minimum coverage) in two two-column format (13,254 characters total). Bootstrap analyses with 100 replicates used heuristic searches, random sequence addition and tree-bisection and reconnection (TBR) for both parsimony analyses.

Maximum likelihood trees of COI sequence data constructed using the online web service PhyML 3.0 (Guindon and Gascuel 2003) used the GTR+i+g model of evolution (COI sequence) based on output from jModelTest v2.1.4 (Darriba et al. 2012). Maximum likelihood trees of concatenated SNPs were constructed in the online web services GARLI 2.1 (Genetic Algorithm for Rapid Likelihood Inference; Zwickl 2006) hosted on Molecularevolution.org (Bazinet et al. 2014) and PhyML using the GTR+g (concatenated SNPs), also based on the output of jModelTest. Branches were swapped through nearest neighbor interchange (NNI) and 100 bootstrap replicates were sampled. Maximum likelihood analyses performed in PhyML calculated aBayes (Bayesian-like transformation of approximate likelihood ratio test) (Anisimova et al. 2011) branch support values due to their faster computational speed than Bayesian posterior probabilities and comparable accuracy.

Bayesian analysis and estimates of COI sequence divergence used BEAST2 (Bouckaert et al. 2014) under the GTR+i+g model of evolution, as determined by jModelTest v2.1.4 (Darriba et al. 2012). Divergences were estimated under a lognormal relaxed molecular clock, using the Yule process as a tree prior, and the age of the root was constrained to 33 million years with a standard deviation of 5 million years, as per Wahlberg et al. (2009 and 2006). Major clades were constrained to monophyly and dated according to results from previous studies (*Polygonia*, 19 Ma; *Kaniska* + *Nymphalis* + *Polygonia* clade, 27 Ma) (Wahlberg et al. 2009). Two independent runs of 100 million generations each were used to estimate parameters, sampling every 100,000 generations; all ESS values in Tracer v1.6 (Rambaut et al. 2014) were well above 200, indicating that the runs had reached convergence.

Parsimony networks were built in TCS (Clement et al. 2000) using 228 COI sequences amplified in this study, in addition to 89 COI sequences from GenBank (NCBI; accession numbers available in appendix). I did not include sequences that were amplified only with MilaIV (*P. faunus* – 5 sequences removed, *P. gracilis* – 1, *P. progne* – 3), as these were significantly shorter than the full 1450bp, nor did I include sequences with long strings of missing data (*P. faunus* – 2 sequences removed). Approximately 100 base pairs (the exact range varies between datasets) were excluded from this analysis due to poor sequence quality where the two fragments, the barcode region and JerryI – PatII, aligned at approximately 650 bp from the 5' end. Gaps were treated as missing data to accommodate fragments that varied slightly in length.

SNP genotypes from a matrix of 827 SNP loci, those with no more than 20% missing data, were used to assess the population structure of 239 *Polygonia* specimens from Alberta and British Columbia using STRUCTURE (Pritchard et al. 2000). *Polygonia interrogationis* and *P. comma* were not included in this analysis, as they were only represented by one specimen each in the data matrix. A burn-in of 50 000 replicates was applied, followed by 500 000 MCMC iterations replicated 10 times at each K value from 1-6. A second analysis without *P. faunus* and *P. satyrus* was executed using the same parameters, from K 1-4, to increase the opportunity for resolution within and between *P. gracilis* and *P. progne*.

I performed a DAPC using the statistical software R (R core group) and the package 'adegenet' (Jombart 2008). This multivariate analysis identifies genetic clusters from large

amounts of genetic data without the use of evolutionary models and does not require computationally intense Bayesian clustering algorithms (Jombart et al. 2010). The same 827 SNP loci used in the STRUCTURE analysis were used here, for the same four putative species; *P. faunus*, *P. gracilis*, *P. progne* and *P. satyrus*. The optimal number of genetic clusters was 6, as was the optimal number of principal components.

2.3 RESULTS

2.3.1 PHYLOGENETIC RELATIONSHIPS

In both the maximum parsimony (MP) and maximum likelihood (ML) trees for COI, the mitochondrial lineages associated with *P. satyrus, P. faunus* and *P. progne* all form well-supported monophyletic groups, although the relationships among them are not completely resolved (Figures 2.2, 2.3, 2.4, 2.5). The COI sequences associated with *P. gracilis*, however, are divided into two major lineages, one of which is sister to *P. satyrus*, the other of which is sister to *P. progne* and the *P. satyrus* + *P. gracilis* clade (Figures 2.4, 2.5). The latter haplotype lineage, indicated as G-X, was found in three individuals out of the 19 used in this study. The more common of the two COI lineages is itself divided into two well-supported clades (Figure 2.4), although these clades were not distinguished by geographic distance or habitat, and several localities were represented in both clades. The low number of non-synonymous mutations, base pair substitutions causing an amino acid change (two in total), in the G-X lineage supports that the sequence was of mitochondrial origin and not a nuclear insertion of a mitochondrial gene. Additionally, the two fragments of COI that were sequenced separately before alignment and analysis were not chimeras, or sequences composed of multiple individuals, as both fragments shared the similar pattern of elevated mutations as compared to other conspecifics and produced similar phylogenies when analyzed separately.

Well-supported, monophyletic clades were also produced by the Bayesian analysis that included GenBank entries for COI from most of the remaining species within the genus. The exceptions were *P. gracilis* and *P. oreas,* which did not form separate COI clades, and *P. progne*, whose clade was poorly supported and was therefore collapsed (Figure 2.6). Phylogenetic trees from both datasets were rooted using members of *Nymphalis* and *Aglais* as outgroups (Wahlberg et al. 2003, 2005).

Divergence times estimated in this study (Figure 2.6) were generally much older than in previous studies; however, the error estimates in this study overlapped with those provided by Wahlberg et al. (2009). For example, Wahlberg et al. (2009) placed the divergence of *P. gracilis* from *P. satyrus* at approximately 2.2 Ma, while this study dates the origin of this clade at 7.9 Ma. Similarly, *P. faunus* was dated as diverging from *P. c-album* nearly 6.2 Ma in Wahlberg et al.'s study, while the present study estimated approximately 10 Ma.

The maximum likelihood tree constructed using SNPs depicted most species in this study as well supported and monophyletic; however, the placement of *P. faunus* was not consistent between the maximum likelihood programs PhyML and GARLI, nor was the monophyly of *P. progne* (Figure A.3, A.4). The overall topology produced by PhyML did not correspond that of the maximum parsimony or distance trees using the same data (neither included in this thesis), placing the *P. satyrus* clade sister to the *P. progne* clade; however, species level relationships were unresolved in the tree produced by GARLI, and so PhyML topologies were used to illustrate intraspecific relationships (Figure 2.7, 2.8, 2.9).

Polygonia gracilis is divided into two separate lineages in the trees constructed with COI, yet the same result is not observed in the SNPs. In fact, the three individuals that formed G-X do not form a clade within *P. gracilis* (Figure 2.8). In addition, the placement of *P. gracilis* is not as the sister group to *P. satyrus*, but instead to *P. progne* (Figures 2.8, 2.9).

2.3.2 HAPLOTYPE NETWORKS

Of the three haplotype networks created from the mitochondrial gene COI, *P. faunus* had the highest number of separate haplotypes (40, Figure 2.10). Of these haplotypes, 33 were found in only a single individual, four of them from two individuals, and the remainder of the specimens were split among three main haplotypes. The main haplotype in *P. faunus*, identified by Kodandaramaiah et al. (2009) as A-Ida, was found in 62 of the 123 specimens in the analysis. The other two haplotypes, identified as F-NHa3 and A-WS3 in Kodandaramaiah et al. (2009), were found in 13 and 7 specimens, respectively. Roughly half of the individuals in A-Ida were from the Rocky Mountains, and a third of them were collected from central Alberta.

Polygonia progne contained the next highest number of haplotypes, with 26, of which 19 were unique to single individuals. Three haplotypes were found in two individuals, and three other haplotypes were in 3, 4 and 5 individuals each. The remaining 40 individuals shared the same haplotype, named P-A here. Although 22 of the 69 *P. progne* specimens were collected from the same locality in southern Alberta, only 7 of those had haplotype P-A. Most of the individuals with the P-A haplotype were from central Alberta, although that haplotype occurred as far as Quebec and New Hampshire.

Polygonia satyrus and *P. gracilis* haplotypes were contained within one network due to similarities among COI sequences of the two species. *Polygonia satyrus* exhibited 22 unique haplotypes and *P. gracilis* had 14; the two species were separated by as few as five nucleotide substitutions. Of the 22 haplotypes in *P. satyrus*, 17 were represented by single individuals, two were found in two specimens, one by three specimens, and another two by five specimens each. The other two haplotypes, named S-A and S-B here, are nearly evenly split among the remaining 36 specimens (19 and 17, respectively).

There are 14 haplotypes in *P. gracilis*, 8 of which are only represented by one specimen, 3 are observed in two specimens, and one has 3 representatives. The two main haplotypes in *P. gracilis* contain 14 (G-A) and 7 individuals, which largely correspond to the two minor clades in *P. gracilis*. The haplotype G-X is also shown on this network, but only by extending the connection limit to 30 in TCS parameters; there are 22 mutations between this group and the haplotypes contained by the other specimens of the species. While there are 3 specimens in the G-X lineage, one of these was amplified with MilaIV, and was removed from this analysis due to sequence length.

2.3.3 POPULATION STRUCTURE

According to STRUCTURE analysis, the optimal number of populations was k=3, as calculated by the Evanno method in Structure Harvester (Earl and VonHoldt 2012). *Polygonia faunus* and *P. satyrus* were genetically distinct, but *P. progne* and *P. gracilis* were indistinguishable from each other in the first run (Figure 2.11). A second STRUCTURE run on a restricted dataset under the same parameters was used to assess any genetic differentiation between *P. gracilis* and *P. progne*, and K=2 was the optimal number of populations. After removing *P. faunus* and *P. satyrus* from the data set, *P. progne* and *P. gracilis* appear much more genetically distinct.

There was little evidence of population structure within species, even when comparing G-X to *P. gracilis* (Figure 2.11). Any genetic differentiation within *P. gracilis* found in mtDNA (Figure 2.10) was not observed in SNPs. It should be noted, however, that distinct groups were visible within species that corresponded to the date of GBS processing (Figures 2.7, 2.8, 2.9). This indicates that there may be observable differences that result from the variation in library preparation between GBS runs. SNP loci occurring in the last 5 base pairs on the 3' end of a read were removed prior to this analysis; however false SNPs from mispriming on reads shorter than 64 bases may not have been filtered. More stringent filtering is required to minimize this artifact.

The DAPC shows clear species boundaries and tight clusters of individuals (Figure 2.12); there was no overlap between any of the species. There were no discernable clusters of *P. g. gracilis* and *P. g. zephyrus*; the two groups were not distinguished from each other. One *P. faunus* individual appeared to be intermediate between *P. gracilis* and *P. faunus*; however, this result was not observed clearly in the STRUCTURE results.

2.4 DISCUSSION

2.4.1 THE *POLYGONIA* SPECIES TREE

Relationships within the genus *Polygonia* have been the subject of several studies that have demonstrated strong phylogenetic discordance between mitochondrial and nuclear markers, specifically in the relative positions of *P. satyrus, P. comma*, *P. interrogationis* and *P. oreas* (Nylin et al. 2001; Wahlberg and Nylin 2003; Wahlberg et al. 2003, 2009). The use of genomic SNPs for the first time in this group has now produced a topology that better reflects the relationships observed in the ecology (*e.g.* host plant use, larval nest building) and morphology (*e.g.* wing pattern traits, genitalia) of *Polygonia* species; however, many clades are not well supported.

Using 1450 base pairs of COI, this study found a similar topology to Wahlberg et al. (2009), with *P. satyrus*, *P. gracilis* and *P. progne* in a clade together, while the topology constructed from 6627 genomic SNPs placed *P. satyrus, P. interrogationis* and *P. comma* in one clade, and *P. gracilis* and *P. progne* in another. The SNP dataset analyzed in GARLI produced a topology that is similar to the tree constructed with morphological, ecological and

behavioural characters (Nylin et al. 2001, Figure 3; Wahlberg and Nylin 2003, Figure 1), nuclear gene trees (Wahlberg and Nylin 2003, Figure 1; Wahlberg et al. 2009, Figure 2), and also coincides with host plant use in the genus (Weingartner et al. 2006, Figure 2). *Polygonia interrogationis* and *P. comma* both feed on the plant families Urticaceae, Ulmaceae and Cannabidaceae which are all "urticalean rosids") (Weingartner et al. 2006). *P. satyrus* feeds on Urticaceae and Cannabidaceae (and Salicaceae for one population in Arizona). *Polygonia progne, P. gracilis* and *P. faunus* all feed on a broader taxonomic range of host plants, consisting of Grossularaceae, Betulaceae and Ericaceae, and *P. faunus* also feeds on Salicaceae. The COI topology either requires three independent host plant shifts to urticalean rosids by *P. interrogationis*, *P. comma* and *P. satyrus* or three independent shifts to broader host plant use by *P. progne*, *P. gracilis* and *P. faunus* (Weingartner et al. 2006). The SNP topology provides a more parsimonious account of host plant use, requiring one shift to a broader use of hosts in the ancestor of North American *Polygonia* and one return to urticalean rosids by the *P. interrogationis*, *P. comma* and *P. satyrus* clade, providing evidence that this tree may better reflect the evolutionary history of the group.

The utility of RAD-seq generated SNPs in systematics is not restricted to this study (Combosch and Vollmer 2015; Longo and Bernardi 2015; Pante et al. 2015); in fact, many studies have found that genomic NGS data produces highly supported monophyletic clades and fully resolved nodes, even in groups whose evolutionary history has been tenuous (Wagner et al. 2013). Across many taxa, species boundaries and evolutionary relationships are being revealed through genomic data – researchers working on cichlids (Wagner et al. 2013), geckos (Leaché et al. 2014), swordtail fish (Jones et al. 2013), octocorals (Pante et al. 2015), flowing plants (Eaton and Ree 2013), ground beetles (Cruaud et al. 2014) and other nymphalid butterflies (Nadeau et al. 2013) have all found unprecedented results – delineation of species, resolution of trees, high topological support indices – using genome wide RAD-seq data. The use of genomic SNPs is an appealing alternative to sequence data, given the absence of a consensus in animal systematics on the use of standard molecular markers (Caterino et al. 2000; DeSalle et al. 2005; Dupuis et al. 2012) and the criticisms of DNA barcoding (Meyer and Paulay 2005; Will et al. 2005; Meier et al. 2006).

2.4.2 ASSESSMENT OF SPECIES BOUNDARIES AND TAXONOMIC RANKS IN *POLYGONIA*

All four *Polygonia* species investigated in this study exhibited clear boundaries at the species level using SNPs, and all species except *P. gracilis* (sensu lato) demonstrated well defined species boundaries in mtDNA. Clades in species level phylogenies were largely well supported and monophyletic in the maximum parsimony and maximum likelihood analyses of both SNPs and COI, although not all relationships were fully resolved. *Polygonia progne* is not monophyletic in the SNP maximum likelihood analysis (Figure 2.8) and the position of *P. faunus* is poorly supported in maximum parsimony analysis (Figures 2.2, 2.7). This inconsistency of the topological support while other studies have reported high support using similar methods could indicate an underestimate of the necessary data filtering thresholds, or conservatism when it comes to the inclusion of missing data (Huang and Knowles 2014). In contrast, the STRUCTURE analysis and DAPC both showed very little overlap in SNP genotypes and clearly resolved clusters corresponding to species. Using the genomic integrity species concept, the evidence in this study supports the species designation of all four species sampled, given the tight, isolated clusters produced by genomic SNPs in DAPC; however, if the phylogenetic species concept were applied in a mtDNA analysis, the species status of *P. gracilis* would be in question. Similarly, *P. progne* may or may not be considered a species depending on the analysis, as it was inconsistently monophyletic in both COI and SNPs (Figures 3.5, 3.8), violating a central tenet of the phylogenetic species concept. This dissonance in species boundaries between concepts and methods illustrates the importance of careful application of species concepts and the need for multiple lines of evidence.

Subspecific boundaries were not clearly delineated by either genetic marker type in this study. The phylogenetic relationships observed between members of *P. g. zephyrus* and *P. g. gracilis* sampled outside of the sympatric zone support the current hypothesis that they are conspecific; however, their status within the species remains unclear. To assess the genetic distinctness of two populations, specimens should be sampled across the sympatric region extending past the zone of interaction between the two populations (*i.e.* in a transect) (Hare and Avise 1996). If the two populations are genetically distinct but hybridizing, then a stepped cline pattern should be observed in individual genotypes across the transect; a steep slope indicates that reproductive barriers are present, while a gradual cline indicates that

there are few barriers to reproduction (Toews et al. 2011; Krajmerová et al. 2015). Due to insufficient sampling within and outside of the area of interaction in Alberta, a confident assessment of *P. gracilis gracilis* and *P. g. zephyrus* status within the species is not possible in this study.

The issue of subspecies is contentious – many authors view them as entirely unnecessary, or even damaging to conservation efforts (Wilson and Brown 1953; Zink 2004; Mallet 2007), others view them as important evolutionary units (Phillimore and Owens 2006; Remsen 2010) while others think that they are valid taxonomic units if treated correctly (Mayr 1982b; Patten 2010). From a conservation perspective, the designation of taxa below the species level is also fraught with debate; the term evolutionary significant unit (ESU) was developed in an attempt to facilitate the conservation of these subspecific units (Ryder 1986), but the number of operational definitions has increased substantially since its inception (Proshek et al. 2015, and references therein). From Darwinian times to today, many have viewed subspecies as incipient species (Wahlberg et al. 2009), and therefore valid evolutionary units; others disagree, viewing subspecies as geographically associated phenotypes without evolutionary implications (Zink 2004). Previously, many studies have tried to validate or invalidate the rank of subspecies on the basis of reciprocal monophyly (O'Brien and Mayr 1991; Moritz 1994; Zink 2004; Phillimore and Owens 2006); however, regardless of whether the units are incipient species or geographically separated phenotypes, gene flow will occur where these populations meet, and as such, monophyly will not likely be observed (Kodandaramaiah et al. 2012b).

The boundaries between subspecies, as well as their utility, are also unclear, since, by definition, subspecies introgress readily where their ranges overlap, confounding the genetic structure therein (Kodandaramaiah et al. 2012b). As a result, the definition of a subspecies relies more on morphological dissimilarity than genomic integrity. The consensus is that subspecies must be 1) morphologically diagnosable, and 2) geographically associated (Wilson and Brown 1953; Mayr and Ashlock 1991; Patten and Unitt 2002). The validity of a subspecies designation, when fronted with appropriate sampling density and breadth, can be addressed with morphological data through the use of the 75% rule (Amadon 1949). This rule states that 75% of the individuals in one population must be morphologically diagnosable from all members of the opposing population (Amadon 1949; Haig et al. 2006).

In the case of *P. gracilis gracilis* and *P. g. zephyrus*, their ranges meet in Alberta, so many individuals sampled in this study are likely intermediates between the two if they are valid taxa, and as such will be expected to have more overlap in morphological traits than those from their exclusive ranges. In addition, two colour variants are observed in both taxa (Royer 1988, Bird et al. 1995) further confusing the issue of diagnosability. A few localities outside of the border between the two ranges were sampled; however, only four specimens were collected. These four specimens were not morphologically differentiable from each other. Broader and denser geographic sampling outside of Alberta is required to answer the question of the validity of the subspecies rank of *P. g. gracilis* and *P. g. zephyrus* either by molecular or morphological methods.

2.4.3 EVOLUTIONARY AND PHYLOGEOGRAPHIC HISTORY OF *POLYGONIA*

While phylogenetic discordance is generally seen as an obstacle to uncovering a group's evolutionary history, it can often point to significant events that have shaped the genetic structure of a genus or species (Maddison 1997). Genetic incongruence can be caused by the retention of ancestral polymorphisms by random assortment, introgression of haplotypes, *Wolbachia* infections, ancient or current introgression, gene duplication and extinction, incomplete lineage sorting, and a suite of other mechanisms (Hoelzer 1997; Maddison 1997; Sperling and Roe 2009). Close comparisons between gene histories can provide insights into which of these mechanisms created that particular discordance, elucidating the true phylogeny of the group. This information, in concert with Bayesian divergence estimates, paleogeography and paleoclimatology can provide a more complete evolutionary history of the group beyond phylogenetic relationships.

The genus *Polygonia* has experienced at least two independent colonization events of North America from the Palearctic (Wahlberg et al. 2005, 2009). Based on mitochondrial divergence time estimates in this study, which are sensitive to appropriate selection of priors and outgroups, these two events occurred at approximately 17 Ma (million years ago) in the Tortonian age of the Miocene epoch (the lineage containing *P. progne*, *P. gracilis*, *P. interrogationis*, *P. comma*, *P. satyrus*, *P. g-argenteum*, *P. oreas* and *P. haroldii*), and 11 Ma in the Missinian age of the Miocene (*P. faunus* lineage) (Figure 2.6). During this time ancestral populations likely migrated across the Bering land bridge (Marincovich and Gladenkov 1999), perhaps following the expansion of their host plant range across North America,

which may have then contributed to the broadening of host plant use in the genus (Weingartner et al. 2006; Janz and Nylin 2008).

The Miocene was a time of great environmental change –the climate had transitioned to an "ice house world" (Miller et al. 1991; Zachos et al. 2001), open vegetation such as tundra and grasslands were expanding, and a period of warming and rapid cooling in the mid-Miocene transformed ecosystems, expanding the Northern coniferous forests (Behrensmeyer 1992). These changes may have drastically affected the range of *Polygonia* and their expansion across North America. The Pleistocene glaciations followed c14.5 Ma after the colonization of and expansion of *Polygonia* across North America, covering much of the continent in ice during glacial periods (Ehlers and Gibbard 2004). During this time, *Polygonia* populations may have been fragmented on opposite sides of ice sheets or in refugia, isolated areas not covered in ice (Bennett et al. 1991), until the glaciers retreated. Some populations may have been isolated on or near the Bering land bridge, while others may have been isolated to the south of the ice sheets, only meeting after the last deglaciation when the corridor through the glaciers opened 12.5 Ka (thousand years ago) (Dyke et al. 2013) through what is now Alberta. The frequent and drastic change in their ancestral geographic ranges has molded the genetic structure we see today.

Consideration of many lines of evidence can contribute to a broader range of hypotheses for observed patterns in the data. For example, *Polygonia faunus* and *P. progne* both exhibit strikingly low haplotype diversity in COI, with 67% and 58% of the individuals studied sharing the same haplotype, respectively (Figure 2.10). This general lack of genetic diversity in these two species could be explained by a bottleneck in the Wisconsinan glaciation from which they have not yet recovered. Alternatively – or additionally – since these haplotype networks were generated from COI sequences, the maternal inheritance of mitochondrial genes could provide another explanation for the lack of genetic diversity in *P. faunus* and *P. progne* – *Wolbachia* infection. This cytoplasmic parasite uses male-killing, feminization and cytoplasmic incompatibility between infected and non-infected individuals to provide a reproductive advantage to infected populations, which could result in a reduction of mitochondrial genetic diversity similar to what is seen in these two species (Werren et al. 2008).

2.4.4 INTROGRESSION IN *P. SATYRUS* AND *P. GRACILIS*

The strongly supported discordance between mtDNA and nDNA in the genus can be seen at several nodes in this study, but it is particularly noticeable in the position of *Polygonia satyrus* – adjacent to *P. gracilis* in mtDNA, but to *P. comma* in SNPs. In addition to the position of *P. satyrus*, another notable difference between the mtDNA and nDNA topologies is the appearance of a new mitochondrial lineage in *P. gracilis,* referred to as G-X. One explanation for the unexpected divergence of a mitochondrial lineage is gene transmission from the mitochondrion into the nucleus – a NUMT (nuclear mitochondrial gene) (Lopez et al.; Zhang and Hewitt 1996). This type of mutation is well described in various mammal species and these insertions behave much differently from genes under functional constraint because many of them are not transcriptionally active (Lopez et al.). These differences result in a higher ratio of non-synonymous mutations, those that would result in an amino acid change during transcription, to the total number of base pair substitutions when compared to other sequences of the same gene. The absence of stop codons and the infrequency of non-synonymous mutations (2 changes total out of 483 amino acids in the COI sequence) provide evidence for the mitochondrial origin of these sequences as opposed to a gene insertion.

Further evidence of discordance is found in the haplotype diversity in the group (Figure 2.6). *P. satyrus* and *P. gracilis* share a parsimony network and some individuals are only five base pair substitutions apart, producing a sequence divergence of only 0.8% between the two species (Table 2.2). Interestingly, the two G-X individuals were separated from their conspecifics by 23 nucleotide substitutions, with 2.2% sequence divergence between them.

This discordance can be caused by a great many events, although the most likely in this case seem to be either ancient introgression with the ancestor of *P. gracilis*, or the retention of an ancestral mitochondrial haplotype (Wahlberg et al. 2009; Barker et al. 2015; Naciri and Linder 2015). If this were the result of retention of an ancestral haplotype, this haplotype should be observed throughout the range of *P. satyrus*; if ancient hybridization were the cause, only individuals from areas where the two species overlapped historically would possess this haplotype. Considering that the current study did not substantially sample outside of Alberta and British Columbia, it is not possible to confidently assign one or the other as the cause of this discord; therefore, further sampling is required outside of the

sympatric region of *P. gracilis* and *P. satyrus*. Even in larger studies, samples of *P. satyrus* have been collected from the western part of its range, generally no further east than Alberta and not in great number (Nylin et al. 2001; Wahlberg and Nylin 2003; Wahlberg et al. 2009). Similarly, *P. gracilis* has been collected in eastern North America, but no further west than Quebec, Canada, leaving much of its range unsampled (Wahlberg et al. 2009).

The G-X mtDNA places it as the sister group to that of *P. gracilis* + *P. satyrus* at an unresolved node with *P. progne,* as compared to the SNP topology that indicates a monophyletic *P. gracilis* as the sister group (or nested within, depending on the monophyly of *P. progne*) to *P. progne*, and *P. satyrus* + *P. comma* as the sister group to *P. interrogationis* (Figure 2.2). Bayesian divergence estimates date this ancient G-X mitochondrial lineage to 11.17 Ma, while the other *P. gracilis* lineage diverged from *P. satyrus* 7.76 Ma (Figure 2.6). This could indicate that mitochondrial introgression took place approximately 7.76 Ma between the ancestors of *P. gracilis* and *P. satyrus*, causing very similar mitochondrial haplotypes. While the divergence estimates produced through Bayesian inference are sensitive to prior selection, these dates place the introgression event well before the Pleistocene, even at the lower bound of the 95% confidence interval (5.17 – 10.45 Ma), which would place it in the Pliocene to late Miocene. If this event is responsible for the phylogenetic discordance in *Polygonia*, then it is possible that the members of the G-X lineage have retained their ancestral mitochondrial haplotype, placing them further from *P. satyrus* and nearer to *P. progne*, as the SNPs and nDNA indicate for the other members of the species (Wahlberg et al. 2009).

Genetic discordance like those seen in *Polygonia* can have an enormous effect on how we interpret phylogenetic relationships and delimit species. Multisource approaches such as integrative and iterative taxonomy identify areas of discordance and provide insight into the evolutionary history of the group to prevent errors that could be caused by relying on a single line of evidence (Schlick-Steiner et al. 2010; Yeates et al. 2011). In this study, *P. gracilis* was polyphyletic when examined with the mitochondrial gene COI, and monophyletic using SNPs. If DNA barcoding alone had been used to assess species boundaries in this group, a new species may have been posited for the lineage. Under the genomic integrity species concept, *P. gracilis* and G-X are undoubtedly one species, as they cluster tightly together in SNP genotypes and maintain their core genomic identity, despite contact with other species. Species concepts can have a substantial effect on our treatment of populations and lineages

in a study – careful consideration of all available evidence and broad genetic sampling are essential when delineating species and assessing species designations.

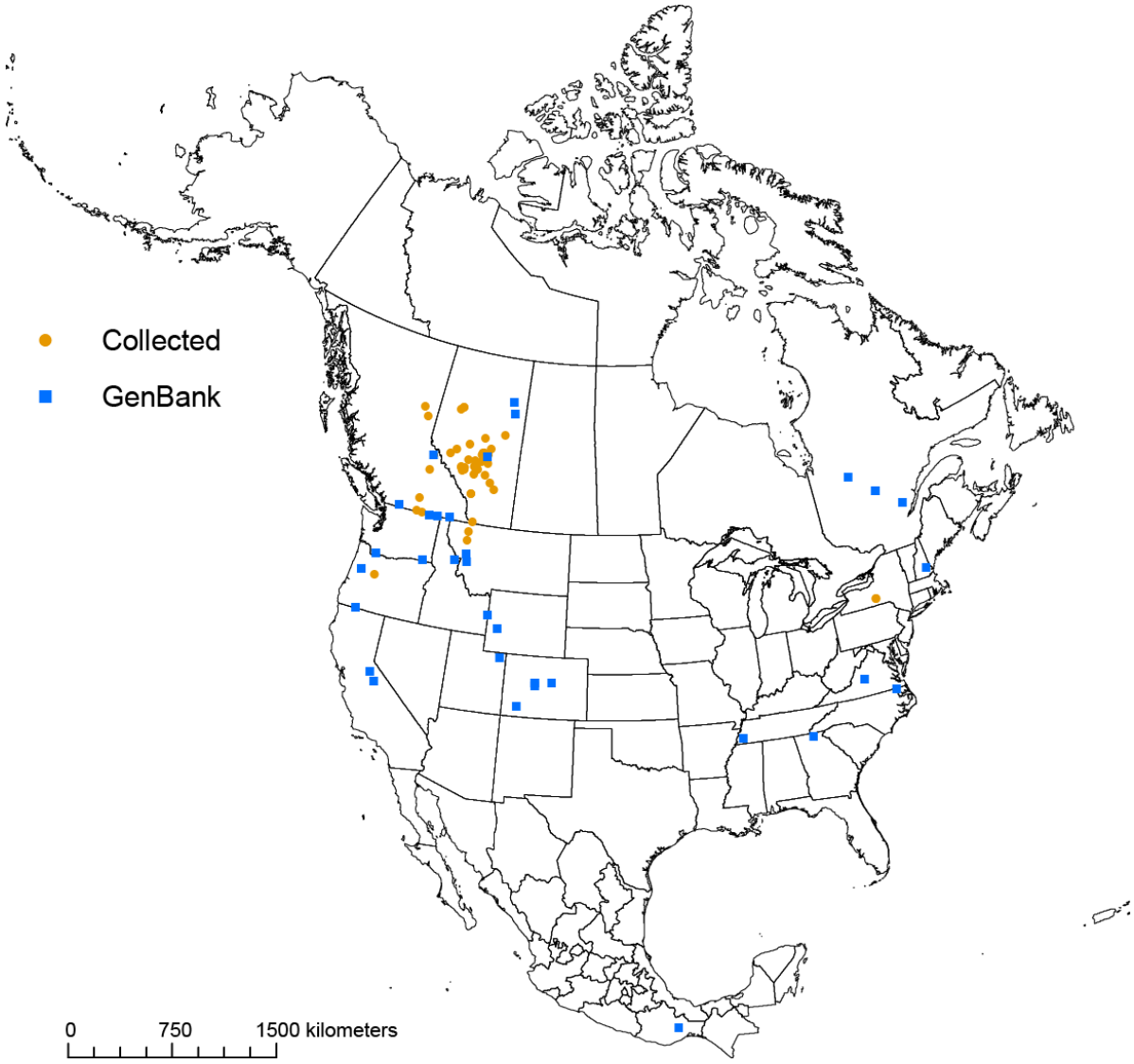


Figure 2.1: Localities of *Polygonia* and outgroup specimens extracted for use in this study (orange circles). This dataset was supplemented with COI sequences from GenBank (blue squares) (Accession numbers; Table A.1).

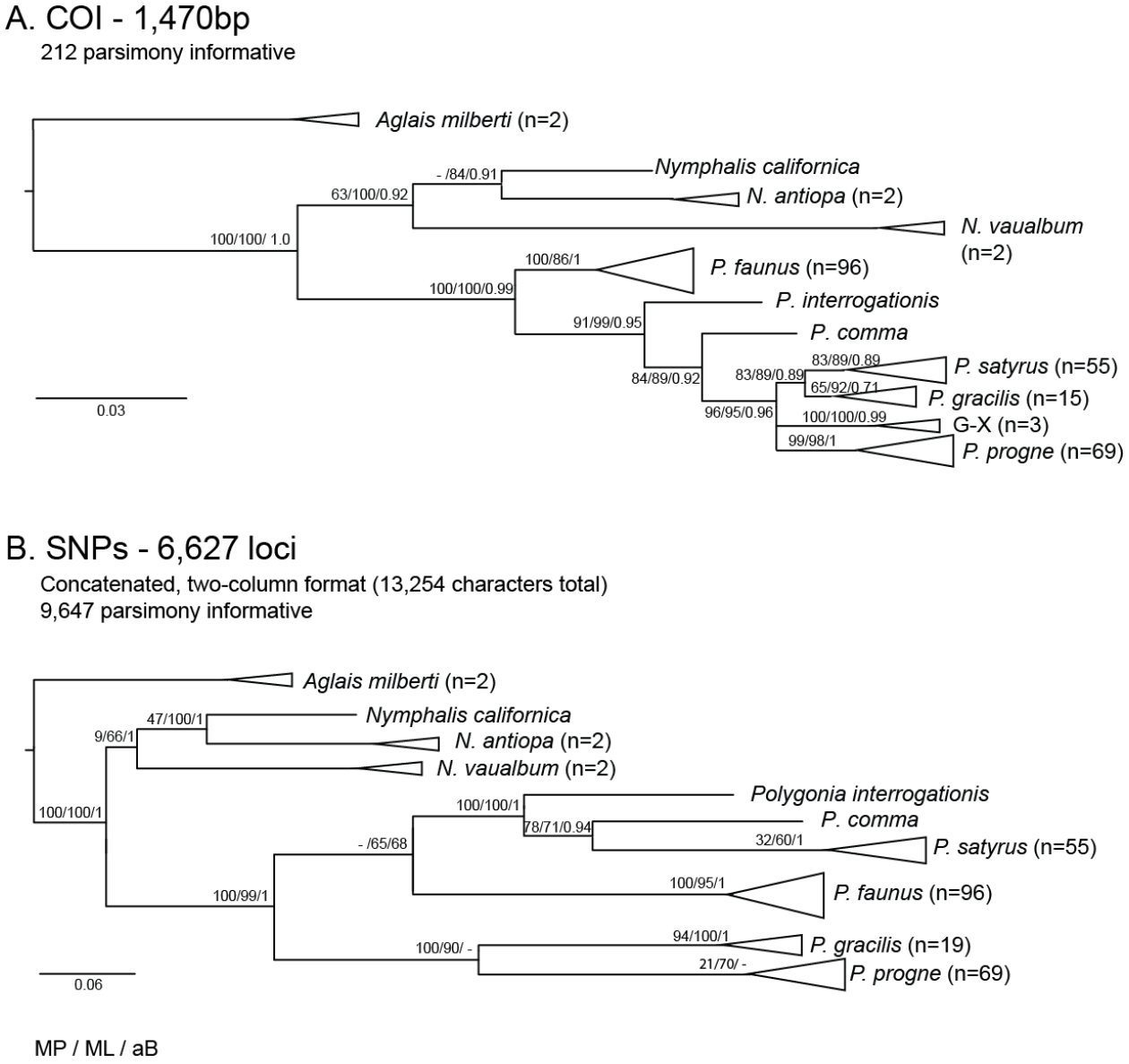


Figure 2.2: (**A.**) Maximum likelihood tree constructed from COI (1470bp) in PhyML, maximum parsimony bootstrap values calculated in PAUP*. (**B.**) Maximum likelihood tree constructed using SNPs (Minimum 20% coverage of individuals per locus, concatenated in two column format, 13,254bp) in GARLI, maximum parsimony bootstrap values calculated in TNT. aBayes support values were calculated in PhyML (Anisimova et al. 2011). Wedges represent collapsed terminal branches; the size of the wedge is approximately proportionate to the number of branches collapsed. Expanded branches can be seen in figures 2.3, 2.4, 2.5, 2.7, 2.8, and 2.9. *Nymphalis californica*, *Polygonia interrogationis* and *P. comma* are represented by one specimen each.

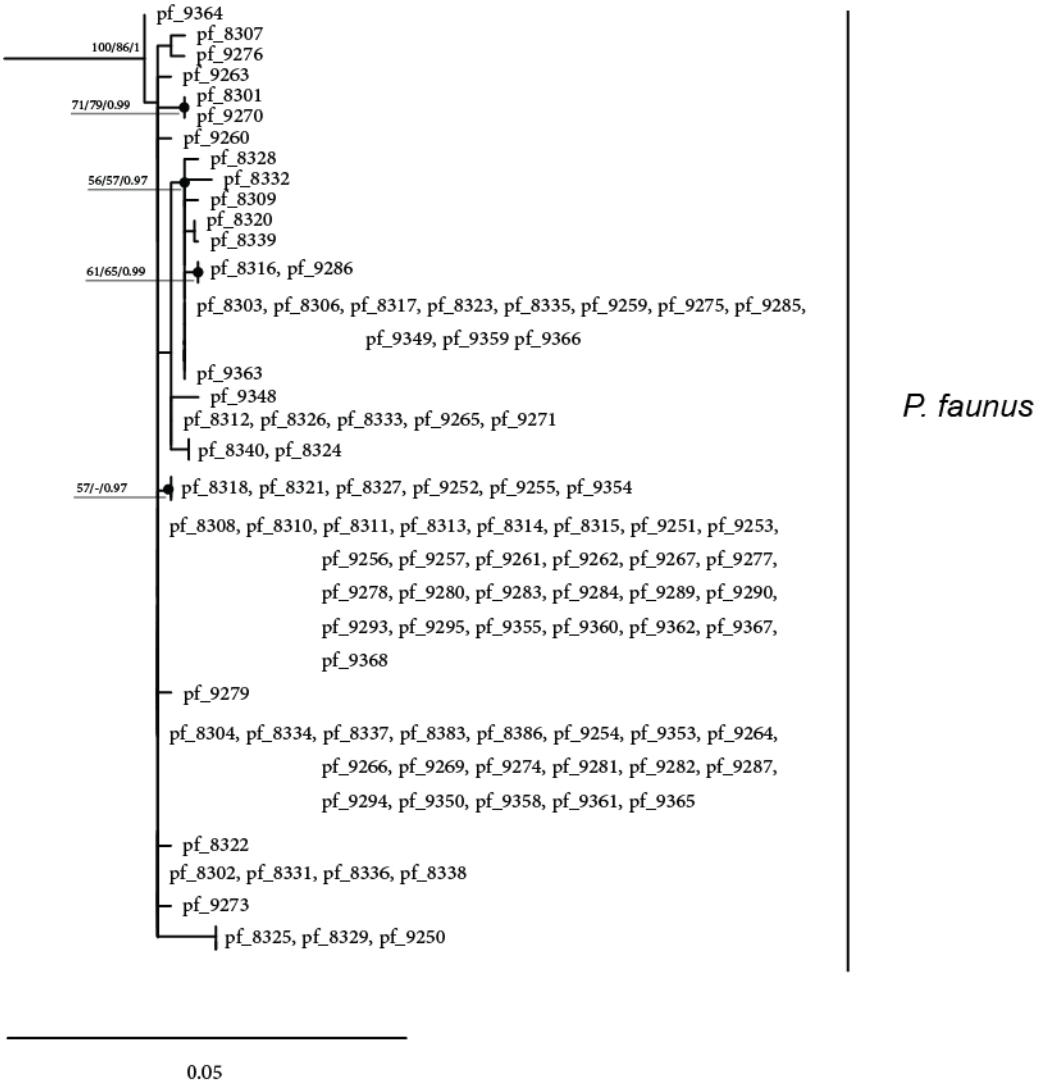
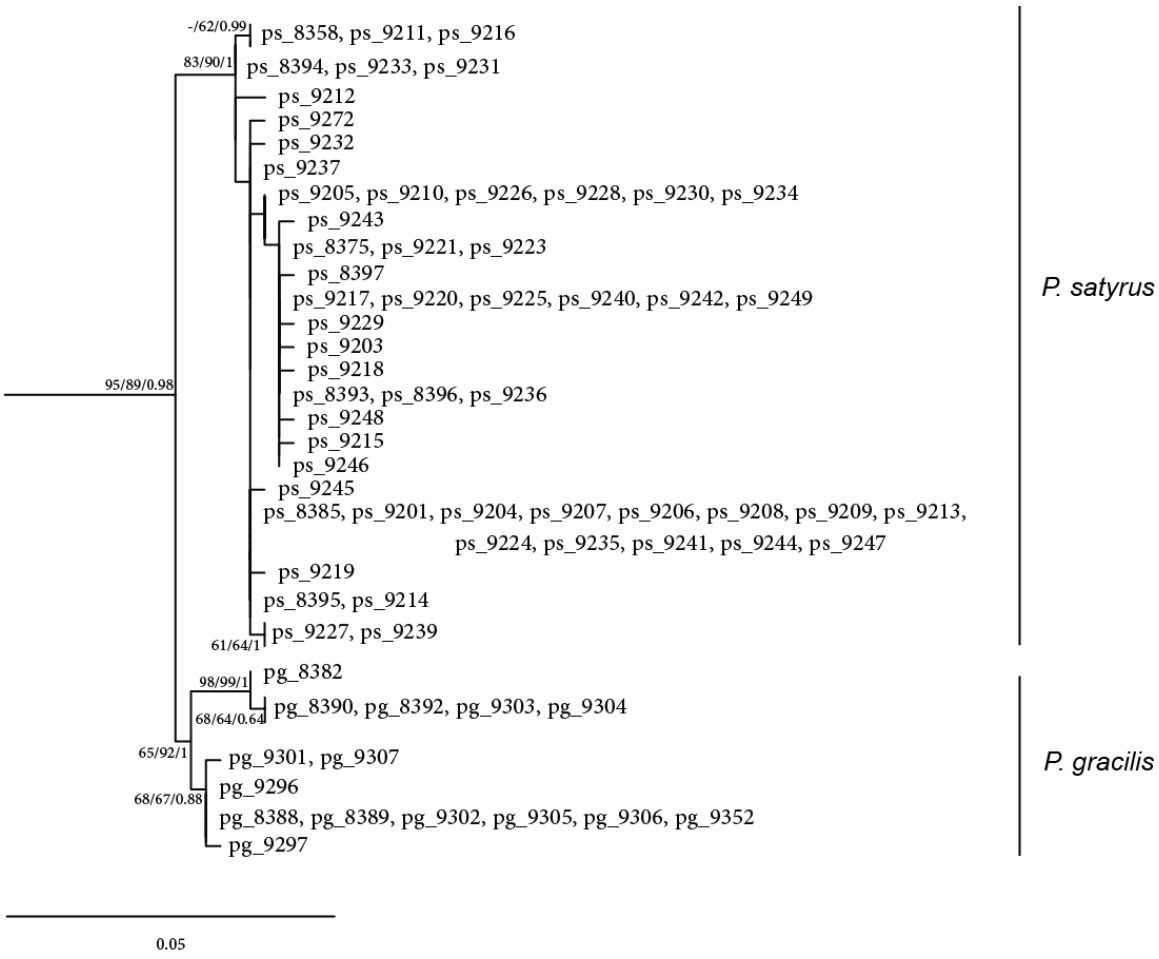
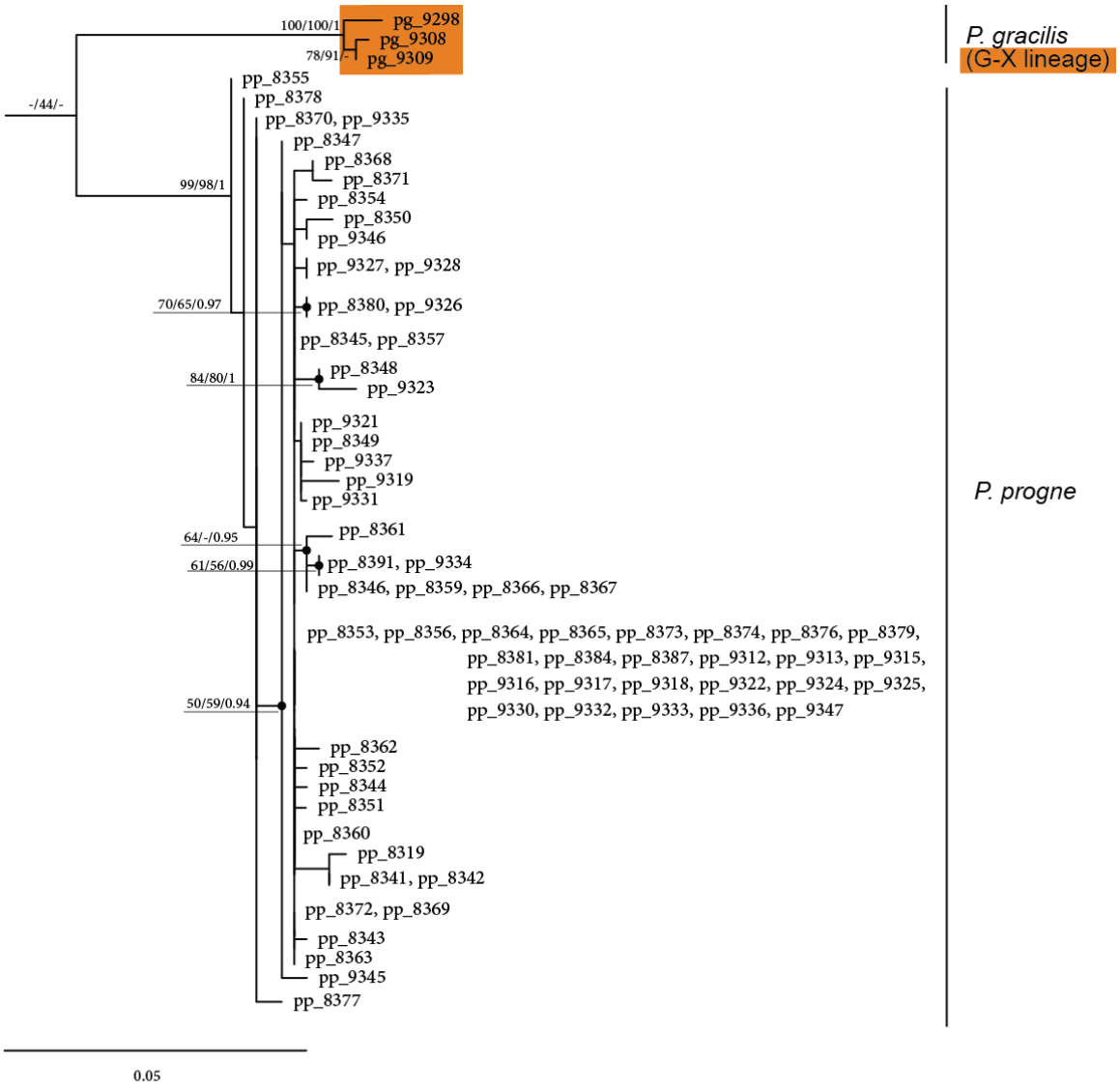


Figure 2.3: Maximum likelihood tree constructed from 1470bp COI, clade containing all *Polygonia faunus* specimens, expanded from Figure 2.2A. All maximum parsimony bootstrap, likelihood bootstrap, and aBayes support values are given in clades that are present in more than one analysis.



MP/ML/aB

Figure 2.4: Maximum likelihood tree constructed from 1470bp COI, clade containing all *Polygonia satyrus* and *P. gracilis* specimens, expanded from Figure 2.2A. All maximum parsimony bootstrap, likelihood bootstrap, and aBayes support values are given in clades that are present in more than one analysis.



MP/ML/aB

Figure 2.5: Maximum likelihood tree constructed from 1470bp COI, clade containing all *Polygonia progne and* G-X specimens, expanded from Figure 2.2A. All maximum parsimony bootstrap, likelihood bootstrap, and aBayes support values are given in clades that are present in more than one analysis.

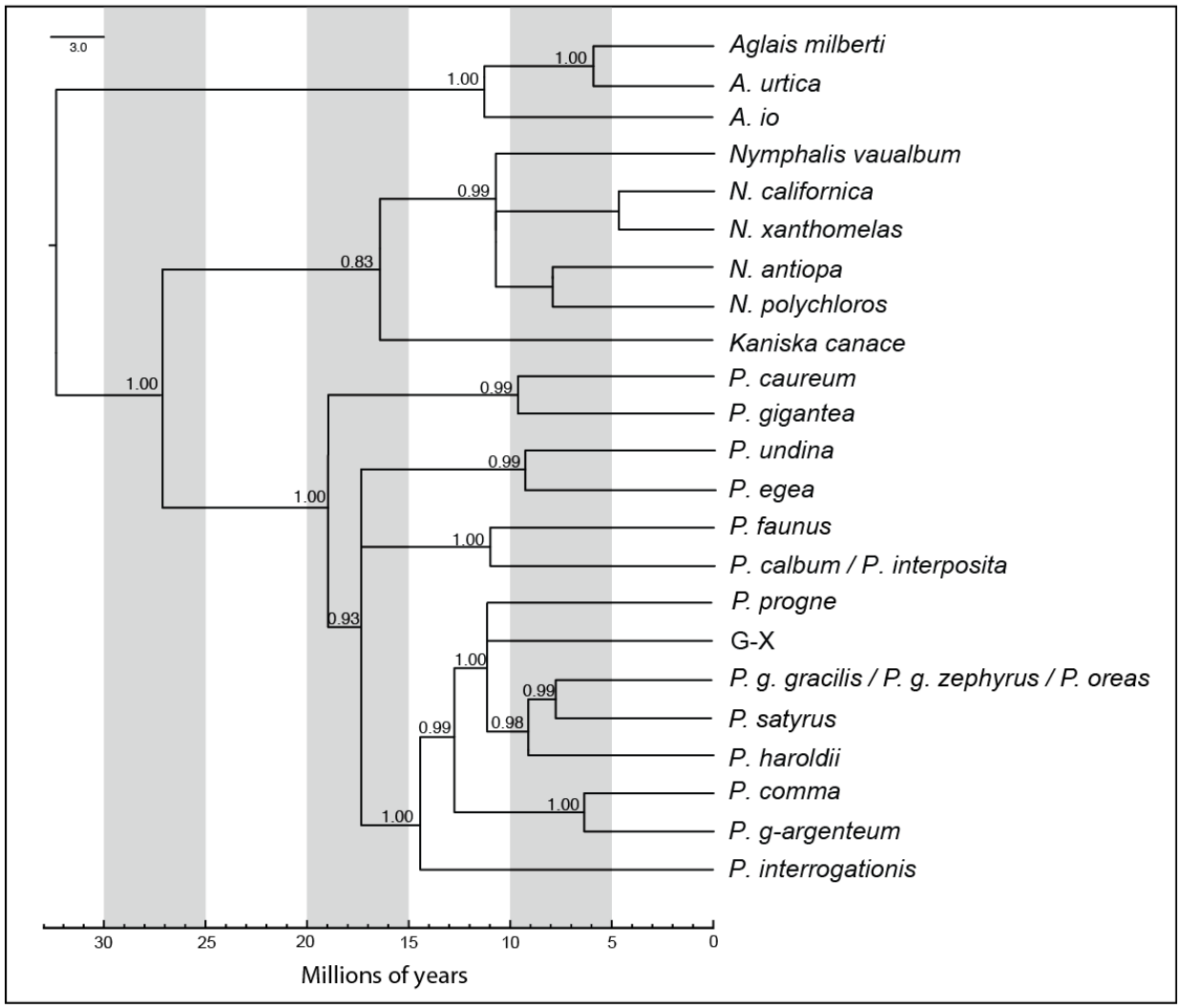


Figure 2.6: Bayesian chronogram of mitochondrial COI (1470 bp) in BEAST 2 (Kearse et al. 2012) showing time from present. Support indices are posterior probabilities; branches with support values lower than 0.70 have been collapsed.

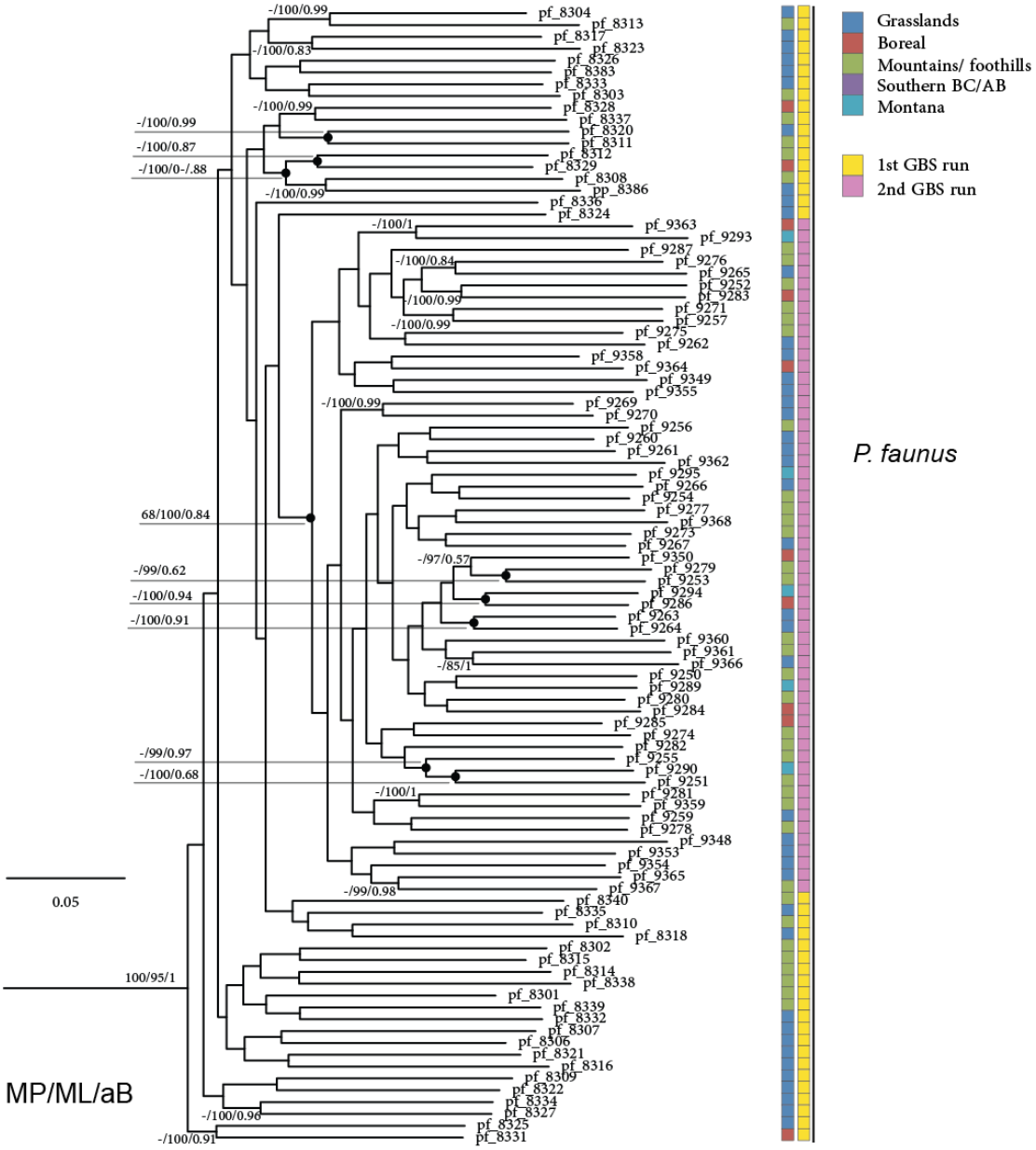


Figure 2.7: Maximum likelihood phylogeny constructed from 6627 concatenated SNP loci in two-column format (13,254 characters total) in PhyML, clade containing all *Polygonia faunus* specimens, expanded from Figure 2.2B. All maximum parsimony bootstrap, likelihood bootstrap, and aBayes support values are given in clades that are present in more than one analysis. Regional colour coding corresponds to the legend map in Figure 2.10.

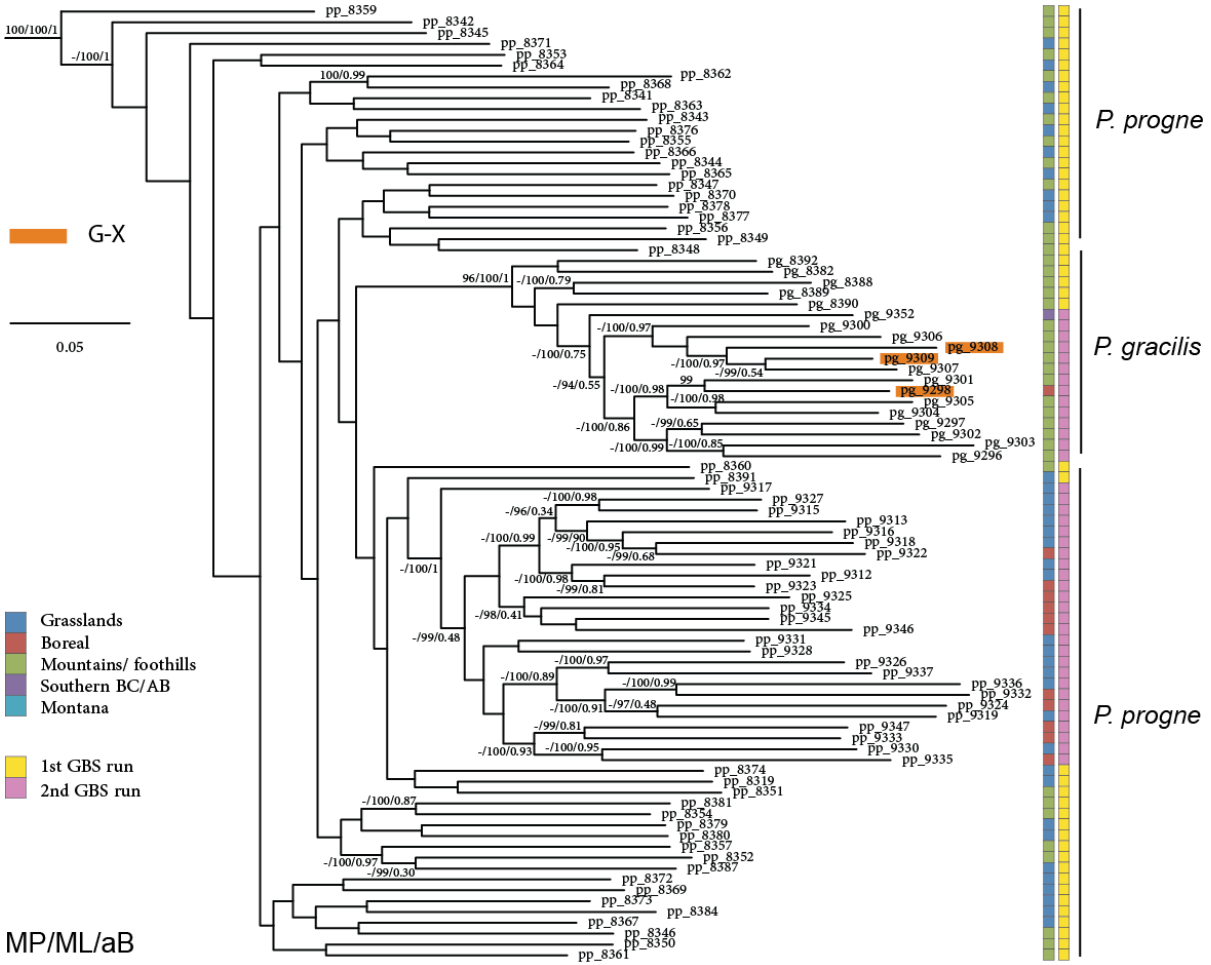


Figure 2.8: Maximum likelihood tree constructed from 6627 concatenated SNP loci in two-column format (13,254 characters total) in PhyML, clade containing all *Polygonia progne and P. gracilis* specimens, expanded from Figure 2.2B. Specimens corresponding to the G-X mitochondrial lineage are highlighted in orange. All maximum parsimony bootstrap, likelihood bootstrap, and aBayes support values are given in clades that are present in more than one analysis. Regional colour coding corresponds to the legend map in Figure 2.10.

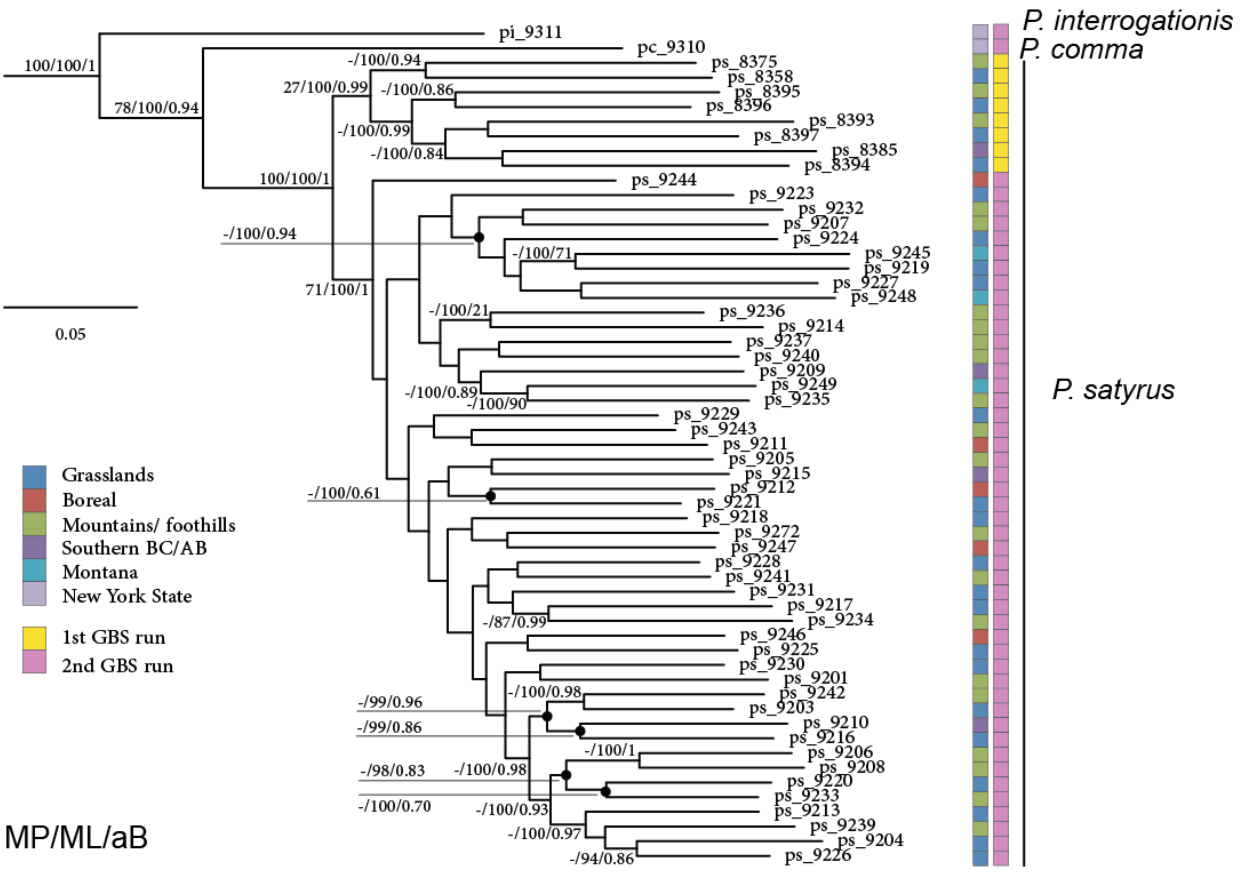


Figure 2.9: Maximum likelihood tree constructed from 6627 concatenated SNP loci in two-column format (13,254 characters total) in PhyML, clade containing all *Polygonia satyrus* specimens, expanded from Figure 2.2B. All maximum parsimony bootstrap, likelihood bootstrap, and aBayes support values are given in clades that are present in more than one analysis. Regional colour coding corresponds to the legend map in Figure 2.10.

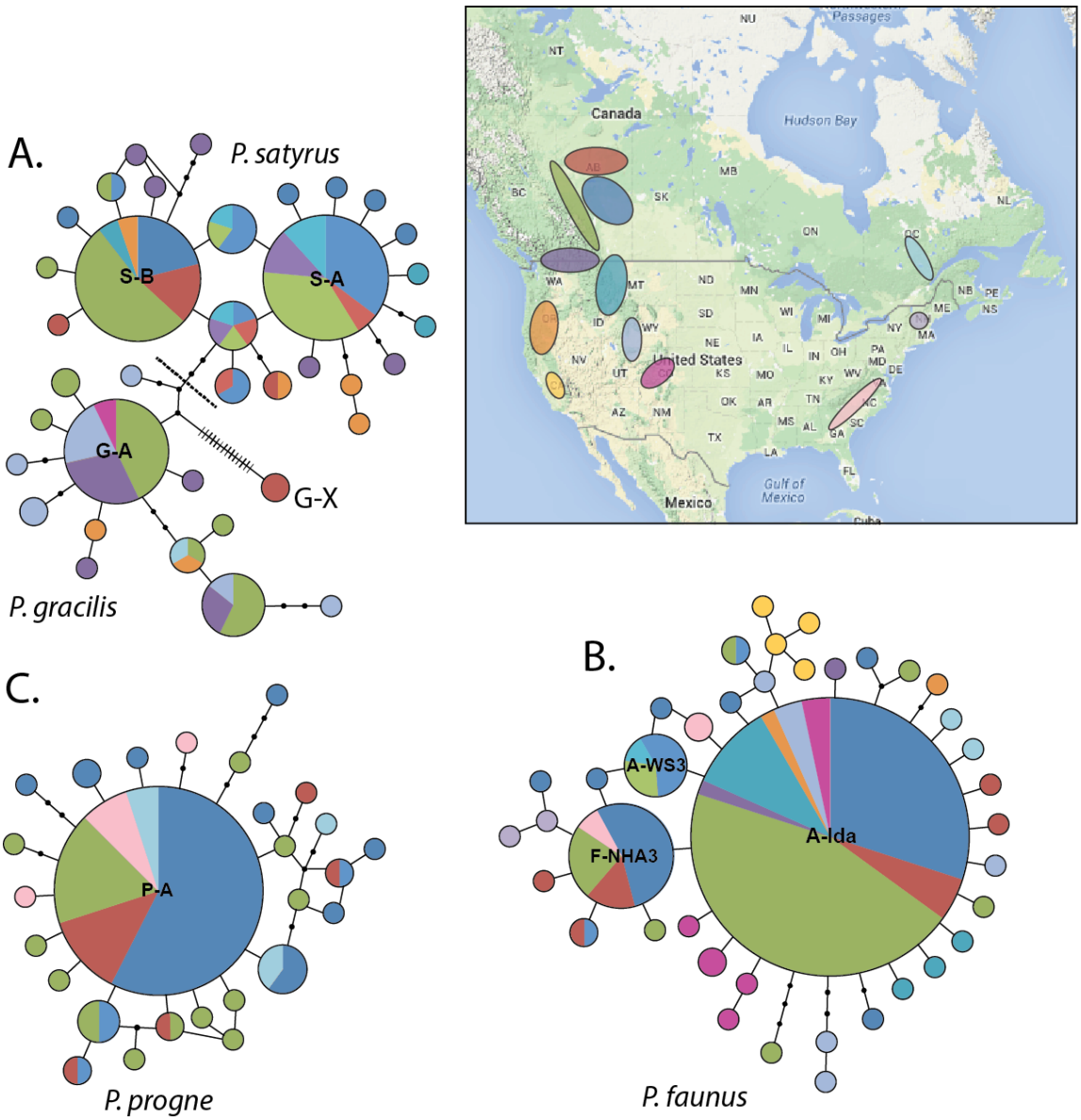


Figure 2.10: Parsimony networks created in TCS using 1350bp of COI for (**A**.) *Polygonia satyrus* and *P. gracilis* (dotted line indicates the boundary between the two; *P. satyrus* above, *P. gracilis* below) (**B.**) *P. faunus* and (**C.**) *P. progne*. Circles represent individual haplotypes; the diameter is proportional to the number of individuals with that haplotype. Small black dots represent missing haplotypes, while the lines connecting each haplotype represent base pair substitutions between haplotypes. The hatched line indicates 22 base pair substitutions between *P. gracilis* and G-X. Individuals are colour coded by region (see legend map above) and the proportion of the haplotypes from each region are indicated as pie charts. The major haplotypes of *P. satyrus*, *P. progne* and *P. gracilis* have been named in this study; those within *P. faunus* were named in Kodandaramaiah et al. 2012b.

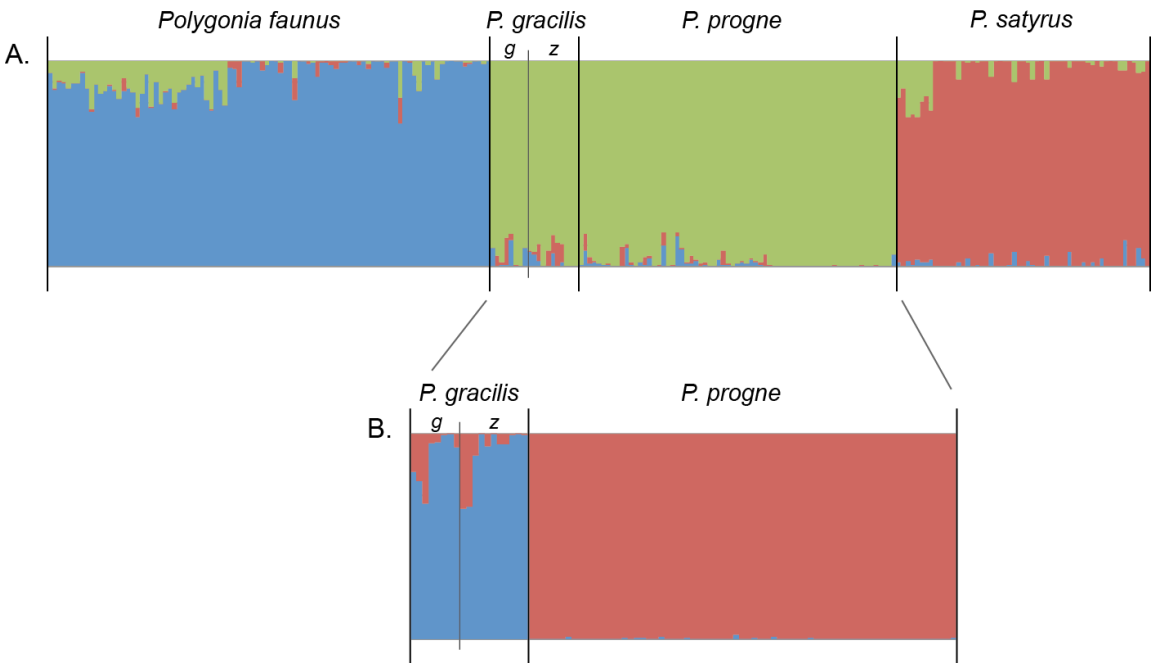


Figure 2.11: Structure analysis (Pritchard et al. 2000) of 239 *Polygonia* specimens in 4 species, *Polygonia faunus* (n=96), *P. gracilis* (n=19; *P. g. gracilis* n=8 (g), *P. g. zephyrus* n=11 (z)), *P. progne* (n=69) and *P. satyrus* (n=55) using 827 SNP loci. The optimal number of populations was k=3 (**A.**) and k=2 (**B.**), as determined by the Evanno method in Structure Harvester (Earl and VonHoldt 2012).

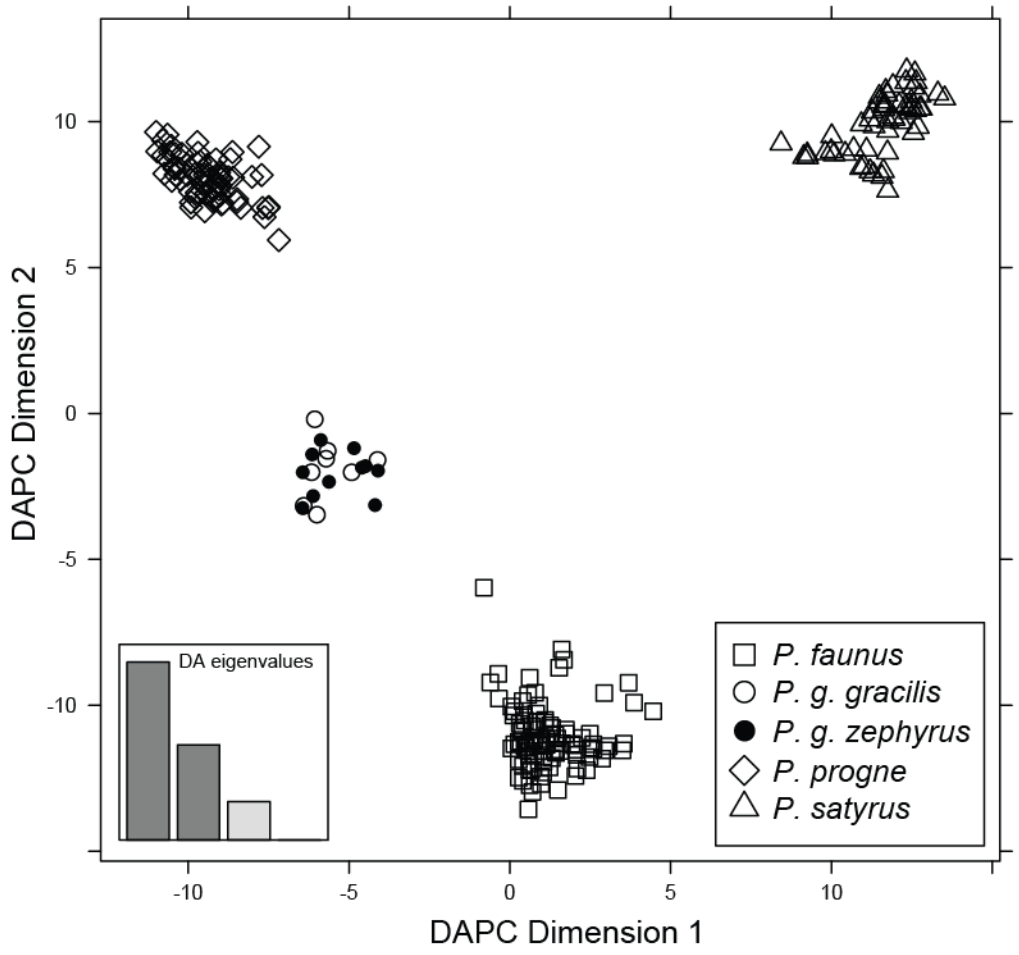


Figure 2.12: Discriminant analysis of principal components of 827 SNP loci completed in R (R cores group) using the statistical package 'adegenet' (Jombart 2008).

Table 2.1: Primers used in this study. Primer positions shown are in reference to the *Drosophila yakuba* mitochondrial genome (Clary and Wolstenholme 1985).

Primer name (direction)	Position	Sequence
LepF1 (forward)	1490	5' - ATT CAA CCA ATC ATA AAG ATA TTG G -3'
HCO (reverse)	2198	5' - TGA TTT TTT GGT CAC CCT GAA GTT TA - 3'
JerryI (forward)	2183	5' - CAA CAT TTA TTT TGA TTT TTT GG - 3'
PatII (reverse)	3013	5' - AGG TAA TGT ATA TTA GAC GGT ATA ACT - 3'
MilaIV (reverse)	2659	5' - AAA TTA GGA CAT TTA TTA CC - 3'

Table 2.2: Mean sequence divergence (base pair difference per site) within (bolded, on the diagonal) and between species, calculated in MEGA 6. Sites with missing data were excluded; 1270 sites remained in the final dataset of the original 1477.

	P. faunus	*P. progne*	*P. satyrus*	*P. gracilis* (all)	*P. gracilis*	G-X
P. faunus (n=96)	**0.000789**					
P. progne (n=69)	0.0435	**0.0011**				
P. satyrus (n=55)	0.0408	0.0169	**0.00085**			
P. gracilis (n=18)	0.0417	0.0162	0.0089	**0.0079**		
P. gracilis (n=15)	0.0417	0.0157	0.0061	--	**0.0018**	
G-X (n=3)	0.0417	0.0191	0.0217	--	0.0196	**0.00067**

CHAPTER 3

3. MORPHOLOGY-BASED DISCRIMINATION OF ALBERTA SPECIES OF *POLYGONIA* HÜBNER (1819) (LEPIDOPTERA: NYMPHALIDAE)

3.1 INTRODUCTION

Delimiting and identifying species is a crucial first step in any biological study, and missteps can have cascading effects in any field (Bortolus 2008). A thorough understanding of species interactions and boundaries is essential to medicine (e.g. Lyme disease: Sperling and Sperling 2009; malaria vectors: Müller et al. 2013) and can have an enormous impact on agriculture and forestry (e.g. cutworms: Sperling, Byers, & Hickey 1996; spruce budworms: Lumley & Sperling 2011; fall armyworm: Dumas et al. 2015). Species and subspecies are also important units of conservation; without the ability to confidently separate one species from another, it is impossible to effectively determine the status of a species and to allocate resources appropriately to those that are threatened (e.g. Proshek et al. 2015; Hedin 2015).

The nature of species has been an ongoing debate among biologists and philosophers for centuries (Mayr 1991; Atran 1993; De Queiroz 2007). Some species concepts use reproduction as the criterion for species delimitation (biological species concept: Dobzhansky 1940; Mayr 1942), others use the niche or adaptive zone (ecological species concept: Andersson 1990; van Valen 1976), monophyly (apomorphy: Rosen 1979; monophyly detected by apomorphy: Mishler and Donoghue 1982), phenetic clustering (phenetic species concept: Michener 1970; Sokal & Crovello 1970) or genetic clusters (genotypic cluster species concept: Mallet 1995; genomic integrity concept: Sperling 2003). This diversity of views shows that the nature of species defies definition as a unitary concept.

Variability between reproductive systems, the complex nature of speciation and the ever-changing face of each population over time also make it difficult to assign necessary and sufficient criteria to species as a group concept (Wheeler 2012). Some of these qualities point to species being ontological individuals, that is, entities with a beginning and an end, as

opposed to classes, which are abstract collections of objects that are constrained by strict criteria for membership in the group and not limited in time and space (Hull 1978). These fundamental philosophical debates in biology continue to contribute to the difficulty of defining species for all situations.

Yet another philosophical view is that there are many species concepts that apply to different situations depending on modes of speciation, reproduction and other life history traits (Wheeler 2012). This perspective, known as pluralism, is opposed by monoism – the view that there is only one species concept and we have yet to discover it (Wheeler 2012). If species are individuals, as Hull (1978) suggests, then it is not possible to ascribe necessary and sufficient criteria to the group; therefore, a single species concept is not attainable, and a suitable concept should be selected for each biological scenario. Many such species concepts have criteria that are incompatible with each other, yet one criterion unites them: populations are recognized as species if they evolve independently of each other (De Queiroz 2007). But since independently evolving lineages can be difficult to detect, it may take effort to find reliable observable characters to delimit species.

This raises a deeper question: are we, as taxonomists, discovering the underlying species when we delimit species boundaries, or are we identifying criteria for arbitrary categories of individuals? This question can be considered in two philosophical frameworks: nominalism and realism (Ereshefsky 2010). From a metaphysical perspective, realists consider the existence of "universals" to be the essence or commonality that joins objects (Ereshefsky 2010). For example, all chairs possess a "chairness" that is key to their identity as chairs. In the context of species, a realist would argue that species are natural categories that share an essence – in other words, taxonomy is the task of discovering natural species boundaries (Bird and Tobin 2015). Nominalists, on the other hand, deny the existence of these universals, and view unifying traits of a group as being human constructs (Rodriguez-Pereyra 2015). An extreme nominalist would consider species identifications and boundaries to be arbitrary labels applied to groups of similar organisms out of convenience (Wheeler 2012). Regardless of whether species are considered to be real categories or human constructs, they can be recognized as objectively delimited clusters by characteristics like their DNA, and it is instructive to consider whether other cognitive categories or labels assigned to them correspond to these units.

3.1.1 COGNITIVE CATEGORIZATION OF SPECIES

Cognitive science is the study of mental processes, including image and word recognition, memory, problem solving, and – most applicable to taxonomy - categorization (Payne and Wenger 1998). There have been many proposed models of categorization in cognitive science, including gestalt, rule-based, prototype, and exemplar models (Goldstein 2014) and biologists appear to use all of these. The term gestalt, defined as the perception of an object as a whole as opposed to breaking it into components, is used by taxonomists who identify specimens without reference to particular diagnostic characters. On the other hand, rule-based categorization is used by those who rely on diagnostic field marks to assign specimens to species. Rule-based approaches, however, often encounter specimens that meet some, but not all, of the necessary criteria for membership in a group. Natural categories may not have fixed boundaries, making rules for categorization problematic (Payne and Wenger 1998). A family resemblance model is more amenable to natural categories, permitting variation between members beyond what strict boundaries would allow, so long as there is overlap of characters (Wilson et al. 2007). Prototype and exemplar models both require biologists to hold a number of images in mind when making identifications. Prototype models involve a "running average" of previous specimens, which changes as more are encountered (Payne and Wenger 1998). New specimens are compared to the evolving mental "prototype." Exemplar models, on the other hand, assume that a mental image of each and every specimen is retained and recalled when a new specimen is encountered, which allows the biologist to conceptualize a greater degree of variation within the group (Galotti 2013).

All of these models seem to correspond to the practice of taxonomy. When learning a new group, most biologists begin with dichotomous keys (Acorn 2013), which may fall under rule-based categorization approaches (Payne and Wenger 1998). Prototype and exemplar models are both used regularly in categorization, including taxonomic descriptions (Payne and Wenger 1998; Goldstein 2014). Exemplars are more informative than prototypes, since actual category members are remembered, as opposed to an average (Goldstein 2014). For this same reason, prototypes may be better suited for large groups such as species, but the variation within the group is better treated by exemplars (Galotti 2013; Goldstein 2014). The comparison of new specimens to exemplars is also reflected in the importance of

holotype specimens in the International Code of Zoological Nomenclature; however, despite common taxonomic practice, it is not necessary for the holotype to act as a phenotypic representative of the species (ICZN 1999).

3.1.2 SPECIES DELIMITATION

Species delimitation, the act of drawing boundaries between species, is a central task of taxonomy (Wiens 2007). This task has been approached in many ways with many tools, including morphological characters (Will and Rubinoff 2004; Schwarzfeld and Sperling 2014), allozymes (Francuski et al. 2009), DNA sequence data (Hebert et al. 2004; Blaxter et al. 2005), molecular markers such as microsatellites and single nucleotide polymorphisms (Lumley and Sperling 2011; Pante et al. 2015), niche association (Rissler and Apodaca 2007), and many others.

Quantitative morphological analysis, also known as morphometrics, is yet another valuable tool for delimiting species (Lumley and Sperling 2010; Schwarzfeld and Sperling 2014). Despite the popularity of molecular analyses for delimitation (Blaxter 2004; Hebert et al. 2004; Ward et al. 2005; Dupuis et al. 2012), morphometrics has often been shown to be equally effective in demarcating even cryptic species complexes (Tofilski 2008; Lumley and Sperling 2010; Buck et al. 2012). By quantifying characters such as colour and shape, a new suite of powerful statistical analyses become available for characters that otherwise might not be used because of perceived subjectivity. In this study, wing colour patterns are quantified by measuring the mean red, green and blue luminance values of a selection of wing pattern features, as per Lumley and Sperling (2010) and compared to an independent dataset of single nucleotide polymorphisms (SNPs) for the same specimens, to determine the utility of these color characters for delimiting species.

3.1.3 IDENTIFICATION KEYS

Most biologists would use a dichotomous key to identify species if they are unable to confidently identify specimens directly through their experience in recognizing these species (Osborne 1963; Walter and Winterton 2007). However, there are several shortcomings to typical dichotomous keys that severely inhibit their ease of use (Acorn 2013). Keys are often written for precision rather than ease of use, so the terminology tends to be exact and may require an anatomical glossary to complete. Additionally, if a couplet is answered incorrectly,

the user will reach the wrong identification (Osborne 1963; Walter and Winterton 2007). This can be particularly challenging if each couplet only contains one character, and it is absent or obscured on the specimen in hand, making a correct identification nearly impossible. Finally, the characters used in diagnostic keys, while they are selected and tested by experienced taxonomists, may not delimit genetic clusters that correspond to species, especially in cases of morphologically similar groups and cryptic species complexes (Jugovic and Koren 2014).

In this chapter I ask, first, whether the characters commonly used in keys and field guides are maximally effective for assigning specimens to species and, second, whether traditional field marks actually circumscribe species boundaries. Finally, I assess whether identification tools can be made more reliable using an evidence-based integration of traditional morphology, digital image analysis and molecular taxonomy.

3.1.4 *POLYGONIA* AS A STUDY GROUP

Polygonia Hübner 1819 is a genus of nymphalid butterflies in the tribe Nymphalini (Wahlberg et al. 2005, 2009). They are commonly known as commas, after the silver comma-shaped marking on their ventral hind wing, or anglewings because of their angular wing margins. Of the 15 species described worldwide (Wahlberg; Pelham 2008) 5 are resident in Alberta, Canada (*P. faunus*, *P. satyrus*, *P. progne*, *P. gracilis*, and *P. oreas*) and many are challenging to identify in the field (Bird et al. 1995; Acorn and Sheldon 2006; Pohl et al. 2010). One taxon, *P. zephyrus*, has been synonymized with *P. gracilis*, due to genetic and morphological continuity (Scott 1984), although this arrangement has not been followed by all authors (*e.g*, Bird 1995, Layberry 2011). Most recent publications recognize *P. g. gracilis* and *P. g. zephyrus* as subspecies (Weingartner et al. 2006; Wahlberg et al. 2009).

Polygonia is an ideal genus in which to study the utility of morphological characters for species recognition because, while the species look very similar to each other, the four species selected for this study form well supported genetic clusters (Chapter 2). The goals of this study are to assess the effectiveness of visually and digitally scored wing characters for delimiting species, and to serve as a study case for the assessment and improvement of field guides for identification. This work has ramifications for professional biology, amateur natural history, and citizen science.

3.2 MATERIALS AND METHODS

3.2.1 SPECIMENS EXAMINED, AND SPECIMEN PREPARATION

A total of 241 *Polygonia* butterflies were collected between May 2012 and September 2013 (96 *Polygonia faunus*, 55 *P. satyrus*, 69 *P. progne*, 8 *P. gracilis gracilis,* 11 *P. g. zephyrus,* 1 *P. comma* and 1 *P. interrogationis*; Table A.1). Sampling was concentrated in Alberta; however, not all resident species were collected. *Polygonia oreas*, which inhabits the southwest corner of Alberta, was not sampled. The majority of specimens were collected by aerial net, although nearly a dozen were collected in traps baited with banana. Some specimens were collected as larvae in the field and reared in the lab. Specimens were kept alive in a fridge or cooler until they could be stored at -60°C, where they remained until DNA extraction, pinning and spreading. These same specimens were characterized by a next-generation sequencing technique, genotyping-by-sequencing, and Sanger sequencing of the mitochondrial gene COI (Chapter 2). A significant portion of the thorax of each specimen was removed for DNA extraction and thus, after an insect pin was placed through the thorax of each butterfly, a drop of Lineco Archival Quality Neutral pH Adhesive was placed in the thoracic void to prevent the body from sliding off the pin after drying. Specimens with various degrees of wear were used in this study, to test the effectiveness of diagnostic characters on individuals typically found in the field. Reared specimens were the least worn, while overwintered individuals were generally the most worn. All specimens, pinned and papered, will be deposited in the E. H. Strickland Entomological Museum, University of Alberta.

3.2.2 SPECIMEN IMAGING

Images of all 241 specimens were captured using a Canon Powershot A650 IS camera in macro setting, using aperture priority mode at f 22 to optimize depth of focus, and an exposure compensation of $+1^{2/3}$ to optimize the brightness of the images. The camera was manually white-balanced between photography sessions, and the same lighting was used throughout the imaging process, although with slight variations that may have been caused by changes in ambient light from a nearby window. The camera was mounted on a copy stand, pointed downward. A consistent distance from specimen to the camera lens was maintained by the use of a foam base on which the specimens were mounted, fitted with a

pinned piece of paper for standardization of wing height. Ventral images were obtained by inverting the specimen and mounting the top of the pin in adhesive putty. Image files were then converted from JPEG to TIFF format in Photoshop to preserve their resolution, and renamed. These image files will be deposited in MorphBank (http://www.morphbank.net).

3.2.3 CHARACTER SELECTION AND VISUAL SCORING

I selected wing characters for this study based on their use in field guides and species descriptions (Table 3.1), in addition to my own observations of the variation between species in *Polygonia*. I recorded diagnostic characters in the literature, especially noting those mentioned in three or more publications. One character was only reported in one publication (ventral forewing pre-submarginal green spots, Character 9, Figure 3.1) (Pyle 2002), and another is new based on my observations (ventral forewing submarginal lunules angled, Character 6, Figure 3.1). The most prevalent diagnostic character in the literature was the row of submarginal green spots on the ventral hindwing of *P. faunus*. This character was not selected because it was present in similar states in other species (*i.e.* similar spots in *P. satyrus* that were yellowish-green, or *P. progne* that were slightly smaller and less obvious). Colour descriptions were less consistent in the literature, although some such as "tawny", and "golden" were common for *P. satyrus*; RGB characters were largely selected based on my observation of specimens and images. After selecting characters, I surveyed my pinned specimens to assess whether these characters could be observed and confidently scored. A visual character needed to be scorable in all taxa, whether present or absent, in some measurable way. Wing colour characters required landmarks or markings that were present across all taxa; *e.g.* the colour of a spot could not be measured if the spot was absent in one species. Ten discrete characters for visual scoring were selected, each with between 2 and 5 character states. A scoring guide was created with images of each character state to ensure consistent scoring between sessions (Figure 3.2). Specimens were visually compared to images in the scoring guide, and all specimens were scored by Christianne McDonald.

Female *P. faunus* and *P. satyrus* exhibit two morphotypes that can be distinguished by the presence or absence of markings on the ventral side of the wings (Figure A.6). Some females, referred to as smeared morphs in this study, generally lack striations and contrast on the ventral side of the wings, while the contrasted females appear banded or striated, much

like their conspecific males. These two morphs were treated separately in this study and identified visually by Christianne McDonald.

Red green and blue (RGB) luminance values were measured in six wing areas to assess the accuracy of colour descriptions reported in field guides. Mean RGB luminance values for each pixel were measured and averaged across the area selected (as per Lumley and Sperling 2010) in ImageJ (Schneider et al. 2012) using the Colour Histogram package (Prodanov 2010). I re-examined all outlying data points to identify possible typographical errors before further analysis of the data matrix.

3.2.4 MULTIVARIATE DATA ANALYSIS

Multivariate analyses are used to consider more than one statistical variable at a time, creating ordination plots in multidimensional space. In this study, four multivariate analyses were used to describe the data – multiple correspondence analysis (MCA), discriminant correspondence analysis (DCA), principal component analysis (PCA), and linear discriminant analysis (LDA). Multiple correspondence analysis maximizes the variance among all individuals scored for discrete characters (in this study, visually scored characters), while PCA performs the same function on continuous characters (RGB data). The first dimension of each analysis corresponds to the axis with the greatest variance. Successive dimensions that are at right angles to each other maximize the remaining variance at each cycle in the analysis. Scatterplots of data points on each axis allow distances and clusters to be easily visualized. Discriminant correspondence analysis (DCA) maximizes separation previously defined classes of data points (here corresponding to species) scored for discrete character data, and LDA performs a similar analysis using continuous characters.

Species identifications for each specimen were based on clusters derived from genetic data, specifically 827 single nucleotide polymorphisms (SNPs) from genotyping-by-sequencing analyses (Chapter 2), and delimited by discriminant analysis on principal components (DAPC) (Figure 2.12). Discrete clusters of SNP data were then used as the *a priori* categories for subsequent DCA and LDA analyses. *Polygonia faunus* and *P. satyrus* were further divided into morphological types (smeared versus contrasted morphs). *P. gracilis* was further divided into the putative subspecies *P. g. gracilis* and *P. g. zephyrus,* based on diagnostic descriptions in field guides and collection location (Bird et al. 1995; Pyle 2002; Shepard and Guppy 2011) and labeled as such in the multivariate analyses.

Multiple correspondence analysis was run on the ten discrete, visually scored character states of *P. faunus* (smeared n=23, and contrasted n=67), *P. gracilis gracilis* (n=6), *P. g. zephyrus* (n=11), *P. progne* (n=62) and *P. satyrus* (smeared n=11, and contrasted n=37) in R (R Core Team 2013), using the 'ade4' statistical package (Thioulouse and Dray 2007). Rows containing missing data were removed from the analyses (20 individuals). Discriminant correspondence analyses was then performed on the same data to separate the *a priori* species groups determined by genetic clusters in Chapter 2 (Thioulouse and Dray 2007; R Core Team 2013).

Principal component analysis was run on 18 continuous RGB characters for 6 areas on the wing of *P. faunus* (smeared n=28, and contrasted n=67), *P. gracilis gracilis* (n=8), *P. g. zephyrus* (n=11), *P. progne* (n=68), and *P. satyrus* (smeared n=17, and contrasted n=38), and the same *a priori* identifications above were used to separate species using LDA. Both of these analyses were executed in the ade4 package in R (Thioulouse and Dray 2007).

3.2.5 KEY CONSTRUCTION

After performing the DCA, I ranked the 10 visually scored characters by their canonical weights, or loadings, which are quantitative evaluations of a character's power to separate *a priori* clusters in the dimensions examined. This was done by summing the absolute value of each character's loadings across all dimensions. I inspected the character with the highest loading magnitude and compared it to the proportion of character states in each species (Table 3.5, Figures 3.1, 3.2) to assess whether this character would distinguish one or more species from the other taxa. If the dataset reflected a substantial overlap of character states between species, then this character was not used for the couplet, and the next highest loading was examined. A second key character was selected for use in the same couplet, using the same procedure as the first. This second character needed to separate the same taxon as the first, and as such was usually not the second highest loading value. I removed the species that was identified by the first two characters (*P. faunus*: Characters 9 and 3, Figure 3.3) from the data file and then performed another DCA on the remaining dataset. Changes in the taxon pool affected the diagnostic ability of each character; some characters that were not effective for the whole dataset were effective once *P. faunus* was removed (*e.g.* Character 10, Table 3.3). In the second DCA, character with the highest loading, character 9, was uninformative as all specimens shared the same character state

(Figure 3.3), so the two characters with the next highest loadings were used (Characters 6 and 10). Character 9 had the highest loading in the third DCA as well, but could not be used in this couplet. The second highest loading, character 2, was selected for use in the key in addition to character 5, despite it being only the fifth highest loading.

3.3 RESULTS

3.3.1 VISUALLY SCORED CHARACTERS

Multiple correspondence analysis of visually scored characters produced several clusters with relatively little overlap, and these corresponded to putative species, as determined using genetic data (Figure 3.3A). Dimensions 1 and 2 accounted for 47.4% and 29.4% of the total variation in the data set, respectively. The third dimension accounted for only 8.42%, so only 2 dimensions are shown in Figure 3.3A. Three *P. faunus* specimens clustered more closely with *P. gracilis*, and although some of these were worn individuals, others showed obvious morphological similarity to *P. gracilis*. Morphological type (*i.e.* smeared or contrasted based on the wing patterning on the ventral side) had a small effect on the character scoring of *P. satyrus*, as there was some weak separation of the two morphs in the cluster for that species. *P. faunus* clustered less by morphological type.

Discriminant correspondence analysis indicated that the visually scored characters were effective for separating *a priori* categories (Figure 3.3B), although the clusters were not as tight as in MCA (Figure 3.3A). The shape of the submarginal chevrons on the ventral forewing and the size and colour of the row of spots basal to them (Characters 6 and 9) (Figure 3.1) were most effective for separating groups on Dimension 1, as indicated by the magnitude of their canonical weights, or loadings (Table 3.2). Diffuseness of submarginal yellow spots on the dorsal hindwing (Characters 2 and 3) and the presence of submarginal chevrons on the ventral hindwing (Character 7) were also very effective for separating groups in Dimension 2.

3.3.2 RGB CHARACTERS

Linear discriminant analysis showed moderate separation of *a priori* identifications using RGB data (Figure 3.3D). Character 5 green luminescence values contributed the most to the variance in Dimension 1 (Table 3.3), indicating that it was the most effective for

separating groups along that axis. Principle component analysis did not cluster individuals by species using the RGB data (Figure 3.3C). Despite 54.8% of the variation in the data being captured by Dimension 1 and 12.4% by Dimension 2, there was still substantial overlap among all species.

The red, green and blue luminance values of wing areas distinguished some of the species in this study; however, there was a high degree of overlap in values for many species (Figure 3.4). Of the seven groupings, *P. progne* and *P. g. zephyrus* were the most differentiated, based on pairwise comparisons of their central tendencies (Figure 3.4, Table 3.4). The luminance means of *P. progne* and *P. g. zephyrus* were significantly different from each other in 11 out of 18 possible pairwise combinations (Table 3.4), as determined by non-overlapping 95% confidence intervals (Figure 3.4). The area basal to the submarginal yellow spots on the dorsal hindwing, and the area on the marginal half of the silver comma on the ventral side (Characters 4 and 5) showed the most significant difference between the four species, followed by the overall colour of the entire ventral hindwing (Character 6) (Figure 3.4, Figure 3.5). The R, G and B values of the area near the base of the dorsal forewing (Character 1) of *P. progne* were significantly different from that of the other species (Figure 3.4); however, none of the other species means were significant different from each other.

3.3.3 KEY TO THE *POLYGONIA* OF ALBERTA

Characters and character states

Character 1: Dorsal forewing discal spot across vein CuA2

0 – Two subequal spots, on either side of vein

1 – Two spots, posterior spot at least twice as large as anterior spot.

2 – One large spot, spanning >70% of distance between CuA2 and adjacent vein

3 – One small spot spanning <70% of distance between CuA2 and adjacent vein

Character 2: Dorsal hindwing submarginal yellow spots

0 – Spots spread across veins, or blending together to form a contiguous band

1 – Spots restricted to cells, not contiguous

Character 3: Dorsal hindwing submarginal yellow spots

0 – Spot boundaries vague or diffuse

1 – Spot boundaries well defined and clear

Character 4: Dorsal forewing, basal spots below costal vein

0 – Two small spots visible and separated by nearly the width of one spot

1 – Two large spots, clearly separated

2 – One spot, bilobed, contiguous

Character 5: Dorsal hindwing, spot in basal region of cell M3

0 – Spot is large, broad and triangular; may nearly or completely fill corner of cell M3

1 – Spot is small, may have indistinct borders, may or may not touch veins of cell M3

2 – Spot is absent, or diffuse dark patch may be present

Character 6: Ventral forewing submarginal dark markings

0 –crescent shaped

1 –sharply angled or triangular

Character 7: Ventral forewing submarginal pale chevrons

0 – absent

1 – present

Character 8: Ventral hindwing discal silver "comma"

0 – "L" shaped, both arms equal in length, 90^{o} angle

1 – "l" shaped, horizontal arm clearly shorter, angle $>90^{o}$

2 – "C" shaped with one or both ends barbed, angle $>90^{o}$

3 – "L" shaped with barb on apex of vertical arm, angle $>90^{o}$

Character 9: Ventral forewing pre-submarginal spots

0 – Small and black, if visible

1 – Large, with dark outline surrounding coloured spot

Character 10: Ventral hindwing contrast between basal and distal halves

0 – Strong contrast is readily observed

1 – Regions weakly contrasting or uniform in colour

Diagnostic key

1a.	Ventral forewing pre-submarginal spots green (Figure 3.2: 9, 1); dorsal forewing submarginal spot boundaries well defined (Figure 3.2: 3, 1).	*Polygonia faunus*
1b.	Ventral forewing pre-submarginal spots black (Figure 3.2: 9, 0); dorsal hindwing submarginal spot boundaries diffuse (Figure 3.2: 3, 0).	 2
2a.	Ventral forewing submarginal dark markings angular (Figure 3.2: 6, 1); ventral hindwing distal and basal halves only slightly contrasting (Figure 3.2: 10, 1).	*P. progne*
2b.	Ventral forewing submarginal dark markings crescent-shaped (Figure 3.2: 6, 0); ventral hindwing distal and basal halves strongly contrasting (Figure 3.2: 10, 0)	 3
3a.	Dorsal hindwing submarginal spots contiguous (Figure 3.2: 2, 0); dorsal hindwing discal spot triangular, black (Figure 3.2: 5, 0).	*P. satyrus*
3b.	Dorsal hindwing submarginal spots discrete (Figure 3.2: 2, 1); dorsal hindwing discal spot absent (Figure 3.2: 5, 2).	*P. gracilis*

3.4 DISCUSSION

3.4.1 MORPHOLOGY

Overall, the visually scored morphological characters examined in this study grouped *Polygonia* individuals into the same clusters that were revealed by their genetic clusters, despite significant overlap in character states and a large amount of variation within species. Some characters were not as diagnostic as expected from their prevalence in the literature, such as the division of the dorsal forewing spot across vein CuA2 (Character 1, Figure 3.1) and the shape of the silver comma (Character 8) based on their loadings from the DCA (Table 3.2). One explanation for this result is that diagnostic characters in field guides are written to allow the reader to distinguish between similar species, not all species in the group. An overlap in character states between *P. satyrus* and *P. progne* would not affect the species diagnosis, since these species are otherwise very different from each other. In fact,

many of these characters would be diagnostic if *P. faunus* were removed from the taxon pool, as the variation in character states in this species was broad (Figure 3.1). As an example, *P. faunus* in this study exhibited every character state in character 8, the shape of the ventral hindwing comma, several did not fit neatly into any of the character states, and one specimen even exhibited a marking that was upside-down.

Two novel characters, 9 and 6, were found to be highly diagnostic for *P. faunus* and *P. progne*, respectively, despite their absence or scarcity in the literature (Table 3.1, 3.3). It remains to be seen whether they are as diagnostic with larger sample sizes. Character 9, the presence of small green spots basal to the larger green lunules, was highly effective for distinguishing *P. faunus* from all other species (Figure 3.1). Similarly, the submarginal chevrons on the ventral forewing were effective for diagnosing *P. progne* from most other species.

3.4.2 THE IMPORTANCE OF COLOUR

Colour can be a valuable character in assessing diversity within and between species, but is under-utilized in science due to its subjectivity to personal perception (Grose 2012). The use of colour names in science is standardized, as described in catalogues that graphically depict common colour descriptions such as "russet", "tawny", and "ferruginous" in swatches to ensure consistent application of these terms (Smithe 1975; Köhler 2012). However, even when using a colour catalogue, the interpretation of "russet" as opposed to "cinnamon brown" is still in the eye of the beholder. It is due to this difference in perception that the colour characters in this study could not be used for diagnoses in the construction of the dichotomous key; however, due to their correspondence with genetic clusters, they should continue to be used as descriptive characters in field guides. As a taxonomic tool, the digital measurement of colour characters can allow the assessment of colour in an empirical, unbiased way, and also makes it possible for those who are colourblind to compare colour accurately.

Digital measurement of wing-pattern colour was ineffective at clustering species in this study (Figure 3.3C), but was marginally successful at separating species by *a priori* clusters (Figure 3.3D). Many wing areas appear very similar in colour across the species (Figure 3.5), which could be due to similar ratios of R G and B values, producing the same hue at a slightly different brightness. Biologically, this could be explained by the presence of

the same pigments in the genus (Ford 1941). Some characters that are frequently described in species diagnoses, such as warm brown ventrally for *P. satyrus*, were highly informative in this analysis, including the submarginal region of the ventral hindwing near the comma, and the overall colour of the ventral hind wing. Other characters were not as useful, such as the shade of orange on the dorsal forewing, despite being listed in many character descriptions in the literature (Bird et al. 1995; Pyle 2002). Since these characters are often described in terms of colour relative to other species, they are rarely included in empirical studies, and are never tested except against the memory of experienced taxonomists.

Additional wing characters could be added to test other common diagnostic characters in *Polygonia*, such as the contrast between the basal and distal halves of the ventral hindwing, or the colour of the marginal border on the dorsal forewings (Bird et al. 1995; Shepard and Guppy 2011).

One application of the use of RGB characters is the detection of subspecies, due to the abundance of data, the versatility of analyses and reliance on morphology instead of molecular methods to cluster groups. Since subspecies must be morphologically differentiable in addition to being geographically associated (Amadon 1949; Wilson and Brown 1953), an objective system to compare colour would be a valuable tool for subspecies delimitation. By measuring the colour of homologous characters across specimens, subtle differences in colour values could be detected between taxonomic units (Lumley and Sperling 2010).

The definition of a subspecies, unlike that of a species, is not under significant debate, but the utility of this taxonomic unit is under scrutiny (Zink 2004; Acorn 2007; Kodandaramaiah et al. 2012b). Some authors believe that subspecies as a unit are useful for conservation, although they may be over-designated and under-regulated (Patten and Unitt 2002; Phillimore and Owens 2006). Others think that since there are no barriers to reproduction, these taxonomic units are not biologically informative (Zink 2004). While the subspecies *P. g. gracilis* and *P. g. zephyrus* were treated as independent units in this study, there is little agreement in the literature regarding their morphological distinction. To illustrate this disagreement, some examples of diagnostic characters for *P. g. zephyrus* include soft, subdued dorsal features such as diffuse submarginal yellow spots on the hindwing, a clearly contrasted – but not hoary – basal and marginal regions of the ventral fore- and hindwings, and the

presence of two morphs: a grey and a brown morph on the ventral side (Bird et al. 1995; Pyle 2002). In contrast, some examples of diagnoses for *P. g. gracilis* from *P. g. zephyrus* include the presence of diffuse submarginal hindwing spots that may merge into a band (Hooper and Long 1973; Dornfeld 1980; Opler 1999; Fisher 2006), (or small submarginal spots, depending on the source, Klassen et al., 1989) starkly contrasting basal and marginal regions of the ventral fore- and hindwing, sometimes hoary, other times brown (Bird et al. 1995). Interestingly, many field guides list *P. g. gracilis* as being easily confused with *P. progne* as opposed to *P. g. zephyrus*, some stating that the hoary colour of the marginal areas of the ventral wings is unique to *P. g. gracilis* in the genus (e.g. Pyle 2002). One possible cause for the discrepancies among these descriptions is the geographic region for which each book was written – *P. g. gracilis* has a broad range across North America, and varies greatly across it (Bird et al. 1995; Layberry et al. 1998). The geographic range and sampling density of these subspecies were limited to Alberta in this study; broadening the sampling should provide more resolution of the appropriate taxonomic designations for these units and diagnostic characters for their distinction.

Subspecies of *P. faunus* were not addressed in this study due to the lack of empirical evidence suggesting that they should be treated otherwise (Kodandaramaiah et al. 2012a). However, two subspecies are currently recognized in Alberta: *P. f. arcticus* and *P. f. hylas* (Scott 1992; Pohl et al. 2010; Kodandaramaiah et al. 2012a).

3.4.3 INTEGRATIVE DIAGNOSTIC KEYS

Diagnostic keys are an important part of a taxonomist's work. Whether they are using a key or constructing one, taxonomists are required to be familiar with them. Keys can come in many formats – matrix, illustrated, even PowerPoint-based – but the standard key is constructed in a dichotomous branching format, with each couplet dividing the taxon pool (Voss 1952; Pankhurst 1970; Walter and Winterton 2007). To construct a dichotomous key, an experienced taxonomist studies specimens carefully and selects characters that he or she feels are appropriate for splitting and diagnosing taxa (Acorn 2013). This character selection is generally based on previous literature and the expert's judgment, with the goal of making the average path shorter. This can be approached using many strategies, such as dividing the taxon pool in half with each couplet, separation of distinct species first, or organizing by probability of encounter. While this method may provide an accurate key that corresponds

to groupings seen in nature, it is important to determine whether these groupings correspond to real biological groups or to phenotypic clusters. I examined the utility of the integration of discriminant correspondence analysis into traditional diagnostic key building methods as a way to ensure that the identified taxa correspond to biological species determined by genetic analysis.

In this study, the dichotomous key could also have been constructed using only the bar graph showing the proportion of character states in each taxon (Figure 3.1), but the application of this approach may be limited to small taxon pools. Larger datasets would make it far more challenging to select diagnostic characters by eye. The most efficient diagnostic keys separate more than one species at each couplet – dividing the taxon pool in half when possible – shortening the average path to identification (Walter and Winterton 2007), but further complicating the selection of characters. Using DCA loadings in addition to bar graphs of character proportions could give the taxonomist a place to start if the raw dataset is too large to parse mentally.

The integration of DCA loading data into key construction in this study produced unintuitive results when compared to previously published keys and diagnoses in field guides. For example, *P. satyrus* is the most readily distinguishable of all the Alberta *Polygonia* species (Acorn and Sheldon 2006), indicated by the species' clear separation in the second axis from the others using RGB values in the LDA and in the first axis in the DCA (Figure 3.1D), and yet the first taxon separated in this key is *P. faunus*. Despite the utility of visually scored characters for separating off *P. satyrus* (Figure 3.1B), no single character is sufficient for diagnosing the species (Figure 3.3). Similarly, the second DCA performed on the data set after removal of *P. faunus* did not separate off *P. satyrus* next. The character with the highest loading (character 10) separated *P. progne*, although *P. gracilis* was more readily distinguished based on the separation of groups in the first two dimensions of the DCA, and *P. progne* was only distinguishable in the second dimension (Figure 3.6B). Since the ordinance plots describe the distinctness of species using all the characters, and the use of loadings identifies the strength of each character to separate the groups, I relied more on loadings in the construction of this key than the plots. As a result, I chose to separate off *P. progne* next in the key using characters 6 and 10.

Depending on the type of continuous data, LDA loadings could be used in constructing diagnostic keys; however, the use of RGB luminance values would not be useful as key characters. Red, green and blue luminance values are difficult to quantify quickly to run a specimen through a key, and qualitative colour characters would be inconsistently scored due to individual perception of colour, colourblindness in some users, and the effects of ambient light and computer monitors on the appearance of colour. If colour variation is distinct enough for its use in a key, quantitative measurement is not likely necessary. Even if a continuous character were to be a different kind of morphological character, for example wing aspect ratio, the central tendencies of the data would need to have very little overlap in order for a single character to be useful in a diagnostic key.

The results from the MCA suggest that *P. progne* is the most distinct, based on the visually scored characters, as the plot of the combination of the first and dimensions separated this species from the other by a decisive gap (Figure 3.1A). The loadings from this analysis may be informative in determining which characters are responsible for the most variance between individuals in the data set, however these values do not necessarily reflect a character's ability to create spread between species clusters, and as such this would not be an appropriate interpretation of the data. This same logic can be applied to the use of PCA loadings for integration into diagnostic keys.

Selection and scoring of characters for morphological analysis is subject to the judgment of taxonomists, which has the potential to introduce bias if the selection process is not performed carefully. Although the use of expert judgment alone in constructing dichotomous keys may be problematic, the selection of key characters based solely on DCA loadings is even more so. Integration of many lines of evidence to construct diagnostic keys is appropriate for the same reason that many genes should be used to reconstruct phylogenetic histories: a single line of evidence may be misleading, and the use of multiple independent data sources highlights incongruences between datasets or characters and mitigates the adoption of faulty results. Integration of morphological and molecular characters through multivariate analysis provides one approach to ensure that phenotypic clusters identified by a key also represent biological species.

Table 3.1: (On following three pages) Diagnostic characters of *Polygonia* butterflies published in North American field guides, with an emphasis on Western North America. If an author designated a character as being primary, or highly diagnostic, it is indicated by a grey background. VHW – ventral hindwing; VFW – ventral forewing; DHW – dorsal hindwing; DFW – dorsal forewing.

	Polygonia faunus (W.H. Edwards, 1862)								*P. satyrus* (W.H. Edwards, 1869)						*P. progne* (Cramer, 1776)							
Publication	Row of greenish spots on VFW	**Two rows green spots**	**DHW row of subM spots, separated**	Very irregular margins	**2 blk spots, inner DFW margin**	**dark spot on DHW**	**comma reduced, barbed or not**	**DFW inner spots fused**	**DHW enlarged subM spots; merge**	**Comma curved, barbed**	**Triangular spot on DHW**	**2 blk spots, inner DFW margin**	Golden orange dorsally	Fine, dark striations VWs	DHW black margin expanded	**DHW small subM spots**	VHW one tone, VFW two-toned	VWs striated dark grey	**Comma is wide angle "L", thin**	1 blk spot on DFW margin	**dark spot on DHW**	**V subMar light grey "v"s**
Alberta Butterflies (Bird et al. 1995)	+		+						+				+		+	+	+					
Butterflies of Manitoba (Klassen et al. 1989)	+		+	+						+			+	+		+		+	+			
Butterflies through Binoculars - The West (Glassberg 2001)	+		+	+	+	+			+		+	+				+	+	+	+	+	-	
Butterflies of Cascadia (Pyle 2002)		+		+			+		+	+	+		+	+								
Butterflies of Saskatchewan (Hooper 1973)	+				+	+	+		+	+		+	+				+	+	+			
Butterflies of North America (Brock and Kaufman 2003)	+		+	+		+			+		+	+				+	+	+	+	+	-	
Butterflies of British Columbia (Acorn and Sheldon 2006)	+		+			+				+		+	+									+
Western Butterflies (Opler 1999)	+			+								+		+			+					
The Butterflies of Oregon (Dornfeld 1980)		+	+				+			+				+								
Butterflies of British Columbia (Guppy and Shepard 2001)	+												+				+		+			
The butterflies of North America (Scott 1992)	+		+	+					+		+	+	+			+				+		
The Butterflies of Colorado: Nymphalidae (Fisher 2006)	+			+					+				+									
The Guide to Butterflies of Oregon and Washington (Neill 2001)	+													+			+					
Butterflies of Rocky Mountain National Park (Angel 2005)	+					+																
Butterflies of North Dakota (Royer 1988)	+		+	+					+			+				+						
Butterflies of Grand Teton and Yellowstone National Parks (Poole 2009)	+		+	+			+	+	+	+	+	+	+			+	+		+			

	P. gracilis gracilis (Grote and Robinson, 1867)								*P. g. zephyrus* (W.H. Edwards, 1870)											
Publication	VHW high contrast b/w Bas/Mar	Two morphs, grey and brown	**Comma tapered at both ends**	**DHW enclosed subM spots**	**DHW subM spots run together**	**Triangular DHW spot**	**DHW two black spots, faint third**		VHW contrast b/w Bas/Mar	Two morphs, grey and brown	**DHW subM spots merge**	**Triangular DHW spot**	Broken row of green spots	**Comma unbarbed curve, "L"**	**2 blk spots basal front edge DFW**	**1 blk patch DHW edge spot below**	**3 spots inward from DFW mar**	**Comma abruptly curved**	**V subMar spots sml, blk, ylw ring**	
Alberta Butterflies (Bird et al. 1995)	+	+							+	+	+									
Butterflies of Manitoba (Klassen et al. 1989)	+		+	+																
Butterflies through Binoculars - The West (Glassberg 2001)	+				+	-														
Butterflies of Cascadia (Pyle 2002)									+		+	-	+	+						
Butterflies of Saskatchewan (Hooper 1973)	+		+				+		+		+	-		+	+	+	+			
Butterflies of North America (Brock and Kaufman 2003)	+				+	-														
Butterflies of British Columbia (Acorn and Sheldon 2006)	+		+																	
Western Butterflies (Opler 1999)									+									+		
The Butterflies of Oregon (Dornfeld 1980)											+						+		+	
Butterflies of British Columbia (Guppy and Shepard 2001)	+								+				+							
The butterflies of North America (Scott 1992)	+								+		+									
The Butterflies of Colorado: Nymphalidae (Fisher 2006)									+		+									
The Guide to Butterflies of Oregon and Washington (Neill 2001)									+		+								+	
Butterflies of Rocky Mountain National Park (Angel 2005)									+			-		+						
Butterflies of North Dakota (Royer 1988)	+	+			+															
Butterflies of Grand Teton and Yellowstone National Parks (Poole 2009)	+		+																	

+/- denotes presence/absence; **Bold** = Visually scored character selected for analysis

	P. oreas (W.H. Edwards, 1869)										*P. interrogationis* (Fabricius, 1798)							*P. comma* (Harris, 1842)							
Publication	DHW with white marginal hair	VHW slight separation b/w basal/mar	Heavily striated VW	Mar band not darker near Mar	Triangular spot on DHW	subMar spots form band	V subMar light grey "v"s	VHW very dark	comma sharp 90°, tapered		Longer tails	Comma is "?"	Small horizontal bar VFW	FW hooked	DFW 2 inner spot	DFW trailing spots divided		VHW postmedian line very jagged	VFW apex two-toned	Small horizontal bar VFW	Comma hooked at both ends	DFW top spot reduced	DFW spots enclosed	DHW spot	
Alberta Butterflies (Bird et al. 1995)	+	+									+														
Butterflies of Manitoba (Klassen et al. 1989)																					+	+			
Butterflies through Binoculars - The West (Glassberg 2001)		+		+		+						+	+					+	+	-					
Butterflies of Cascadia (Pyle 2002)																									
Butterflies of Saskatchewan (Hooper 1973)											+	+			+	+					+	+	+	+	
Butterflies of North America (Brock and Kaufman 2003)		+	+	+	<						+	+	+					+		-					
Butterflies of British Columbia (Acorn and Sheldon 2006)							+																		
Western Butterflies (Opler 1999)								+																	
The Butterflies of Oregon (Dornfeld 1980)			+			+		+	+																
Butterflies of British Columbia (Guppy and Shepard 2001)	+	+						+																	
The butterflies of North America (Scott 1992)						+		+				+		+				+			+		+		
The Butterflies of Colorado: Nymphalidae (Fisher 2006)						+		+				+													
The Guide to Butterflies of Oregon and Washington (Neill 2001)																									
Butterflies of Rocky Mountain National Park (Angel 2005)																									
Butterflies of North Dakota (Royer 1988)												+		+							+	+	+		
Butterflies of Grand Teton and Yellowstone National Parks (Poole 2009)						+		+																	

+/- = presence/absence; < = faint character

Table 3.2: Canonical weights, or loadings, of ten visually scored characters in three successive discriminant correspondence analyses (DCA) as calculated by discriminant correspondence analysis (DCA) in R (R core group) using the package 'ade4' (Thioulouse and Dray 2007). The first DCA included all four species in this study, *Polygonia faunus*, *P. gracilis*, *P. progne*, *P. satyrus*. *Polygonia gracilis gracilis* and *P. g. zephyrus* were treated as separate *a priori* groups.

	DCA 1			**DCA 2**		**DCA 3**	
	Dim. 1	Dim. 2	Dim. 3	Dim. 1	Dim. 2	Dim. 1	Dim. 2
Character	*91.5%*	*78.2%*	*49.3%*	*96.7%*	*77.6%*	*91.5%*	*36.4%*
1	-3.228	-3.170	9.072	2.612	4.101	6.249	5.256
2	-0.766	-11.636	-4.622	6.880	-6.989	7.485	-66.580
3	1.991	-11.333	-19.196	3.212	-10.087	0.049	-15.682
4	1.594	7.999	9.673	-0.635	-1.636	-1.316	-9.350
5	-2.563	-3.115	17.388	5.038	15.333	7.373	6.469
6	-12.296	8.944	-6.887	4.577	-11.061	1.350	-9.421
7	-2.720	12.248	-1.228	-0.028	-9.711	-2.931	12.596
8	2.158	5.940	0.352	-2.394	-5.347	-3.947	0.118
9	16.734	-19.437	9.238	-38.502	70.611	-5.673	75.990
10	-0.202	8.769	-27.409	2.663	-27.455	-6.477	5.397

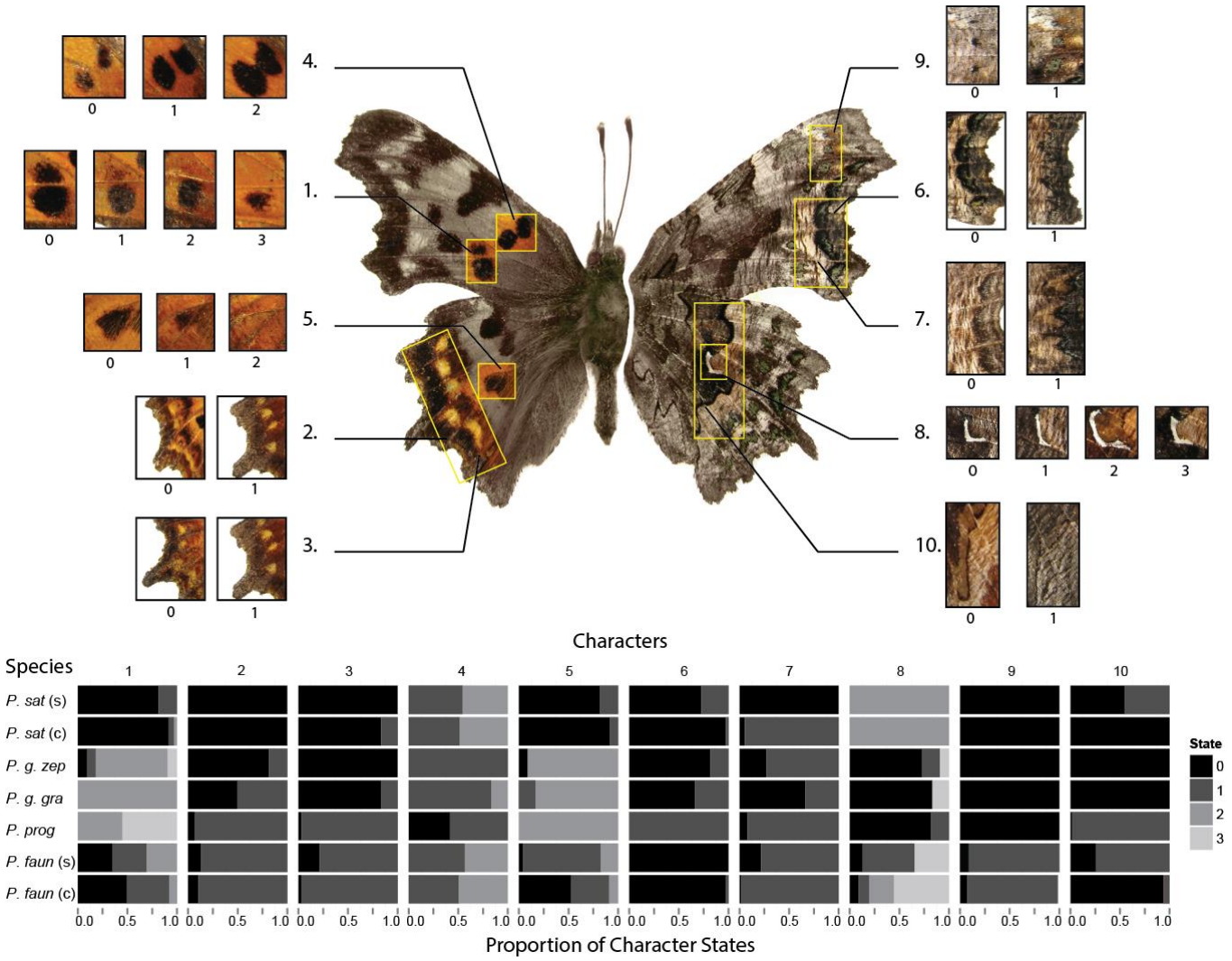
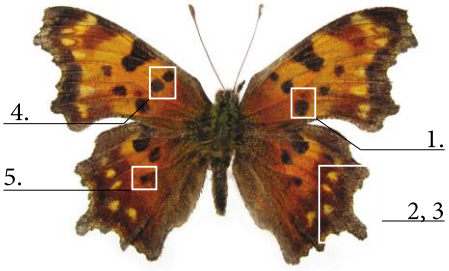


Figure 3.1: The proportion of visually scored character states across seven groups of *Polygonia* butterflies (*P. faunus* (smeared and contrasted morphs; n=23 and 67 respectively), *P. progne* (n=62), *P. gracilis gracilis* (n=6), *P. g. zephyrus* (n=11) and *P. satyrus* (smeared and contrasted morphs; n=11 and 37, respectively).

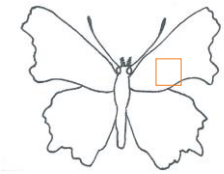
Figure 3.2: (Following four pages) Scoring guide used during the visual morphological character scoring process, with images taken from specimens in this study.

Scoring guide for *Polygonia* morphological characters

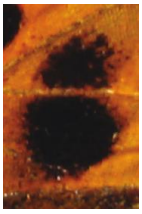
Dorsal characters



1. Discal forewing spots across vein CuA2; Ordered



0



Two subequal spots across vein CuA2

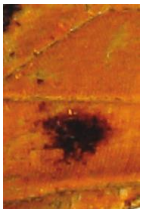
1

Two spots, one at least twice as large as the other

2

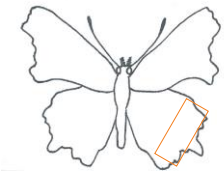
One large spot spanning >70% of the distance between the veins

3

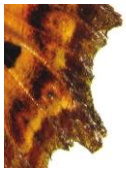


One small spot spanning <70% of the distance beteen the veins

2. Submarginal yellow spots on hindwing, Unordered



0

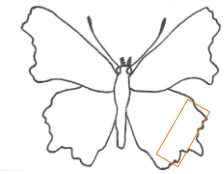


Spots spread beyond veins, at times blending together

1

Spots are distinct from each other and remain in their cell

3. Submarginal yellow spots on hindwing; Unordered



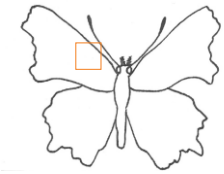
0

Spot boundaries are vague or dif use.

1

Spot boundaries are well-def ned and clear

4. Basal spots below costal vein on the forewing; Ordered



0

Two small spots are visible and separated by nearly the width of one spot.

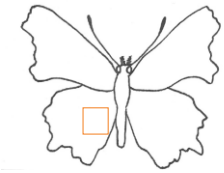
1

Two large spots are clearly separated

2

One bilobed spot, no clear division

5. Triangular spot in basal region of cell M3; Ordered



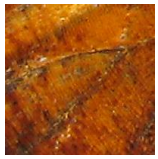
0

Spot is large and broad, may nearly or completely f ll the corner of cell M3.

1

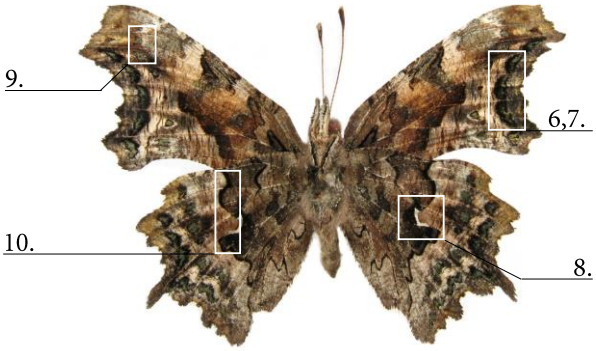
Spotis small and may have indistinct borders, may or may not touch the veins bordering cell M3

2

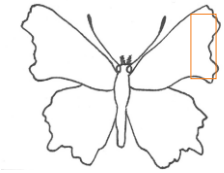


Spot is absent, or small dark patch may be present

Ventral Characters



6. Submarginal chevrons near the outer margin of the ventral forewing; Unordered



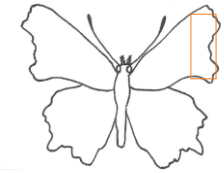
0

Crescent shaped

1

Angular

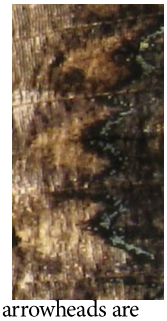
7. Pale submarginal arrowheads basal to chevrons on ventral forewing; Unordered



0

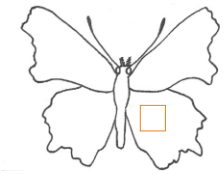
Area basal to chevrons is largely uniform in colour.

1



Pale arrowheads are visible basal to the sub-marginal chevrons

8. Silver “comma” in the discal region of the ventral hindwing; Unordered



0

“L” shaped, both arms are equal in length, 90 degrees

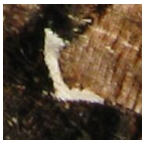
1

“I” shaped, the horizontal arm clearly shorter, >90 degrees

2

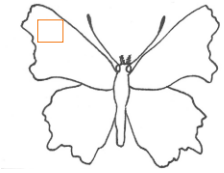
“C” shaped with one or two barbs, >90 degrees

3



“L” shaped with a barb on the vertical arm, >90 degrees.

9. Spots basal to the submarginal chevrons on the ventral forewing; Unordered



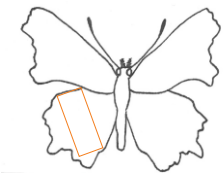
0

Small, if visible. No outline present.

1

Large with a dark outline surrounding a lighter spot

10. Contrast between basal and distal halves of the ventral hindwing; Unordered



0

Strong contrast is readily observed

1

Regions are weakly contrasting or similar in color

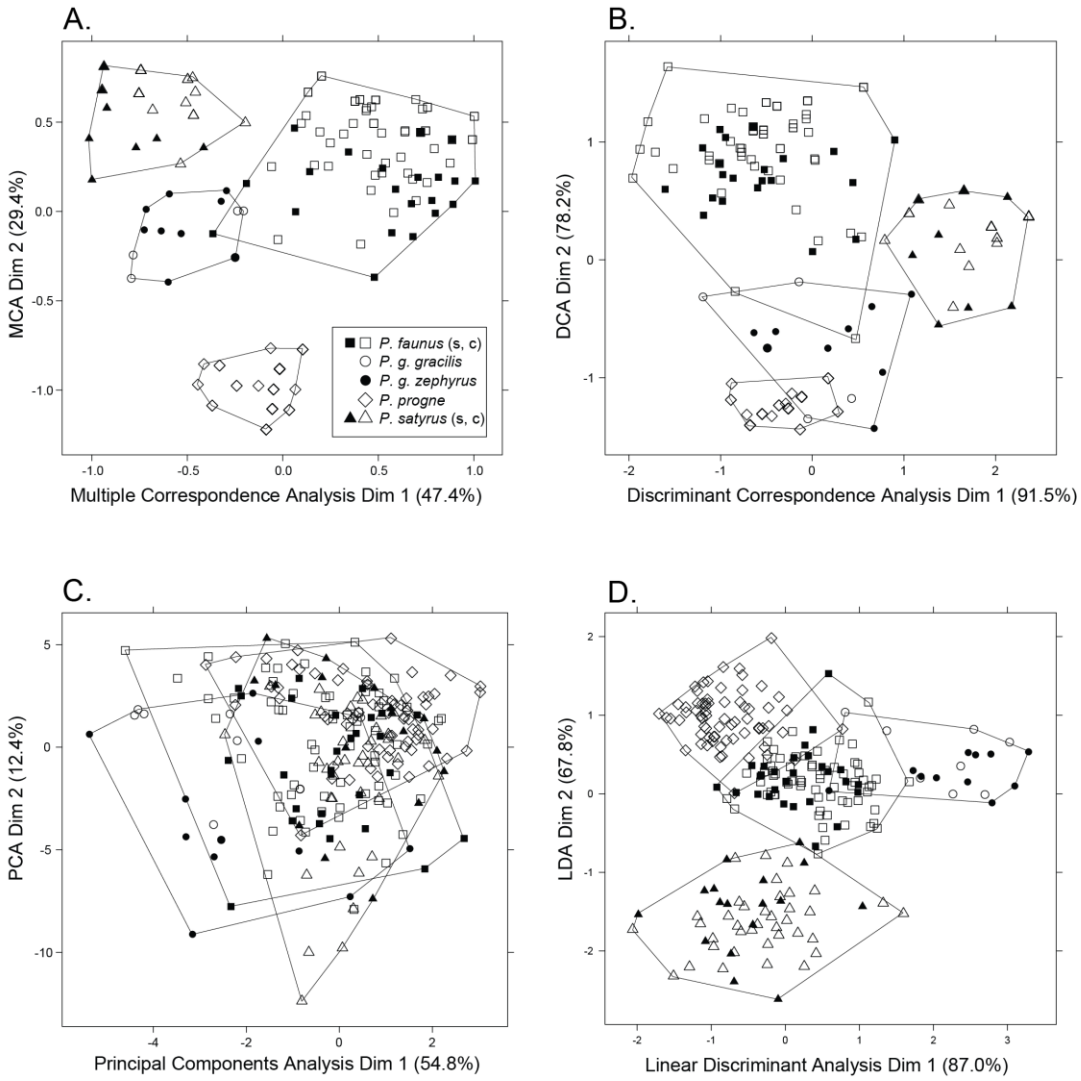


Figure 3.3: Multivariate analyses of wing characters in *Polygonia*. MCA(**A**.) and DCA (**B**.) of 10 visually scored characters. PCA (**C.**) and LDA (**D.**) of 18 RGB characters on 6 areas on the wing. Percentages on axis labels represent the proportion of total variation (MCA, PCA) and the remaining proportion of variation (DCA, LDA).

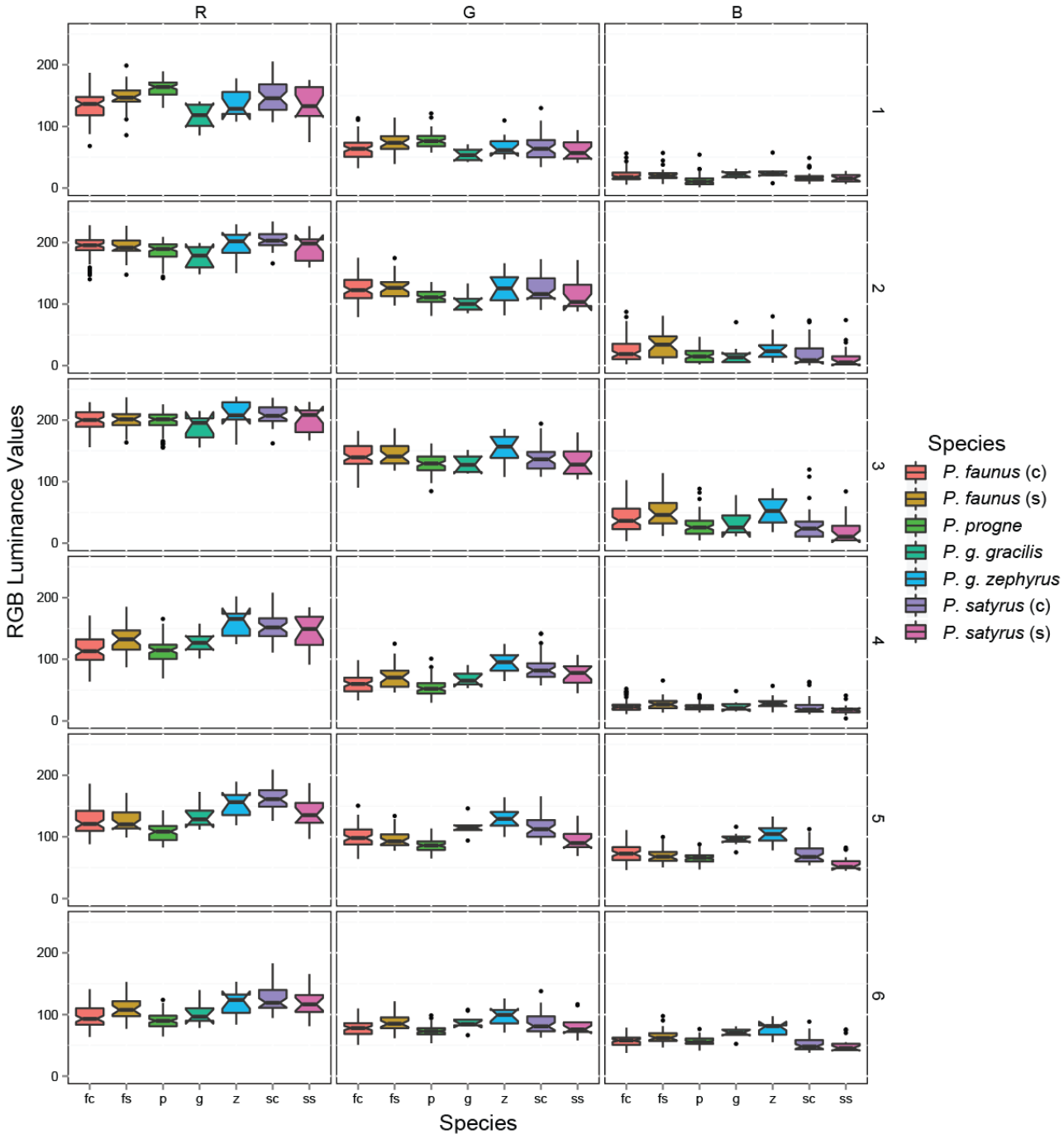


Figure 3.4: Central tendency statistics of mean R, G and B values of 6 wing characters across 248 *Polygonia* butterfies measured by ImageJ (Schneider et al. 2012). The upper and lower boundaries of each box are 3rd and 1st quartiles, respectively; vertical span of the notches correspond to the upper and lower 95% confidence bounds; the mean is represented by black lines bisecting each box. The upper and lower fences (whiskers) indicate the maximum and minimum values in the dataset, while outliers - those outside 1.5 times the interquartile range above the 4th quartile or below the first quartile - are shown as black points.

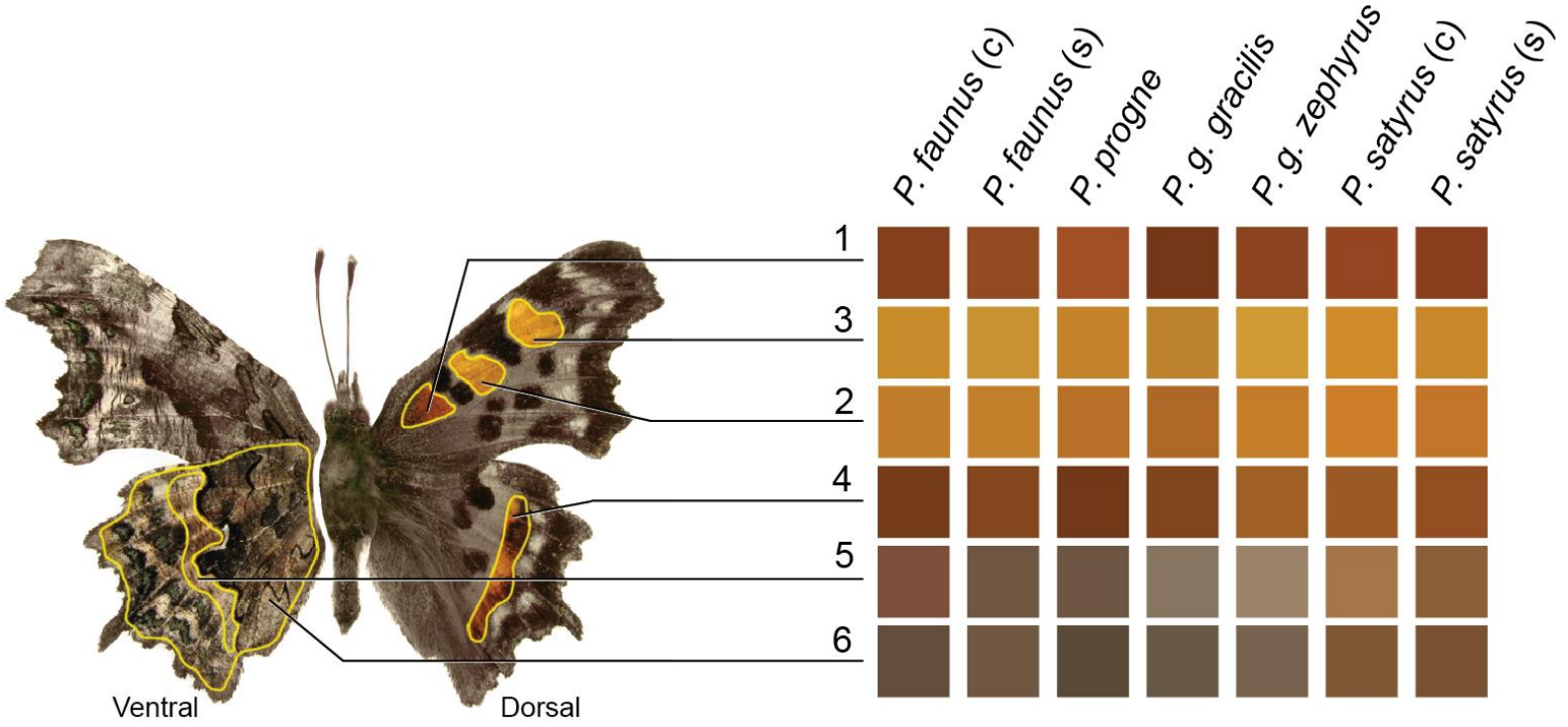


Figure 3.5: The average RGB luminance values visualized as colour swatches across 7 groups of *Polygonia* butterflies - *Polygonia faunus* (smeared and contrasted morphs)(n=28 and 67, respectively), *P. gracilis gracilis* (n=8), *P. g. zephyrus* (n=11), *P. progne* (n=68) and *P. satyrus* (smeared and contrasted morphs)(n=17 and 38, respectively).

Table 3.3: Canonical weights (loadings) for each of the 18 mean red, green and blue luminance values four species in the genus Polygonia (*P. faunus*, n=96; *P. gracilis* (sensu lato), n=18; *P. progne*, n=69; *P. satyrus*, n=55) as calculated by linear discriminant analysis (LDA) in R (R core group) using the package 'ade4' (Thioulouse and Dray 2007).

Char.	Color Comp.	Dim. 1	Dim. 2
1	R	0.056	-0.174
1	G	0.328	-0.531
1	B	-0.132	0.387
2	R	0.067	0.481
2	G	-0.386	-0.237
2	B	0.181	0.005
3	R	0.110	-0.233
3	G	0.162	0.437
3	B	0.021	-0.075
4	R	0.279	-0.102
4	G	-0.464	0.637
4	B	0.175	-0.360
5	R	0.419	0.435
5	G	-1.045	-1.313
5	B	0.589	1.275
6	R	-0.834	-0.485
6	G	0.013	0.152
6	B	0.574	0.243

Table 3.4: Tally of pairwise luminance means that were significantly different between groups: *Polygonia faunus* – contrasted (c) and smeared (s) morph, *P. progne*, *P. gracilis gracilis*, *P. g. zephyrus* and *P. satyrus* – contrasted (c) and smeared (s) morphs. Total number of significant comparisons in each species is in bold on the diagonal.

Species	*P. faunus* (c)	*P. faunus* (s)	*P. progne*	*P. g. gracilis*	*P. g. zephyrus*	*P. satyrus* (c)	*P. satyrus* (s)
P. faunus (c)	**26**						
P. faunus (s)	2	**24**					
P. progne	7	4	**40**				
P. g. gracilis	3	5	7	**24**			
P. g. zephyrus	7	6	11	2	**37**		
P. satyrus (c)	4	4	6	5	4	**26**	
P. satyrus (s)	3	3	5	2	7	3	**23**

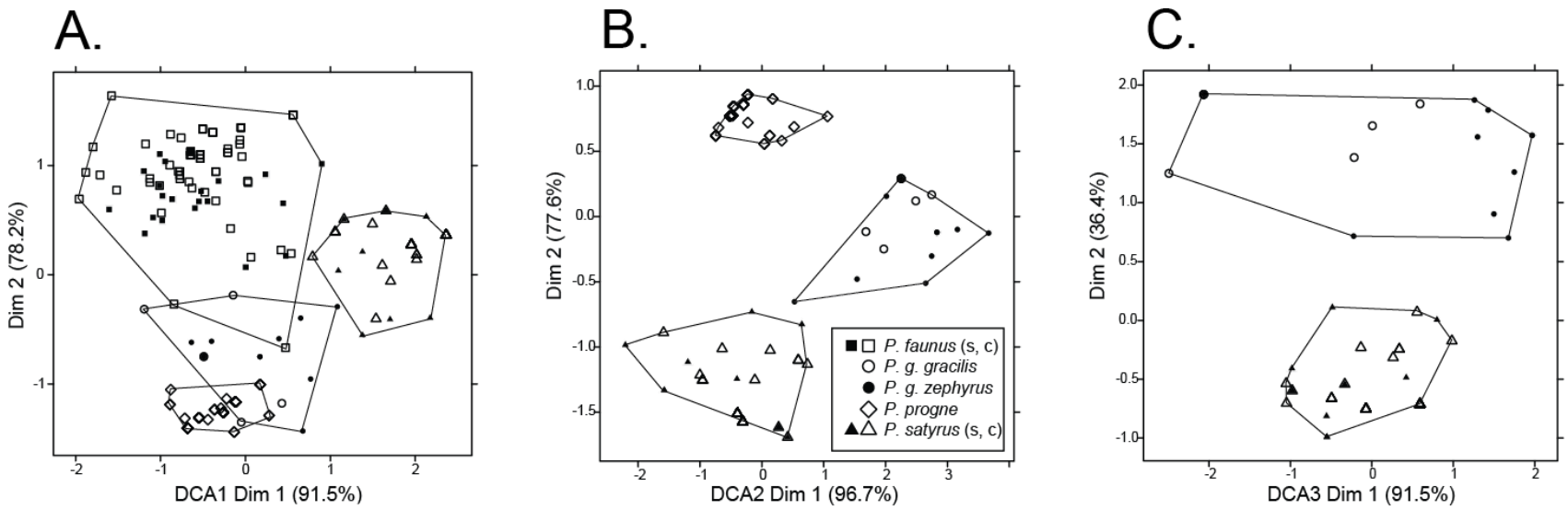


Figure 3.6: Three successive discriminant correspondence analyses (DCA) used in the construction of the dichotomous key. Analyses were performed in the ade4 package in R (Thioulouse and Dray 2007; R Core Team 2013).

CHAPTER 4

4. GENERAL CONCLUSIONS

4.1 INTRODUCTION

Many species of *Polygonia* are similar in appearance and can be difficult to identify in the field, especially where their ranges overlap. This can make identifications even more challenging in Alberta, Canada, which is home to five species of *Polygonia*: *P. faunus* (Edwards, 1870), *P. satyrus* (Edwards, 1869), *P. progne* (Cramer, 1775), *P. gracilis* (Grote and Robinson, 1867) and *P. oreas* (Edwards, 1869) (not sampled in this study). Studies have been conducted on the phylogenetic relationships within the genus; however, sampling in Alberta, a biodiversity hot spot (Acorn 1993; Bird et al. 1995; Layberry et al. 1998), has been limited (Wahlberg and Nylin 2003; Wahlberg et al. 2003, 2009).

The goals of this thesis were to assess species boundaries in the nymphalid butterfly genus *Polygonia* using a combination of mitochondrial sequence data and genomic SNPs obtained through the next-generation sequencing (NGS) technique genotyping-by-sequencing (GBS), and then to use multivariate analyses to assess the diagnostic utility of field markings provided in field guides when compared to genetic species identities.

4.2 DISCORDANCE AS AN OPPORTUNITY

Speciation is a complex process that is crucial to understand for all fields within biology (Bortolus 2008). Phylogenetic discordance, topological incongruence between gene trees, is observed in many taxonomic groups and poses both an obstacle to reconstructing evolutionary relationships and an opportunity to identify evolutionary events to better understand mechanisms of evolution (Maddison 1997; Ebel et al. 2015).

Phylogenies for COI and SNP data illustrated the genetic discordance observed in previous studies (Wahlberg and Nylin 2003; Wahlberg et al. 2009), such as the positions of *P.*

satyrus relative to *P. gracilis*. I also identified new discordances in the COI gene tree when compared to SNPs, specifically, the discovery of a new mitochondrial lineage in *P. gracilis*, which I have called G-X, positioned outside the mitochondrial *P. gracilis*+*P. satyrus* clade. I determined that this new lineage was not the result of nuclear insertion of a mitochondrial gene (NUMT) or two fragments spliced together from two individuals (chimera) by evaluating the total number of non-synonymous substitutions and constructing phylogenetic trees from both fragments independently. If not produced by either of these two means, these results could have been caused by the retention of an ancestral polymorphism in one of the two groups, or widespread ancient hybridization between the ancestors of these two species, transferring mitochondria from one species to another, with the transferred mtDNA from *P. satyrus* eventually displacing the original mtDNA of *P. gracilis* over most but not all of its range. The presence of the new mitochondrial lineage of G-X could thus represent a population that was sheltered from the introgression of *P. satyrus* and *P. gracilis*, or a stochastic retention of an ancient haplotype in one population.

The cytoplasmic parasite *Wolbachia* affects reproduction through male-killing, cytoplasmic incompatibility between infected and non-infected individuals, parthenogenesis and feminization (Werren 1997). Being maternally inherited, mitochondrial DNA is sensitive to this manipulation of reproduction in a species (Hurst and Jiggins 2005), so it is possible that *Wolbachia*-infected populations may have contributed to the discordance observed in *Polygonia* (Kodandaramaiah et al. 2013). Typically, Wolbachia infections spread rapidly through a population and negatively affect the genetic variation in the population; the mitochondrial genetic diversity in *P. satyrus* and *P. gracilis* are not nearly as low as *P. faunus* (Figure 2.10), suggesting that this may not be the source of the discordance in *P. gracilis*. Amplification of *wsp* (*Wolbachia*-specific gene) would confirm the presence of an infection in *Polygonia* populations (Kodandaramaiah et al. 2013; Roehrdanz and Wichmann 2013).

Although the geographic sampling in this study was not broad enough, nor was it dense enough to thoroughly investigate these hypotheses, sampling genomic SNPs and COI of specimens outside of the sympatric range of the two species could help elucidate the cause of the discord. If the G-X lineage were geographically isolated to regions allopatric to the range of *P. satyrus* or near the boundary, then the introgression hypothesis would be

supported. If, however, the new lineage were distributed throughout the range of *P. gracilis*, then the retention of an ancient polymorphism would be more likely.

4.3 UTILITY OF GENOMIC SNPS

In Chapter 2, I reconstructed phylogenetic relationships and assessed species boundaries of *P. faunus*, *P. gracilis*, *P. progne* and *P. satyrus* using genomic SNPs sequenced using GBS, a multiplexed double digested restriction-site-associated (RAD) sequencing method. Additionally, I used mitochondrial sequence data in the COI gene to investigate a well-documented case of genetic discordance between nuclear and mitochondrial gene trees, and compared the mtDNA topology to that obtained from genomic SNPs (Nylin et al. 2001; Wahlberg et al. 2003, 2009).

The GBS data was filtered through the UNEAK de novo pipeline (Lu et al. 2013) of the program TASSEL (Bradbury et al. 2007), which discovered over 40,000 SNP loci. Of these, 6627 loci were used for phylogenetic analyses (including all loci genotyped in a minimum 20% the total individuals), and 827 were applied to population structure analyses (minimum 80% individuals genotyped per locus). STRUCTURE and DAPC (discriminant analysis on principal components) analyses are Bayesian and multivariate methods, respectively, used here to assess similarities within and between species in their genetic structure (Pritchard et al. 2000; Jombart et al. 2010). The purpose of this was to test species boundaries and determine *a priori* species clusters for further integrated multivariate analysis in chapter 3. For mitochondrial DNA, 1450 base pairs of COI were compared with these genomic SNPs to detect genetic clusters and reconstruct the phylogenetic history of mtDNA. The use of mtDNA in addition to genomic SNPs allowed us to test the phylogenetic utility of SNPs against a molecular marker with well-documented challenges (Wahlberg and Nylin 2003; Wahlberg et al. 2009) using the same set of specimens.

Many studies have reported that SNPs are more effective than sequence data for reconstructing phylogenies of complex groups and estimating species trees (Nadeau et al. 2013; Wagner et al. 2013; Cruaud et al. 2014; Leaché et al. 2014). Broad genomic sampling of SNPs produces a tree representing the summed evolutionary history of all of the genes sampled, reconstructing phylogenetic relationships on a genomic scale, and approximating

the species tree far better than the use of sequence data for a limited selection of genes (Cariou et al. 2013; Wagner et al. 2013). The SNP-based phylogeny in this study was aligned with morphological, ecological and behavioural analysis performed by Nylin et al. (2001) and is also supported by patterns of host plant use across the genus (Weingartner et al. 2006). Not all clades were fully resolved, nor were they well supported in all analyses. The position of *P. faunus* and the monophyly of *P. progne* were not consistent in all SNP topologies, greatly affecting overall support values. This lack of topological support, despite the widespread success of SNP loci in other phylogenetic studies, could be due to underestimates of the necessary filtering thresholds, or conservatism when it comes to the inclusion of missing data (Huang and Knowles 2014).

Next generation sequencing data has powerful applications in systematics and many other biological fields (Davey et al. 2010). The multitude of SNP loci produced by RAD-seq and the ever-increasing ability to handle such computationally demanding data sets provide unprecedented resolution of species boundaries (Wagner et al. 2013), phylogeography (Lexer et al. 2013), phylogenetics (Nadeau et al. 2013), and population genetics (Martin et al. 2015). In addition to the SNP loci, the flanking sequences – the entire length of each 64 – 100 bp read – can be used to map synapomorphic or autapomorphic loci to known genes, and allow researchers to hypothesize which selective pressures have acted on populations to prompt diversification (Bird 2013; Janes et al. 2014). By understanding how selection is currently acting on species, we can better appreciate the nuances of evolutionary processes such as speciation with gene flow (Gagnaire et al. 2013), forest pest irruptions in species such as the spruce budworm complex (Bird 2013) and mountain pine beetle (Janes et al. 2014), and complex mimicry systems such as long-winged butterflies (Nadeau et al. 2013).

4.4 Integrative alpha taxonomy

STRUCTURE analysis and DAPC both produced distinct genetic clusters that corresponded with the four *Polygonia* species sampled in Alberta. I used these well-delimited clusters to serve as *a priori* hypotheses for specimen identifications, which were then compared to the clusters produced by multivariate analyses of diagnostic wing pattern features and digital wing colour morphometrics. I found that the visually scored wing pattern

characters delimited species boundaries, although some characters that have been described in field guides as highly diagnostic were not effective for distinguishing groups in the data. More importantly, two characters that were undescribed (Character 6) or scarcely reported (Character 9) in previously published identification resources were found to be highly diagnostic. In contrast, digital analysis of RGB luminance value in wing characters was ineffective for clustering individuals in principal components analysis (PCA) and only marginally effective for delimiting clusters based on prior genetic species in linear discriminant analysis (LDA). Despite promising results in other taxa (Lumley and Sperling 2010), in *Polygonia* this method may be sensitive to character selection and may require a different assemblage of characters.

Integration of discriminant correspondence analysis (DCA) of morphological characters into the construction of dichotomous keys should support identifications that better correspond to biological species, as opposed to phenotypic clusters. In large datasets, DCA loadings might highlight for the taxonomist those characters that may be the most diagnostic, reducing time spent parsing data matrices. This method employs iterative multivariate analysis of pools of taxa as they move through the key, helping the taxonomist select characters that are quantitatively diagnostic based on canonical weights, or loadings. These characters are then investigated for their ability to separate species, or groups of species, by the character states exhibited by different species. It is important to note that large canonical weights did not always identify informative or diagnostic characters, and that judgment calls based on the presence of multiple diagnostic characters were often made in the construction of the dichotomous key in this study. Integrative keys that are based on species boundaries from genomic data could provide insights into the diagnosis of morphologically challenging taxa and elucidate discrepancies between diagnostic characters and genetic clusters, as has been found in the genus *Melitaea* (Jugovic and Koren 2014); however, the same goal can be achieved in small to moderately sized datasets through genetically identifying specimens prior to key construction. More work is needed to weigh the benefits of using this method against the time-consuming process that it entails.

Discriminant analysis could also be applied to the assessment of identification resources such as field guides and dichotomous keys, supporting accurate identification of species in all fields of biology to avoid the cascading effects of bad taxonomy (Bortolus

2008). Identification accuracy is also important for threatened species (Vogler and Desalle 1994; Wiens 2007). With the recent and continuing decline of the field of taxonomy (Godfray 2002), many researchers have turned to the public to increase their geographic study area and sampling density (citizen science), which is especially important for assessing the conservation status of species.

By harnessing the public's enthusiasm for the local biota, a great number of person-hours can be mustered and the amount of data increases dramatically (Ladle 2008; Devictor et al. 2010). Citizen science projects such as e-Butterfly (www.e-butterfly.org), the UK Butterfly Monitoring Scheme (www.ukbms.org) and eBird (www.ebird.org) rely on the participation and enthusiasm of the public for the collection of widespread data on the distribution of species regionally, nationally or even worldwide (Devictor et al. 2010). The tradeoff of using data collected by the public is that it must be vetted to ensure accuracy before analysis. Logically, the more accurate the original identification is, the less strain on experts in the vetting process, especially if no photograph or field notes are provided by the observer. Some projects incorporate data quality filters to flag suspicious entries based on the likelihood of occurrence at that time of year or in that region in an effort to improve the quality of data being received (*e.g.* eBird: Sullivan et al. 2014). This substantial effort illustrates the importance of accuracy and quality of taxonomic data in citizen science.

Key construction plays an integral role in utility, as the taxonomist building the key has complete control over the characters, terminology and illustrations, if any. An attempt should be made by the taxonomic community to increase the accessibility and usability of diagnostic materials to the general public. By selecting characters that are unambiguous and quantitatively diagnostic, the utility of a key increases as the number of ambiguous judgment calls decreases, making diagnoses more accurate and less onerous on the user, especially if they lack experience in identifying that taxon (Walter and Winterton 2007). Keys should also use more than one character in a couplet to avoid the impossible identification of worn specimens (Walter and Winterton 2007), but not so many characters that the user loses track of which character states are exhibited in their specimen. By making small changes in the wording and construction of keys like reduction of jargon, more conversational wording and the inclusion of images or illustrations to describe character states, keys can be made useful for people other than experienced taxonomists.

4.5 CONCLUSION

In this thesis, I analyzed genomic SNPs produced by GBS, mitochondrial gene sequences, traditional diagnostic characters and wing area RBG data in Alberta *Polygonia* in an attempt to resolve the phylogenetic relationships in the genus and assess the accuracy of identification resources for this group. Genomic SNPs produced a phylogeny that more closely reflects morphological, ecological and behavioural traits and host plant use, suggesting that this topology may be a better approximation of the species tree. Additionally, a new mitochondrial lineage was discovered in *P. gracilis*, G-X, inviting further research into the causes of the phylogenetic discordance between mtDNA and nDNA in the genus. Traditional morphological characters from field guides delimited species boundaries that corresponded to genetic clusters by SNPs; however, RGB wing area analysis was not successful in the same task. Further research in this group is required to assess the validity of subspecies rank in *P. gracilis gracilis* and *P. g. zephyrus* and to find other populations exhibiting the G-X mitochondrial haplotype, which can be investigated through broader and denser geographic sampling. Similarly, the discordance in the positions of *P. satyrus*, *P. comma* and *P. interrogationis* in relation to *P. gracilis* and *P. progne* should be further examined through broader sampling to Eastern Canada and the United States. The genus *Polygonia* also provides opportunities to investigate mechanisms of evolution and speciation, and the effects of glacial periods on geographic range, interspecies interactions and genetic structure.

LITERATURE CITED

Acorn J.H., Sheldon I. 2006. Butterflies of British Columbia. Vancouver: Lone Pine Publishing.

Acorn J.H. 1993. Butterflies of Alberta. Vancouver: Lone Pine Publishing.

Acorn J.H. 2007. Sublogical Subspecificity. Am. Entomol. 53:255–256.

Acorn J.H. 2013. Diabolical Dichotomies. Am. Entomol. 59:190–192.

Amadon D. 1949. The seventy-five per cent rule for subspecies. Condor. 51:250–258.

Andersson L. 1990. The driving force: species concepts and ecology. Taxon. 39:375–382.

Angel L. 2005. Butterflies of Rocky Mountain National Park: An Observer's Guide. Boulder: Big Earth Publishing.

Anisimova M., Gil M., Dufayard J.-F., Dessimoz C., Gascuel O. 2011. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. Syst. Biol. 60:685–99.

Arnold B., Corbett-Detig R.B., Hartl D., Bomblies K. 2013. RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. Mol. Ecol. 22:3179–90.

Atran S. 1993. Cognitive Foundations of Natural History: Towards an Anthropology of Science. Cambridge: Cambridge University Press.

Avise J.C., Arnold J., Ball R.M., Bermingham E., Lamb T., Neigel J.E., Reeb C.A., Saunders N.C. 1987. Intraspecific Phylogeography: The Mitochondrial DNA Bridge Between Population Genetics and Systematics. Annu. Rev. Ecol. Syst. 18:489–522.

Avise J.C., Shapira J.F., Daniel S.W., Aquadro C.F., Lansman R.A. 1983. Mitochondrial DNA differentiation during the speciation process in *Peromyscus*. Mol. Biol. Evol. 1:38–56.

Baird N.A., Etter P.D., Atwood T.S., Currey M.C., Shiver A.L., Lewis Z.A., Selker E.U., Cresko W.A., Johnson E.A. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One. 3:e3376.

Barker D.G., Barker T.M., Davis M.A., Schuett G.W. 2015. A review of the systematics and taxonomy of Pythonidae: an ancient serpent lineage. Zool. J. Linn. Soc.175:1–19

Barraclough T.G., Vogler A.P. 2000. Detecting the Geographical Pattern of Speciation from Species-Level Phylogenies. Am. Nat. 155:419–434.

Bazinet A.L., Zwickl D.J., Cummings M.P. 2014. A gateway for phylogenetic analysis powered by grid computing featuring GARLI 2.0. Syst. Biol. 63:812–8.

Behrensmeyer A.K. 1992. Terrestrial Ecosystems Through Time: Evolutionary Paleoecology of Terrestrial Plants and Animals. Chicago: University of Chicago Press.

Beltran M., Jiggins C.D., Bull V., Linares M., Mallet J., McMillan W.O., Bermingham E. 2002. Phylogenetic Discordance at the Species Boundary: Comparative Gene Genealogies Among Rapidly Radiating *Heliconius* Butterflies. Mol. Biol. Evol. 19:2176–2190.

Bennett K.D., Tzedakis P.C., Willis K.J. 1991. Quaternary Refugia of North European Trees. J. Biogeogr. 18:pp. 103–115.

Bennett S. 2004. Solexa Ltd. Pharmacogenomics. 5:433–8.

Bird A., Tobin E. 2015. Natural Kinds. In: Zalta E.N., editor. The Stanford Encyclopedia of Philosophy.

Bird C.D., Hilchie G.J., Kondla N.G., Pike E.M., Sperling F.A.H. 1995. Alberta Butterflies. Edmonton: Provincial Museum of Alberta.

Bird H.M. 2013. Phylogenomics of the Choristoneura fumiferana species complex.

Blaxter M., Mann J., Chapman T., Thomas F., Whitton C., Floyd R., Abebe E. 2005. Defining operational taxonomic units using DNA barcode data. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 360:1935–1943.

Blaxter M.L. 2004. The promise of a DNA taxonomy. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 359:669–679.

Bortolus A. 2008. Error Cascades in the Biological Sciences: The Unwanted Consequences of Using Bad Taxonomy in Ecology. AMBIO A J. Hum. Environ. 37:114–118.

Bouckaert R., Heled J., Kühnert D., Vaughan T., Wu C.-H., Xie D., Suchard M.A., Rambaut A., Drummond A.J. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10:e1003537.

Bradbury P.J., Zhang Z., Kroon D.E., Casstevens T.M., Ramdoss Y., Buckler E.S. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics. 23:2633–2635.

Brock J.P., Brock J.P., Kaufman K. 2003. Kaufman Field Guide to Butterflies of North America. Boston: Houghton Mifflin Harcourt.

Brower A.V.Z., DeSalle R. 1998. Patterns of mitochondrial versus nuclear DNA sequence divergence among nymphalid butterflies: the utility of wingless as a source of characters for phylogenetic inference. Insect Mol. Biol. 7:73–82.

Brunet B.M.T., Hundsdoerfer A.E., Sperling F.A.H. 2013. Identification and ecological characterisation of *Choristoneura occidentalis* (Lepidoptera: Tortricidae) populations in southwestern Alberta, Canada. Can. Entomol. 145:521–528.

Buck M., Cobb T.P., Stahlhut J.K., Hanner R.H. 2012. Unravelling cryptic species diversity in eastern Nearctic paper wasps, *Polistes* (*Fuscopolistes*), using male genitalia, morphometrics and DNA barcoding, with descriptions of two new species (Hymenoptera: Vespidae). Zootaxa.:1–48.

Cariou M., Duret L., Charlat S. 2013. Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. Ecol. Evol. 3:846–52.

Caterino M.S., Cho S., Sperling F.A. 2000. The current state of insect molecular systematics: a thriving Tower of Babel. Annu. Rev. Entomol. 45:1–54.

Clary D.O., Wolstenholme D.R. 1985. The mitochondrial DNA molecular of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. J. Mol. Evol. 22:252–71.

Clement M., Posada D., Crandall K.A. 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. 9:1657–9.

Combosch D.J., Vollmer S. V. 2015. Trans-Pacific RAD-Seq population genomics confirms introgressive hybridization in Eastern Pacific *Pocillopora* corals. Mol. Phylogenet. Evol. 88:154–162.

Cruaud A., Gautier M., Galan M., Foucaud J., Sauné L., Genson G.G., Dubois E., Nidelet S., Deuve T., Rasplus J.-Y.Y. 2014. Empirical assessment of RAD sequencing for interspecific phylogeny. Mol. Biol. Evol. 31:1272–4.

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

Davey J.W., Davey J.L., Blaxter M.L., Blaxter M.W. 2010. RADSeq: next-generation population genetics. Brief. Funct. Genomics. 9:416–23.

Davey J.W., Hohenlohe P.A., Etter P.D., Boone J.Q., Catchen J.M., Blaxter M.L. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat. Rev. Genet. 12:499–510.

Dayrat B. 2005. Towards integrative taxonomy. Biol. J. Linn. Soc. 85:407–415.

Degnan J.H., Rosenberg N.A. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24:332–40.

DeSalle R., Egan M.G., Siddall M. 2005. The unholy trinity: taxonomy, species delimitation and DNA barcoding. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 360:1905–16.

Devictor V., Whittaker R.J., Beltrame C. 2010. Beyond scarcity: citizen science programmes as useful tools for conservation biogeography. Divers. Distrib. 16:354–362.

Dobzhansky T. 1940. Speciation as a Stage in Evolutionary Divergence. Am. Nat. 74:312.

Dornfeld E.J. 1980. Butterflies of Oregon. Portland: Timber Press.

Dumas P., Barbut J., Le Ru B., Silvain J.-F., Clamens A.-L., D'Alençon E., Kergoat G.J. 2015. Phylogenetic molecular species delimitations unravel potential new species in the pest genus *Spodoptera* Guenée, 1852 (lepidoptera, noctuidae). PLoS One. 10:e0122407.

Dupuis J.R., Roe A.D., Sperling F.A.H. 2012. Multi-locus species delimitation in closely related animals and fungi: one marker is not enough. Mol. Ecol. 21:4422–36.

Dyke A.S., Moore A., Robertson L. 2013. Deglaciation of North America.

Earl D.A., VonHoldt B.M. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet Resour.:359–361.

Eaton D.A.R., Ree R.H. 2013. Inferring phylogeny and introgression using RADseq data: an example from flowering plants (*Pedicularis*: Orobanchaceae). Syst. Biol. 62:689–706.

Ebel E.R., DaCosta J.M., Sorenson M.D., Hill R.I., Briscoe A.D., Willmott K.R., Mullen S.P. 2015. Rapid diversification associated with ecological specialization in Neotropical Adelpha butterflies. Mol. Ecol. 24:2392–405.

Edwards S. V., Beerli P. 2000. Perspective: Gene divergence, population divergence, and the variance incoalescence time in phylogeographic studies. Evolution (N. Y). 54:1839.

Ehlers J., Gibbard P.L. 2004. Quaternary Glaciations - Extent and Chronology: Part II: North America. Elsevier.

El-Metwally S., Ouda O.M., Helmy M. 2014. Next Generation Sequencing Technologies and Challenges in Sequence Assembly. New York: Springer New York.

Eldredge N., Cracraft J. 1980. Phylogenetic patterns and the evolutionary process. Method and theory in comparative biology. New York: Columbia University Press.

Elshire R.J., Glaubitz J.C., Sun Q., Poland J.A., Kawamoto K., Buckler E.S., Mitchell S.E. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One. 6:e19379.

Endo K., Masaki T., Kumagai K. 1988. Neuroendocrine regulation of the development of seasonal morphs in the Asian comma butterfly, *Polygonia c-aureum* L.: difference in activity of summer-morph-producing hormone from brain-extracts of the long-day and short-day pupae. Zoolog. Sci. 5:145–152.

Endo K. 1969. Relation between ovarian maturation and activity of the corpora allata in seasonal forms of the butterfly, *Polygonia c-aureum* L. Dev. Growth Differ. 11:297–304.

Endo K. 1973. Hormonal regulation of mating in the butterfly, *Polygonia c-aureum* L. Dev. Growth Differ. 15:1–10.

Ereshefsky M. 2010. Species. In: Zalta E.N., editor. The Stanford Encyclopedia of Philosophy.

Etter P.D., Bassham S., Hohenlohe P.A., Johnson E.A., Cresko W.A. 2011. SNP discovery and genotyping for evolutionary genetics using RAD sequencing. Methods Mol. Biol. 772:157–78.

Fisher M.S.. 2006. Lepidoptera of North America: Nymphalidae, the Subfamily Nymphalinae the Checkerspots and Crescenspots (tribe Melitaeini) the Anglewings, Tortoiseshells, Painted Ladies and Buckeyes (tribes Nymphalini and Junoniini) the Subfamilies Limenitidinae, Apaturin. Fort Collins: C.P. Gillette Museum of Arthropod Diversity, Department of Bioagricultural Sciences and Pest Management, Colorado State University.

Ford E.B. 1941. Studies on the chemistry of pigments in the Lepidoptera, with reference to their bearing on systematics. 1. The Anthoxanthins. Proc. R. Entomol. Soc. London. Ser. A, Gen. Entomol. 16:65–90.

Francuski L., Ludoski J., Vujic A., Milankov V. 2009. Wing Geometric Morphometric Inferences on Species Delimitation and Intraspecific Divergent Units in the Merodon ruficornis Group (Diptera, Syrphidae) from the Balkan Peninsula. Zoolog. Sci. 26:301–308.

Fujita M.K., Leaché A.D., Burbrink F.T., McGuire J.A., Moritz C. 2012. Coalescent-based species delimitation in an integrative taxonomy. Trends Ecol. Evol. 27:480–8.

Fukuda S., Endo K. 2006. Hormonal Control of the Development of Seasonal Forms in the Butterfly, *Polygonia c-aureum* L. Proc. Jpn. Acad. 42:1082–1087.

Gagnaire P.-A., Pavey S.A., Normandeau E., Bernatchez L. 2013. The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by RAD sequencing. Evolution. 67:2483–97.

Galotti K.M. 2013. Cognitive Psychology In and Out of the Laboratory. SAGE Publications.

Gebiola M., Bernardo U., Ribes A., Gibson G.A.P. 2015. An integrative study of *Necremnus* Thomson (Hymenoptera: Eulophidae) associated with invasive pests in Europe and North America: taxonomic and ecological implications. Zool. J. Linn. Soc. 173:352–423.

Gerson L.P. 1999. Aristotle: Logic and metaphysics. London: Taylor & Francis.

Glassberg J. 2001. Butterflies Through Binoculars. Oxford: Oxford University Press.

Godfray H.C.J. 2002. Challenges for taxonomy. Nature. 417:17–9.

Goldstein E. 2014. Cognitive Psychology: Connecting Mind, Research and Everyday Experience. Boston: Cengage Learning.

Goloboff P.A., Farris J.S., Nixon K.C. 2008. TNT, a free program for phylogenetic analysis. Cladistics. 24:774–786.

Grose M.J. 2012. Plant colour as a visual aspect of biological conservation. Biol. Conserv. 153:159–163.

Grote A.H. 1873. The Canadian Entomologist. v. 4-5.

Guindon S., Gascuel O. 2003. A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. Syst. Biol. 52:696–704.

Haig S.M., Beever E.A., Chambers S.M., Draheim H.M., Dugger B.D., Dunham S., Elliott-Smith E., Fontaine J.B., Kesler D.C., Knaus B.J., Lopes I.F., Loschl P., Mullins T.D., Sheffield L.M. 2006. Taxonomic considerations in listing subspecies under the U.S. Endangered Species Act. Conserv. Biol. 20:1584–94.

Hansson C., Smith M.A., Janzen D.H., Hallwachs W. 2015. Integrative taxonomy of New World Euplectrus Westwood (Hymenoptera, Eulophidae), with focus on 55 new species from Area de Conservación Guanacaste, northwestern Costa Rica. Zookeys.:1–236.

Hare M.P., Avise J.C. 1996. Molecular genetic analysis of a stepped multilocus cline in the American oyster (*Crassostrea virginica*). Evolution (N. Y).:2305–2315.

Harrison R.G. 1989. Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. Trends Ecol. Evol. 4:6–11.

Hausdorf B. 2011. Progress toward a general species concept. Evolution (N. Y). 65:923–31.

Hebert P.D.N., Cywinska A., Ball S.L., deWaard J.R. 2003. Biological identifications through DNA barcodes. Proc. Biol. Sci. 270:313–21.

Hebert P.D.N., Penton E.H., Burns J.M., Janzen D.H., Hallwachs W. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. Proc. Natl. Acad. Sci. U. S. A. 101:14812–7.

Hedin M. 2015. High-stakes species delimitation in eyeless cave spiders (Cicurina, Dictynidae, Araneae) from central Texas. Mol. Ecol. 24:346–61.

Hemming F. 1943. The Bulletin of Zoological Nomenclature. Int. Comm. Zool. Nomencl. v.1:pt.1-1.

Hemming F. 1967. The Generic Names of the Butterflies and their Type-Species (Lepidoptera: Rhopalocera). London :BM(NH).

Hoelzer G.A. 1997. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees revisited. Evolution (N. Y). 51:622–626.

Hooper R.R., Long R. 1973. The Butterflies of Saskatchewan. Regina: Department of Natural Resources.

Huang H., Knowles L.L. 2014. Unforeseen Consequences of Excluding Missing Data from Next-Generation Sequences: Simulation Study of RAD Sequences. Syst. Biol.:syu046–.

Hull D.L. 1978. A matter of individuality. Philos. Sci.:335–360.

Hurst G.D.D., Jiggins F.M. 2005. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. Proc. Biol. Sci. 272:1525–34.

ICZN. 1999. International Code of Zoological Nomenclature, 4th Ed.

James D., Nunnallee D. 2011. Life Histories of Cascadia Butterflies. Oregon State University Press.

Janes J.K., Li Y., Keeling C.I., Yuen M.M.S., Boone C.K., Cooke J.E.K., Bohlmann J., Huber D.P.W., Murray B.W., Coltman D.W., Sperling F.A.H. 2014. How the mountain pine beetle (*Dendroctonus ponderosae*) breached the Canadian Rocky Mountains. Mol. Biol. Evol. 31:1803–15.

Janz N., Nylin S. 2008. The oscillation hypothesis of host-plant range and speciation. Spec. speciation, Radiat. Evol. Biol. Herbiv. insects.:203–215.

Janzen D.H., Hallwachs W., Blandin P., Burns J.M., Cadiou J.-M., Chacon I., Dapkey T., Deans A.R., Epstein M.E., Espinoza B., Franclemont J.G., Haber W.A., Hajibabaei M., Hall J.P.W., Hebert P.D.N., Gauld I.D., Harvey D.J., Hausmann A., Kitching I.J., Lafontaine D., Landry J.-F., Lemaire C., Miller J.Y., Miller J.S., Miller L., Miller S.E., Montero J., Munroe E., Green S.R., Ratnasingham S., Rawlins J.E., Robbins R.K., Rodriguez J.J., Rougerie R., Sharkey M.J., Smith M.A., Solis M.A., Sullivan J.B., Thiaucourt P., Wahl D.B., Weller S.J., Whitfield J.B., Willmott K.R., Wood D.M., Woodley N.E., Wilson J.J. 2009. Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. Mol. Ecol. Resour. 9:1–26.

Jombart T., Devillard S., Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. 11:94.

Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics. 24:1403–5.

Jones J.C., Fan S., Franchini P., Schartl M., Meyer A. 2013. The evolutionary history of *Xiphophorus* fish and their sexually selected sword: a genome-wide approach using restriction site-associated DNA sequencing. Mol. Ecol. 22:2986–3001.

Jugovic J., Koren T. 2014. Wing pattern morphology of three closely related <i>Melitaea</I> (Lepidoptera, Nymphalidae) species reveals highly inaccurate external morphology-based species identification. Nota Lepi.:75–90.

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., Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28:1647–9.

Klassen P., Westwood A.R., Preston W.B., McKillop W.B. 1989. The Butterflies of Manitoba. Winnipeg: Manitoba Museum of Man and Nature.

Kodandaramaiah U., Simonsen T.J., Bromilow S., Wahlberg N., Sperling F. 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.

Kodandaramaiah U., Weingartner E., Janz N., Leski M., Slove J., Warren A., Nylin S. 2012a. Investigating concordance among genetic data, subspecies circumscriptions and hostplant use in the Nymphalid butterfly *Polygonia faunus*. PLoS One. 7.

Kodandaramaiah U., Weingartner E., Janz N., Leski M., Slove J., Warren A. 2012b. Validity of the subspecies paradigm - a case study of the nymphalid butterfly *Polygonia faunus*. PLoS One.

Köhler G. 2012. Color Catalogue for Field Biologists. Offenbach: Herpeton.

Krajmerová D., Paule L., Zhelev P., Voleková M., Evtimov I., Gagov V., Gömöry D. 2015. Natural hybridization in eastern-Mediterranean firs: The case of *Abies borisii-regis*. Plant Biosyst. - An Int. J. Deal. with all Asp. Plant Biol.:1–11.

Ladle R.J. 2008. Catching fairies and the public representation of biogeography. J. Biogeogr. 35:388–391.

Layberry R.A., Hall P.W., Lafontaine J.D. 1998. The Butterflies of Canada. Toronto: University of Toronto Press.

Leaché A.D., Fujita M.K., Minin V.N., Bouckaert R.R. 2014. Species delimitation using genome-wide SNP data. Syst. Biol. 63:534–42.

Leaché A.D., Koo M.S., Spencer C.L., Papenfuss T.J., Fisher R.N., McGuire J.A. 2009. Quantifying ecological, morphological, and genetic variation to delimit species in the coast horned lizard species complex (*Phrynosoma*). Proc. Natl. Acad. Sci. U. S. A. 106:12418–23.

Lemmon E.M., Lemmon A.R. 2013. High-Throughput Genomic Data in Systematics and Phylogenetics. Annu. Rev. Ecol. Evol. Syst. 44:99–121.

Lerner H., Fleischer R. 2010. Prospects for the Use of Next-Generation Sequencing Methods in Ornithology. Auk. 127:4–15.

Lexer C., Mangili S., Bossolini E., Forest F., Stölting K.N., Pearman P.B., Zimmermann N.E., Salamin N. 2013. "Next generation" biogeography: towards understanding the drivers of species diversification and persistence. J. Biogeogr. 40:1013–1022.

Lin C.-P., Danforth B.N. 2004. How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets. Mol. Phylogenet. Evol. 30:686–702.

Longo G., Bernardi G. 2015. The evolutionary history of the embiotocid surfperch radiation based on genome-wide RAD sequence data. Mol. Phylogenet. Evol. 88:55–63.

Lopez J. V., Yuhki N., Masuda R., Modi W., O'Brien S.J. Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. J. Mol. Evol. 39:174–190.

Lu F., Lipka A.E., Glaubitz J., Elshire R., Cherney J.H., Casler M.D., Buckler E.S., Costich D.E. 2013. Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. PLoS Genet. 9:e1003215.

Lumley L.M., Sperling F.A.H. 2010. Integrating morphology and mitochondrial DNA for species delimitation within the spruce budworm (*Choristoneura fumiferana*) cryptic species complex (Lepidoptera: Tortricidae). Syst. Entomol. 35:416–428.

Lumley L.M., Sperling F.A.H. 2011. Utility of microsatellites and mitochondrial DNA for species delimitation in the spruce budworm (*Choristoneura fumiferana*) species complex (Lepidoptera: Tortricidae). Mol. Phylogenet. Evol. 58:232–43.

Maddison W.P., Maddison D.R. 2008. Mesquite: A modular system for evolutionary analysis. Evolution (N. Y). 62:1103–1118.

Maddison W.P. 1997. Gene Trees in Species Trees. Syst. Biol. 46:523–536.

Mallet J. 1995. A species definition for the modern synthesis. Trends Ecol. Evol. 10:294–299.

Mallet J. 2005. Hybridization as an invasion of the genome. Trends Ecol. Evol. 20:229–37.

Mallet J. 2007. Subspecies, Semispecies, Superspecies. Encycl. Biodivers.:1–5.

Marincovich L., Gladenkov A.Y. 1999. Evidence for an early opening of the Bering Strait. 397:149–151.

Martin C.H., Cutler J.S., Friel J.P., Dening C.T., Coop G., Wainwright P.C. 2015. Complex histories of repeated gene flow in Cameroon crater lake cichlids cast doubt on one of the clearest examples of sympatric speciation. Evolution (N. Y).

Mayr E., Ashlock P.D. 1991. Principles of systematic zoology. New York: McGraw-Hill.

Mayr E. 1942. Systematics and the Origin of Species. .

Mayr E. 1963. Animal species and evolution. Cambridge: Harvard University Press.

Mayr E. 1982a. The Growth of Biological Thought: Diversity, Evolution, and Inheritance. Harvard University Press.

Mayr E. 1982b. Of What Use Are Subspecies? Auk. 99:pp. 593–595.

Mayr E. 1991. One Long Argument: Charles Darwin and the Genesis of Modern Evolutionary Thought. Cambridge: Harvard University Press.

Mayr E. 1992. A Local Flora and the Biological Species Concept. Available from http://www.jstor.org/discover/10.2307/2445111?uid=2&uid=4&sid=21106425710271.

McCormack J.E., Hird S.M., Zellmer A.J., Carstens B.C., Brumfield R.T. 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. Mol. Phylogenet. Evol. 66:526–38.

McCormack J.E., Maley J.M., Hird S.M., Derryberry E.P., Graves G.R., Brumfield R.T. 2012. Next-generation sequencing reveals phylogeographic structure and a species tree for recent bird divergences. Mol. Phylogenet. Evol. 62:397–406.

Meier R., Shiyang K., Vaidya G., Ng P.K.L. 2006. DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. Syst. Biol. 55:715–28.

Meyer C.P., Paulay G. 2005. DNA barcoding: error rates based on comprehensive sampling. PLoS Biol. 3:e422.

Michener C.D. 1970. Diverse approaches to systematics. Evol. Biol.:1–38.

Miller K.G., Wright J.D., Fairbanks R.G. 1991. Unlocking the Ice House: Oligocene-Miocene oxygen isotopes, eustasy, and margin erosion. J. Geophys. Res. 96:6829.

Mishler B.D., Donoghue M.J. 1982. Species concepts: a case for pluralism. Syst. Biol. 31:491–503.

Moritz C. 1994. Defining "Evolutionarily Significant Units" for conservation. Trends Ecol. Evol. 9:373–5.

Müller P., Pflüger V., Wittwer M., Ziegler D., Chandre F., Simard F., Lengeler C. 2013. Identification of cryptic *Anopheles* mosquito species by molecular protein profiling. PLoS One. 8:e57486.

Mutanen M. 2005. Delimitation difficulties in species splits: a morphometric case study on the *Euxoa tritici* complex (Lepidoptera, Noctuidae). Syst. Entomol. 30:632–643.

Naciri Y., Linder H.P. 2015. Species delimitation and relationships: The dance of the seven veils. Taxon. 64:3–16.

Nadeau N.J., Martin S.H., Kozak K.M., Salazar C., Dasmahapatra K.K., Davey J.W., Baxter S.W., Blaxter M.L., Mallet J., Jiggins C.D. 2013. Genome-wide patterns of divergence and gene flow across a butterfly radiation. Mol. Ecol. 22:814–26.

Nagel L., Schluter D. 1998. Body size, natural selection, and speciation in sticklebacks. Evolution (N. Y).:209–218.

Narita S., Nomura M., Kato Y., Fukatsu T. 2006. Genetic structure of sibling butterfly species affected by *Wolbachia* infection sweep: evolutionary and biogeographical implications. Mol. Ecol. 15:1095–108.

Neill W., Hepburn D. 2001. The Guide to Butterflies of Oregon and Washington. Englewood: Westcliffe Publishers.

Van Nieukerken E.J., Doorenweerd C., Stokvis F.R., Groenenberg D.S.J. 2012. DNA barcoding of the leaf-mining moth subgenus *Ectoedemia* s. str.(Lepidoptera: Nepticulidae) with COI and EF1-α: two are better than one in recognising cryptic species. Contrib. to Zool. 81:1–24.

Nylin S., Nyblom K., Ronquist F., Janz N., Belicek J., Kallersjo M. 2001. Phylogeny of *Polygonia*, *Nymphalis* and related butterflies (Lepidoptera: Nymphalidae): a total-evidence analysis. Zool. J. Linn. Soc. 132:441–468.

Nylin S. 1991. Seasonal plasticity in life history traits: growth and developments in *Polygonia c-album* (Lepidoptera: Nymphalidae). Biol. J. Linn. Soc.:301–323.

O'Brien S.J., Mayr E. 1991. Bureaucratic Mischief, Recognizing Endangered Species and Subspecies. Science (80-.). 251:1187–1188.

Opler P.A. 1999. A Field Guide to Western Butterflies. Houghton Mifflin Harcourt.

Orr H.A. 2001. Some doubts about (yet another) view of species. J. Evol. Biol. 14:870–871.

Osborne D. V. 1963. Some aspects of the theory of dichotomous keys. New Phytol. 62:144–160.

Padial J.M., De La Riva I. 2009. Integrative taxonomy reveals cryptic Amazonian species of *Pristimantis* (Anura: Strabomantidae). Zool. J. Linn. Soc. 155:97–122.

Pankhurst R.J. 1970. A computer program for generating diagnostic keys. Comput. J. 13:145–151.

Pante E., Abdelkrim J., Viricel A., Gey D., France S.C., Boisselier M.C., Samadi S. 2015. Use of RAD sequencing for delimiting species. Heredity (Edinb). 114:450–9.

Patten M. a, Unitt P. 2002. Diagnosability Versus Mean Differences of Sage Sparrow Subspecies. Auk. 119:26–35.

Patten. 2010. Null Expectations in Subspecies Diagnosis. Ornithol. Monogr. 67:35–41.

Payne D.G., Wenger M.J. 1998. Cognitive psychology. Boston: Houghton Mifflin College Division.

Pelham J.P. 2008. A catalogue of the butterflies of the United States and Canada: with a complete bibliography of the descriptive and systematic literature. Beverly Hills: Lepidoptera Research Foundation.

Phillimore A.B., Owens I.P.F. 2006. Are subspecies useful in evolutionary and conservation biology? Proc. Biol. Sci. 273:1049–53.

Pohl G.R., Anweiler G.G., Schmidt B.C., Kondla N.G. 2010. An annotated list of the Lepidoptera of Alberta, Canada. Zookeys. 38:1–549.

Poland J. a., Brown P.J., Sorrells M.E., Jannink J.L. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One. 7.

Poole S. 2009. Butterflies of Grand Teton and Yellowstone National Parks. Moose: Grand Teton Association.

Pritchard J.K., Stephens M., Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics. 155:945–959.

Prodanov D. 2010. Color Histogram. Available from http://imagej.nih.gov/ij/plugins/color-histogram.html.

Proshek B., Dupuis J.R., Engberg A., Davenport K., Opler P.A., Powell J.A., Sperling F.A. 2015. Genetic evaluation of the evolutionary distinctness of a federally endangered butterfly, Lange's Metalmark. BMC Evol. Biol. 15:73.

Pyle R.M. 2002. The Butterflies of Cascadia: A Field Guide to All the Species of Washington, Oregon, and Surrounding Territories. Seattle: Seattle Audubon Society.

De Queiroz K. 2007. Species concepts and species delimitation. Syst. Biol. 56:879–86.

Quine W. 1970. Natural Kinds. In: Rescher N., editor. Essays in honor of Carl G. Hempel. p. 1–23.

R Core Team. 2013. R: A Language and Environment for Statistical Computing. .

Rambaut A., Suchard M.A., Xie D., Drummond A.J. 2014. Tracer v1.6. Available from http://beast.bio.ed.ac.uk/Tracer.

Remsen J. 2010. Subspecies as a Meaningful Taxonomic Rank in Avian Classification. Ornithol. Monogr. 67.

Rissler L.J., Apodaca J.J. 2007. Adding more ecology into species delimitation: ecological niche models and phylogeography help define cryptic species in the black salamander (*Aneides flavipunctatus*). Syst. Biol. 56:924–42.

Rodriguez-Pereyra G. 2015. Nominalism in Metaphysics. In: Zalta E.N., editor. The Stanford Encyclopedia of Philosophy.

Roe A.D., Sperling F.A.H. 2007. Population structure and species boundary delimitation of cryptic *Dioryctria* moths: An integrative approach. Mol. Ecol. 16:3617–3633.

Roehrdanz R.L., Wichmann S.G.S. 2013. *Wolbachia wsp* Gene Clones Detect the Distribution of *Wolbachia* Variants and *wsp* Hypervariable Regions Among Individuals of a

Multistrain Infected Population of Diabrotica barberi (Coleoptera: Chrysomelidae). Ann. Entomol. Soc. Am. 106:329–338.

Rosen D.E. 1979. Fishes from the uplands and intermontane basins of Guatemala: Revisionary studies and comparative geography. Nat. Hist. 162:267–376.

Royer R. 1988. Butterflies of North Dakota: An Atlas and Guide. Minot: Minot State University.

Ryder O.A. 1986. Species conservation and systematics: the dilemma of subspecies. Trends Ecol. Evol. 1:9–10.

Schlick-Steiner B.C., Steiner F.M., Seifert B., Stauffer C., Christian E., Crozier R.H. 2010. Integrative taxonomy: a multisource approach to exploring biodiversity. Annu. Rev. Entomol. 55:421–38.

Schneider C.A., Rasband W.S., Eliceiri K.W. 2012. NIH Image to ImageJ: 25 years of image analysis. Nat. Methods. 9:671–675.

Schwarzfeld M.D., Sperling F.A.H. 2014. Species delimitation using morphology, morphometrics, and molecules: definition of the *Ophion scutellaris* Thomson species group, with descriptions of six new species (Hymenoptera, Ichneumonidae). Zookeys.:59–114.

Scott J.A. 1984. A review of *Polygonia progne* (*oreas*) and *P. gracilis* (*zephyrus*) (Nymphalidae), including a new subspecies from the southern Rocky Mountains. J. Res. Lepid. 23.

Scott J.A. 1992. The Butterflies of North America: A Natural History and Field Guide. Redwood City: Stanford University Press.

Shapiro A.M. 1990. Ball Mountain revisited: Anomalous species richness of a montane barrier zone. J. Res. Lepid. 29:143–156.

Shepard J., Guppy C. 2011. Butterflies of British Columbia: Including Western Alberta, Southern Yukon, the Alaska Panhandle, Washington, Northern Oregon, Northern Idaho, and Northwestern Montana. Vancouver: UBC Press.

Slade R., Moritz C., Heideman A. 1994. Multiple nuclear-gene phylogenies: application to pinnipeds and comparison with a mitochondrial DNA gene phylogeny. Mol. Biol. Evol. 11:341–356.

Smithe F.B. 1975. Naturalist's Color Guide. New York: American Museum of Natural History.

Soderlind L., Nylin S. 2011. Genetics of diapause in the comma butterfly *Polygonia c-album*. Physiol. Entomol. 36:8–13.

Sokal R.R., Crovello T.J. 1970. The Biological Species Concept: A Critical Evaluation. Am. Nat. 104:127–153.

Sperling F. a H., Roe A.D. 2009. Molecular Dimensions of Insect Taxonomy. In: Foottit R., Adler P., editors. Insect Biodiversity: Science and Society. Oxford: Blackwell Publishing. p. 397–415.

Sperling F., Byers R., Hickey D. 1996. Mitochondrial DNA sequence variation among pheromotypes of the dingy cutworm, *Feltia jaculifera* (Gn.) (Lepidoptera: Noctuidae). Can. J. Zool. 74:2109–2117.

Sperling F. 2003. Butterfly Molecular Systematics : From Species Definitions to Higher-Level Phylogenies. Evolution (N. Y). pp:1–22.

Sperling J.L.H., Sperling F.A.H. 2009. Lyme borreliosis in Canada: biological diversity and diagnostic complexity from an entomological perspective. Can. Entomol. 141:521–549.

Sullivan B.L., Aycrigg J.L., Barry J.H., Bonney R.E., Bruns N., Cooper C.B., Damoulas T., Dhondt A.A., Dietterich T., Farnsworth A., Fink D., Fitzpatrick J.W., Fredericks T., Gerbracht J., Gomes C., Hochachka W.M., Iliff M.J., Lagoze C., La Sorte F.A., Merrifield M., Morris W., Phillips T.B., Reynolds M., Rodewald A.D., Rosenberg K. V.,

Trautmann N.M., Wiggins A., Winkler D.W., Wong W.-K., Wood C.L., Yu J., Kelling S. 2014. The eBird enterprise: An integrated approach to development and application of citizen science. Biol. Conserv. 169:31–40.

Swofford D. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.

Thioulouse J., Dray S. 2007. Interactive multivariate data analysis in R with the ade4 and ade4TkGUI packages. J. Stat. Softw. 22.

Thompson J.D., Gibson T.J., Higgins D.G. 2002. Multiple sequence alignment using ClustalW and ClustalX. Curr Protoc Bioinforma. Chapter 2:Unit 2 3.

Toews D.P.L., Brelsford A., Irwin D.E. 2011. Hybridization between Townsend's *Dendroica townsendi* and black-throated green warblers *D. virens* in an avian suture zone. J. Avian Biol. 42:434–446.

Tofilski A. 2008. Using geometric morphometrics and standard morphometry to discriminate three honeybee subspecies. Apidologie. 39:1–6.

Van Valen L. 1976. Ecological species, multispecies, and oaks. Taxon. 25:233–239.

Vogler A.P., Desalle R. 1994. Diagnosing Units of Conservation Management. Conserv. Biol. 8:354–363.

Voss E.G. 1952. The history of keys and phylogenetic trees in systematic biology.

Wagner C.E., Keller I., Wittwer S., Selz O.M., Mwaiko S., Greuter L., Sivasundar A., Seehausen O. 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. Mol. Ecol. 22:787–98.

Wahlberg N., Brower A.V.Z., 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., Nylin S. 2003. Morphology versus molecules: resolution of the positions of *Nymphalis*, *Polygonia*, and related genera (Lepidoptera: Nymphalidae). Cladistics. 19:213–223.

Wahlberg N., Weingartner E., Nylin S. 2003. Towards a better understanding of the higher systematics of Nymphalidae (Lepidoptera: Papilionoidea). Mol. Phylogenet. Evol. 28:473–484.

Wahlberg N., Weingartner E., Warren A.D., 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. 2006. That awkward age for butterflies: insights from the age of the butterfly subfamily Nymphalinae (Lepidoptera: Nymphalidae). Syst. Biol. 55:703–14.

Walter D.E., Winterton S. 2007. Keys and the crisis in taxonomy: extinction or reinvention? Annu. Rev. Entomol. 52:193–208.

Ward R.D., Zemlak T.S., Innes B.H., Last P.R., Hebert P.D.N. 2005. DNA barcoding Australia's fish species. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 360:1847–57.

Weingartner E., Wahlberg N., Nylin S. 2006. Dynamics of host plant use and species diversity in *Polygonia* butterflies (Nymphalidae). J. Evol. Biol. 19:483–91.

Werren J.H., Baldo L., Clark M.E. 2008. Wolbachia: master manipulators of invertebrate biology. Nat. Rev. Microbiol. 6:741–51.

Werren J.H. 1997. Biology of *Wolbachia*. Annu. Rev. Entomol. 42:587–609.

Wheeler W.C. 2012. Systematics. Chichester, UK: John Wiley & Sons, Ltd.

Wiens J.J. 2006. Missing data and the design of phylogenetic analyses. J. Biomed. Inform. 39:34–42.

Wiens J.J. 2007. Species delimitation: new approaches for discovering diversity. Syst. Biol. 56:875–8.

Will K.W., Mishler B.D., Wheeler Q.D. 2005. The perils of DNA barcoding and the need for integrative taxonomy. Syst. Biol. 54:844–51.

Will K.W., Rubinoff D. 2004. Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. Cladistics. 20:47–55.

Wilson E.O., Brown W.L.J. 1953. The subspecies concept and its taxonomic application. Syst. Zool.:97–111.

Wilson R. a., Barker M.J., Brigandt I. 2007. When Traditional Essentialism Fails: Biological Natural Kinds. Philos. Top. 35:189–215.

Winston J.E. 1999. Describing Species: Practical Taxonomic Procedure for Biologists. Columbia University Press.

Yeates D.K., Seago A., Nelson L., Cameron S.L., Joseph L., Trueman J.W.H. 2011. Integrative taxonomy, or iterative taxonomy? Syst. Entomol. 36:209–217.

Zachos J., Pagani M., Sloan L., Thomas E., Billups K. 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. Science. 292:686–93.

Zhang D.-X., Hewitt G.M. 1996. Nuclear integrations: challenges for mitochondrial DNA markers. Trends Ecol. Evol. 11:247–251.

Zink R.M. 2004. The role of subspecies in obscuring avian biological diversity and misleading conservation policy. Proc. Biol. Sci. 271:561–4.

Zwickl D.J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Thesis. University of Texas at Austin.

APPENDIX A

SUPPLEMENTARY MATERIAL

Table A.1: (following 11 pages) Collection information of all specimens and sequences used in this study and a summary of specimen use across analyses. Sequences obtained from GenBank are shown on a grey background. MP = maximum parsimony analysis, ML = maximum likelihood analysis, Bayes = Bayesian analysis, STR = STRUCTURE, Hap. Net. = haplotype (or parsimony) network, DAPC = discriminant analysis on principal components, MCA = multiple correspondence analysis, DCA = discriminant correspondence analysis, PCA = principal components analysis, LDA = linear discriminant analysis. Specimen "am_9341" was not included in the maximum parsimony analysis due to constraints on the number of outgroup specimens in TNT, and as such is represented as "/+" in the MP/ML column.

					COI				SNPs				Visually scored			RGB	
Species	ID#	Accession #	Country	Locality	MP/ML	Bayes	Hap. Net.		MP/ML	STR.	DAPC		MCA	DCA		PCA	LDA
Polygonia faunus	pf_8301	--	CAN: Alberta	Bragg Creek	+	+	+		+	+	+		+	+		+	+
	pf_8302	--	CAN: Alberta	Bragg Creek	+	+	+		+	+	+		+	+		+	+
	pf_8303	--	CAN: Alberta	Bragg Creek	+	+	+		+	+	+		+	+		+	+
	pf_8304	--	CAN: Alberta	Lloyd Creek Nat. area	+	+	+		+	+	+		+	+		+	+
	pf_8306	--	CAN: Alberta	Edmonton, nr Emily Murphy Park	+	+	+		+	+	+		+	+		+	+
	pf_8307	--	CAN: Alberta	Edmonton, nr Emily Murphy Park	+	+	+		+	+	+		+	+		+	+
	pf_8308	--	CAN: Alberta	Aylmer Rec. Area (nr Nordegg)	+	+	+		+	+	+		+	+		+	+
	pf_8309	--	CAN: Alberta	Edmonton, N Sask. Rv Val, U of A	+	+	+		+	+	+		+	+		+	+
	pf_8310	--	CAN: Alberta	Shunda viewpoint (nr Nordegg)	+	+	+		+	+	+		+	+		+	+
	pf_8311	--	CAN: Alberta	Shunda viewpoint (nr Nordegg)	+	+	+		+	+	+		+	+		+	+
	pf_8312	--	CAN: Alberta	Shunda viewpoint (nr Nordegg)	+	+	+		+	+	+					+	+
	pf_8313	--	CAN: Alberta	Shunda viewpoint (nr Nordegg)	+	+	+		+	+	+					+	+
	pf_8314	--	CAN: Alberta	Shunda viewpoint (nr Nordegg)	+	+	+		+	+	+		+	+		+	+
	pf_8315	--	CAN: Alberta	North Ram River (nr Nordegg)	+	+	+		+	+	+						
	pf_8316	--	CAN: Alberta	Edmonton, N Sask. Rv Val, U of A	+	+	+		+	+	+		+	+		+	+
	pf_8317	--	CAN: Alberta	Pigeon Lake, Itaska	+	+	+		+	+	+		+	+		+	+
	pf_8318	--	CAN: Alberta	Pigeon Lake, Itaska	+	+	+		+	+	+					+	+
	pf_8320	--	CAN: Alberta	Lloyd Creek Nat. area	+				+	+	+		+	+		+	+
	pf_8321	--	CAN: Alberta	Lloyd Creek Nat. area	+	+	+		+	+	+		+	+		+	+
	pf_8322	--	CAN: Alberta	Lloyd Creek Nat. area	+	+	+		+	+	+					+	+
	pf_8323	--	CAN: Alberta	Lloyd Creek Nat. area	+	+	+		+	+	+		+	+		+	+
	pf_8324	--	CAN: Alberta	Lloyd Creek Nat. area	+				+	+	+		+	+		+	+
	pf_8325	--	CAN: Alberta	Medicine Lodge Hills	+				+	+	+		+	+		+	+
	pf_8326	--	CAN: Alberta	Buck Mt.	+	+	+		+	+	+		+	+		+	+
	pf_8327	--	CAN: Alberta	Pigeon Lake, Itaska	+	+	+		+	+	+		+	+		+	+
	pf_8328	--	CAN: Alberta	Whitecourt Mt.	+	+	+		+	+	+					+	+
	pf_8329	--	CAN: Alberta	Whitecourt Mt.	+				+	+	+		+	+		+	+
	pf_8331	--	CAN: Alberta	Whitecourt Mt.	+				+	+	+		+	+		+	+
	pf_8332	--	CAN: Alberta	Pigeon Lake, Itaska	+	+	+		+	+	+		+	+		+	+
	pf_8333	--	CAN: Alberta	Edmonton, Edith Ravine	+	+	+		+	+	+		+	+		+	+
	pf_8334	--	CAN: Alberta	Edmonton, Edith Ravine	+	+	+		+	+	+		+	+		+	+
	pf_8335	--	CAN: Alberta	Edmonton, Edith Ravine	+	+	+		+	+	+		+	+		+	+
	pf_8336	--	CAN: Alberta	Edmonton, Edith Ravine	+	+	+		+	+	+		+	+		+	+
	pf_8337	--	CAN: Alberta	North Ram River (nr Nordegg)	+	+	+		+	+	+		+	+		+	+
	pf_8338	--	CAN: Alberta	North Ram River (nr Nordegg)	+	+	+		+	+	+		+	+		+	+
	pf_8339	--	CAN: Alberta	North Ram River (nr Nordegg)	+	+	+		+	+	+		+	+		+	+

					COI				SNPs				Visually scored			RGB	
Species	ID#	Accession #	Country	Locality	MP/ML	Bayes	Hap. Net.		MP/ML	STR.	DAPC		MCA	DCA		PCA	LDA
Polygonia faunus	pf_8340	--	CAN: Alberta	North Ram River (nr Nordegg)	+				+	+	+		+	+		+	+
	pf_8383	--	CAN: Alberta	Lloyd Creek Nat. area	+	+			+	+	+		+	+		+	+
	pf_8386	--	CAN: Alberta	Lloyd Creek Nat. area	+	+	+		+	+	+		+	+		+	+
	pf_9250	--	CAN: British Columbia	E of Valemont	+				+	+	+		+	+		+	+
	pf_9251	--	CAN: British Columbia	E of Valemont	+	+	+		+	+	+		+	+		+	+
	pf_9252	--	CAN: British Columbia	E of Valemont	+	+	+		+	+	+		+	+		+	+
	pf_9253	--	CAN: British Columbia	E of Valemont	+	+	+		+	+	+		+	+		+	+
	pf_9254	--	CAN: British Columbia	N of Gambel Creek	+	+	+		+	+	+		+	+		+	+
	pf_9255	--	CAN: British Columbia	N of Gambel Creek	+	+	+		+	+	+		+	+		+	+
	pf_9256	--	CAN: British Columbia	N of Gambel Creek	+	+	+		+	+	+		+	+		+	+
	pf_9257	--	CAN: British Columbia	N of Gambel Creek	+	+	+		+	+	+		+	+		+	+
	pf_9259	--	CAN: Alberta	W of Rimbey	+	+	+		+	+	+		+	+		+	+
	pf_9260	--	CAN: Alberta	W of Rimbey	+	+	+		+	+	+		+	+		+	+
	pf_9261	--	CAN: Alberta	W of Rimbey	+	+	+		+	+	+		+	+		+	+
	pf_9262	--	CAN: Alberta	W of Rimbey	+	+	+		+	+	+		+	+		+	+
	pf_9263	--	CAN: Alberta	Devonian Bot. Garden	+	+	+		+	+	+		+	+		+	+
	pf_9264	--	CAN: Alberta	Devonian Bot. Garden	+	+	+		+	+	+		+	+		+	+
	pf_9265	--	CAN: Alberta	Chickakoo Lake	+	+	+		+	+	+		+	+		+	+
	pf_9266	--	CAN: Alberta	Chickakoo Lake	+	+	+		+	+	+		+	+		+	+
	pf_9267	--	CAN: Alberta	Opal Natural Area	+	+	+		+	+	+		+	+		+	+
	pf_9269	--	CAN: Alberta	Opal Natural Area	+	+	+		+	+	+		+	+		+	+
	pf_9270	--	CAN: Alberta	Open creek dam	+	+	+		+	+	+		+	+		+	+
	pf_9271	--	CAN: British Columbia	S of Tumbler Ridge	+	+	+		+	+	+		+	+		+	+
	pf_9273	--	CAN: British Columbia	S of Tumbler Ridge	+	+	+		+	+	+		+	+		+	+
	pf_9274	--	CAN: British Columbia	S of Tumbler Ridge	+	+	+		+	+	+		+	+		+	+
	pf_9275	--	CAN: Alberta	S of Hinton (Gregg River)	+	+	+		+	+	+		+	+		+	+
	pf_9276	--	CAN: Alberta	S of Hinton (Gregg River)	+	+	+		+	+	+		+	+		+	+
	pf_9277	--	CAN: Alberta	S of Hinton (Gregg River)	+	+	+		+	+	+		+	+		+	+
	pf_9278	--	CAN: Alberta	S of Hinton (Gregg River)	+	+	+		+	+	+		+	+		+	+
	pf_9279	--	CAN: British Columbia	S of Chetwynd	+	+	+		+	+	+		+	+		+	+
	pf_9280	--	CAN: British Columbia	S of Chetwynd	+	+	+		+	+	+		+	+		+	+
	pf_9281	--	CAN: British Columbia	S of Chetwynd	+	+	+		+	+	+		+	+		+	+
	pf_9282	--	CAN: British Columbia	S of Chetwynd	+	+	+		+	+	+		+	+		+	+
	pf_9283	--	CAN: Alberta	N of Lac la Biche (lake)	+	+	+		+	+	+		+	+		+	+
	pf_9284	--	CAN: Alberta	N of Lac la Biche (lake)	+	+	+		+	+	+		+	+		+	+
	pf_9285	--	CAN: Alberta	N of Lac la Biche (lake)	+	+	+		+	+	+		+	+		+	+

					COI				SNPs				Visually scored			RGB	
Species	ID#	Accession #	Country	Locality	MP/ML	Bayes	Hap. Net.		MP/ML	STR.	DAPC		MCA	DCA		PCA	LDA
Polygonia faunus	pf_9286	--	CAN: Alberta	N of Lac la Biche (lake)	+	+	+		+	+	+		+	+		+	+
	pf_9287	--	CAN: British Columbia	N. Kelowna	+	+	+		+	+	+		+	+		+	+
	pf_9289	--	USA: Montana	Flathead Lk Bio. Res.	+	+	+		+	+	+		+	+		+	+
	pf_9290	--	USA: Montana	NW of Blankenship brdge	+	+	+		+	+	+		+	+		+	+
	pf_9293	--	USA: Montana	NW of Blankenship brdge	+	+	+		+	+	+		+	+		+	+
	pf_9294	--	USA: Montana	Flathead Lk Bio. Res.	+	+	+		+	+	+		+	+		+	+
	pf_9295	--	USA: Montana	Flathead Lk Bio. Res.	+	+	+		+	+	+		+	+		+	+
	pf_9348	--	CAN: Alberta	Opal Natural Area	+	+	+		+	+	+		+	+		+	+
	pf_9349	--	CAN: Alberta	Chickakoo Lake	+	+	+		+	+	+		+	+		+	+
	pf_9350	--	CAN: Alberta	N of Fawcett HWY 44	+	+	+		+	+	+		+	+		+	+
	pf_9353	--	CAN: Alberta	Lloyd Creek Nat. area	+	+	+		+	+	+		+	+		+	+
	pf_9354	--	CAN: Alberta	Lloyd Creek Nat. area	+	+	+		+	+	+		+	+		+	+
	pf_9355	--	CAN: Alberta	Pigeon Lake, Itaska	+	+	+		+	+	+		+	+		+	+
	pf_9358	--	CAN: Alberta	W of Rimbey	+	+	+		+	+	+		+	+		+	+
	pf_9359	--	CAN: Alberta	S of Hinton (Gregg River)	+	+	+		+	+	+		+	+		+	+
	pf_9360	--	CAN: Alberta	S of Hinton (Gregg River)	+	+	+		+	+	+		+	+		+	+
	pf_9361	--	CAN: British Columbia	S of Chetwynd	+	+	+		+	+	+		+	+		+	+
	pf_9362	--	CAN: British Columbia	S of Chetwynd	+	+	+		+	+	+		+	+		+	+
	pf_9363	--	CAN: Alberta	N of Lac la Biche (lake)	+	+	+		+	+	+		+	+		+	+
	pf_9364	--	CAN: Alberta	N of Lac la Biche (lake)	+	+	+		+	+	+		+	+		+	+
	pf_9365	--	CAN: Alberta	Buck Mt.	+	+	+		+	+	+		+	+		+	+
	pf_9366	--	CAN: Alberta	NW of Delburne, RR 251	+	+	+		+	+	+		+	+		+	+
	pf_9367	--	CAN: British Columbia	N of Gambel Creek	+	+	+		+	+	+		+	+		+	+
	pf_9368	--	CAN: British Columbia	N of Gambel Creek	+	+	+		+	+	+		+	+		+	+
Polygonia gracilis zephyrus	pz_8388	--	CAN: Alberta	Shunda viewpoint	+	+	+		+	+	+		+	+		+	+
	pz_8389	--	CAN: Alberta	Shunda viewpoint	+	+	+		+	+	+					+	+
	pz_9297	--	CAN: British Columbia	S of Chetwynd	+	+	+		+	+	+		+	+		+	+
	pz_9300	--	CAN: British Columbia	E of Valemont					+	+	+		+	+		+	+
	pz_9301	--	CAN: British Columbia	E of Valemont	+	+	+		+	+	+		+	+		+	+
	pz_9302	--	CAN: British Columbia	N of Gambel Creek	+	+	+		+	+	+		+	+		+	+
	pz_9303	--	CAN: British Columbia	N of Gambel Creek	+	+	+		+	+	+		+	+		+	+
	pz_9304	--	CAN: British Columbia	N of Gambel Creek	+	+	+		+	+	+		+	+		+	+
	pz_9306	--	CAN: Alberta	S of Hinton (Gregg River)	+	+	+		+	+	+		+	+		+	+
	pz_9307	--	CAN: Alberta	S of Hinton (Gregg River)	+	+	+		+	+	+		+	+		+	+
	pz_9308	--	CAN: Alberta	S of Hinton (Gregg River)	+	+	+		+	+	+		+	+		+	+
	pz_9352	--	CAN: British Columbia	Kettle Valley	+	+	+		+	+	+					+	+

Species	ID#	Accession #	Country	Locality	MP/ML	Bayes	Hap. Net.		MP/ML	STR.	DAPC		MCA	DCA		PCA	LDA
Polygonia gracilis gracilis	pg_8382	--	CAN: Alberta	Nordegg	+	+	+		+	+	+		+	+		+	+
	pg_8390	--	CAN: Alberta	Shunda viewpoint	+	+	+		+	+	+		+	+		+	+
	pg_8392	--	CAN: Alberta	Ram Lookout	+	+	+		+	+	+		+	+		+	+
	pg_9296	--	CAN: British Columbia	S of Tumbler Ridge	+	+	+		+	+	+		+	+		+	+
	pg_9298	--	CAN: Alberta	N of Lac la Biche (lake)	+				+	+	+		+	+		+	+
	pg_9305	--	CAN: British Columbia	N of Gambel Creek	+	+	+		+	+	+		+	+		+	+
	pg_9309	--	CAN: Alberta	S of Hinton (Gregg River)	+	+	+		+	+	+		+	+		+	+
Polygonia progne	pp_8319	--	CAN: Alberta	Buck Mt.	+				+	+	+		+	+		+	+
	pp_8341	--	CAN: Alberta	Bragg Creek	+				+	+	+		+	+		+	+
	pp_8342	--	CAN: Alberta	Bragg Creek	+				+	+	+		+	+		+	+
	pp_8343	--	CAN: Alberta	Bragg Creek	+	+	+		+	+	+		+	+		+	+
	pp_8344	--	CAN: Alberta	Bragg Creek	+	+	+		+	+	+		+	+		+	+
	pp_8345	--	CAN: Alberta	Bragg Creek	+	+	+		+	+	+		+	+		+	+
	pp_8346	--	CAN: Alberta	Bragg Creek	+	+	+		+	+	+		+	+		+	+
	pp_8347	--	CAN: Alberta	Bragg Creek	+	+	+		+	+	+		+	+		+	+
	pp_8348	--	CAN: Alberta	Bragg Creek	+	+	+		+	+	+		+	+		+	+
	pp_8349	--	CAN: Alberta	Bragg Creek	+	+	+		+	+	+		+	+		+	+
	pp_8350	--	CAN: Alberta	Bragg Creek	+	+	+		+	+	+		+	+		+	+
	pp_8351	--	CAN: Alberta	Bragg Creek	+	+	+		+	+	+		+	+		+	+
	pp_8352	--	CAN: Alberta	Bragg Creek	+	+	+		+	+	+		+	+		+	+
	pp_8353	--	CAN: Alberta	Bragg Creek	+	+	+		+	+	+		+	+		+	+
	pp_8354	--	CAN: Alberta	Bragg Creek	+	+	+		+	+	+		+	+		+	+
	pp_8355	--	CAN: Alberta	Bragg Creek	+	+	+		+	+	+		+	+		+	+
	pp_8356	--	CAN: Alberta	Bragg Creek	+	+	+		+	+	+		+	+		+	+
	pp_8357	--	CAN: Alberta	Bragg Creek	+	+	+		+	+	+		+	+		+	+
	pp_8359	--	CAN: Alberta	Bragg Creek	+	+	+		+	+	+		+	+		+	+
	pp_8360	--	CAN: Alberta	Bragg Creek	+	+	+		+	+	+		+	+		+	+
	pp_8361	--	CAN: Alberta	Bragg Creek	+	+	+		+	+	+		+	+		+	+
	pp_8362	--	CAN: Alberta	Bragg Creek	+	+	+		+	+	+					+	+
	pp_8363	--	CAN: Alberta	Pigeon Lake, Itaska	+	+	+		+	+	+		+	+		+	+
	pp_8364	--	CAN: Alberta	Lloyd Creek Nat. area	+	+	+		+	+	+		+	+		+	+
	pp_8365	--	CAN: Alberta	Lloyd Creek Nat. area	+	+	+		+	+	+		+	+		+	+
	pp_8366	--	CAN: Alberta	Lloyd Creek Nat. area	+	+	+		+	+	+		+	+		+	+
	pp_8367	--	CAN: Alberta	Buck Mt.	+	+	+		+	+	+		+	+		+	+
	pp_8368	--	CAN: Alberta	Peace Pk. N Wetaskewin	+	+	+		+	+	+		+	+		+	+
	pp_8369	--	CAN: Alberta	Peace Pk. N Wetaskewin	+	+	+		+	+	+		+	+		+	+

					COI				SNPs				Visually scored			RGB	
Species	ID#	Accession #	Country	Locality	MP/ML	Bayes	Hap. Net.		MP/ML	STR.	DAPC		MCA	DCA		PCA	LDA
Polygonia progne	pp_8370	--	CAN: Alberta	Hand Hills	+	+	+		+	+	+		+	+		+	+
	pp_8371	--	CAN: Alberta	Buck Mt.	+	+	+		+	+	+					+	+
	pp_8372	--	CAN: Alberta	Buck Mt.	+	+	+		+	+	+		+	+		+	+
	pp_8373	--	CAN: Alberta	Lloyd Creek Nat. area	+	+	+		+	+	+		+	+		+	+
	pp_8374	--	CAN: Alberta	Hand Hills	+	+	+		+	+	+					+	+
	pp_8376	--	CAN: Alberta	Lloyd Creek Nat. area	+	+	+		+	+	+		+	+		+	+
	pp_8377	--	CAN: Alberta	Buck Mt.	+	+	+		+	+	+		+	+		+	+
	pp_8378	--	CAN: Alberta	Hand Hills	+	+	+		+	+	+		+	+		+	+
	pp_8379	--	CAN: Alberta	Medicine Lodge Hills	+	+	+		+	+	+		+	+		+	+
	pp_8380	--	CAN: Alberta	Edith Ravine	+	+	+		+	+	+		+	+		+	+
	pp_8381	--	CAN: Alberta	Bragg Creek	+	+	+		+	+	+		+	+		+	+
	pp_8384	--	CAN: Alberta	Lloyd Creek Nat. area	+	+	+		+	+	+					+	+
	pp_8387	--	CAN: Alberta	Lloyd Creek Nat. area	+	+	+		+	+	+		+	+		+	+
	pp_8391	--	CAN: Alberta	Edith Ravine	+	+	+		+	+	+		+	+		+	+
	pp_9312	--	CAN: Alberta	Edmonton, Windsor Pk.	+	+	+		+	+	+		+	+		+	+
	pp_9313	--	CAN: Alberta	Edmonton, Windsor Pk.	+	+	+		+	+	+		+	+		+	+
	pp_9315	--	CAN: Alberta	Lloyd Creek Nat. area	+	+	+		+	+	+		+	+		+	+
	pp_9316	--	CAN: Alberta	Pigeon Lake, Itaska	+	+	+		+	+	+		+	+		+	+
	pp_9317	--	CAN: Alberta	W. of Stauffer	+	+	+		+	+	+		+	+		+	+
	pp_9318	--	CAN: Alberta	W of Rimbey	+	+	+		+	+	+		+	+		+	+
	pp_9319	--	CAN: Alberta	W of Rimbey	+	+	+		+	+	+					+	+
	pp_9321	--	CAN: Alberta	Edmonton	+	+	+		+	+	+		+	+		+	+
	pp_9322	--	CAN: Alberta	NE side of Peace River	+	+	+		+	+	+		+	+		+	+
	pp_9323	--	CAN: Alberta	Kaufman Hill	+	+	+		+	+	+		+	+		+	+
	pp_9324	--	CAN: Alberta	nr Shaftsbury Ferry	+	+	+		+	+	+		+	+		+	+
	pp_9325	--	CAN: Alberta	Opal Natural Area	+	+	+		+	+	+		+	+		+	+
	pp_9326	--	CAN: Alberta	Hasse Lake	+	+	+		+	+	+		+	+		+	+
	pp_9327	--	CAN: Alberta	W of Rimbey	+	+	+		+	+	+		+	+		+	+
	pp_9328	--	CAN: Alberta	Muir Lake	+	+	+		+	+	+		+	+		+	+
	pp_9330	--	CAN: Alberta	Hasse Lake	+	+	+		+	+	+		+	+		+	+
	pp_9331	--	CAN: Alberta	Pigeon Lake, Itaska	+	+	+		+	+	+		+	+		+	+
	pp_9332	--	CAN: Alberta	N of Lac la Biche (lake)	+	+	+		+	+	+						
	pp_9333	--	CAN: Alberta	N of Lac la Biche (lake)	+	+	+		+	+	+		+	+		+	+
	pp_9334	--	CAN: Alberta	N of Lac la Biche (lake)	+	+	+		+	+	+		+	+		+	+
	pp_9335	--	CAN: Alberta	N of Lac la Biche (lake)	+	+	+		+	+	+		+	+		+	+
	pp_9336	--	CAN: Alberta	Buck Mt.	+	+	+		+	+	+					+	+

					COI				SNPs				Visually scored			RGB	
Species	ID#	Accession #	Country	Locality	MP/ML	Bayes	Hap. Net.		MP/ML	STR.	DAPC		MCA	DCA		PCA	LDA
Polygonia progne	pp_9337	--	CAN: Alberta	NW of Delburne, RR 251	+	+	+		+	+	+		+	+		+	+
	pp_9345	--	CAN: Alberta	N of Lac la Biche (lake)	+	+	+		+	+	+		+	+		+	+
	pp_9346	--	CAN: Alberta	N of Lac la Biche (lake)	+	+	+		+	+	+		+	+		+	+
	pp_9347	--	CAN: Alberta	N of Lac la Biche (lake)	+	+	+		+	+	+		+	+		+	+
Polygonia satyrus	ps_8358	--	CAN: Alberta	Pigeon Lake, Itaska	+	+	+		+	+	+		+	+		+	+
	ps_8375	--	CAN: Alberta	Bragg Creek	+	+	+		+	+	+					+	+
	ps_8385	--	CAN: Alberta	Pigeon Lake, Itaska	+	+	+		+	+	+					+	+
	ps_8393	--	CAN: Alberta	Bragg Creek	+	+	+		+	+	+		+	+		+	+
	ps_8394	--	CAN: Alberta	Waterton area	+	+	+		+	+	+		+	+		+	+
	ps_8395	--	CAN: Alberta	Ram Lookout	+	+	+		+	+	+		+	+		+	+
	ps_8396	--	CAN: Alberta	Edith ravine	+	+	+		+	+	+		+	+		+	+
	ps_8397	--	CAN: Alberta	Rowley	+	+	+		+	+	+		+	+		+	+
	ps_9201	--	CAN: Alberta	Bragg Creek	+	+	+		+	+	+					+	+
	ps_9203	--	CAN: Alberta	Edmonton, Glenora	+	+	+		+	+	+		+	+		+	+
	ps_9204	--	CAN: Alberta	Edmonton, Glenora	+	+	+		+	+	+		+	+		+	+
	ps_9205	--	CAN: Alberta	Ram Lookout	+	+	+		+	+	+		+	+		+	+
	ps_9206	--	CAN: Alberta	Ram Lookout	+	+	+		+	+	+		+	+		+	+
	ps_9207	--	CAN: Alberta	Ram Lookout	+	+	+		+	+	+					+	+
	ps_9208	--	CAN: Alberta	Ram Lookout	+	+	+		+	+	+		+	+		+	+
	ps_9209	--	CAN: Alberta	Waterton area	+	+	+		+	+	+		+	+		+	+
	ps_9210	--	CAN: Alberta	Waterton area	+	+	+		+	+	+		+	+		+	+
	ps_9211	--	CAN: Alberta	Whitecourt Mt.	+	+	+		+	+	+		+	+		+	+
	ps_9212	--	CAN: Alberta	Whitecourt Mt.	+	+	+		+	+	+		+	+		+	+
	ps_9213	--	CAN: Alberta	Lloyd Creek Nat. area	+	+	+		+	+	+		+	+		+	+
	ps_9214	--	CAN: British Columbia	E of Valemont	+	+	+		+	+	+		+	+		+	+
	ps_9215	--	CAN: British Columbia	Oliver	+	+	+		+	+	+					+	+
	ps_9216	--	CAN: Alberta	Lloyd Creek Nat. area	+	+	+		+	+	+		+	+		+	+
	ps_9217	--	CAN: Alberta	W of Rimbey	+	+	+		+	+	+					+	+
	ps_9218	--	CAN: Alberta	W of Rimbey	+	+	+		+	+	+		+	+		+	+
	ps_9219	--	CAN: Alberta	W of Rimbey	+	+	+		+	+	+		+	+		+	+
	ps_9220	--	CAN: Alberta	Lloyd Creek Nat. area	+	+	+		+	+	+		+	+		+	+
	ps_9221	--	CAN: Alberta	Opal Natural Area	+	+	+		+	+	+		+	+		+	+
	ps_9223	--	CAN: Alberta	Opal Natural Area	+	+	+		+	+	+		+	+		+	+
	ps_9224	--	CAN: Alberta	S of Dev. Bot. Garden	+	+	+		+	+	+		+	+		+	+
	ps_9225	--	CAN: Alberta	S of Dev. Bot. Garden	+	+	+		+	+	+		+	+		+	+
	ps_9226	--	CAN: Alberta	S of Dev. Bot. Garden	+	+	+		+	+	+		+	+		+	+

					COI				SNPs				Visually scored			RGB	
Species	ID#	Accession #	Country	Locality	MP/ML	Bayes	Hap. Net.		MP/ML	STR.	DAPC		MCA	DCA		PCA	LDA
Polygonia satyrus	ps_9227	--	CAN: Alberta	Brazeau Co, RR 72	+	+	+		+	+	+		+	+		+	+
	ps_9228	--	CAN: Alberta	Hasse Lake	+	+	+		+	+	+		+	+		+	+
	ps_9229	--	CAN: Alberta	Hasse Lake	+	+	+		+	+	+		+	+		+	+
	ps_9230	--	CAN: Alberta	Hasse Lake	+	+	+		+	+	+		+	+		+	+
	ps_9231	--	CAN: Alberta	Hasse Lake	+	+	+		+	+	+					+	+
	ps_9232	--	CAN: British Columbia	S of Tumbler Ridge	+	+	+		+	+	+		+	+		+	+
	ps_9233	--	CAN: British Columbia	S of Tumbler Ridge	+	+	+		+	+	+		+	+		+	+
	ps_9234	--	CAN: British Columbia	S of Tumbler Ridge	+	+	+		+	+	+		+	+		+	+
	ps_9235	--	CAN: British Columbia	S of Tumbler Ridge	+	+	+		+	+	+		+	+		+	+
	ps_9236	--	CAN: Alberta	S of Hinton (Gregg River)	+	+	+		+	+	+		+	+		+	+
	ps_9237	--	CAN: Alberta	S of Hinton (Gregg River)	+	+	+		+	+	+		+	+		+	+
	ps_9239	--	CAN: Alberta	E. Cougar Creek camp	+	+	+		+	+	+		+	+		+	+
	ps_9240	--	CAN: British Columbia	S of Chetwynd	+	+	+		+	+	+		+	+		+	+
	ps_9241	--	CAN: British Columbia	S of Chetwynd	+	+	+		+	+	+		+	+		+	+
	ps_9242	--	CAN: British Columbia	S of Chetwynd	+	+	+		+	+	+		+	+		+	+
	ps_9243	--	CAN: British Columbia	S of Chetwynd	+	+	+		+	+	+		+	+		+	+
	ps_9244	--	CAN: Alberta	N of Lac la Biche (lake)	+	+	+		+	+	+		+	+		+	+
	ps_9245	--	CAN: Alberta	N of Lac la Biche (lake)	+	+	+		+	+	+		+	+		+	+
	ps_9246	--	CAN: Alberta	N of Lac la Biche (lake)	+	+	+		+	+	+		+	+		+	+
	ps_9247	--	CAN: Alberta	N of Lac la Biche (lake)	+	+	+		+	+	+		+	+		+	+
	ps_9248	--	USA: Montana	Flathead Lk Bio. Res.	+	+	+		+	+	+		+	+		+	+
	ps_9249	--	USA: Montana	NW of Blankenship brdge	+	+	+		+	+	+		+	+		+	+
	ps_9272	--	CAN: British Columbia	S of Tumbler Ridge	+	+			+	+	+		+	+		+	+
Polygonia comma	pc_9310	--	USA: New York	Ithaca, Mundey Flwr Grdn	+	+			+								
Polygonia interrogationis	pi_9311	--	USA: New York	Ithaca, Monkey Run	+	+			+								
Aglais milberti	am_9341	--	CAN: Alberta	Ram Trail	+	+			/+								
	am_9342	--	CAN: Alberta	Moose Mt. nr Bragg Ck	+	+			+								
Nymphalis antiopa	na_9343	--	CAN: Alberta	Hasse Lake	+	+			+								
	na_9344	--	CAN: Alberta	Lloyd Creek Nat. area	+	+			+								
Nymphalis californica	nc_9340	--	USA: Oregon	Santiam Pass HWY 20	+	+			+								
Nymphalis vaualbum	nv_9338	--	CAN: Alberta	W of Rimbey	+	+			+								
	nv_9339	--	CAN: Alberta	N of Lac la Biche (lake)	+	+			+								

					COI				SNPs				Visually scored			RGB	
Species	ID#	Accession #	Country	Locality	MP/ML	Bayes	Hap. Net.		MP/ML	STR.	DAPC		MCA	DCA		PCA	LDA
Aglais io	--	EF683674.1	--	--		+											
Aglais urticae	--	EF683676.1	--	--		+											
	--	HM999894.1	--	--		+											
	NW63-3	AY248786.1	SWED: Stockholm Co.	Stockholm		+											
Kaniska canace	EW19-10	FJ639396.1	JAP: Okinawa	Mt. Urabu, Yonaguni is.		+											
	NW164-1	FJ639397.1	LAOS: Phongsaly	Mt. Phoufa		+											
Nymphalis polychloros	NW62-2	AY248788.1	SWED: Öland	Tävelsrum		+											
Nymphalis xanthomelas	NW84-1	AY248790.1	RUS: Sakha	Ust Nera, Indigirka Rv, Yakutia		+											
Polygonia c-album	EW47-5	JN093252.1	--	--		+											
	EW47-6	JN093253.1	--	--		+											
	EW47-7	JN093254.1	--	--		+											
	EW47-8	JN093255.1	--	--		+											
	EW47-9	JN093256.1	--	--		+											
Polygonia c-aureum	EW13-12	FJ639402.1	RUS: Amur Oblast	Blagoveschensk, Amur River		+											
	EW13-13	FJ639403.1	RUS: Amur Oblast	Blagoveschensk, Amur River		+											
	NW65-8	AY248799.1	JAP:	--		+											
Polygonia comma	EW21-10	FJ639404.1	USA: North Carolina	Great Dismal swamp		+											
	EW21-11	FJ639405.1	USA: North Carolina	Great Dismal swamp		+											
	NW65-6	AY248794.1	USA: Tennesee	Shelby Co.		+											
Polygonia egea	NW120-7	FJ639406.1	IRAN: Yazd	Shir Kuh, Deh Bala		+											
	NW77-15	AY248800.1	GREC:	--		+											
Polygonia faunus	EW12-7	JX134831	CAN: British Columbia	Santa Rosa road nr Rossland			+										
	EW18-13	JX134773	USA: Washington	Silver Star Mtn.			+										
	EW21-12	JX134760	CAN: Quebec	Reserve faunique de Laurentides			+										
	EW22-2	JX134756	CAN: Quebec	Reserve faunique de Assinica			+										
	EW22-3	JX134759	USA: New Hampshire	Scott's Bog (Pittsfield)			+										
	EW22-4	JX134758	USA: New Hampshire	Scott's Bog (Pittsfield)			+										
	EW22-5	JX134785	USA: New Hampshire	Scott's Bog (Pittsfield)			+										
	EW23-17	JX134794	CAN: Alberta	South of Ft. McMurray			+										
	EW23-9	JX134786	CAN: Alberta	Opal			+										
	EW29-8	JX134757	USA: California	Dorrington			+										
	EW39-11	JX134798	USA: California	Sugar Pine			+										
	EW39-13	JX134803	USA: Colorado	Roaring River			+										
	EW39-15	JX134811	USA: Utah	Red Cloud Loop			+										
	EW40-10	JX134806	USA: Utah	Red Cloud Loop			+										
	EW40-11	JX134835	USA: Utah	Red Cloud Loop			+										

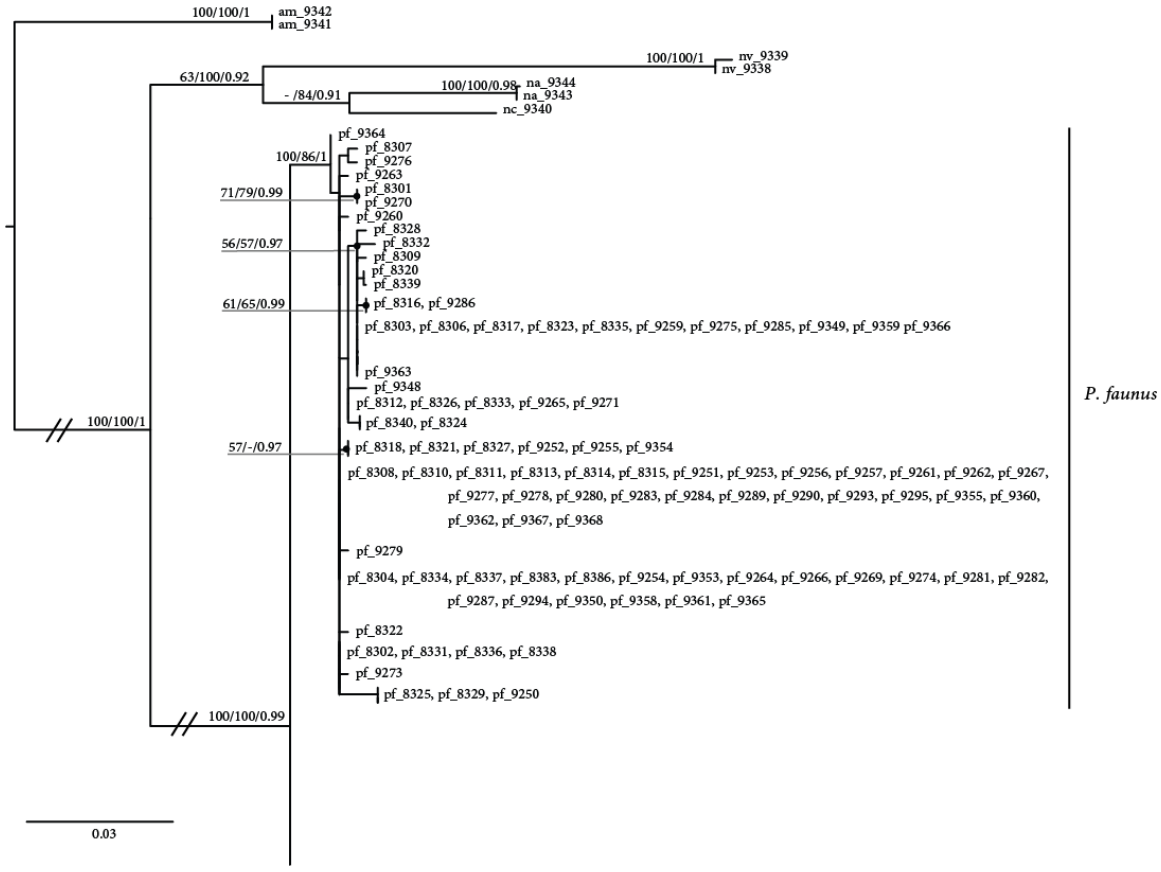
					COI				SNPs				Visually scored			RGB	
Species	ID#	Accession #	Country	Locality	MP/ML	Bayes	Hap. Net.		MP/ML	STR.	DAPC		MCA	DCA		PCA	LDA
Polygonia faunus	EW40-13	JX134807	USA: Utah	Red Cloud Loop			+										
	EW40-14	JX134838	USA: Utah	Red Cloud Loop			+										
	EW40-18	JX134836	USA: Colorado	Crested Butte			+										
	EW40-19	JX134860	USA: Colorado	Roaring River			+										
	EW40-23	JX134837	USA: Colorado	Turquoise Lake			+										
	EW40-4	JX134804	USA: Montana	Granite Co.			+										
	EW40-8	JX134805	USA: California	Sugar Pine			+										
	EW40-9	JX134834	USA: California	Sugar Pine			+										
	EW41-13	JX134884	USA: Utah	Red Cloud Loop			+										
	EW41-6	JX134840	USA: Colorado	Roaring River			+										
	EW41-7	JX134849	USA: Colorado	Crested Butte			+										
	EW42-16	JX134821	USA: Colorado	Rampart Range			+										
	EW42-3	JX134883	USA: Montana	Trail to Stuart Peaks			+										
	EW42-9	JX134830	USA: Montana	Trail to Stuart Peaks			+										
	EW43-15	JX134847	USA: Washington	Blue Mountains			+										
	EW43-19	JX134858	USA: Washington	Blue Mountains			+										
	EW43-5	JX134856	USA: Montana	Crazy mts., Half moon Camp area			+										
	EW43-6	JX134853	USA: Montana	Crazy mts., Half moon Camp area			+										
	EW48-13	JX134888	USA: Idaho	Idaho co.			+										
	EW50-7	JX134891	USA: Georgia	Cooper Creek			+										
Polygonia g-argenteum	NW165-2	FJ639412	MEX: Oaxaca	Mpio. Sta. Ma. Papaio		+											
Polygonia gracilis zephyrus	EW10-10	FJ639457	USA: Wyoming	Big Piney, Forest Road 10046			+										
	EW10-11	FJ639458	USA: Wyoming	Big Piney, Forest Road 10046			+										
	EW10-12	FJ639459	USA: Wyoming	Big Piney, Forest Road 10046			+										
	EW10-13	FJ639460	USA: Wyoming	Big Piney, Forest Road 10046			+										
	EW10-14	FJ639461	USA: Wyoming	Big Piney, Forest Road 10046			+										
	EW10-9	FJ639462	USA: Wyoming	Big Piney, Forest Road 10046			+										
	EW11-4	FJ639463	USA: Wyoming	Big Piney, Forest Road 10046			+										
	EW11-9	FJ639464	USA: Oregon	Jackson Co. Mt. Ashland Rd			+										
	EW14-10	FJ639465	CAN: British Columbia	Skagit Valley, SE of Hope			+										
	EW14-11	FJ639466	CAN: British Columbia	Skagit Valley, SE of Hope			+										
	EW14-12	FJ639467	CAN: British Columbia	Goatfel FSR nr Yahk		+	+										
	EW14-14	FJ639468	CAN: British Columbia	Goatfel FSR nr Yahk		+	+										
	EW14-15	FJ639469	CAN: British Columbia	Goatfel FSR nr Yahk		+	+										
	EW15-1	FJ639470	CAN: British Columbia	Skagit Valley, SE of Hope		+	+										
	EW15-15	FJ639471	CAN: British Columbia	Santa Rosa rd nr Rossland			+										

					COI				SNPs				Visually scored			RGB	
Species	ID#	Accession #	Country	Locality	MP/ML	Bayes	Hap. Net.		MP/ML	STR.	DAPC		MCA	DCA		PCA	LDA
Polygonia gracilis zephyrus	EW15-4	FJ639473	USA: Washington	Silver Star Mtn.			+										
	NW63-11	FJ639474	CAN: British Columbia	Vale Mountain (Valemont?)			+										
	NW74-5	FJ639475	USA: Wyoming	Big Piney, Forest Road 10046			+										
	NW74-6	AY248797	USA: Wyoming	Big Piney, Forest Road 10046		+	+										
Polygonia gracilis gracilis	EW19-7	FJ639456	USA: Colorado	San Juan Co, Cascade Creek			+										
	EW21-14	FJ639414.1	CAN: Quebec	Assinica		+											
	EW22-8	FJ639415.1	USA: New Hampshire	Scott's Bog (Pittsfield)		+											
	EW22-9	FJ639416	CAN: Quebec	Reserve faunique de Laurentides		+	+										
Polygonia gigantea	NW166-5	FJ639413.1	CHIN: Sichuan	Lushan		+											
Polygonia haroldi	NW165-1	FJ639417.1	MEX: Guanajuato	Santa Rosa		+											
Polygonia interposita	NW166-6	FJ639418.1	KYR: Chuy	Zailiysky Mts. Chon-Kemin R.		+											
Polygonia interrogationis	EW22-12	FJ639419.1	USA: Virginia	--		+											
	EW22-14	FJ639420.1	USA: New Hampshire	Scott's Bog (Pittsfield)		+											
	NW77-12	AY248793.1	USA: Tennesee	Fayette Co.		+											
Polygonia oreas	EW14-7	FJ639421.1	CAN: British Columbia	Skagit Valley, SE of Hope		+											
	EW14-8	FJ639422.1	CAN: British Columbia	Skagit Valley, SE of Hope		+											
	EW14-9	FJ639423.1	CAN: British Columbia	Skagit Valley, SE of Hope		+											
	NW74-11	FJ639424.1	USA: Oregon	Benton Co.		+											
Polygonia progne	EW19-9	FJ639425	CAN: Alberta	Edmonton, McTaggart Sanctuary			+										
	EW21-16	FJ639476	CAN: Quebec	Reserve faunique de Ashuapmushuan			+										
	EW21-2	FJ639426	USA: New Hampshire	Scott's Bog (Pittsfield)			+										
	EW21-3	AY248795	USA: Virginia	THR (?)			+										
	EW21-4	FJ639427	CAN: Quebec	Reserve faunique de Ashuapmushuan			+										
	EW22-15	FJ639428	USA: New Hampshire	Scott's Bog (Pittsfield)			+										
	EW22-16	FJ639429	USA: New Hampshire	Scott's Bog (Pittsfield)			+										
	EW22-17	FJ639430	CAN: Quebec	Reserve faunique de Ashuapmushuan			+										
	EW22-18	FJ639431	CAN: Quebec	Reserve faunique de Assinica			+										
	EW22-19	FJ639432	CAN: Quebec	Reserve faunique de Assinica			+										
	EW22-20	FJ639433	USA: New Hampshire	Scott's Bog (Pittsfield)			+										
	EW22-21	FJ639434	USA: New Hampshire	Scott's Bog (Pittsfield)			+										
Polygonia satyrus	EW11-10	FJ639436	USA: Idaho	Bonneville Co. Little Elk Creek			+										
	EW11-11	FJ639437	USA: Oregon	Benton Co., MacDonald Forest			+										
	EW11-12	FJ639438	USA: Oregon	Benton Co., MacDonald Forest			+										
	EW15-11	FJ639440	CAN: British Columbia	Pend-d'Oreille FSR, SE of Trail			+										
	EW15-13	FJ639442	CAN: British Columbia	Pend-d'Oreille FSR, SE of Trail			+										
	EW15-14	FJ639443	CAN: British Columbia	Santa Rosa rd nr Rossland			+										

					COI				SNPs				Visually scored			RGB	
Species	ID#	Accession #	Country	Locality	MP/ML	Bayes	Hap. Net.		MP/ML	STR.	DAPC		MCA	DCA		PCA	LDA
Polygonia satyrus	EW15-16	FJ639444	CAN: British Columbia	Santa Rosa rd nr Rossland			+										
	EW15-9	FJ639445	CAN: British Columbia	Skagit Valley, SE of Hope			+										
	EW16-6	FJ639446	CAN: British Columbia	Skagit Valley, SE of Hope			+										
	EW16-7	FJ639447	CAN: British Columbia	Skagit Valley, SE of Hope			+										
	EW23-13	FJ639448	CAN: Alberta	South of Ft. McMurray			+										
	EW23-14	FJ639449	CAN: Alberta	Ft. McKay			+										
	NW65-9	FJ639450	USA: Washington	Blue Mountains			+										
	NW74-8	FJ639451	USA: Oregon	Benton Co.			+										
	NW74-9	AY248796	USA: Oregon	Benton Co.			+										
Polygonia undina	EW30-1	FJ639452.1	TADJ:	Tadjikalad Ganishou		+											
	EW33-24	FJ639453.1	UZB: Kashkadarya	Gissar Mts.		+											
	EW43-1	FJ639454.1	KYR: Dzhalal-Abadskaya	Chapchyma Pass		+											
	EW43-2	FJ639455.1	KYR: Dzhalal-Abadskaya	Chapchyma Pass		+											

CAN Canada
CHIN China
GREC Greece
IRAN Iran
KYR Kyrgyzstan
LAOS Laos
MEX Mexico
RUS Russia
SWE Sweden
TADJ Tadjikistan
USA United States of America
UZB Uzbekistan

Figure A.1: (Following two pages) Maximum likelihood topology of 1450 bp of COI constructed in PhyML. Maximum parsimony bootstrap, maximum likelihood bootstraps and aBayes support values shown in clades present in more than one analysis.


100/100/1
am_9342
am_9341
100/100/1
nv_9339
nv_9338
63/100/0.92
100/100/0.98
na_9344
na_9343
- /84/0.91
nc_9340
pf_9364
100/86/1
pf_8307
pf_9276
pf_9263
71/79/0.99
pf_8301
pf_9270
pf_9260
pf_8328
56/57/0.97
pf_8332
pf_8309
pf_8320
pf_8339
61/65/0.99
pf_8316, pf_9286
pf_8303, pf_8306, pf_8317, pf_8323, pf_8335, pf_9259, pf_9275, pf_9285, pf_9349, pf_9359 pf_9366
pf_9363
pf_9348
pf_8312, pf_8326, pf_8333, pf_9265, pf_9271
100/100/1
pf_8340, pf_8324
57/-/0.97
pf_8318, pf_8321, pf_8327, pf_9252, pf_9255, pf_9354
pf_8308, pf_8310, pf_8311, pf_8313, pf_8314, pf_8315, pf_9251, pf_9253, pf_9256, pf_9257, pf_9261, pf_9262, pf_9267, pf_9277, pf_9278, pf_9280, pf_9283, pf_9284, pf_9289, pf_9290, pf_9293, pf_9295, pf_9355, pf_9360, pf_9362, pf_9367, pf_9368
pf_9279
pf_8304, pf_8334, pf_8337, pf_8383, pf_8386, pf_9254, pf_9353, pf_9264, pf_9266, pf_9269, pf_9274, pf_9281, pf_9282, pf_9287, pf_9294, pf_9350, pf_9358, pf_9361, pf_9365
pf_8322
pf_8302, pf_8331, pf_8336, pf_8338
pf_9273
pf_8325, pf_8329, pf_9250
100/100/0.99
0.03
P. faunus

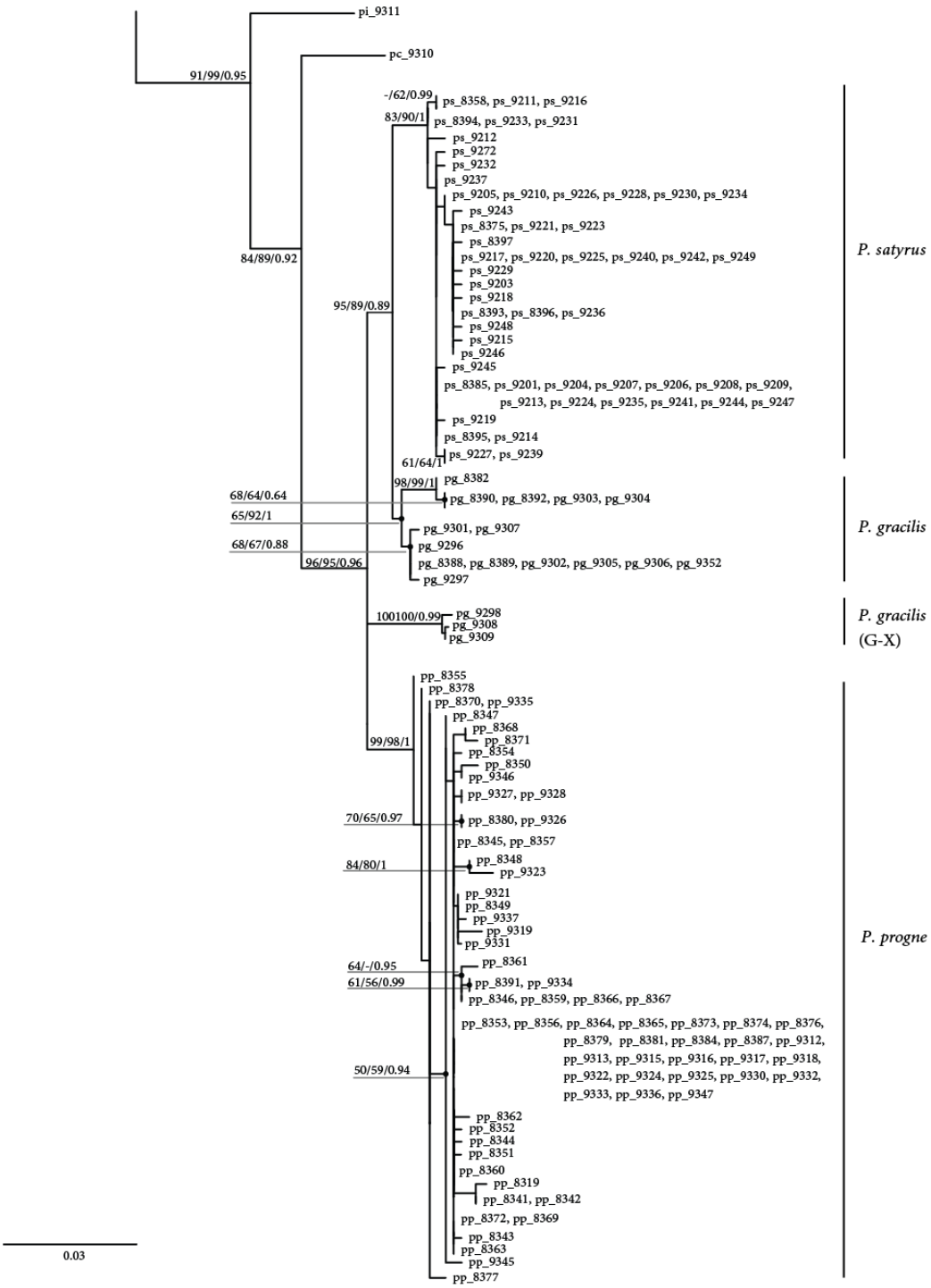
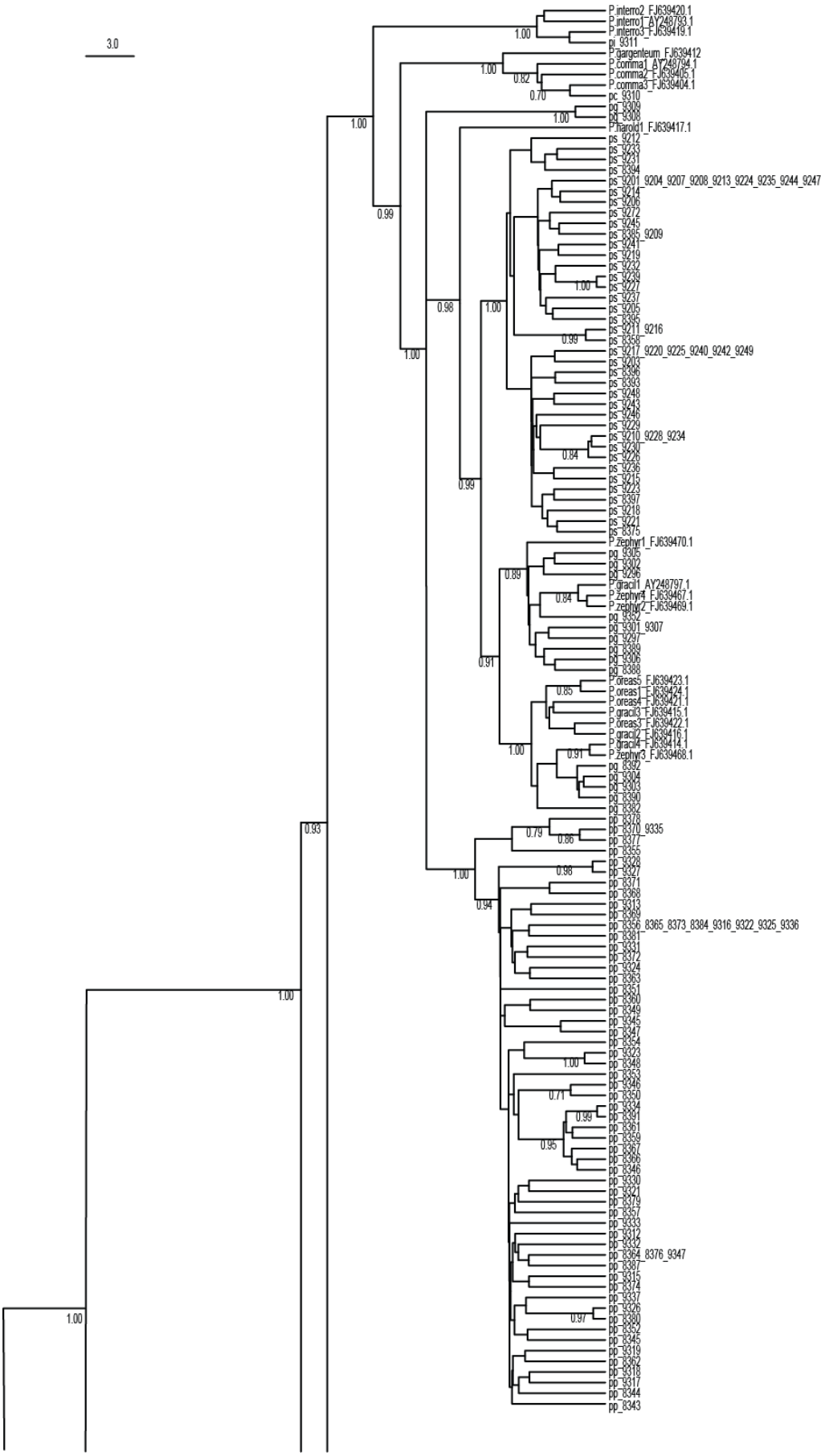
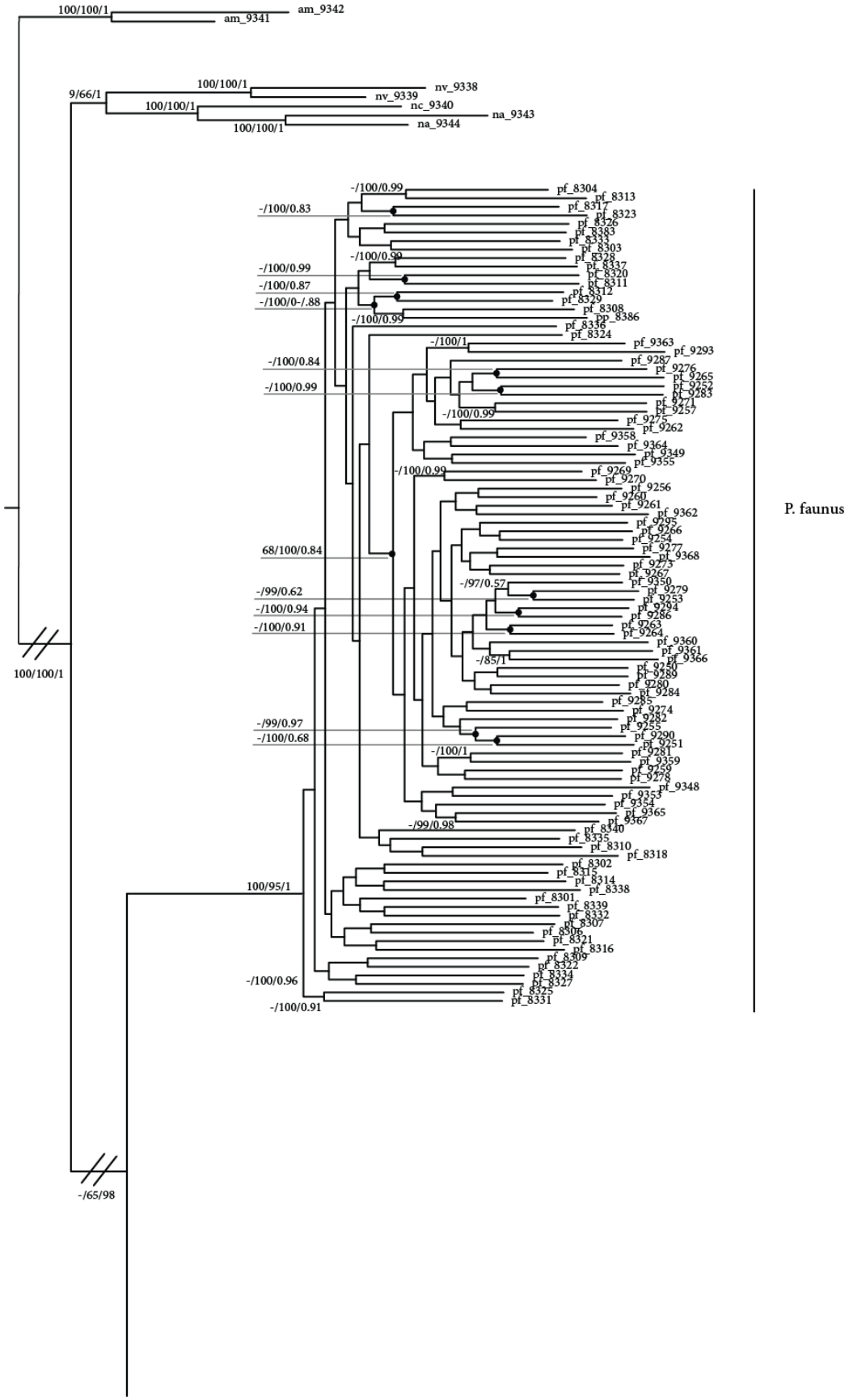

pi_9311
pc_9310
91/99/0.95
-/62/0.99
ps_8358, ps_9211, ps_9216
83/90/1
ps_8394, ps_9233, ps_9231
ps_9212
ps_9272
ps_9232
ps_9237
ps_9205, ps_9210, ps_9226, ps_9228, ps_9230, ps_9234
ps_9243
ps_8375, ps_9221, ps_9223
ps_8397
84/89/0.92
ps_9217, ps_9220, ps_9225, ps_9240, ps_9242, ps_9249
ps_9229
ps_9203
ps_9218
95/89/0.89
ps_8393, ps_8396, ps_9236
ps_9248
ps_9215
ps_9246
ps_9245
ps_8385, ps_9201, ps_9204, ps_9207, ps_9206, ps_9208, ps_9209,
ps_9213, ps_9224, ps_9235, ps_9241, ps_9244, ps_9247
ps_9219
ps_8395, ps_9214
61/64/1
ps_9227, ps_9239
P. satyrus
98/99/1
pg_8382
68/64/0.64
pg_8390, pg_8392, pg_9303, pg_9304
65/92/1
pg_9301, pg_9307
68/67/0.88
pg_9296
96/95/0.96
pg_8388, pg_8389, pg_9302, pg_9305, pg_9306, pg_9352
pg_9297
P. gracilis
100100/0.99
pg_9298
pg_9308
pg_9309
P. gracilis
(G-X)
pp_8355
pp_8378
pp_8370, pp_9335
pp_8347
pp_8368
99/98/1
pp_8371
pp_8354
pp_8350
pp_9346
pp_9327, pp_9328
70/65/0.97
pp_8380, pp_9326
pp_8345, pp_8357
pp_8348
84/80/1
pp_9323
pp_9321
pp_8349
pp_9337
pp_9319
pp_9331
P. progne
pp_8361
64/-/0.95
61/56/0.99
pp_8391, pp_9334
pp_8346, pp_8359, pp_8366, pp_8367
pp_8353, pp_8356, pp_8364, pp_8365, pp_8373, pp_8374, pp_8376,
pp_8379, pp_8381, pp_8384, pp_8387, pp_9312,
pp_9313, pp_9315, pp_9316, pp_9317, pp_9318,
pp_9322, pp_9324, pp_9325, pp_9330, pp_9332,
pp_9333, pp_9336, pp_9347
50/59/0.94
pp_8362
pp_8352
pp_8344
pp_8351
pp_8360
pp_8319
pp_8341, pp_8342
pp_8372, pp_8369
pp_8343
pp_8363
pp_9345
pp_8377
0.03

Figure A.2: (Following two pages) Bayesian analysis of 1450 bp of COI performed in BEAST 2. Posterior probabilities greater than 70% are shown to the left of the node. Specimens were supplemented with GenBank sequences; accession numbers available in Table A.1.


3.0
P.interro2_FJ639420.1
P.interro1_AY248793.1
P.interro3_FJ639419.1
pi_9311
P.argenteum_FJ639412
P.comma1_AY248794.1
P.comma2_FJ639405.1
P.comma3_FJ639404.1
pc_9310
pg_9309
pg_9308
P.harold1_FJ639417.1
ps_9212
ps_9233
ps_9232
ps_9231
ps_8394
ps_9201_9204_9207_9208_9213_9224_9235_9244_9247
ps_9214
ps_9206
ps_9272
ps_9245
ps_8385_9209
ps_9241
ps_9219
ps_9202
ps_9239
ps_9227
ps_9237
ps_9205
ps_8395
ps_9211_9216
ps_8358
ps_9217_9220_9225_9240_9242_9249
ps_9203
ps_8396
ps_8393
ps_9248
ps_9243
ps_9246
ps_9229
ps_9210_9228_9234
ps_9230
ps_9226
ps_9236
ps_9215
ps_9222
ps_8397
ps_9218
ps_9221
ps_8375
P.zephyr1_FJ639470.1
pg_9305
pg_9302
pg_9306
pg_9296
P.graci1_AY248797.1
P.zephyr4_FJ639467.1
P.zephyr2_FJ639469.1
pg_9352
pg_9301_9307
pg_9297
pg_8389
pg_8386
pg_8388
P.oreas5_FJ639423.1
P.oreas1_FJ639424.1
P.oreas4_FJ639421.1
P.graci3_FJ639415.1
P.oreas3_FJ639422.1
P.graci2_FJ639416.1
P.graci4_FJ639414.1
P.zephyr3_FJ639468.1
pg_8392
pg_9304
pg_9303
pg_8390
pg_8382
pp_8378
pp_8370_9335
pp_8377
pp_8355
pp_9228
pp_9227
pp_8371
pp_8368
pp_9313
pp_8369
pp_8356_8365_8373_8384_9316_9322_9325_9336
pp_8381
pp_9331
pp_8372
pp_9324
pp_8363
pp_8351
pp_8360
pp_8349
pp_9345
pp_8347
pp_8354
pp_9223
pp_8348
pp_8353
pp_8346
pp_8350
pp_9334
pp_8391
pp_8361
pp_8359
pp_8367
pp_8366
pp_8346
pp_9330
pp_9321
pp_8379
pp_8357
pp_9333
pp_9312
pp_9332
pp_8364_8376_9347
pp_8387
pp_9315
pp_8374
pp_9337
pp_9326
pp_8380
pp_8352
pp_8345
pp_9319
pp_8362
pp_8385
pp_9318
pp_9317
pp_8344
pp_8343

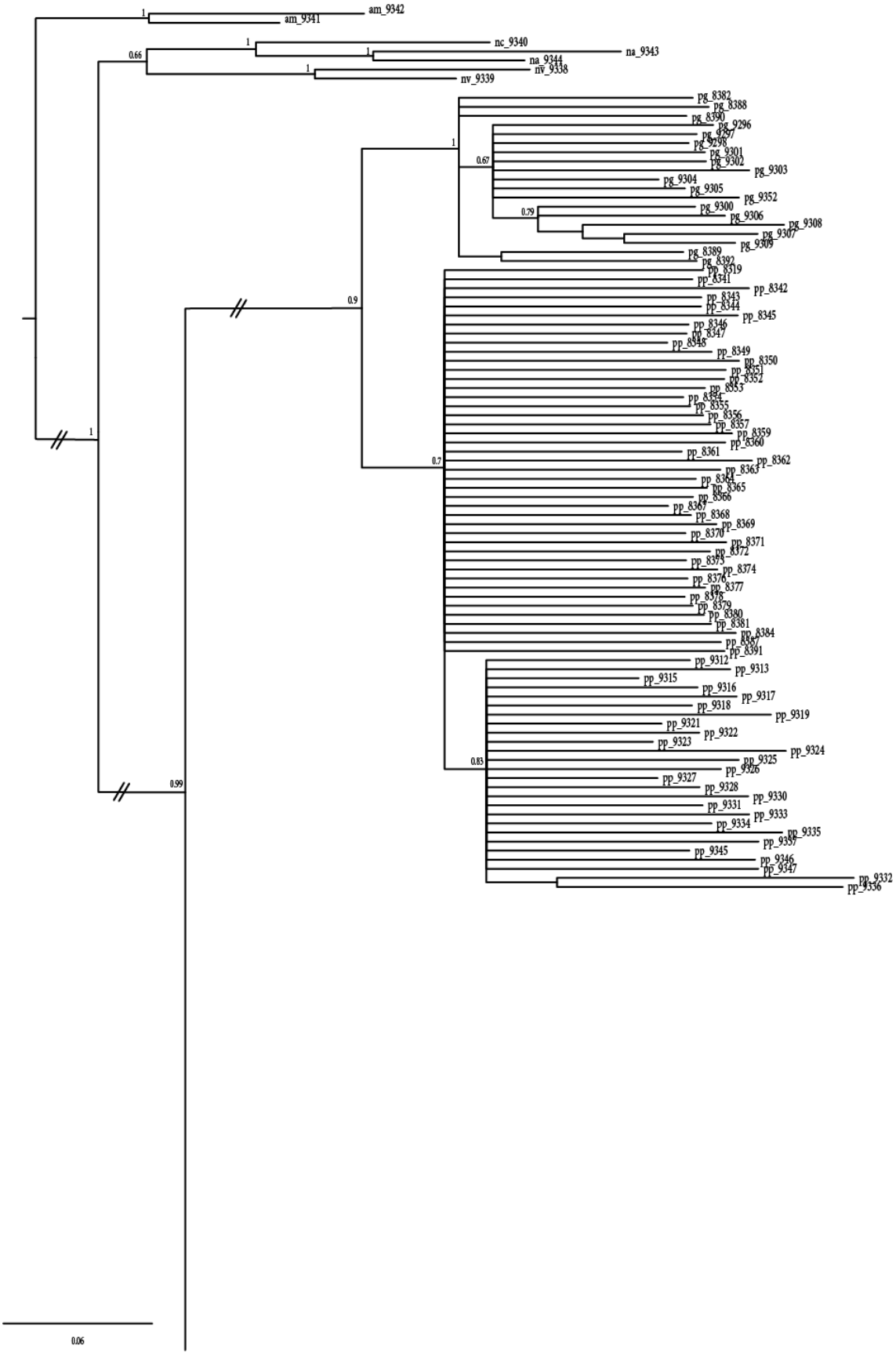

P.egea_FJ639406.1
P.egea_AY248800.1
P.undin4_FJ639452.1
P.undin3_FJ639453.1
P.undin2_FJ639454.1
P.undin1_FJ639455.1
P.calbu5_JN093252.1
P.calbu3_JN093254.1
P.interposita_FJ639418.1
P.calbu4_JN093253.1
P.calbu2_JN093255.1
P.calbu1_JN093256.1
pf_8312_9265
pf_9271
pf_8333
pf_8326
pf_9348
pf_9285
pf_8332
pf_9266
pf_8316
pf_9359
pf_8303_8306_9259
pf_8335
pf_8328
pf_9353
pf_9289
pf_8338
pf_8306
pf_8317_8323_9275_9349
pf_8309
pf_8308_8310_8313_8314_8315_9251_9253_9256_9257_9261_9267_9277_
9280_9284_9289_9290_9293_9295_9355_9360_9368
pf_9282
pf_9365
pf_9362_9367
pf_9268
pf_9274
pf_9354
pf_8321_9252_9255
pf_8327
pf_8318
pf_8358
pf_9294
pf_8337
pf_9263
pf_9281
pf_8307
pf_9254_9350_9353
pf_8322
pf_9296
pf_8334
pf_8336
pf_8311
pf_9260
pf_9269
pf_9267
pf_9279
pf_9276
pf_9264
pf_9266
pf_9361
pf_8303
pf_9278
pf_9262
pf_8338
pf_8304
pf_9263
pf_8334
pf_9273
pf_8302
pf_9270
pf_8301
P.caur1_AY248799.1
P.caur4_FJ639402.1
P.caur2_FJ639403.1
P.gigantea_FJ639413.1
K.canace2_FJ639396.1
K.canace1_FJ639397.1
N.polychloros_AY248788.1
na_9344
na_9343
N.xanthomelas_AY248790.1
nc_9340
nv_9339
nv_9338
A.io_EF683674.1
A.urticae_AY248786.1
A.urtica2_EF683676.1
A.urtica1_HM999894.1
am_9342
am_9341
1.00
0.99
1.00
0.94
0.81
1.00
0.82
1.00
0.99
1.00
0.98
0.99
1.00
0.99
0.83
0.99
1.00
1.00
1.00
3.0

Figure A.3: (Following two pages) Maximum likelihood phylogeny of 13,254 concatenated SNPs (6627 loci in two-column format) constructed in PhyML. Maximum parsimony bootstraps, maximum likelihood bootstraps and aBayes support values shown in clades present in more than one analysis.


100/100/1
am_9342
am_9341
9/66/1
100/100/1
nv_9338
nv_9339
100/100/1
nc_9340
na_9343
100/100/1
na_9344
-/100/0.99
pf_8304
pf_8313
-/100/0.83
pf_8317
pf_8323
pf_8326
pf_8383
pf_8331
pf_8303
-/100/0.99
pf_8328
-/100/0.99
pf_8337
pf_8320
-/100/0.87
pf_8311
pf_8312
-/100/0-/.88
pf_8329
pf_8308
pp_8386
-/100/0.99
pf_8336
pf_8324
-/100/1
pf_9363
pf_9293
-/100/0.84
pf_9287
pf_9276
pf_9265
pf_9252
-/100/0.99
pf_9283
pf_9271
pf_9257
-/100/0.99
pf_9275
pf_9262
pf_9358
pf_9364
pf_9349
pf_9355
-/100/0.99
pf_9269
pf_9270
pf_9256
pf_9260
pf_9261
pf_9362
pf_9295
pf_9266
pf_9254
68/100/0.84
pf_9277
pf_9368
pf_9273
pf_9267
-/97/0.57
pf_9350
-/99/0.62
pf_9279
pf_9253
-/100/0.94
pf_9294
pf_9286
-/100/0.91
pf_9263
pf_9264
pf_9360
pf_9361
-/85/1
pf_9366
pf_9250
pf_9289
pf_9280
pf_9284
pf_9285
pf_9274
-/99/0.97
pf_9282
pf_9255
-/100/0.68
pf_9290
pf_9251
-/100/1
pf_9281
pf_9359
pf_9259
pf_9278
pf_9348
pf_9353
pf_9354
pf_9365
pf_9367
-/99/0.98
pf_8340
pf_8335
pf_8310
pf_8318
pf_8302
pf_8315
pf_8314
100/95/1
pf_8338
pf_8301
pf_8339
pf_8332
pf_8307
pf_8306
pf_8321
pf_8316
pf_8309
pf_8322
pf_8334
-/100/0.96
pf_8327
pf_8325
pf_8331
-/100/0.91
100/100/1
-/65/98
P. faunus


P. progne
P. gracilis
P. progne
P. satyrus
0.08

Figure A.4: (Following two pages) Maximum likelihood bootstrap consensus tree of 13,254 concatenated SNPs (6627 loci in two-column format) constructed in GARLI. Bootstrap values greater than 70% are shown to the left of the nodes.


am_9342
am_9341
nc_9340
na_9343
na_9344
nv_9338
nv_9339
pg_8382
pg_8388
pg_8390
pg_9296
pg_9298
pg_9301
pg_9302
pg_9303
pg_9304
pg_9305
pg_9352
pg_9300
pg_9306
pg_9308
pg_9307
pg_9309
pg_8389
pg_8392
pp_8319
pp_8341
pp_8342
pp_8343
pp_8344
pp_8345
pp_8346
pp_8347
pp_8348
pp_8349
pp_8350
pp_8351
pp_8352
pp_8353
pp_8354
pp_8355
pp_8356
pp_8357
pp_8359
pp_8360
pp_8361
pp_8362
pp_8363
pp_8364
pp_8365
pp_8366
pp_8367
pp_8368
pp_8369
pp_8370
pp_8371
pp_8372
pp_8373
pp_8374
pp_8376
pp_8377
pp_8378
pp_8379
pp_8380
pp_8381
pp_8384
pp_8387
pp_8391
pp_9312
pp_9313
pp_9315
pp_9316
pp_9317
pp_9318
pp_9319
pp_9321
pp_9322
pp_9323
pp_9324
pp_9325
pp_9326
pp_9327
pp_9328
pp_9330
pp_9331
pp_9333
pp_9334
pp_9335
pp_9339
pp_9345
pp_9346
pp_9347
pp_9332
pp_9336
1
0.66
1
1
1
1
0.67
0.79
0.9
1
0.7
0.83
0.99
0.06

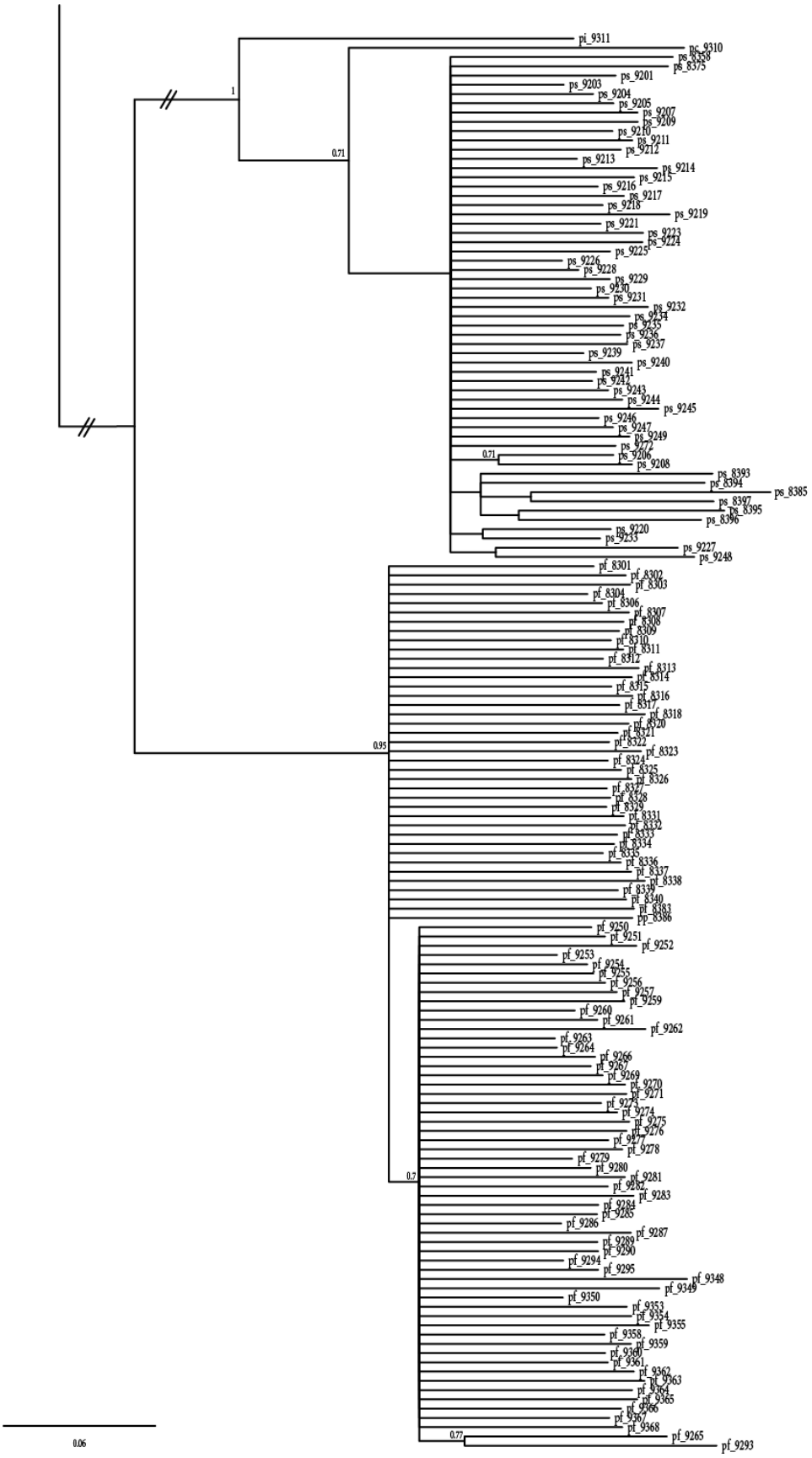

pi_9311
ps_9310
ps_8375
ps_9201
ps_9203
ps_9204
ps_9205
ps_9207
ps_9209
ps_9210
ps_9211
ps_9212
ps_9213
ps_9214
ps_9215
ps_9216
ps_9217
ps_9218
ps_9219
ps_9221
ps_9223
ps_9224
ps_9225
ps_9226
ps_9228
ps_9229
ps_9230
ps_9231
ps_9232
ps_9234
pf_9235
ps_9236
ps_9237
ps_9239
ps_9240
ps_9241
ps_9242
ps_9243
ps_9244
ps_9245
ps_9246
ps_9247
ps_9249
ps_9272
ps_9206
ps_9208
ps_8393
ps_8394
ps_8385
ps_8397
ps_8395
ps_8396
ps_9220
ps_9233
ps_9227
ps_9248
pf_8301
pf_8302
pf_8303
pf_8304
pf_8306
pf_8307
pf_8308
pf_8309
pf_8310
pf_8311
pf_8312
pf_8313
pf_8314
pf_8315
pf_8316
pf_8317
pf_8318
pf_8320
pf_8321
pf_8322
pf_8323
pf_8324
pf_8325
pf_8326
pf_8327
pf_8328
pf_8329
pf_8331
pf_8332
pf_8333
pf_8334
pf_8335
pf_8336
pf_8337
pf_8338
pf_8339
pf_8340
pf_8383
pp_8386
pf_9250
pf_9251
pf_9252
pf_9253
pf_9254
pf_9255
pf_9256
pf_9257
pf_9259
pf_9260
pf_9261
pf_9262
pf_9263
pf_9264
pf_9266
pf_9267
pf_9269
pf_9270
pf_9271
pf_9273
pf_9274
pf_9275
pf_9276
pf_9277
pf_9278
pf_9279
pf_9280
pf_9281
pf_9282
pf_9283
pf_9284
pf_9285
pf_9286
pf_9287
pf_9289
pf_9290
pf_9294
pf_9295
pf_9348
pf_9349
pf_9350
pf_9353
pf_9354
pf_9355
pf_9358
pf_9359
pf_9360
pf_9361
pf_9362
pf_9363
pf_9364
pf_9365
pf_9366
pf_9367
pf_9368
pf_9265
pf_9293
1
0.71
0.71
0.95
0.7
0.77
0.06

Figure A.5: Examples of smeared (**A.**) and contrasted (**B.**) morphs of *Polygonia faunus* and *P. satyrus* used in this study.

BIOGRAPHY

I was born on January 19, 1988 in Halifax, Nova Scotia, the second child to Donald and Susan McDonald. I don't remember much of Halifax, since we moved away when I was 3 years old to a small town outside of Kingston, Ontario: Gananoque. My love of nature was born on the banks of the St. Lawrence River. My sister and I would catch frogs, collect "sea shells" (more like pond snail shells) and investigate the thicket behind our house every minute we could. When I was 8 years old, we moved across the country to St. Albert, Alberta. My fascination with nature picked up where we left off, just at a new body of water – Big Lake. Being the queen of creepy-crawlies (at least in my own mind), I leapt at the opportunity study invertebrates at the University of Alberta during my undergraduate degree from 2006 – 2011 (B.Sc/ B.Ed combined degree). My plan was to become the greatest high school biology teacher that ever was, but I wanted to learn everything about invertebrates along the way.

One day, my Zoology 351 class was going on a collection trip, and a familiar face from my childhood hopped aboard the bus – John Acorn, the Nature Nut himself. My journey to entomology had officially begun. John encouraged me to take another entomology course towards the end of my degree; it was another collection course, which I loved. This course was taught by a certain Dr. Felix Sperling – I had no idea at that time that this course would shape my future the way it has.

After my undergraduate degree, I tried my hand at teaching in the public school system. I hated it - marking, lesson planning, the other teachers – I hated everything about being a teacher except teaching. After sharing my woeful, newfound lack of direction with Felix, he casually suggested that there was space in his lab for another graduate student, and that I should apply. In September 2012, I started my MSc in the Sperling lab, co-supervised by John Acorn.

I didn't let a little thing like graduate studies stop me from living my life, so I decided that I should buy a house, adopt a cat, get married, and start a full-time job in the Department of Biological Sciences – all during my MSc. I can see how that may have spread me a little thin, but I wouldn't do it any differently if I could.

Grad school is the hardest thing I've done in my life so far, but definitely the most rewarding. The skills I have learned under the supervision of Felix and John can be used in all areas of my personal and professional life – even PCR troubleshooting.