Molecular systematics of the greater fritillary butterflies (Nymphalidae: *Speyeria*): reduced representation sequencing, phylogeny, and incipient speciation

by

Erin O'dell Campbell

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

Doctor of Philosophy
in
Systematics and Evolution

Department of Biological Sciences
University of Alberta

Abstract

Species delimitation and phylogenetic reconstruction are essential for understanding and categorizing lineage divergences, and ultimately species. Yet, because of the complexities underlying speciation, inferring these patterns and characterizing their sources has been a consistent challenge for systematists. Recent advances in genomics have increased our understanding of how speciation may be influenced by ecology, biology, geography, and demography. However, the expansion of methods for producing genomic data warrants continued assessment to ensure they produce robust systematic inferences. Here, I conceptually and empirically assess the technical compatibility of some of the most commonly-used reduced representation sequencing (RRS) approaches for producing genomic single nucleotide polymorphism (SNP) data for systematic and population genomic inference in non-model taxa. I then apply these RRS approaches to re-assess the molecular systematic relationships of species in the butterfly genus *Speyeria* Scudder, 1872. These butterflies have a long history of taxonomic instability that is largely due to morphological variability and insufficient taxon sampling in previous systematic studies, which has impeded the assessment of species limits despite growing conservation concern for several taxa.

In a literature review of studies that describe new RRS techniques, I found that several published methods are only subtly different from one another, and often have convergent technical features despite being published with distinct names. I then empirically assessed the compatibility of two commonly used RRS techniques using several *Speyeria* species. Despite large intial differences between methods in sequencing depth and coverage, the genotyping consistency between methods was high and improved the recovery of monophyletic clades of *Speyeria*. In an expanded taxonomic and geographic phylogenetic study of *Speyeria*, I recovered

extensive mito-nuclear discordance between the SNP-based species phylogeny and the mitochondrial *COI* gene phylogeny, and recovered genomic divergences in two species that likely warrant taxonomic revision. Finally, I re-assessed the species limits in the *S. atlantis-hesperis* species complex, which has become infamous for its morphological variability. My results uphold *S. atlantis* (Edwards, 1862) as a species distinct from *S. hesperis* (Edwards, 1864), and additionally indicate that *S. hesperis* should be taxonomically recognized as two species.

This work presents the most geographically comprehensive taxonomic sampling of *Speyeria* to date, and the first use of genomic SNPs for elucidating the systematics of these enigmatic butterflies. These studies build conceptually on methodological advances to investigate their impact on species detection, and adds to a growing body of research that has demonstrated the utility of genomic SNPs for clarifying long-standing evolutionary problems.

Preface

This thesis represents original work by Erin O Campbell, however it would not have been possible without the collaboration and support of several other people. As such, I use the term "we" to signify their contributions throughout.

A version of Chapter 2 has been published as Campbell EO, Brunet BTM, Dupuis JR, and Sperling FAH (2018) "Would an RRS by any other name sound as RAD?", Methods in Ecology and Evolution volume 9, pages 1920-1927. EOC and BTMB conducted the meta-analysis and literature review, and drafted the manuscript. All authors contributed to the study design and to manuscript draft revisions.

A version of Chapter 3 has been published as Campbell EO, Davis CS, Dupuis JR, Muirhead K, and Sperling FAH (2017) "Cross-platform compatibility of *de novo*-aligned SNPs in a non-model butterfly genus", Molecular Ecology Resources volume 17, pages e84-e93. EOC conducted data analysis, and drafted the manuscript. KM provided custom bioinformatic scripts. EOC, CSD, JRD, and FAHS contributed to the study design and to manuscript draft revisions.

A version of Chapter 4 has been published as Campbell EO, Gage EV, Gage RV, and Sperling FAH (2019) "Single nucleotide polymorphism-based species phylogeny of greater fritillary butterflies (Lepidoptera: Nymphalidae: *Speyeria*) demonstrates widespread mito-nuclear discordance", Systematic Entomology, DOI:10.1111/syen.12393. EOC, EVG, RVG, and FAHS contributed to specimen collection. EOC conceived the study design, with input from EVG, RVG, and FAHS, and EOC conducted analyses, and drafted the manuscript. FAHS provided conceptual guidance. All others contributed to manuscript draft revisions.

Acknowledgements

First, thank you Felix, for taking a chance on a student with an English degree and no prior entomological, wet lab, or bioinformatic experience. Your leap of faith and mentorship made this thesis possible. Thank you also to Heather Proctor and Paul Stothard for your support and input as committee members, to Yan Boucher and Jocelyn Hall for participating in my candidacy examination, and to Toby Spribille and Niklas Wahlberg for acting as my arms-length and external defense examiners.

This work was facilitated by so many people - thanks to Lien Luong for your early support and for helping me through my first experience with the academic publishing process. Thanks to Sophie Dang and Cheryl Nargang from the MBSU, and Brian Boyle from IBIS for their sequencing support, and to the many avid *Speyeria* enthusiasts I've met along the way that have contributed specimens or insight to this work: Thomas Simonsen, one of the first friendly faces I met at LepSoc, Chris Schmidt, Ted Pike, Bob Parsons, Dave Lawrie, Gary Anweiler, Norbert Kondla, and John Acorn. And of course, a huge thanks to Ed and Randy Gage - your cross-country butterfly-hunting expeditions over the past 5 years have been an amazing contribution and commitment to this work, and the reason this project was able to grow from a regionally-focused M.Sc into a Ph.D.

Thanks to Corey Davis for your ongoing advice and support - our chats helped me to navigate the deep waters of molecular biology and NGS. Thanks also to the many fellow students and lab members I've had the privilege of crossing paths with over the years - Giovanny Fagua, Christianne McDonald, Stephen Trevoy, Gwylim Blackburn, Fabien Condamine, Kurt Yakimovich, Emily Dong, Marnie Wright, Zac MacDonald, Federico Riva, Kyle Snape, Phil Batista, Lisa Lumley, Caroline Whitehouse, Rowan French, Janet Sperling, Tiago Simões, Bryan

Brunet, Jasmine Janes, and Oksana Vernygora. And to the “core four” and the “elite” - Tyler, Victor, Brittany, Ronald, and Julian - your friendship and support over the years has meant so much.

Enormous thanks are also due to my family, who have put up with my incessant talking about work, manuscript writing on Christmas Eve, and the co-opting of family vacations to collect butterflies: Mom, Dad, Sheila, Grandma, Kevin, Danny, Hunter, William, Daniel, Kirby, Guy, and Luisa - I won the jackpot with all of you. Mom: thank you for your unwavering support and your cowboy attitude towards fieldwork and international travel. Kirby and Rusty, thank you for braving ornery cattle and murderous grouse with me. And to the triplets, a veritable gang of ant-stompers and moth-catchers - thank you for reminding me of the magic of “bug science” seen through the eyes of a child (or three).

And James, none of this would have been possible without you. Your unconditional love and support during grad school and over the past fourteen years has brought so much to my life that words just don’t do it justice.

Funding for this research was provided by an Alberta Conservation Association Grant in Biodiversity to myself, and by a Natural Sciences and Engineering Research Council of Canada Discovery Grant to Felix Sperling.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

COI - cytochrome oxidase 1 gene

ddRAD/ddRADseq - double digest restriction site-associated DNA sequencing

ESU - evolutionarily significant unit

GBS - genotyping-by-sequencing

ILS - incomplete lineage sorting

MAF - minor allele frequency

MSC - multispecies coalescent

NGS - next generation sequencing

RAPD - random amplified polymorphic DNA

RRS - reduced representation sequencing

SBG - sequence-based genotyping

SNP - single nucleotide polymorphism

List of Appendices

Chapter 1

Introduction and thesis objectives

1.1 General introduction

Phylogenetic reconstruction and species taxonomy are the cornerstones of systematics (Weins 2007), but the complexities of lineage divergence and the proliferation of various species concepts have consistently complicated these endeavors (Mayden 1997; Sites & Marshall 2001; Weins 2007). Early species concepts generally sought to categorize biological entities using discrete criteria, such as clear morphological "types" or the assessment of reproductive isolation (Mayr 1942; Ruse 1969; and summarized in Mayden 1997 and de Queiroz 2007). However, as the field of systematics has progressed, it has become increasingly apparent that the traditional perspective of species as discrete, dichotomously branching lineages is overly simplified (Mallet 1995; Mayden 1997; Sites & Marshall 2003; de Queiroz 1998, 2005, 2007). Instead, it is now commonly acknowledged that the trajectory of speciation is highly influenced through time by interactions with other species, the environment, and geography, and can result in conflicting, often reticulate, patterns of divergence among data (Sousa & Hey 2013; Galtier 2018).

Recent methodological and analytical advances, particularly in molecular systematics and population genomics, have made it increasingly feasible to disentangle these evolutionary patterns, and has rapidly advanced our understanding of diversification and speciation (Degnan & Rosenberg 2006; Knowles & Carstens 2007; Hohenlohe *et al.* 2012; Sousa & Hey 2013). This conceptual shift amidst technological innovation demonstrates how our changing perceptions of species are inextricably tied to the methods that we use to detect and characterize lineage divergences. However, the expansion of new methods has also created a landscape of technical jargon and lab protocols that has made it difficult to distinguish methods from one another and

which may have serious implications for data compatibility, the introduction of methodological biases, and ultimately, the reliability of taxonomic and phylogenetic assessments.

Systematic uncertainties can have repercussions for essentially all fields of biology, much of modern medicine, and is of growing social and economic importance amidst anthropogenic climate change and increased global trade (Galtier 2018). For instance, effective conservation and management strategies for endangered species often rely upon the delimitation of evolutionarily significant units (ESUs, Ryder 1986; Moritz 1994) or on the identification of adaptive potential of species (Stanton *et al.* 2019). International importing and exporting has led to the spread of invasive species that are not always easily distinguished from native species, but which can have devastating impacts on local biodiversity and industry (Armstrong & Ball 2005). Reliable delimitation and identification of human pathogens is also necessary for accurate medical diagnoses and effective treatment (Gast *et al.* 1996; Sperling *et al.* 2017).

Therefore, continued investigation into how species form, are related to one another, and are reliably detected amidst rapid technological shifts has myriad conceptual and practical ramifications that extend far beyond the preoccupations of evolutionary biologists.

Butterflies in the genus *Speyeria* Scudder, 1872 are infamous for their morphological variability, which has been a source of fascination and frustration for generations of enthusiasts, and has led to a long history of taxonomic instability (dos Passos & Grey 1947; Dunford 2009; Pelham 2019). Variable wing colouring and patterning, close relationships, similar ecologies, and large regions of sympatry between named species have made morphological species delimitation difficult, and limited taxon sampling in phylogenetic reconstructions of *Speyeria* has hindered clarification of broad-scale phylogenetic relationships in this genus (Scott *et al.* 1998; Hammond 1978, 1990; Dunford 2007, 2009). Therefore, much of *Speyeria* evolution remains

unknown or only speculated upon. But like many other insects, several species of *Speyeria* are experiencing major habitat loss and population declines (Williams 2002; Shepherd & Debinski 2005; Powell *et al.* 2006; Sanford 2010; Breed *et al.* 2012; McHugh *et al.* 2013; Wells & Tonkyn 2014; Wells *et al.* 2015; Sims 2017). *Speyeria* thus represent an ideal system with which to assess the utility of new genomic approaches to systematics, particularly amidst mounting urgency to clarify the relationships in this genus so that effective conservation strategies may be deployed.

1.2 A shifting perspective of the "species problem"

Systematics has long been preoccupied with species concepts, which have sought to crystallize the biological tendencies and patterns that distinguish species (Mayr 1942; Mayden 1997; Sites & Marhsall 2003; de Queiroz 2007). In general, species concepts identify sets of criteria that are used to assess species limits, however these criteria vary considerably among concepts (of which there are nearly 30), and may alternatively emphasize the significance of morphology, ecology, genetics, general evolutionary tendencies, or simply "diagnosability" (Mayr 1942; Simpson 1951; Sokal & Crovello 1970; Van Valen 1976; Templeton 1989; Mallet 1995). The proliferation of, and, often, the incompatibility between concepts has become known as the "species problem", and has contributed to extensive discussion about how to recognize species (Ghiselin 1974; Mayden 1997; de Queiroz 2005).

Empirical work has repeatedly demonstrated that ecological, demographic, and historical variables can greatly influence species assessments using different criteria. For instance, cryptic species can form due to ecological or genetic drivers that promote speciation in the absence of morphological change (Bickford *et al.* 2007; Fišer *et al.* 2018); mimicry rings can result in

morphological convergence between even distantly related taxa (Turner 1976; Marek & Bond 2009); ring species exhibit a spectrum of reproductive isolation proportionate to the geographic distance between chains of populations (Irwin *et al.* 2001; Devitt *et al.* 2011); and both hybridization and incomplete lineage sorting after recent divergence can result in pronounced morphological and genetic discontinuities (Linnen & Farrell 2007; Li *et al.* 2016). These processes can also impact the accuracy of phylogenetic inference and result in weak or conflicting phylogenetic signals (Lavretsky *et al.* 2014; Leaché *et al.* 2014).

Accordingly, approaches to species delimitation have recently shifted to acknowledge speciation as a variable, often reticulate process rather than a linear one that enables the binning of taxa based on universal criteria (Mayden 1997; de Queiroz 1998, 2007; Mallet 2008; Fišer *et al.* 2018; Stanton *et al.* 2019). The general lineage species concept (de Queiroz 1998) largely formalized this perspective by recognizing that: (i) all species concepts fundamentally agree on the theoretical concept of species as independently evolving lineages, only differing in *how* species are assessed; and (ii) that alternate species criteria are not contradictory, but rather indicative of speciation as a continuum. Therefore, species may exhibit diverse properties at different points in time (the "gray zone" of de Queiroz), which may be variably detected by alternative systematic approaches (Mayden 1997; de Queiroz 2007; Carstens *et al.* 2013).

Assessments of genetic differences between taxa are now commonly used as criteria for recognizing species (Bradley & Baker 2001; Pons *et al.* 2006; Weins 2007). These methods generally incorporate some assessment of gene flow as a proxy for reproductive isolation (Mayr 1942; Mallet 1995). However, this has some limitations for taxa that exist within de Queiroz's gray zone. Logically, populations separated in space are reproductively isolated, but whether this is due to inherent genomic incompatibilities or is a mere consequence of their distribution is an

important issue currently not addressed by other more well-known species concepts, such as Mayr's biological species concept (Mayr 1942). The genomic integrity species concept (Sperling 2003) integrates the advantages of the genetic cluster concept of Mallet (1995) or cohesion concept of Templeton (1989) when divergent populations contact each other, and calibrates species rank based on relative genomic divergence in comparison to unambiguous sister taxa for populations that are not in contact. This provides us with a method of delimiting populations and species with varying degrees of contact in the context of high-resolution genomic data.

1.3 Systematics amidst a molecular revolution

While morphology and ecology are, and will continue to be, essential sources of information for species delimitation, the strength of molecular systematics lies in its ability to assess and characterize the genomic patterns that underlie lineage divergence and contribute to discordant assessments of species, even at fine scales (Hillis *et al.* 1996; Maddison & Knowles 2006). However, it has also become apparent amidst this revolution that the use of alternate molecular techniques can greatly impact systematic inferences (Caterino *et al.* 2000; Carstens *et al.* 2013; Leaché *et al.* 2015).

The early 2000s saw the development of DNA barcoding for phylogenetic reconstruction and species delimitation (Hebert *et al.* 2003). DNA barcoding has been appealing for close to two decades now because it purports to identify cryptic or morphologically variable species using molecular markers that, theoretically, exhibit low intraspecific divergence and high interspecific divergence (in animals, this is often the mitochondrial *COI* gene, Hebert *et al.* (2003)). This approach is highly operational and has led to the creation and widespread use of online data repositories such as the Barcode of Life Data System (BOLD, Ratnasingham &

Hebert 2007) and NCBI GenBank (Benson *et al.* 2007). These data repositories facilitate global data sharing and have been a significant boon to molecular systematics, which is often constrained by taxon sampling.

However, many empirical studies and meta-analyses have demonstrated that DNA barcoding is susceptible to several of the same shortcomings as more traditional types of data used in systematic inference, and that the single-locus approach advocated by DNA barcoding often obstructs or misleads the inference of species (Will & Rubinoff 2004; Brower 2006; Dasmahapatra *et al.* 2010; Dupuis *et al.* 2012). Multi-locus approaches are now largely preferred to single-locus techniques because they recognize that individual gene trees may not align with the overall species tree (Maddison 1997; Edwards 2009; Heled & Drummond 2010), and additionally permit the characterization of gene tree/species tree discordance (Degnan & Rosenberg 2006; Knowles & Carstens 2007).

Multi-locus methods of systematic inference are thus more compatible with the general lineage concept, as they can reveal the processes intrinsic to the de Queiroz "gray zone" of lineage divergence (1998, 2007). In addition, the distinction of species "conceptualization" from "delimitation" made by de Queiroz emphasizes the link between species criteria and the methods we use to measure them. Methodological advancements, then, not only accompany conceptual shifts in systematics, they also influence them. This is especially apparent in the current era of molecular systematics, which is experiencing a paradigm shift facilitated by rapid advances in next-generation sequencing (NGS) techniques and the adoption of a genomic perspective of speciation (Campbell *et al.* 2018; Stanton *et al.* 2019).

1.4 The rise of next-generation sequencing (NGS)

NGS is a catch-all term that encompasses over 400 distinct methods, and generally refers to high-throughput sequencing platforms that can genotype hundreds to thousands of genomic loci in many individuals at once (Hadfield & Retief 2018). In systematics, the most common NGS methods use restriction enzymes to divide the genome into small subsets, which are then sequenced and genotyped for single nucleotide polymorphisms, commonly abbreviated as SNPs (Baird *et al.* 2008; Davey *et al.* 2011; Peterson *et al.* 2012; Andrews *et al.* 2016). Methods that use restriction enzymes to produce SNPs are often collectively referred to as reduced-representation sequencing, or RRS, techniques (Altshuler *et al.* 2000). These methods have since become extensively used in population genomics (Andrews & Luikart 2014), and are now being integrated into systematic analyses where they have demonstrated the pervasiveness of incomplete lineage sorting and hybridization among taxa during lineage divergence (McCormack *et al.* 2013; Escudero *et al.* 2014; Vargas *et al.* 2017; Hinojosa *et al.* 2019)

In less than ten years there has been a proliferation of these methods that, owing to the technical complexities inherent in the production of these data, can introduce distinct sources of bias which can be hard to identify or distinguish from biological processes (van Dijk *et al.* 2014; Andrews *et al.* 2016; Flanagan & Jones 2018). This has not only the potential to introduce methodological artefacts into data analyses, but also to limit the consistency and compatibility of data produced using these methods. There is thus a growing need to clarify the meaningful differences between these methods and assess their compatibility if we are to ensure that systematic research in the genomics era is robust, and continues to benefit from the data-sharing practices already in place.

1.5 *Speyeria* morphology and ecology

Speyeria butterflies (Lepidoptera: Nymphalidae), colloquially known as greater fritillaries or silverspot fritillaries, are endemic to North America and widespread throughout Canada and the United States where they inhabit grasslands, boreal, and mixed-wood habitats of various elevations. *Speyeria* differs in distribution from their closest relatives in the genera *Argynnis* Fabricius, 1807 and *Fabriciana* Reuss, 1920 (referred to hereafter as *Argynnis* s. l., following de Moya *et al.* (2017)), which are restricted to Eurasia, and it is likely that *Speyeria* colonized eastern North America from Asia via the Bering Land Bridge within the last six million years (Simonsen *et al.* 2006; de Moya *et al.* 2017). Most of the current taxonomic diversity of *Speyeria* is located in the western portion of North America, particularly in and around the Rocky Mountains where populations probably experienced periods of isolation in glacial refugia before subsequent re-colonization of newly unglaciated habitat during the Pleistocene. This likely promoted the current species and subspecies diversity in western regions (Hammond 1990; de Moya *et al.* 2017).

Speyeria are univoltine and overwinter as first instar larvae (James & Nunnallee 2011). Like *Argynnis* s. l., larvae are nocturnal and feed on a variety of *Viola* Linnaeus (Violaceae) species. It is likely that the initial radiation of *Speyeria* across North America was facilitated by existing *Viola* diversity (Hammond 1981; de Moya *et al.* 2017), but they appear to exhibit little host-plant specialization among species (Bird *et al.* 1995; Layberry *et al.* 1998; Guppy & Shepard 2001; Brock & Kaufman 2003; Dunford 2007). Larvae of one species, *S. mormonia* (Biosduval, 1869), has been observed feeding on *Bistorta bistortoides* (Pursh) Small, 1906 (Polygonaceae) (Wolfe 2017).

Speyeria are medium to large-sized butterflies, and are generally recognizable by a bright orange dorsal wing surface that is marked with black spots and banding. The general patterning of the dorsal wing surface is highly consistent within *Argynnis* s. l. and has remained similarly conserved within *Speyeria*, however there are differences in the size of these markings between several species, and often between sexes of the same species. The ventral wing surface typically bears metallic silver spots, though they are sometimes a flat cream colour, and the ground colour (the background or base colour) of the ventral hindwing is particularly variable: in many *Speyeria* the ventral hindwing ground colour exhibits clinal variation between populations and/or convergence with other species that inhabit the same region, and may be buff or pale cream, iridescent green, reddish-purple, reddish-brown, tawny brown, or chocolate brown (Moeck 1975; Hammond 1978). Two species deviate from the typical morphology of other *Speyeria*: *S. diana* (Cramer, 1779) females exhibit striking sex-limited mimicry of the pipevine swallowtail, *Battus philenor* (Linnaeus, 1771), and are blue with black markings, though the males are orange with a solid dark brown disc that spans from the base of both wings to the median band, covering approximately half of the dorsal surface on each wing (Dunford 2007; Wells & Tonkyn 2014); and both sexes of *S. idalia* (Drury, 1773) possess hindwings that are mostly black or dark brown dorsally, and sometimes tinged with blue scales. The dorsal spots of *S. idalia* may be black or white on the forewing, and white or orange on the hindwing (Williams 2001).

In comparison to wing morphology, *Speyeria* genitalia are much less variable and mostly uninformative for species delimitation and identification (dos Passos & Grey 1947; Dunford 2009; Hammond *et al.* 2013). Some species are differentiated by the presence of an accessory bursa sac in females and by differences in the size and shape of the uncus in males, which has led to the designation of two groupings - *Semnopsyche,* representing mostly basal lineages, and

Callippe, containing the remaining, typically more derived, species (dos Passos & Grey 1947; Hammond 1978). In addition, some species exhibit intermediate genitalic characteristics (dos Passos & Grey 1947; Dunford 2007). Except for a few basal species, most *Speyeria* do not exhibit substantial post-zygotic reproductive isolation between species and can be successfully hybridized and back-crossed under laboratory conditions (Hammond *et al.* 2013). The extent of natural hybridization in *Speyeria*, however, is largely unknown.

1.6 A systematic summary of *Speyeria*

Extensive variability in the wing pattern and colouring of *Speyeria* has contributed to a history of taxonomic revisions that has produced close to 200 species and subspecies names, many of which are now regarded as invalid synonyms (Elwes 1889; Holland 1898, 1913; Snyder 1900; dos Passos & Grey 1947; Dunford 2009). Current taxonomy recognizes 16 species and 124 subspecies (Opler & Warren 2005; Pelham 2019) which form multiple species complexes marked by morphological ambiguities and range overlap (Moeck 1975; Scott *et al.* 1981; Dunford 2007, 2009), occasional regional hybridization (Hammond *et al.* 2013), and recent or ongoing species divergences (de Moya *et al.* 2017; Thompson *et al.* 2019).

The most well-known of these is the *S. atlantis-hesperis* complex, which contains 26 subspecies and exhibits extensive regional sympatry between species and high rates of morphological variability within and among subspecies (Dunford 2009, Scott *et al.* 1998). *Speyeria atlantis* (Edwards, 1862) is generally recognized as an eastern species that inhabits open woodlands and bogs, while *S. hesperis* (Edwards, 1864) tends to inhabit drier habitats in the west and is commonly found in prairies and grasslands in and around the Rocky Mountains at various elevations. *Speyeria hesperis* is also more morphologically variable than *S. atlantis*

(Edwards, 1862), and includes 21 of the 26 subspecies in this species complex (Dunford 2009; Pelham 2019). These taxa broadly overlap in the Canadian prairies, boreal plains, and the Great Plains of the United States where they often fly together, but the designation of both species and subspecies in this complex has been controversial, as there are several delimitations that appear to represent clinal populations rather than morphologically distinct forms (Moeck 1975; Scott *et al.* 1998). These taxonomic ambiguities are typical of *Speyeria* systematics.

Phylogenetic reconstructions using morphology and gene sequences have largely validated the higher-level taxonomic relationship of *Speyeria* to *Argynnis* s. l. (Simonsen 2006a,b; Simonsen *et al.* 2006; de Moya *et al.* 2017), however work focused specifically on the relationships within *Speyeria* has been less comprehensive and often regionally constrained, frequently omitting several species and/or sampling only a small proportion of the recognized subspecific diversity of this genus, usually one or two subspecies per species (Brittnacher *et al.* 1978; Dunford 2007; McHugh *et al.* 2013; de Moya *et al.* 2017; Hill *et al.* 2018, Thompson *et al.* 2019). More often than not, these phylogenetic studies are discordant with one another, and traditional lepidopteran DNA barcoding genes (Hebert *et al.* 2003; Wahlberg & Wheat 2008) frequently recover non-monophyletic genetic clusters that do not consistently align with morphological species delimitations (Dunford 2007; McHugh *et al.* 2013; de Moya *et al.* 2017).

A genomic re-assessment of species relationships in *Speyeria* will not only provide much-needed systematic clarity for this genus, but also presents an opportunity to assess the utility of new molecular methods for characterizing incipient speciation in a group that has been difficult to resolve using traditional systematic approaches.

1.7 Thesis Objectives

This thesis re-assesses the systematics of *Speyeria* butterflies amidst several conceptual issues that have been raised by recent sequencing advancements and their application to molecular systematics and evolutionary biology.

RRS methods have been a windfall to molecular systematic and population genomic studies, particularly in systems lacking reference genomes (Baird *et al.* 2008; Davey *et al.* 2011; Andrews and Luikart 2014), but their proliferation has led to conceptual ambiguities that result from the many technical subtleties between methods, as well as a growing abundance of acronyms to refer to them (NUAP 2001; Andrews 2016; Hadfield & Retief 2018). These differences are not inconsequential, and can introduce distinct sources of bias that may seriously impact genotyping as well as downstream analyses (Linnen & Farrell 2007; Puritz *et al.* 2014; Li *et al.* 2016). As the number of RRS methods continue to expand and their usage in systematics increases, there is growing impetus to characterize the methodological nuances between suites of RRS methods, clarify the relationships between them, and inform their appropriate usage in molecular ecology and systematic research. This is the focus of **Chapter 2**, in which we conceptually map 36 RRS methods based on their nomenclatural and methodological similarities. In addition, we conducted a literature review of papers that have cited published RRS techniques and evaluated the incidence of improper attribution of these methods in the literature, which we suggest is primarily due to confusion among researchers regarding the differences between methods and the many colloquialisms used to refer to them. This study is now published as Campbell, Brunet, Dupuis, and Sperling (2018).

The development of global data repositories for sequence data has greatly facilitated molecular systematic studies through dataset augmentation. While these resources have, until

recently, been largely restricted to gene sequence data, the rapid adoption of RRS techniques for systematic research presents an opportunity for sharing genomic data in the same way. However, these techniques are predisposed to many methodological quirks, and the use of different RRS protocols to generate genomic SNPs may produce differences in locus recovery and genotyping consistency, particularly when loci are assembled *de novo* (Davey *et al.* 2011; Arnold *et al.* 2013; Puritz *et al.* 2014). These artefacts may result in potentially large differences between datasets, reducing the potential for data sharing among researchers. Yet, there has been no formal assessment of whether *de novo* RRS data produced in different ways can be made compatible via bioinformatic processing. In **Chapter 3**, we determine the consistency of genotyping and locus recovery in SNP data produced for several *Speyeria* species using two commonly used RRS techniques: double digest restriction site-associated sequencing (ddRAD, Peterson *et al.* 2012) and two-enzyme genotyping by sequencing (two-enzyme GBS, Poland *et al.* 2012). Assessing whether data produced using different RRS methods introduces any significant differences in genotyping will have implications for the cross-platform compatibility and longevity of these RRS methods for systematic research. This study has been published as Campbell, Davis, Dupuis, Muirhead, and Sperling (2017).

Phylogenetic inference in *Speyeria* has been historically impeded by pronounced morphological inconsistencies, insufficient taxon sampling, and the use of only one or a few genetic markers (Hammond 1990; Dunford 2009; McHugh *et al.* 2013; de Moya *et al.* 2017; Hill *et al.* 2018). Yet, a robust assessment of the species phylogeny in *Speyeria* will contribute to an essential foundation for future species delimitation and conservation work, and additionally presents an opportunity to assess the efficacy of genomic SNPs for clarifying the species relationships in otherwise hard-to-delimit taxa. In **Chapter 4**, we present the most

geographically comprehensive taxonomic sampling of *Speyeria* to date, consisting of 15 species and 46 subspecies, or about 47% of the current taxonomic diversity. Using genome-wide SNPs and the mitochondrial *COI* gene to re-assess the species phylogeny and reveal mito-nuclear discordance in *Speyeria*, we demonstrate the ability of SNPs to clarify long-standing ambiguities in the species-level relationships of this genus, and identify current species delimitations that likely warrant revision. This study is now published as Campbell, Gage, Gage, and Sperling (2019).

In **Chapter 5**, we build on the systematic foundation provided by Chapter 4 and use genomic SNPs to characterize the historical interactions between the extremely morphologically variable *S. atlantis-hesperis* species complex and *S. zerene* (Boisduval, 1852), a species that is experiencing serious population declines in several regions (McHugh *et al.* 2013; Hill *et al.* 2018). Using new analytical approaches for the assessment of species limits with genomic data (Degnan & Rosenberg 2009; Reich *et al.* 2009), we explore how introgressive hybridization and rapid range expansion contribute to patterns of mito-nuclear discordance in the genome and impacts the inference of species limits (Papakostas *et al.* 2016; Razkin *et al.* 2016; MacGuigan *et al.* 2017; Weigand *et al.* 2017). This work not only clarifies long-standing systematic ambiguities in these taxa, but also has direct implications for the recognition of biologically meaningful conservation units in *Speyeria* amidst rapid habitat loss and climate change (Stanton *et al.* 2019).

In summary, this thesis evaluates the technical similarities among suites of methods for producing genome-wide SNPs, and additionally determines the level of compatibility between two methods that are among the most commonly used in systematic research. We then re-assess the phylogenetic relationships and species delimitations of *Speyeria* using genomic SNPs, and characterize many of the demographic, evolutionary, and ecological factors that have contributed

to longstanding systematic challenges in this genus. This work conceptually demonstrates how the recognition of species limits is tightly linked to the methods we use to detect them, and adds to the growing body of work that validates the use of genomic data for systematic studies of taxa that have not been satisfactorily resolved using more traditional approaches.

1.8 Literature cited

Altshuler, D., Pollara, V. J., Cowles, C. R., Van Etten, W. J., Baldwin, J., Linton, L., and Lander, E. S. (2000) A SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature*, **407**, 513-516.

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., and Hohenlohe, P. A. (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, **17**, 81-92.

Andrews, K. R., and Luikart, G. (2014) Recent novel approaches for population genomics data analysis. *Molecular Ecology*, **23**, 1661-1667.

Armstrong, K. F., and Ball, S. L. (2005) DNA barcodes for biosecurity: invasive species identification. *Philosophical Transactions of the Royal Society B*, **10**, **S10.**

Arnold, B., Corbett-Detig, R. B., and Bomblies, K. (2013) RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology*, **22**, 3179-3190.

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

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Wheeler, D. L. (2007) Genbank. *Nucleic Acids Research*, **35**, D21-D25.

Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., Ingram, K. K., and Das, I. (2007) Cryptic species as a window on diversity and conservation. *TRENDS in Ecology and Evolution*, **22**, 148-155.

Bird, C. D., Hilchie, G. J., Kondla, N. G., Pike, E. M., and Sperling, F. A. H. (1995) *Alberta Butterflies*, The Provincial Museum of Alberta, Edmonton, Alberta, Canada.

Bradley, R. D., and Baker, R. J. (2001) A test of the genetic species concept: cytochrome-*b* sequences and mammals. *Journal of mammology*, **82**, 960-973.

Breed, G. A., Stichter, S., and Crone, E. E. (2012) Climate-driven changes in northeastern US butterfly communities. *Nature Climate Change*, **3**, 142-145.

Brittnacher, J. G., Sims, S. R., and Ayala, F. J. (1978) Genetic Differentiation Between Species of the Genus Speyeria (Lepidoptera: Nymphalidae). *Evolution.* 32: 199-210.

Brock, J. P. and Kaufman, K. (2003) *Kaufman Field Guide to Butterflies of North America*. ftrHoughton Mifflin Co., New York.

Brower, A. V. Z. (2006) Problems with DNA barcodes for species delimitation: 'ten species' of *Astraptes fulgerator* reassessed (Lepidoptera: Hesperiidae). *Systematics and Biodiversity*, **4**, 127-132.

Campbell, C. R., Poelstra, J. W., and Yoder, A. D. (2018) What is speciation genomics? The roles of ecology, gene flow, and genomic architecture in the formation of species. *Biological Journal of the Linnean Society*, **124**, 561-583.

Campbell, E. O., Brunet, B. T. M., Dupuis, J. R., and Sperling, F. A. H. (2018) Would an RRS by any other name sound as RAD? *Methods in Ecology and Evolution*, **9**, 1920-1927.

Campbell, E. O., Davis, C. S., Dupuis, J. R., Muirhead, K., and Sperling, F. A. H. (2017) Cross-platform compatibility of *de novo*-aligned SNPs in a non-model butterfly genus. *Molecular Ecology Resources*, **6**, e84-e93.

Campbell, E. O., Gage, E. V., Gage, R. V., and Sperling, F. A. H. (2019) Single nucleotide polymorphism-based phylogeny of greater fritillary butterflies (Lepidoptera:

Nymphalidae: *Speyeria*) reveals widespread mito-nuclear discordance. *Systematic Entomology*, DOI:10.1111/syent.12393

Carstens, B. C., Pelletier, T. A., Reid, N. M., and Satler, J. D. (2013) How to fail at species delimitation. *Molecular Ecology*, **22**, 4369-4383.

Caterino, M. S., Cho, S., and Sperling, F. A. H. (2000) The current state of insect molecular systematics: a thriving tower of Babel. *Annual Review of Entomology*, **45**, 1.54.

Dasmahapatra, K. K. Elias, M., Hill, R. I., Hoffman, J I., and Mallet, J. (2010) Mitochondrial DNA barcoding detects some species that are real, and some that are not. *Molecular Ecology Resources*, **10**, 264-273.

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., and Blaxter M. L. (2011) Genome-wide marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, **12**, 499-510.

Degnan, J. H., and Rosenberg, N. A. (2006) Discordance of species trees with their most likely gene trees. *PLoS Genetics*, **2**, e68.

Degnan, J. H., and Rosenberg, N. A. (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *TRENDS in Ecology and Evolution*, **24**, 332-340.

de Moya, R. S., Savage, W. K., Tenney, C., Bao, X., Wahlberg, N., and Hill, R. I. (2017) Interrelationships and diversification of *Argynnis* Fabricius and *Speyeria* Scudder butterflies. *Systematic Entomology*, **42**, 635-649.

de Queiroz, K. (1998) The general lineage concept of species, species criteria, and the process of speciation. Pp. 57-75. IN Howard, D. J., Berlocher, S. H. (Eds.) *Endless forms: species and speciation*. Oxford University Press, New York.

de Queiroz, K. (2005) Ernst Mayr and the modern concept of species. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 6660-6607.

de Queiroz, K. (2007) Species Concepts and Species Delimitation. *Systematic Biology*, **56**, 879-886.

Devitt, T. J., Baird, S. J. E., and Moritz, C. (2011) Asymmetric reproductive isolation between terminal forms of the salamander ring species *Ensatina eschscholtzii* revealed by fine-scale genetic analysis of a hybrid zone. *MBC Evolutionary Biology*, **11**, 245.

dos Passos, C. F. and Grey, L. P. (1947) Systematic catalogue of *Speyeria* (Lepidoptera, Nymphalidae) with designations of types and fixations of type localities. *American Museum Novitates*, **1370**, 1-30.

Dunford, J. C. (2007) *The genus Speyeria and the Speyeria atlantis/hesperis complex: species and subspecies accounts, systematics, and biogeography (Lepidoptera: Nymphalidae)*. Unpublished Dissertation, University of Florida.

Dunford, J. C. (2009) Taxonomic overview of the greater fritillary genus *Speyeria* Scudder and the atlantis-hesperis species complexes, with species accounts, type images, and relevant literature (Lepidoptera: Nymphalidae). *Insecta Mundi*, **90**, 1-74

Dupuis, J. R., Roe, A. D., and Sperling, F. A. H. (2012) Multi-locus species delimitation in closely related animals and fungi: one marker is not enough. *Molecular Ecology*, **21**, 4422-4436.

Edwards, S. V. (2009) Is a new and general theory of molecular systematics emerging? *Evolution*, **63**, 1-19.

Elwes, H.J., (1889) Revision of *Argynnis*. Transactions of the Royal Entomological Society of London. 37: 535-575.

Escudero, M., Eaton, D. A. R., Hahn, M., Hipp, A. L. (2014) Genotyping-by-sequencing as a tool to infer phylogeny and ancestral hybridization: a case study in *Carex* (Cyperaceae). *Molecular Phylogenetics and Evolution*, **79**, 359-367.

Fišer, C., Robinson, C. T., and Malard, F. (2018) Cryptic species as a window into the paradigm shift of the species concept. *Molecular Ecology*, **27**, 613-635.

Flanagan, S. P. and Jones, A. G. (2018) Substantial differences in bias between single-digest and double-digest RAD-seq libraries: a case study. *Molecular Ecology Resources*, **18**, 264-280.

Galtier, N. (2018) Delineating species in the speciation continuum: a proposal. *Evolutionary Applications*, **12**, 657-663.

Gast, R. J., Ledee, D. R., Fuerst, P. A., and Byers, T. J. (1996) Sybgenus systematics of *Acanthamoeba:* four nuclear 18S rDNA sequence types. *Eukaryotic Microbiology*, **43**, 498-504.

Ghiselin, M. T. (1974) A radical solution to the species problem. *Systematic Zoology*, **23**, 536-544.

Guppy, C. S., and Shepard, J. H. (2001) *Butterflies of British Columbia*, UBC Press, Vancouver, British Columbia.

Hadfield, J. and Retief, J. (2018) A profusion of confusion in NGS methods naming. *Nature Methods*, **15**, 7-9.

Hammond, P. C. (1978) *Geographic variation and speciation in the Nymphalid butterfly genus Speyeria.* PhD Dissertation, Virginia Polytechnic Institute and State University, Blacksburg.

Hammond, P. C. (1981) The colonization of violets and *Speyeria* butterflies on the ash-pumice fields deposited by Cascadian volcanoes. *Journal of the Research on Lepidoptera*, **20**, 179-191.

Hammond, P. C. (1990) Patterns of geographic variation and evolution in polytypic butterflies. *Journal of Research on the Lepidoptera*, **29**, 54-76.

Hammond, P. C., McCorkle, D., and Bergman, W. (2013) Hybridization studies of genomic compatibility and phenotypic expression in the greater fritillary butterflies (Nymphalidae: Argynnini). *Journal of the Lepidopterists' Society*, **67**, 263-273.

Hebert, P. D. N., Cywinska, A., Ball, S. L., and deWaard, J. R. (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B.* **270**, 313-321.

Heled, J. and Drummond, A. J. (2010) Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, **27**, 570-580.

Hillis, D. M., Moritz, C., and Mabel, B. K. (Eds.) (1996) *Molecular Systematics* (2nd edition), Sinauer, Sunderland, Massachusetts.

Hinojosa, J. C., Koubínová, D., Szenteczki, M. A., Pitteloud, C., Dincă, V., Alvarez, N., and Vila, R. (2019) A mirage of cryptic species: genomics uncover striking mitonuclear discordance in the butterfly *Thymelicus sylvestris*. *Molecular Ecology*. DOI: https://doi.org/10.1111/mec.15153

Hohenlohe, P. A., Day, M. D., Amish, S. J., Miller, M. R., Kamps-Hughes, N., Boyer, M. C., Muhlfeld, C. C., Allendorf, F. W., Johnson, E. A., and Luikart, G. (2013) Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. *Molecular Ecology*, **22**, 3002-3013.

Holland, W.J., (1898) *The Butterfly Book: A Popular Guide to a Knowledge of theButterflies of North America.* Doubleday, Page, and Company, New York. Reprinted 1907.

Holland, W. J. (1931) Notes on some American butterflies mainly relating to classification and nomenclature. Part 3. *Annals of the Carnegie Museum,* **20**, 255-265.

Irwin, D. E., Irwin, J. H., and Price, T. D. (2001) Ring species as bridges between microevolution and speciation. *Genetica*, **112-113**, 223-243.

James, D. G., and Nunnallee, D. (2011) *Life Histories of Cascadia Butterflies*, Oregon State University press, Corvallis, Oregon.

Knowles, L. L., and Carstens, B. C. (2007) Delimiting species without monophyletic gene trees. *Systematic Biology*, **56**, 887-895.

Lavretsky, P., Hernández-Baños, B. E., and Peters, J. L. (2014) Rapid radiation and hybridization contribute to weak differentiation and hinder phylogenetic inference in the New World Mallard complex (*Anas* spp.). *BioOne Complete*, **131**, 524-538.

Layberry, R. A., Hall, P. W., and Lafontaine, J. D. (1998) *The Butterflies of Canada*. University of Toronto Press, Toronto.

Leaché, A. D., Banbury, B. L., Felsentein, J., Nieto-Montes de Oca, A., and Stamatakis, A. (2015) Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. *Systematic Biology*, **64**, 1032-1047.

Leaché, A. D., Harris, R. B., Rannala, B., and Yang, Z. (2014) The influence of gene flow on specie tree estimation: a simulation study. *Systematic Biology*, **63**, 17-30.

Li, G., Davis, B. W., Eizirik, E., and Murphy, W. J. (2016) Phylogenomic evidence for ancient hybridization in the genomes of living cats (Felidae). *Genome Resources*, **26**, 1-11.

Linnen, C. R., and Farrell, B. D. (2007) Rampant mitochondrial introgression in *Neodiprion* (Hymenoptera: Diprionidae) sawflies. *Evolution*, **61**, 1417-1438.

MacGuigan, D. J., Geneva, A. J., and Glor, R. E. (2017) A genomic assessment of species boundaries and hybridization in a group of highly polymorphic anoles (*distichus* species complex). *Ecology and Evolution*, **7**, 3657-3671.

Maddison, W. P. (1997) Gene trees in species trees. *Systematic Biology*, **46**, 523-536.

Maddison, W. P. and Knowles, L. L. (2006) Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology*, **55** 21-30.

Mallet, J. (1995) A species definition for the modern synthesis. *TRENDS in Ecology and Evolution*, **10**, 294-299

Mallet, J. (2008) Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society B*, **363**, 2971-2986.

Marek, P. E., and Bond, J. E. (2009) A Müllerian mimicry ring in Appalachian millipedes. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 9755-9760.

Mayden, R. L. (1997) A hierarchy of species concepts: The denouement in the saga of the species problem. Pp. 381-424. IN Claridge, M.F., Dawah, H. A., and Wilson, M.R (Eds.) *Species: the units of biodiversity*, Chapman and hall, London.

Mayr, E. (1942) *Systematics and the origin of species*, Columbia University Press, New York.

McCormack, J.E., Hird, S. M., Zellmer, A. J., Carstens, B. C., and Brumfield, R. T. (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, **66**, 526-538.

McHugh, A. Greever, C., and Van Buskirk, R. (2013) A molecular phylogenetic analysis of *Speyeria* and its implications for the management of the threatened *Speyeria zerene hippolyta. Journal of Insect Conservation,* **17**, 1237-1253.

Moeck, A. H. (1975) Geographic variability in *Speyeria*: comments, records, and description of a new subspecies. Los Angeles: Entomological Reprint Specialists.

Moritz, C. (1994) Defining 'evolutionarily significant units' for conservation. *TRENDS in Ecology and Evolution*, **9**, 372-375.

NUAP (no unnecessary acronyms please). (2011) *Nature Methods*, **8**, 521.

Opler, P. A., and Warren, A. D. (2005) *Lepidoptera of North America. 4. Scientific names list for butterfly species of North America, north of Mexico.* Contributions of the C. P. Gillette Museum of Arthropod Diversity, Colorado State University, Fort Collins, CO.

Papakostas, S., Michaloudi, E., Proios, K., Brehm, M., Verhage, L., Rota, J., Peña, C., Stamou, G. Prithcard, V. L., Fontaneto, D., and Declerck, S. A. J. (2016) Integrative taxonomy recognizes evolutionary units despite widespread mitonuclear discordance: evidence from a Rotifer cryptic species complex. *Systematic Biology*, **65**, 508-524.

Pelham, J. P. (2019) A catalogue of the butterflies of the United States and Canada; with a complete bibliography of the descriptive and systematic literature. *The Lepidoptera Research Foundation/Journal of Research on the Lepidoptera,* Volume 40, Beverly Hills, CA, USA.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., and Hoekstra, H. E. (2012) Double digest RADseq: an inexpensive method for *de novo* discovery and genotyping in model and non-model species. *PLoS ONE*, **7**, e37135.

Poland, J. A., Brown, P. J., Sorrells, M. E., and Jannink, J. (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE*, **7**, e32253.

Pons, J., Barraclough, T. G., Gomez-Zurita, J., Cardoso, A., Duran, D. P., Hazell, S., Kamoun, S., Sumlin, W. D., and Vogler, A. P. (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, **55**, 595-609.

Powell, A. F. L. A., Busby, W. H., and Kindscher, K. (2006) Status of the regal fritillary (*Speyeria idalia*) and effects of fire management on its abundance in northeastern Kansas, USA. *Journal of Insect Conservation*, **11**, 299-308.

Puritz, J. B., Matz, M. V., Toonen, R. J., Weber, J. N., Bolnick, D. I., and Bird, C. E. (2014) Demystifying the RAD fad. *Molecular Ecology*, **23**, 5937-5942.

Ratnasingham, S., and Hebert, P. D. (2007) BOLD: the barcode of life data system. *Molecular Ecology Notes*, **7**, 355-364.

Razkin, O., Gómez-Moliner, B. J., Vardinoyannis, K., Martínez-Ortí, A., and Madeira, M. J. (2016) Species delimitation for cryptic species complexes: case study of *Pyramidula* (Gastropoda: Pulmonata). *Zoological Scripta*, **46**, 55-72.

Reich, D. Thangaraj, K., Patterson, N., Price, A. L., and Singh, L. (2009) Reconstructing Indian population history. *Nature*, **461**, 489-494.

Ruse, M. (1969) Definitions of species in biology. *British Journal for the Philosophy of Science*, **20**, 97-119.

Ryder, O. A. (1986) Species conservation and systematics: the dilemma of subspecies. *TRENDS in Ecology and Evolution*, **1**, 9-10.

Sanford, M. P. (2011) Improving conservation and management of the imperiled Carson valley silverspot butterfly *Speyeria nokomis carsonensis* (Lepidoptera: Nymphalidae) based on rapid assessments of distribution, habitat, and threats. *Journal of Insect Conservation*, **15**, 715-725.

Scott, J. A., Kondla, N. G., and Spomer, S. M. (1998) *Speyeria hesperis* and *Speyeria atlantis* are distinct species. *Papilio New Series*, **8**, 1-32.

Shepherd, S., and Debinski, D. M. (2005) Reintroduction of regal fritillary (*Speyeria idalia*) to a restored prairie. *Ecological Restoration*, **23**, 244 - 250.

Simonsen, T. J. (2006a) Fritillary phylogeny, classification, and larval host plants: reconstructed mainly on the basis of male and female genitalic morphology (Lepidoptera: Nymphalidae: Argynnini). *Biological Journal of the Linnaean Society*, **89**, 627-673.

Simonsen, T. J. (2006b) Glands, muscles and genitalia. Morphological and phylogenetic implications of histological characters in the male genitalia of Fritillary butterflies (Lepidoptera: Nymphalidae: Argynnini). *Zoologica Scripta*, **35**, 231-241.

Simonsen, T. J., Wahlberg, N., Brower, A. V. Z., and de Jong, R., (2006) Morphology, molecules and fritillaries: approaching a satble phylogeny for Argynnini (Lepidoptera: Nymphaliade). *Insect Systematics and Evolution*, **37**, 405-418.

Simpson, G. G. (1951) The species concept. *Evolution*, **5**, 285-298.

Sims, S. R. (2017) *Speyeria* (Lepidoptera: Nymphalidae) conservation. *Insects*, **8**, 45.

Sites, J. W., Jr., and Marshall, J. C. (2003) Delimiting species: a Renaissance issue in systematic biology. *TRENDS in Ecology and Evolution*, **18**, 462-470.

Snyder, A. J. (1900) The argynnids of North America. *Occasional Memoirs of the Chicago Entomological Society*, **1**, 27-38.

Sokal, R. R., and Crovello, T. J. (1970) The biological species concept: a critical evaluation. *The American Society of Naturalists*, **104**, 127-153.

Sousa, V., and Hey, J. (2013) Understanding the origin of species with genome-scale data: modelling gene flow. *Nature Reviews Genetics*, **14**, 404-414.

Sperling, F. A. H. (2003) Butterfly molecular systematics: from species definitions to higher level phylogenies. Pp. 431-458. IN Boggs, C., Erlich, P. and Watt, W. (Eds). *Ecology and Evolution Taking Flight: Butterflies as model study systems*. University of Chicago Press, Chicago.

Sperling, J. L., Silva-Brandao, K. L., Brandao, M. M., Lloyd, V. K., Dang, S., Davis, C. S., Sperling, F. A. H., and Magor, K. E. (2017) Comparison of bacterial 16S rRNA variable regions for microbiome surveys of ticks. *Ticks and Tick-borne Diseases*, **8**, 453-461.

Stanton, D. W.G., Frandsen, P., Waples, R. K., Heller, R., Russo, I-R. M., Orozco-terWengel, P. A., Pedersen, C-E. T., Siegismund, H. R., and Bruford, M. W. (2019) More grist for the mill? Species delimitation in the genomic era and its implcations for conservation. *Conservation Genetics*, **20**, 101-113.

Templeton, A. R. (1989) The meaning of species and speciation: A genetic perspective. Pp. 3-27. IN Howard, D. J., and Berlocher, S. H., (Eds.). *Endless forms: species and speciation*, Oxford University Press, New York.

Thompson, E., Baumsteiger, J., and Hill, R. I. (2019) Phylogenomic analyses clarify true species within the butterfly genus *Speyeria* despite evidence of a recent adaptive radation. *Insects*, **10**, 209-224.

Turner, J. R. G. (1976) Adaptive radiation and convergence in subdivisions of butterfly genus *Heliconius* (Lepidoptera: Nymphalidae). *Zoological Journal of the Linnean Society,* **58**, 297-308.

van Dijk, E. L., Jaszczyszyn, Y., and Thermes, C. (2014) Library preparation methods for next-generation sequencing: tone down the bias. *Experimental Cell Research*, **32**, 12-20.

Van Valen, L. (1976) Ecological species, multispecies, and oaks. *Taxon*, **25**, 233-239.

Vargas, O. M., Ortiz, E. M., Simpson, B. B. (2017) Conflicting phylogenomic signals reveal a pattern of reticulate evolution in a recent high-Andean diversification (Asteraceae: Astereae: *Diplostephium*). *New Phytologist*, **214**, 1736-1750.

Wahlberg, N. and Wheat, C. W. (2008) Genomic outposts serve the phylogenomic pioneers: designing novel nuclear markers for genomic DNA extractions of Lepidoptera. *Systematic Biology*, **57**, 231-242.

Weigand, H., Weiss, M., Cai, H., Li, Y., Yu, L., Zhang, C., and Leese, F. (2017) Deciphering the origin of mito-nuclear discordance in two sibling caddisfly species. *Molecular Ecology*, **26**, 5705-5715.

Weins, J. J. (2007) Species delimitation: new approaches for discovering diversity. *Systematic Biology*, **56**, 87-878.

Wells, C. N., Marko, P. B., and Tonkyn, D. W. (2015) The phylogeographic history of the threatened Diana fritillary, *Speyeria diana* (Lepidoptera: Nymphalidae): with implications for conservation. *Conservation Genetics*, **16**, 703-716.

Wells, C. N. and Tonkyn, D. W. (2014) Range collapse in the Diana fritillary, *Speyeria diana* (Nymphalidae). *Insect Conservation and Diversity*, **7**, 365-380.

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

Williams, B. L. (2001) Patterns of morphological variation in *Speyeria idalia* (Lepidoptera: Nymphalidae) with implications for taxonomy and conservation. *Annals of the Entomological Society of America*, **94**, 239-243.

Williams, B. L. (2002) Conservation genetics, extinction, and taxonomic status: a case history of the regal fritillary. *Conservation Biology*, **16**, 148-157.

Wolfe, K. (2017) Natural non-*Viola* hostplant of a Sierran *Speyeria* (Nymphalidae, Heliconiinae) and an associated parasite (Diptera, Tachinidae). *Journal of the Lepidopterists' Society*, **71**, 129-131.

Chapter 2

Would an RRS by any other name sound as RAD?

A version of this chapter has been published as Campbell EO, Brunet BTM, Dupuis JR, and Sperling FAH (2018) "Would an RRS by any other name sound as RAD?", Methods in Ecology and Evolution volume 9, pages 1920-1927.

2.1 Summary

Sampling markers throughout a genome with restriction enzymes emerged in the 2000s as reduced representation shotgun sequencing (RRS). Rapid advances in sequencing technology have since spurred modifications of RRS, giving rise to many derivatives with unique names, such as restriction site-associated DNA sequencing (RADseq). But naming conventions have often been more creative than consistent and criteria for recognizing unique methods have been unclear, resulting in a proliferation of names characterized by ambiguity. We give an overview of methodological and etymological relationships among 36 restriction enzyme-based methods, and survey the consistency of references to five prominent methods in the literature. We identified several instances of methodological convergence, and note that many published derivatives have modified only minor elements of parent protocols. Misattribution through ambiguous or inconsistent literature references was observed in 8.4% of journal articles citing the original one and two-enzyme RADseq and GBS, as well as SBG publications. The rapid expansion of names associated with derivative protocols is confusing and, in many cases, unwarranted. We urge greater restraint in naming derivative methods and suggest general

guidelines for naming that promote a balance between clarity, descriptiveness, and recognition of scientific innovation.

2.2 Introduction

Recent advances in next-generation sequencing (NGS) have given researchers access to unprecedented amounts of genomic data. The versatility of NGS, exemplified by its myriad applications to biology (Andrews *et al.* 2016), is arguably one of its greatest assets and has in turn led to more than 400 published methods that use this technology (Hadfield & Retief 2018). While NGS has indisputably spurred rapid innovation across biology, associated names have also proliferated. These names are commonly acronyms meant to clearly identify a methodology or application, but, due to their sheer numbers, are now themselves a source of confusion. Suggested guidelines for the use of such acronyms were published several years ago (NUAP 2011), and Hadfield and Retief (2018) recently reignited discussion on the excess of names for NGS methods, but they did not characterize how this name usage is reflected in the literature.

Methods that use restriction enzymes to sample genomes represent a common subset of NGS techniques. These methods provide diverse options for reducing genomic complexity and surveying large numbers of loci across populations or species, and are widely employed across ecology and evolutionary biology (Baird *et al.* 2008; Davey *et al.* 2011). Early approaches include reduced representation shotgun sequencing (RRS, Altshuler *et al.* 2000) and Complexity Reduction of Polymorphic Sequences (CRoPs™, van Orsouw *et al.* 2007). These have since served as springboards for derivative techniques, most of them published with unique names. At least 36 named methods have been published as of December 2017 (Appendix 2.1). Two methods in particular, restriction site-associated DNA sequencing (RADseq, Baird *et al.* 2008)

and genotyping-by-sequencing (GBS, Elshire *et al.* 2011), have been modified for diverse research on association and genetic mapping, population structure, and shallow-scale phylogenetics (Poland *et al.* 2012; Baird *et al.* 2013; Narum *et al.* 2013; Eaton 2014). In fact, they are so popular that 26 of the 36 methods have been explicitly modified from either RADseq or GBS (Appendix 2.1). The increasing importance of these methods has been extensively documented in recent reviews that have primarily focused on methodological differences between techniques and the applications of the data (Davey *et al.* 2011; Andrews *et al.* 2016; Jiang *et al.* 2016), as well as debate (Andrews & Luikart 2014; Puritz *et al.* 2014; Andrews *et al.* 2014; Lowry *et al.* 2016; Lowry *et al.* 2017; McKinney *et al.* 2017; and Catchen *et al.* 2017). Although their authors have tried to distinguish among approaches, it is clear from the reviews and debate, as well as from informal discussion on online forums (Appendix 2.2), that the differences between many techniques are not always obvious.

Naming conventions for derivatives have been variable and inconsistent, and literature discussing or employing these techniques has been ambiguous about the origins of techniques as well as which names to use as "catch-all" terms. For instance, "GBS" is sometimes used to refer to all restriction-based methods collectively (e.g. Narum *et al.* 2013; Franchini *et al.* 2017), while other authors use "RADseq" as a generic term because of its descriptive qualities (e.g. Andrews *et al.* 2016; Hoffberg *et al.* 2016). Some authors have even given a common acronym a new meaning. NextRAD (Fu *et al.* 2017) was developed by authors of the original RADseq papers (Miller *et al.* 2007; Baird *et al.* 2008) but uses transposomes to cut DNA rather than restriction enzymes; in this instance the RAD acronym stands for reductively-amplified DNA rather than restriction site-associated DNA. Thus, reduced representation genome sampling methods exemplify the naming problem that is now typical of NGS methods. Here we refer to these

techniques as RRS methods following Altshuler *et al*. (2000). While this method was pre-NGS and used Sanger sequencing, it was also the first name associated with these methods and is inclusive of the wide array of methodologies considered. The term also avoids favoring a derived name to refer to earlier methods (see Discussion).

The proliferation of new names for RRS methods and ambiguity of their naming conventions (Jiang *et al.* 2016) raises three questions about the use of these names. First, what are the current trends or criteria for naming new methods? Second, are researchers citing and referring to these methods consistently? Finally, when should we be naming such methods? To answer these questions, we summarize the methodological and etymological relationships of these techniques with a concept map, then characterize the patterns of literature citations for five prominent methods, followed by suggestions for the deployment of new names for RRS methods.

2.3 Methods

2.3.1 *RRS conceptual map*

We compiled a list of RRS methods published on or before 31 December 2017 (N = 36), and evaluated approaches based on their methodological characteristics (Appendix 2.1). We then created a conceptual map of all methods, linking each derived technique to the main protocol that served as the basis for the modification, as specified by the authors (Fig. 2.1). In several cases a parent protocol was not directly specified, and in these instances we linked methods based on overall methodological similarity. The subjective construction of this map reflects our experience as typical arms-length users of several of these approaches. Any technique that explicitly altered a protocol was considered a direct modification, and in this conceptualized map, a separate node.

We plotted defining characteristics for each derivative along the connecting branches to assess distinctiveness or methodological convergence. Defining characteristics were generally those considered by the authors of the protocol to distinguish the derived method from its parent. To preserve clarity, characteristics that were highly variable across methods (for instance barcode and adaptor design and the overall order of methodological steps in each protocol) were not plotted on the map unless they were definitive for the method(s). We also downloaded complete citation data from Web of Science® (Clarivate Analytics, Philadelphia, PA) for the 36 methods, and determined the average number of citations per year for each publication. The size of ellipses in Fig. 2.1 reflects these numbers.

2.3.2 *Literature citation patterns for RADseq, GBS, and SBG*

To assess whether methods are recognized and attributed accurately in the literature, we compiled a list of all journal articles citing one- and two-enzyme RADseq and GBS methods (RADseq, Baird *et al.* 2008; GBS, Elshire *et al.* 2011; two-enzyme GBS, Poland *et al.* 2012; and ddRADseq, Peterson *et al.* 2012), as well as Sequence-based Genotyping (SBG, Truong *et al.* 2012), and searched for instances of inconsistent name usage. RADseq, GBS, two-enzyme GBS and ddRADseq are four of the most widely cited RRS approaches, and have been extensively modified to form the basis of many derivative methods. While the SBG protocol of Truong *et al.* (2012) is far less frequently cited, we included it for its methodological and etymological similarity to these methods, as well as its date of publication which occurred between that of Poland *et al.* (2012) and Peterson *et al.* (2012). It is also the subject of U.S. patent 8,815,512 B2, owned by KeyGene, which claims legal ownership and protection of all methods that

simultaneously discover and genotype single nucleotide polymorphisms, including RADseq, GBS, two-enzyme GBS and ddRADseq (KeyGene, 2016).

Complete citation lists were downloaded from Web of Science® on 6 February 2018 for the period up to and including 31 December 2017 for each of Baird *et al.* (2008), Elshire *et al.* (2011), Poland *et al.* (2012), Truong *et al.* (2012), and Peterson *et al.* (2012). Only articles with titles, abstracts, or keywords containing "GBS", "SBG" or "RAD" (and all variant search strings in Appendix 2.3) were retained for further analysis. Inconsistent name usage was defined as any case of an alternate name being used to refer to a technique (e.g. "GBS" to describe the RADseq protocol of Baird *et al.* (2008)). Strings for "two-enzyme GBS" and "ddRADseq" were not searched separately since these were treated as variants of "GBS" and "RAD", respectively. The results were graphed using the ggplot2 library (v. 2.2.1, Wickham, H. (2009)) in R (v. 3.4.1, R Core Team, 2017). A complete description of the methods used is in Appendix 2.4, and a full list of the papers included in the literature analysis can be found in the online supplemental data for Campbell *et al.* (2018).

2.4 Results

2.4.1 *Trends and criteria for naming new methods*

Of the 36 RRS methods we examined, those of Baird *et al.* (2008), Peterson *et al.* (2012), and Elshire *et al.* (2011) are the precursors of the greatest number of directly derived methods (Fig. 2.1), and the most highly cited (Appendix 2.1). "RAD" was used in 18 named techniques, while "GBS" was used in six and the remaining 12 methods had names that lacked "RAD", "GBS", or any specified name at all (see Sonah *et al.* 2013 and Mascher *et al.* 2013). Many derived methods were named after the protocol they modified (e.g. ddRADseq (Peterson *et al.*

2012) from RADseq (Baird *et al.* 2008)), but there were several exceptions (e.g. SBG (Truong *et al.* 2012) from CRoPS™ (van Orsouw *et al.* 2007)). We observed multiple occurrences of methodological convergence across methods, including the use of paired restriction enzymes in double digest methods, sequence capture, bisulfite sequencing, and the use of PCR amplification to create reduced representation libraries, which we discuss below.

2.4.2 *Consistency of citations and referencing*

The number of journal articles that refer to "GBS", "SBG", or "RAD" within their title, abstract or keywords and uniquely cite either Baird *et al.* (2008), Elshire *et al.* (2011), Poland *et al.* (2012), Truong *et al.* (2012), or Peterson *et al.* (2012) has increased rapidly since 2010, with the greatest number of citations occurring in 2017. Of a total of 788 journal articles, 335 (42.5%) refer only to GBS, two (0.2%) refer only to SBG, and 418 (53.1%) refer only to RAD (Fig. 2.2; Appendix 2.5). Two or more of these names ("Multiple(≥2)") are used in 33 (4.3%) journal articles and these refer only to GBS and RAD, not SBG (Fig. 2.2; Appendix 2.5).

Each name has been used inconsistently to refer to methods described by their parent publications, but to varying degrees: 8% (28/349) of publications that uniquely cite Elshire *et al.* (2011) or Poland *et al.* (2012) refer to SBG or RAD alone or in combination with GBS; 66.7% (2/3) of publications that uniquely cite Truong *et al.* (2012) refer to GBS or RAD alone or in combination with SBG; and 8.3% (36/436) of publications that uniquely cite Baird *et al.* (2008) or Peterson *et al.* (2012) refer to GBS or SBG alone or in combination with RAD (Fig. 2.2; Appendix 2.5). Thus, use of ambiguous or inconsistent names is apparent in about 8.4% of journal articles citing these five papers.

2.5 Discussion

We have characterized the relationships of RRS methods with a concept map showing that RAD-based methods are more numerous than other RRS methods. Although derived methods are often given unique names, most follow some of the etymological elements of the parent technique that was modified, even when derived protocols from different camps converge methodologically (Fig. 2.1). We also identified a rate of ~8.4% ambiguous or inconsistent citation for the five methods illustrated in Fig. 2.2. The RAD acronym is leading the popularity race when considering citations for RAD-based methods as well as the number of derivative protocols bearing this term; GBS-based methods, on the other hand, have fewer overall citations and methodological offspring. By comparison, references to SBG in the literature are virtually non-existent even though its authors currently hold the patent over most RRS methods.

2.5.1 *"What's in a name?"* (Shakespeare 1594-98)

While the original RADseq or modified ddRADseq methods may simply be more methodologically attractive, linguistic factors, both conscious and unconscious, may be contributing to this trend. Acronyms that form simple, recognizable words are more likely to be remembered (NUAP 2011). This might explain the more prevalent use of the RAD acronym, despite citation of a GBS or SBG publication, than either reciprocal configuration (Fig. 2.2). RAD is also easy to incorporate into memorable titles that improve name recognition and visibility in a rapidly expanding field (e.g. "Demystifying the RAD fad" (Puritz *et al.* 2014); "Breaking RAD" (Lowry *et al.* 2016); present study). However, rates of misattribution do not appear to be substantially biased toward one of these methods over another (Fig. 2.2), so researchers who are unclear on or unconvinced of the distinctions between methods may simply

be using these terms synonymously, or choosing one to function as a general, catch-all term due to personal preference. Finally, the methodological convergence of several GBS- and RAD-based techniques (Fig. 2.1) could also be contributing to ambiguity among protocols.

Publication of new methods implies that their authors consider them to be substantively different from other published methods, thereby warranting a separate name. But for RRS methods, many differences between techniques are minor, often implementing streamlined library preparation and cost reduction (e.g. GGRS (Chen *et al.* 2013), ezRAD (Toonen *et al.* 2013)) or the use of specific restriction enzymes and adaptors designed to optimize sequencing depth, coverage, and multiplexing capacity (e.g. MSG (Andolfatto *et al.* 2011); two-enzyme GBS (Poland *et al.* 2012); SLAF-seq (Sun *et al.* 2013); quaddRAD (Franchini *et al.* 2017)). Thus, many published methods, although prone to distinct biases or technical difficulties and subject to a myriad of downstream bioinformatic considerations (van Dijk *et al* 2014; Flanagan & Jones 2017), arguably do not meet proposed criteria for publication with a unique name (e.g. NUAP 2011). The recently upheld US KeyGene patent covering these methods also seems to suggest that, from a legal standpoint, they are not significantly different from one another (US Patent 8,815,512 B2).

Applying a new name to a method may also be problematic because naming implies ownership over the innovations underlying the technique (NUAP 2011). However, in cases where only minor changes are made to an existing protocol, the authors of the new method profit from advances made by others that may comprise the bulk of the methodology. The broad convergence of several methods in Fig. 2.1 makes this issue more complex, as two or more separate groups of authors are essentially claiming ownership over similar techniques that have different names. For instance, ddRADseq-ion (Recknagel *et al.* 2015) and GBS for

semiconductor sequencing platforms (Mascher *et al.* 2013) have both incorporated double digests and modified adaptors for Ion Torrent sequencing, epiGBS (van Gurp *et al.* 2016) and bs-RADseq (Trucchi *et al.* 2016) both incorporate bisulfite sequencing, and several methods have employed some form of sequence capture (Spiked GBS (Rife *et al.* 2016); RADcap (Hoffberg *et al.* 2016); HyRAD (Suchan *et al.* 2016); Rapture (Ali *et al.* 2016); 3RAD (Graham *et al.* 2015); hyRADx (Schmid *et al.* 2017)). Despite this, new protocols continue to be published with unique names.

These factors complicate our ability to distinguish between methods, and therefore how we refer to and attribute them. An important consideration for discussing methods is what name to use, but this is not always straightforward. Ideally, methods should be referred to by their original, published names, but we recognize that the use of general terms for discussing groups of methods is often necessary. In these instances, we recommend that researchers should first assess how protocols are related methodologically and chronologically before proposing an appropriate general term. This should improve clarity when discussing RRS (and NGS) methods, though it does not directly address the continued expansion of names associated with modified protocols.

2.5.2 *"Action is eloquence"* (Shakespeare 1605-08)

Rapid sequencing advances may have unwittingly created a sense of momentum among researchers, thereby fostering the proliferation of names for genome-sampling techniques. The overabundance of names may be partly influenced by how new NGS technology is; for instance, the developers of many derivative RRS techniques may genuinely think their modifications represent a significant deviation from existing protocols and thus warrant unique names.

However, we argue that this alone is not a justification for naming a new method, and further point out that there are clear patterns in the development and publication of derivative RRS techniques that suggest additional motivations for naming methods. Almost half of the methods in Fig. 2.1 and all five of the key methods in Fig. 2.2 were published in *PLoS ONE*. Other journals have also published multiple derivative methods, although to a lesser degree. Several researchers have also been involved in the naming and publication of more than one method, indicating research groups focusing on the development of suites of techniques.

Taken together, this suggests a preoccupation with name recognition as a means to increase the visibility of research. We recognize that catchy titles are not inherently negative or irresponsible, and that this practice can beneficially increase the impact of research. But the recent "modify, name, and publish" trend seems more likely to be driven by efforts to increase citations, which dilutes the eloquence of acronyms. This system further confers risk to researchers who choose not to name an adapted technique by leaving a door open for someone else to employ the same change and name it, taking the credit. While some journals exercise final discretion over whether a submitted protocol should be accompanied with a unique name (e.g. *Nature* (NUAP 2011)), this is not an explicit policy of all journals. And because academic and journalistic success is so closely tied to citation metrics and impact factors, there is little incentive to take the high road.

It is instructive to compare our reliance on easy-to-digest acronyms to online clickbait headlines in academic publishing and research. An example may be the recent publication of an incendiary essay, presumably to increase the impact of a journal, despite the article not passing peer review (Flaherty 2017). Academic metrics do not distinguish between "good" and "bad" citations (Gallien & Roelofs 2017), and we are incentivized to market our research beyond the

merit of the research itself. Sequencing technology will undoubtedly continue to advance (Goodwin *et al.* 2016), and new RRS approaches will continue to evolve. Exploring the utility and limitations of these approaches has resulted in a wealth of biological knowledge that has been hitherto out of reach. At this level, Shakespeare's immortal phrase got it right: "a rose by any other name would smell as sweet" (Shakespeare 1591-94). But Linnaeus may have disagreed with this sentiment – names *do* matter because they serve to communicate and organize the world around us.

There are undoubtedly differences between many published RRS methods that require careful consideration by researchers before they invest in any one approach (for instance, access to equipment, the type of data produced, and informatic requirements or constraints). Despite these considerations, we suggest that in many instances, published RRS modifications do not warrant unique names, and that this has had an overall negative impact on our ability to understand and discuss these methods.

Accordingly, we propose a set of guidelines (Fig. 2.3) that we hope may guide researchers when deciding when and how to name new techniques that modify existing methods. We caution against naming modifications that primarily serve to streamline or reduce the cost of existing protocols (Stage 1, Fig. 2.3), on the basis that it creates a conflict of intellectual ownership, as discussed above. Modifications that change or increase the functionality of an existing method are more likely to necessitate larger methodological changes (Stage 2, Fig. 2.3). In these cases, if it is determined that a new name is necessary, we recommend that names be informative and descriptive of the technique (Stage 3, Fig. 2.3). We also acknowledge that incorporating a reference to the parent technique in the new name is useful for recognizing the originators of the core technique (e.g. ddRADseq, Peterson *et al.* (2012)) and is sometimes

appropriate, but authors should ensure that new names are sufficiently different from the names of unrelated methods. These guidelines are designed to emphasize descriptiveness, innovation, and due diligence by researchers, to ideally minimize redundancy between methods and improve the clarity of new names.

Thus, we add our voices to those of Hadfield and Retief (2018); deciding which innovations are substantial enough to warrant new names is subjective, but our scientific community would be better served by greater restraint in naming new techniques, except for indisputably large methodological innovations. Continued adaptation of methods is clearly beneficial, but the publication of new names for minor changes in existing NGS methodologies is a symptom of a larger cultural shift in academia. And the responsibility for righting that course lies with us as researchers, reviewers, and editors.

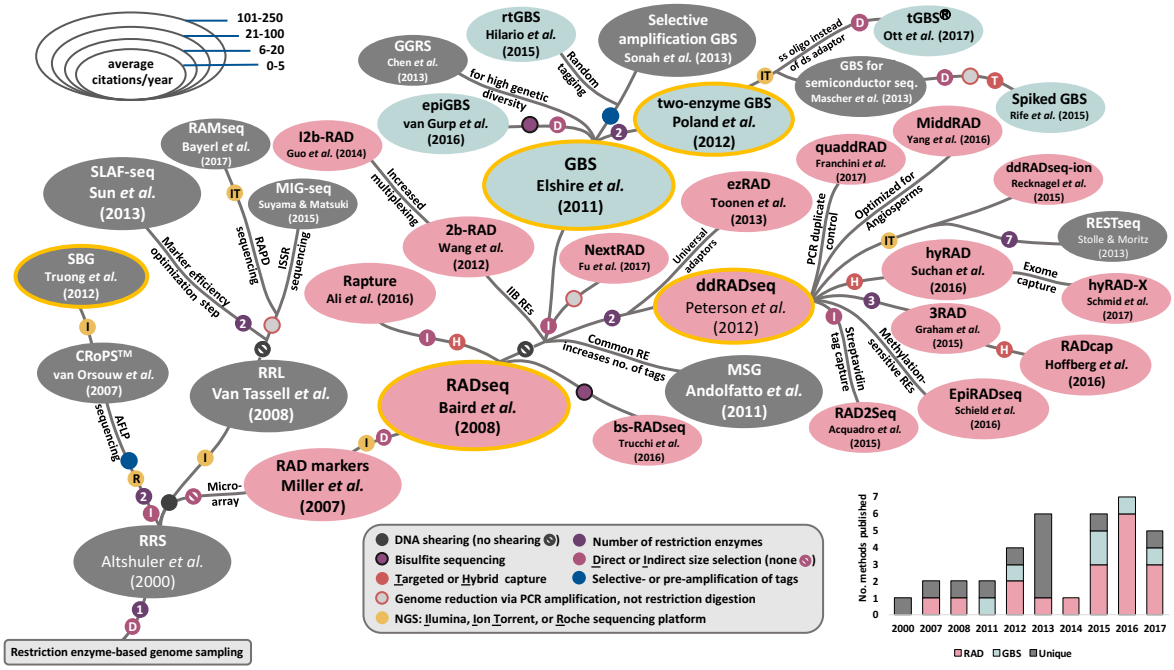


Figure 2.1 Concept map displaying methodological and etymological relationships among 36 reduced representation genome-sampling methods and their derivatives. Branches connect derived methods to their inferred parent protocols, and significant differences between protocols are indicated by coloured circles on branches. Variations that originate only once are indicated by text along branches. Red ellipses indicate named methods using the "RAD" acronym, blue ellipses indicate names derived from "GBS", and methods with unique names, or lacking names altogether, are in grey. The five methods used to assess attribution rates (Fig. 2.2) are indicated by ellipses with a gold outline. Inset histogram shows the accumulation of methods by year. Ellipse size indicates the average number of citations per year (total number of citations divided by the number of years since publication) for each method. See Appendix 2.1 for a summary of each method.

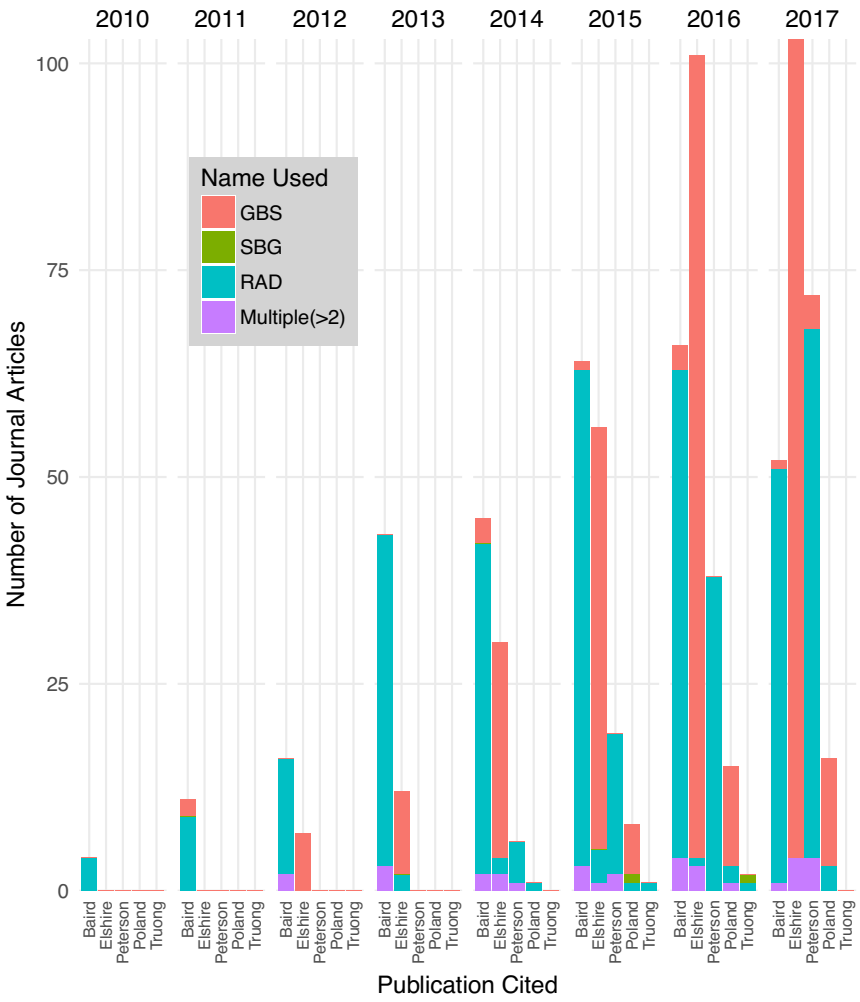


Figure 2.2 Trends in the use of the names "GBS" (from Elshire *et al.* 2011) and Poland *et al.* (2012)), "SBG" (from Truong *et al.* (2012)), and "RAD" (from Baird *et al.* (2008) and Peterson *et al.* (2012)) in the title, abstract, or keywords of journal articles that cite either Baird *et al.* (2008), Elshire *et al.* (2011), Poland *et al.* (2012), Truong *et al.* (2012) or Peterson *et al.* (2012). Bars indicate the number of journal articles citing each publication, while colours indicate the number referring to each name. About 8.4% of papers use an ambiguous or inconsistent name in reference to the cited method (e.g. ~4% of papers uniquely citing Baird *et al.* (2008) in 2017 refer specifically to GBS or SBG alone or in combination with RAD, despite neither name being used in that paper.

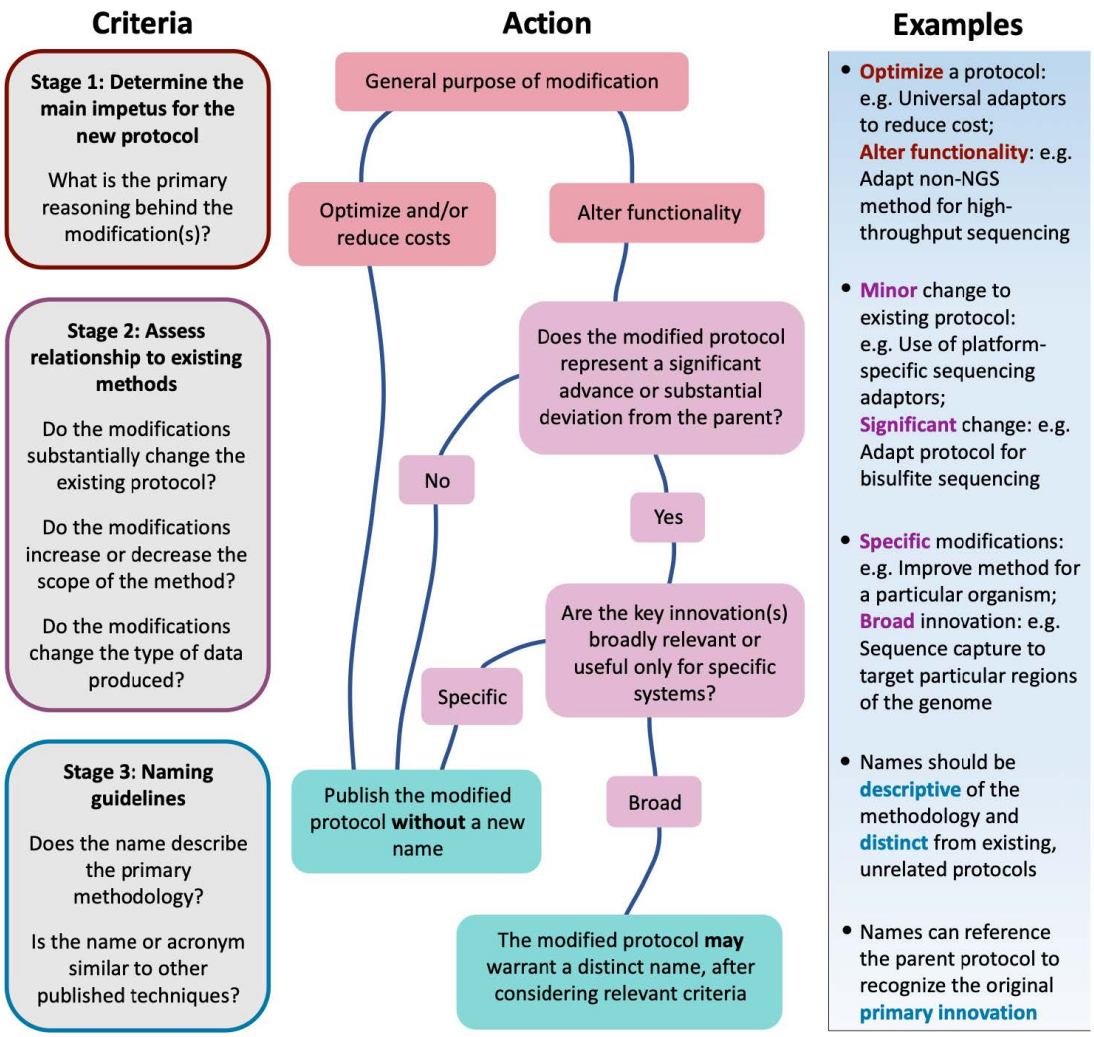


Figure 2.3 To name or not to name? A proposed decision tree for naming modified RRS protocols. To reduce the expansion of names, we recommend restraint when naming modifications of existing protocols. New names should only be used if the modification represents a significant methodological advance or deviation from the parent protocol. Any name given to a modified protocol should be unique, descriptive, and, if appropriate, reference the parent protocol

2.6 Literature cited

Acquadro, A., Barchi, L., Portis, E., Carrasquilla-Garcia, N., Tirone, M., Lanteri, S., & Comino, C. (2016) RAD2seq: an efficient protocol for plant genotyping by sequencing. Acta Horticulturae, **1147**, 1-8.

Ali, O. A., O'Rourke, S. M., Amish, S. J., Meek, M. H., Luikart, G., Jeffres, C., & Miller, M. R. (2016) RAD Capture (Rapture): Flexible and Efficient sequence-based genotyping. Genetics, **202**, 389-400.

Altshuler, D., Pollara, V. J., Cowles, C. R., Van Etten, W. J., Baldwin, J., Linton, L., & Lander, E. S. (2000) An SNP map of the human genome generated by reduced representation shotgun sequencing. Nature, **407**, 513-516.

Andolfatto, P., Davison, D., Erezyilmaz, D., Hu, T. T., Mast, J., Sunayama-Morita, T., & Stern, D. L. (2011) Multiplexed shotgun genotyping for rapid and efficient genetic mapping. Genome Research, **21**, 610-617.

Andrews, K. R., Good, J. M, Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. Nature Reviews Genetics, **17**, 81-92.

Andrews, K. R., & Luikart, G. (2014) Recent novel approaches for population genomics data analysis. Molecular Ecology, **23**, 1661-1667.

Andrews, K. R., Hohenloe, P. A., Miller, M. R., Hand, B. K., Seeb, J. E., & Luikart, G. (2014). Trade-offs and utility of alternative RADseq methods: Reply to Puritz et al. 2014. Molecular Ecology, **23**, 5943-5946.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., … Johnson, E. A. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One, **3**, e3376. https://doi.org/10.1371/journal.pone.0003376

Bayerl, H., Kraus, R.H.S., Nowak, C., Foerster, D.W., Fickel, J., & Kuehn, R. (2017) Fast and cost-effective single nucleotide polymorphism (SNP) detection in the absence of a reference genome using semideep next-generation Random Amplicon Sequencing (RAMseq). Molecular Ecology Resources, **18**, 107-117.

Campbell, E. O., Brunet, B. T. M., Dupuis, J. R., and Sperling, F. A. H. 2018. Would an RRS by any other name sound as RAD? *Methods in Ecology and Evolution*, **9**, 1920-1927.

Catchen, J. M., Hohenlohe, P. A., Bernatchez, L., Funk, W. C., Andrews, K. R., & Allendorf, F. W. (2017) Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. Molecular Ecology Resources, **17**, 362-365.

Chen, Q., Ma, Y., Yang, Y., Chen, Z., Liao, R., Xie, X., … Pan, Y. (2013) Genotyping by Genome Reducing and Sequencing for Outbred Animals. PLoS One **8**, e67500.

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. Nature Reviews Genetics, **12**, 499-510

Eaton, D. A. R. (2014) PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics, **30**, 1844-1849.

Elshire, R.J., Glaubitz, J. C., 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.

Flaherty, M. (2017) Resignations at 'Third World Quarterly'. Inside Higher Ed. Retrieved from https://www.insidehighered.com/news/2017/09/20/much-third-world-quarterlys-editorial-board-resigns-saying-controversial-article

Flanagan, S. P., & Jones, A. G. (2017) Substantial differences in bias between single-digest and double-digest RAD-seq libraries: a case study. Molecular Ecology Resources, **2**, 264-280.

Franchini, P. Parera, D. M., Kautt, A. F., & Meyer, A. (2017) quaddRAD: a new high-multiplexing and PCR duplicate removal ddRAD protocol produces novel evolutionary insights in a nonradiating cichlid lineage. Molecular Ecology, **28**, 2783-2795.

Fu, Z., Epstein, B., Kelley, J. L., Zheng, Q., Bergland, A. O., Carrillo, C. I. C., … Snyder, W. E. (2017) Using NextRAD sequencing to infer movement of herbivores among host plants. PLoS One, **12**, e0177742.

Gallien, M., & Roelofs, P. (2017) What the success of a journal article supporting colonialism says about academia in a clickbait era. Scroll.in. Retrieved from https://scroll.in/article/851215/after-the-media-the-lure-of-clickbait-has-taken-over-academia

Goodwin, S., McPherson J. D., & McCombi, W. R. (2016) Coming of age: ten years of next-generation sequencing technologies. Nature Reviews Genetics**, 27**, 333-351.

Graham, C. F., Glenn, T. C., McArthur, A. G., Boreham, D. R., Kieran, T., Lance, S., … Somers, C. M. (2015) Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADseq), Molecular Ecology Resources, **15**, 1304-1315.

Guo, Y., Fang, D., Song, L., Liu, Y., Liu, Y., Wu, L., … Zhang, H. (2014) An improved 2b-RAD approach (I2B-RAD) offering genotyping tested by a rice (Oryza sativa L.) F2 population. BMC genomics, **15**, 956.

Hadfield, J., & Retief, J. (2018) A profusion of confusion in NGS methods naming. Nature Methods, **15**, 7-8.

Hilario, E., Barron, L., Deng, C. H., Datson, P. M., De Silva, N., Davy, M. W., & Storey, R. D. (2015). Random tagging genotyping by sequencing (rtGBS), an unbiased approach to locate restriction enzyme sites across the target genome. PLoS One, **10**, e0143193.

Hoffberg, S. L., Kieran, T. J., Catchen, J. M., Devault, A. Faircloth, B. C., Mauricio, R., & Glenn, T. C. (2016) RADcap: sequence capture of dual-digest RADseq libraries with identifiable duplicates and reduced missing data. Molecular Ecology Resources, **16**, 1264-1278.

Jiang, Z., Wang, H., Michal, J. J., Zhou, X., Liu, B., Solberg Woods, L. C., & Fuchs, R. A. (2016) Genome wide sampling sequencing for SNP genotyping: methods, challenges, and future development. Int. J. Biol. Sci., **12**, 100-108.

KeyGene (2016). KeyGene's SBG patent upheld by the USPTO after ex parte reexamination. [online] Available at: http://www.keygene.com/wp-content/uploads/2016/03/Press-release-KeyGenes-SBG-patent-upheld-by-the-USPTO-after-ex-parte-reexamination.pdf [Accessed 20 Nov. 2017].

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer, A. (2016) Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. Molecular Ecology Resources, **17**, 142-152.

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer, A. (2017) Responsible RAD: Striving for best practices in population genomic studies of adaptation. Molecular Ecology Resources, **17**, 366-369.

Mascher, M., Wu, S., St. Amand, P., Stein, N., & Poland, J. (2013) Application of genotyping-by-sequencing on semiconductor sequencing platforms: a comparison of genetic and reference-based marker ordering in barley. PLoS One, **8**, e76925.

McKinney G. J., Larson, W. A., Seeb, L. W., & Seeb, J. E. (2017) RADseq provides unprecedented insights into molecular ecology and evolutionary insights: comment on Breaking RAD by Lowry et al. (2016). Molecular Ecology Resources, **17**, 356-361.

Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A., & Johnston, E. A. (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Research, **17**, 240-248.

Narum, S. R., Buerkle, C. A., Davey, J. W., Miller, M. R., & Hohenlohe, P. A. (2013) Genotyping-by-sequencing in ecological and conservation genomics. Molecular Ecology, **22**, 2841-2847.

NUAP (no unnecessary acronyms please). (2011) Nature Methods, **8**, 521.

Ott, A., Liu, S., Schnable, J.C., Yeh, C.-T., Wang, K.-S., & Schnable, P.S. (2017) tGBS® genotyping-by-sequencing enables reliable genotyping of heterozygous loci. Nucleic Acids Research, **45**, e178.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One, **7**, e37135. https://doi.org/10.1371/journal.pone.0037135

Poland, J. A., Brown, P. J., Sorrells, M. E., & Jannink, J. (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One, **7**, e32253.

Puritz, J. B., Matz, M. V., Toonen, R. J., Weber, J. N., Bolnick, D. I., & Bird, C. E. (2014) Demystifying the RAD fad. Molecular Ecology, **23**, 5937-5942.

R Core Team (2017) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org

Recknagel, H. Jacobs, A., Herzyk., P., & Elmer, K. R. (2015) Double-digest RAD sequencing using ion proton semiconductor platform (ddRADseq-ion) with nonmodel organisms. Molecular Ecology Resources, **15**, 1316-1329.

Rife, T. W., Wu, S., Bowden, R. L., & Poland, J. A. (2015) Spiked GBS: a unified, open platform for single marker genotyping and whole-genome profiling. BMC Genomics, **16**, 248.

Schield, D.R., Walsh, M.R., Card, D.C., Andrew, A.L., Adams, R.H., Castoe, T.A., & Bunce, M. (2016) EpiRADseq: scalable analysis of genomewide patterns of methylation using next-generation sequencing. Methods in Ecology and Evolution, **7**, 60-69.

Schmid, S., Genevest, R., Gobet, E., Suchan, T., Sperisen, C., Tinner, W., & Alvarez, N. (2017) hyRAD-X, a versatile method combining exome capture and RAD sequencing to extract genomic information from ancient DNA. Methods in Ecology and Evolution, **8**, 1374-1388.

Shakespeare, W. 1591-94. “Romeo and Juliet” The complete works of William Shakespeare comprising his plays and poems, The World Syndicate Publishing Company, 1930, pp. 901.

Shakespeare, W. 1605-08. “Coriolanus” The complete works of William Shakespeare comprising his plays and poems, The World Syndicate Publishing Company, 1930, pp. 703.

Sonah, H., Bastien, M., Iquira, E., Tardivel, A., Légaré, G., Boyle, B., … Belzile, F. (2013) An improved Genotyping by Sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. PLoS One, **8**, e54603.

Stolle, E., & Moritz, R. F. A. (2013) RESTseq - Efficient benchtop population genomics with RESTriction Fragment SEQuencing. PLoS One, **8**, e63960.

Suchan, T., Pitteloud, C., Gerasimova, N. S., Kostikova, A., Schmid, S., Arrigo, N., … Alvarez, N. (2016) Hybridization capture using RAD probes (hyRAD), a new tool for performing genomic analyses on collection specimens. PLoS One, **11**, e0151651.

Sun, X., Liu, D., Zhang, X., Li, W., Liu, H., Hong, W., … Zheng, H. (2013) SLAF-seq: An efficient method of large-scale De Novo SNP discovery and genotyping using high-throughput sequencing. PLoS One, **8**, e58700.

Suyama, Y., & Matsuki, Y. (2015) MIG-seq: an effective PCR-based method for genome-wide single nucleotide polymorphism genotyping using the next-generation sequencing platform. Scientific Reports, **5**, 16963.

Toonen, R. J., Puritz, J. B., Forsman, Z. H., Whitney, J. L., Fernandez-Silva, I., Andrews, K. R., & Bird, C. E. (2013) ezRAD: a simplified method for genomic genotyping in non-model organisms. PeerJ, **l**, e203.

Trucchi, E., Mazzarella, A. B., Gilfillan, G. D., Lorenzo, M. T., Schönswetter, P., & Paun, O. (2016) BsRADseq: screening DNA methylation in natural populations of non-model species. Molecular Ecology, **25**, 1697-1713.

Truong, H. T., Ramos, A. M., Yalcin, F., de Ruiter, M., van der Poel H. J. A., Huvenaars, K. H. J., … van Eijk, J. T. (2012) Sequence-based genotyping for marker discovery and co-dominant scoring in germplasm and populations. PLoS One, **7**, e37565.

van Dijk, E. L., Jaszczyszyn, Y., & Thermes, C. (2014) Library preparation methods for next-generation sequencing: Tone down the bias. Experimental Cell Research, **32**, 12-20.

van Gurp, T. P., Wagemaker, N. C. A. M., Wouters, B., Vergeer, P., Ouborg, J. N. J., & Verhoeven, K. J. F. (2016) epiGBS: reference-free reduced representation bisulfite sequencing. Nature Methods, **13**, 322-328.

van Orsouw, N. J., Hogers, R. C. J., Janssen, A., Yalcin, F., Snoeihers, S., Verstege, E., … van Eijk, M. J. T. (2007) Complexity reduction of polymorphic sequences (CRoPs^{TM}): a novel approach for large-scale polymorphism discovery in complex genomes. PLoS One, **2**, e1172.

Van Tassell, C. P., Smith, T. P. L., Matukumalli, L. K., Taylor, J. F., Schnabel, R. D., Lawley, C. T., … Sonstegard, T. S. (2008) SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nature Methods, **5**, 247-252.

Wang, S., Meyer, E., McKay, J. K., & Matz, M. V. (2012) 2b-RAD: a simple and flexible method for genome-wide genotyping. Nature Methods, **9**, 808-812.

Wickham, H. (2009) ggplot2: Elegant graphics for data analysis. Springer-Verlag, New York, United States of America. https:// ggplot2.org

Yang, G., Chen, Y., Wang, J., Guo, C., Zhao, L., Wang, X., … Guo, Z. (2016) Development of a universal and simplified ddRAD library preparation approach for SNP discovery and genotyping in angiosperm plants. Plant Methods, **12**, 39.

Chapter 3

Cross-platform compatibility of *de novo*-aligned SNPs in a non-model butterfly genus

A version of this chapter has been published as Campbell EO, Davis CS, Dupuis JR, Muirhead K, and Sperling FAH (2017) "Cross-platform compatibility of de novo-aligned SNPs in a non-model butterfly genus", Molecular Ecology Resources volume 17, pages e84-e93.

3.1 Summary

High-throughput sequencing methods for genotyping genome-wide markers are being rapidly adopted for phylogenetics of non-model organisms in conservation and biodiversity studies. However, the reproducibility of SNP genotyping and degree of marker overlap or compatibility between datasets from different methodologies have not been tested in non-model systems. Using double-digest restriction site associated DNA sequencing, we sequenced a common set of 22 specimens from the butterfly genus *Speyeria* Scudder, 1872 on two different Illumina platforms, using two variations of library preparation. We then used a *de novo* approach to bioinformatic locus assembly and SNP discovery for subsequent phylogenetic analyses. We found a high rate of locus recovery despite differences in library preparation and sequencing platforms, as well as overall high levels of data compatibility after data processing and filtering. These results provide the first application of NGS methods for phylogenetic reconstruction in *Speyeria,* and support the use and long-term viability of SNP genotyping applications in non-model systems.

3.2 Introduction

Double-digest restriction site associated DNA sequencing, or ddRAD (Peterson *et al.* 2012) is a next-generation sequencing (NGS) technique that reduces genome complexity and allows sequencing of hundreds or thousands of loci across many individuals at relatively low cost. This method is particularly suited for non-model systems because it requires no *a priori* genomic information for locus development. Instead, ddRAD uses restriction enzymes to obtain thousands or millions of loci across the genome, which are subsequently amplified and sequenced. The number of shared restriction sites between populations and species is correlated to genetic relatedness (Avise *et al.* 1979), making ddRAD an appealing and powerful method for shallow-scale phylogenetic research (DaCosta & Sorenson 2015). Given these characteristics, it is likely that ddRAD will become more widely used in this field. However, despite the methodological advantages of this approach, the lack of a reference genome in non-model systems may impose limitations on these data. In particular, dataset combination may be compromised if there is no reference available to ensure that the same loci are sequenced across platforms. To the best of our knowledge, no research has investigated whether *de novo*-aligned ddRAD data are compatible in this manner.

Traditional molecular phylogenetics (using a small number of genes) has relied heavily on publicly available data repositories (NCBI GenBank) and marker types (single copy ortholog gene sequences) that are easily combined across datasets. Historical accumulation of publicly available data allows researchers to augment datasets and examine more complete phylogenies when their own data collection is limited. Reliability and compatibility of cross-platform data is an advantage here, and markers that do not meet these criteria have limited usefulness. For

instance, the irreproducibility of RAPDs, which are short DNA fragments that identify single nucleotide polymorphisms in randomly amplified regions of the genome (Williams *et al.* 1990), has resulted in researchers abandoning this method in favor of more reliable approaches (Penner *et al.* 1993; Jones *et al.* 1997).

If next-generation sequencing methods like ddRAD are to become a mainstay for phylogenetics, it is important to determine if the same level of data sharing can be achieved using these approaches as for gene sequence data. Online repositories for raw NGS sequence data such as Dryad and the NCBI Sequence Read Archive are already in place as platforms for data sharing, but it remains to be seen whether it is feasible to combine different restriction-based sequencing methodologies, such as ddRAD (Peterson *et al.* 2012) or genotyping-by-sequencing (GBS, Poland *et al.* 2012).

Many factors affect whether restriction enzyme-based datasets can be combined. Library preparation methods, sequencing platforms, and the sampling bias inherent in reduced-representation genomic sequencing approaches may affect which loci are represented in any given dataset (Davey *et al.* 2013; DaCosta & Sorenson 2014). Much like a specific set of primers, the use of the same restriction enzymes ensures that the same sites are cut in each individual across sequencing runs, and is therefore essential for dataset combination in both model and non-model systems. The use of a reference genome can help to standardize other differences such as variation in the number of unique versus common loci by aligning sequence reads to a common reference catalog. This approach has demonstrably high rates of concordance among SNPs sequenced in different laboratories and on different platforms (Hong *et al.* 2012; DaCosta & Sorenson 2014). Studies using non-model systems, which make up the bulk of phylogenetic research, may be disadvantaged by the lack of a reference genome for identifying

and cataloging loci. This can potentially limit the combinability of *de novo*-aligned SNP data generated using different protocols and on different sequencing platforms. Thus, continued investment in ddRAD and other, similar, methods on a large scale will likely be influenced by whether *de novo*-aligned data produced in different ways can be made compatible through bioinformatic data processing.

Here, we test the reproducibility of *de novo*-aligned SNPs generated using different sequencing platforms and library preparation methods, and explore the effect of data combination on phylogenetic inference in a group lacking genomic resources, the butterfly genus *Speyeria* Scudder 1872. Species in this genus exhibit high intraspecific geographic variability in wing coloration and pattern (dos Passos & Grey 1947). Field identification is further complicated by extensive regional sympatry and parapatry between species and an apparent lack of species-specific host plant use, with all *Speyeria* feeding on members of the genus *Viola* Linnaeus (Violaceae) (Brittnacher *et al.* 1978; Hammond 1981). As a result, morphological and ecological delimitation of these species and their relationships has historically been challenging. Allozymes (Brittnacher *et al.* 1978) and a few mitochondrial and nuclear genes (Dunford 2007; de Moya *et al.* 2017) have shown some congruence between molecular and broad-scale morphological species delimitation (but see McHugh *et al.* 2013), but species relationships within the genus remain unclear. Several species of *Speyeria* are at risk of extirpation or extinction due to human-mediated shifts in land use and climate change (Hammond 1981; Breed *et al.* 2012; McHugh *et al.* 2013), providing further incentive to clarify relationships in this genus. A fine-scale, genome-wide approach using ddRAD will help to resolve these relationships.

3.3 Methods

3.3.1 *DNA extraction and NGS sequencing*

Twenty-two *Speyeria* butterflies were collected by aerial net across western North America. Two outgroup specimens of the species *Argynnis aglaja* (Linnaeus, 1758), a close relative of *Speyeria* (Simonsen *et al.* 2006), were collected near Åland, Finland and Banesque, Spain. In total, 24 specimens were processed and sequenced, comprising 10 putative species of *Speyeria* and 1 species of *Argynnis* Fabricius, 1807**.** Species identifications were verified by the authors using wing characters and known species range information (Warren *et al.* 2012). DNA was isolated from legs and thoracic tissue of each specimen using a DNeasy Blood & Tissue DNA purification kit (Qiagen) with RNAse A treatment. The isolated DNA was ethanol precipitated, re-suspended in Millipore water, and kept frozen at -20°C until use. Collection localities for all specimens are listed in Appendix 3.1.

Two NGS library preparation methods and sequencing on two platforms were conducted for each specimen using 200 ng of input DNA. First, a GBS library was prepared by the Institut de biologie integrative et des systèmes (IBIS) at Université Laval in Quebec City, QC, using *PstI/MspI* restriction enzymes in accordance with the protocol of Poland *et al.* (2012). A complexity reduction step was included during library preparation, with a single cytosine added to the reverse primer in order to preferentially amplify only 25% of all double digest restriction fragments (Sonah *et al.* 2013). Individual-specific indexes were 4 to 8 base pairs (bp) in length. Single-end, 100 bp sequencing was then performed on an Illumina HiSeq 2000 at the McGill University-Génome Québec Innovation Centre in Montreal, QC.

Using DNA from the same specimens, a second library was constructed at the Molecular Biology Services Unit (MBSU) at the University of Alberta in Edmonton, AB. Library

construction followed the protocol of Peterson *et al.* (2012), with two exceptions: first, to make this data compatible with the data generated at the McGill University-Génome Québec Innovation Centre following Poland *et al.* (2012), we omitted the flex oligos of Peterson *et al.* (2012) in favor of *MspI*- and *PstI*-specific oligos, ensuring that only fragments containing these sites were sequenced. Secondly, we did not perform a size selection on the data, as was done by Peterson *et al*. (2012). We did not incorporate the complexity reduction step of Sonah *et al.* (2013) in the NextSeq library preparation procedure, thereby generating differences in the number of loci sequenced and the read depth of each locus across datasets. Individual indexes were all 8 bp in length. Single-end, 75 bp sequencing was performed on a single high output flowcell of an Illumina NextSeq 500 housed in the MBSU.

3.3.2 *Data processing*

Raw reads generated on both platforms were demultiplexed and split into individual data files using the *process radtags* program in version 1.35 of Stacks (Catchen *et al.* 2011). The NextSeq 500 reads were truncated to 67 bp, which was the length of each read after removing the 8 bp index sequence. Manual inspection of the data revealed sequencing error in the *PstI* site in some reads, so we used Cutadapt version 1.9.1 (Martin 2011) to trim 5 bp from the 5' end of each read to prevent erroneous SNP calling. Resulting reads were 62 bp in length. The HiSeq 2000-generated 100 bp reads were also truncated to 62 bp in the same way to ensure downstream compatibility in Stack*s*, which requires uniform read length when building loci *de novo*. Reads that failed Illumina's chastity filters and had Phred quality scores below 20 were discarded.

After trimming and quality-filtering raw sequence reads, we used the Stacks programs *ustacks*, *cstacks*, and *sstacks* to build *de novo* catalog loci using various combinations of both the

HiSeq- and NextSeq-generated sequences. We specified a minimum of 5 reads per locus to build primary catalog stacks, and permitted a maximum distance of 2 nucleotide mismatches within these stacks. We allowed 3 mismatches between primary stacks and secondary reads. Because the specimens in our dataset represented multiple species across two genera, we also allowed 1 nucleotide mismatch between final catalog loci; this merged putatively differentially fixed versions of the same locus into a single locus. Finally, we used a chi-square significance value of 1% to call heterozygote and homozygote loci.

After catalog construction, we used the *populations* program in Stacks for final filtering of our data. As our dataset was relatively small, with 1 to 5 specimens per species, we assigned all 24 individuals to a single population. In addition, we specified a minimum stack depth of 5 for each locus. After running *populations*, minimum coverage per locus was set to 80%, and coverage per individual was filtered manually to a minimum of 75% for ingroup specimens and 50% for outgroup specimens, which had fewer restriction sites in common with the ingroup.

To assess compatibility between sequencing platforms, we produced 7 different combinations of the HiSeq and NextSeq data: 1. NextSeq data only (n=24); 2. HiSeq data only (n=24); 3. all HiSeq and NextSeq sequences (n=48), with each individual represented twice (once per sequencing platform); and 4. to 7. four datasets where sequences for each individual were randomly chosen without replacement from either the NextSeq or HiSeq data (datasets R_1-R_4, each n=24). These randomized datasets represent scenarios where additional samples from online repositories might be added to a researcher-generated dataset to increase sampling.

We filtered each of the 7 datasets using 3 different minor allele frequency (MAF) thresholds: 3%, 5%, and 10%. These values span the MAF range commonly used in model systems (Tabangin *et al.* 2009). Testing multiple MAF thresholds allowed us to assess the

downstream effect of allelic representation on phylogenetic inference. In addition to allelic variation, differences in read depth between the sequencing platforms may result in alleles being under-sampled and erroneously removed from the dataset. For instance, if the minor allele of a locus is rare in a population and does not meet the specified MAF threshold, it might then be treated as sequencing error and the locus will be discarded. Furthermore, under-sampled heterozygote loci may have only one allele present in a dataset, and so even if they are retained, insufficient read depth may lead to false homozygote calling (Lynch 2009; Nielsen *et al.* 2012). Disproportionate representation of rare loci and variations in read depth in combined datasets may bias genotyping and impact downstream phylogenetic analyses, but this has not yet been explored in a *de novo* locus-generated context.

de novo catalog construction and filtering in Stacks was done using custom perl wrappers. VCFtool*s* version 1.12b (Danecek *et al.* 2011) was used for filtering each dataset for locus coverage and for calculating mean read depth.

3.3.3 *NGS dataset comparisons*

After generating *de novo* catalogs separately for the HiSeq and NextSeq datasets, we used BLAST+ version 2.3.0 (Camacho *et al.* 2009) to find matches between loci sequenced on either platform. Since the HiSeq dataset had fewer sequenced loci due to the complexity reduction step, we used the *de novo* HiSeq catalog generated in Stacks to build a custom BLAST database. We then queried the *de novo* NextSeq catalog from Stacks against this HiSeq database to determine the number of loci that were sequenced in common on both HiSeq and NextSeq platforms.

We used the online *Ident and Sim* tool from The Sequence Manipulation Suite (Stothard 2000) to determine percent sequence similarity among all individuals via pairwise comparisons

in the combined NextSeq and HiSeq dataset filtered with an intermediate MAF of 5%. We compared HiSeq- and NextSeq-generated sequences for each individual, as well as among individuals and species. We consider high sequence similarity between the two sets of sequences to indicate that our filtering parameters removed variation due to different loci being sequenced between platforms.

3.3.4 *Phylogenetic and population analyses*

Phylogenetic analyses used datasets comprised of SNPs along with the invariant flanking region for each locus. We ran Maximum Likelihood analyses using the IQ-TREE web server version 1.5.0 (Trifinopoulos *et al.* 2016) using the best-fitting DNA substitution model (as determined by IQ-TREE) with 1000 ultrafast bootstrap replicates (Minh *et al.* 2013) and 1000 replicates of the SH-aLRT branch test (Guindon *et al.* 2009).

To further evaluate the delimitation of evolutionary units in *Speyeria*, we performed Bayesian clustering in Structure (Pritchard *et al.* 2000), which assigns individuals to populations or discrete groups (denoted as *K*) based on allele frequencies. This analysis tested for "run effects" where sequences may cluster according to platform or sequencing run rather than according to their species designation. For this analysis, we only used the combined HiSeq and NextSeq dataset with a MAF of 5% (ingroup specimens only, two sequences per individual, n=44). Since there were 10 putative species of *Speyeria* represented in the data, and each species was sequenced on two platforms, we tested values of *K* between 1 and 20. We replicated each value of *K* ten times with 500,000 MCMC reps and a burn-in period of 50,000. We used version 0.6.94 of the Structure Harvester web server (Dent & vonHoldt, 2012) to determine the optimal value of *K* using both the Evanno method (Evanno *et al.* 2005) and Pritchard method (Pritchard

et al. 2000), and then combined and permuted each of the 10 replicates for the optimal *K* into a single file for visualization using the program CLUMPP version 1.1.2 (Jakobsson & Rosenberg, 2007). All file conversions from vcf to other formats used PGDSpider version 2.0.8.2 (Lischer & Excoffier, 2012).

After observing some topological discordance between the resulting phylogenies, we conducted Shimodaira-Hasegawa (SH) topology tests (Shimodaira & Hasegawa, 1999) on the combined HiSeq and NextSeq dataset with parameters following the GTR+I model in version 4.0b of PAUP* (Swofford, 2002). This test compares the likelihood scores of two or more topologies against a provided data matrix, and outputs a *p*-value indicating whether or not the topologies differ significantly ($p \leq 0.05$) from one another. Because we generated trees using three different MAF thresholds, we had three data matrices that contained different loci. Therefore, we conducted pairwise comparisons for each of three trees using each of three data matrices (9 tests in total).

3.4 Results

3.4.1 *NGS sequencing and comparison*

After initial per-base quality filtering and demultiplexing, we retained an average of 1.3 million reads per individual for the HiSeq 2000 data, and an average of 0.7 million reads per individual for the NextSeq 500 data. The *Speyeria* and *Argynnis* specimens were multiplexed and sequenced in both runs with other DNA samples, as 96-plex on the HiSeq 2000 and 192-plex on the NextSeq 500.

The HiSeq dataset had the smallest catalog (Table 3.1). In total there were 69,753 catalog loci constructed *de novo* in Stacks, as compared to 105,503 loci in the NextSeq catalog.

The combined HiSeq and NextSeq dataset had the largest catalog, with 133,330 loci. The four randomized datasets had catalogs that were intermediate in size, ranging from 88,331 to 94,998 loci. In each dataset a lower MAF threshold resulted in the retention of more loci and more SNPs per locus when compared to higher MAF thresholds. There were 35,801 unique, identical locus matches shared between the HiSeq and NextSeq catalogs, which was 51% of the catalog loci in HiSeq dataset and 34% of the total number of catalog loci present in the NextSeq dataset.

Mean filtered read depth per locus varied across each dataset (Table 3.1), but did not vary substantially by MAF threshold: each of the 3 MAF thresholds per dataset differed by only 1-3 reads. While the NextSeq dataset had a large catalog, it had the lowest mean read depth at 50 reads per locus. In contrast, the HiSeq dataset had the smallest catalog, but the highest mean read depth at 171 reads per locus. The four randomized datasets and the combined HiSeq and NextSeq dataset had intermediate mean read depths that ranged from 109 to 125.

Pairwise percent sequence similarity between the HiSeq and NextSeq data for the two outgroup specimens versus ingroup specimens ranged from 28% to 48% (Fig. 3.1). Percent sequence similarity among ingroup specimens ranged from 45% to 85%, and pairwise comparisons within each individual ranged from 80% to 98%. Interspecific percent sequence similarity of ingroup specimens was generally 10% - 25% lower than intraspecific percent sequence similarity. Four sequences, the HiSeq-generated *S. aphrodite 9634* and *S. callippe 9638*, as well as the HiSeq- and NextSeq-generated sequences for *S. zerene 10417*, had lower interspecific percent sequence similarity values than any other ingroup specimens. These specimens also had the highest proportions of missing data among ingroup specimens, ranging between 17% and 22%.

3.4.2 *Phylogenetic and population analyses*

In all 21 maximum likelihood trees (3 trees from each of 7 datasets), *S. nokomis* (Edwards, 1862) was consistently the most basal species within *Speyeria*. *Speyeria cybele* (Fabricius, 1775) and *S. aphrodite* (Fabricius, 1787) were always sister species to one another, as were *S. atlantis* (Edwards, 1862) and *S. hesperis* (Edwards, 1864). *Speyeria hydaspe* (Boisduval, 1869), *S. callippe* (Boisduval, 1852), *S. egleis* (Behr, 1862), and *S. zerene* (Boisduval, 1852) always grouped together in a derived clade (hereafter referred to as the *S. hydaspe* clade). However, we also observed several topological incongruences. One topology occurred 14 out of 21 times, and is referred to as the dominant topology. Three alternate topologies were present in the remaining 7 trees, and are discussed below. All specimens clustered together according to their morphological identifications. Figure 3.2 depicts representative cladograms for the dominant and alternate topologies. A summary of each analysis is found in Table 3.2.

For the 3 alternate topologies, one had *S. mormonia* (Boisduval, 1869) in a derived position near the *S. hydaspe* clade in the R_2 dataset with MAF thresholds of 3% and 10%, rather than the basal position near *S. nokomis* as in the dominant topology. The clade containing *S. hesperis* and *S. atlantis* had a sister relationship to the *S. cybele* and *S. aphrodite* clade in the dominant topology, but the second alternate topology (for the R_3 dataset with MAF thresholds of 3% and 5%, as well as the R_4 dataset with a MAF of 3%) instead had a sister relationship between *S. atlantis/S. hesperis* and the *S. hydaspe* clade. A third alternate topology that contained both of these differences was obtained from the combined HiSeq and NextSeq dataset with MAF thresholds of 3% and 10%. *Speyeria hesperis* was also depicted as paraphyletic with respect to *S. atlantis* in several datasets with a MAF of 10%, and in one dataset with a MAF of 5%. Phylogenetic trees for each analysis are in Appendix 3.2.

Regardless of whether a given tree depicted the dominant or an alternate topology in the combined HiSeq and NextSeq dataset, all maximum likelihood analyses consistently grouped both sequences for each individual together with branch lengths of 0 or nearly 0, and with bootstrap support of 100 (Fig. 3.3). Structure analysis predicted an optimal *K* value of 3, and depicted near identical cluster assignments for both sequences in each individual. The SH tests compared the dominant topology, present in the tree generated using a MAF of 5%, against the alternate topology D, present at MAF thresholds of both 3% and 10% (seen in Fig. 3.2 and Appendix 3.2) in the combined HiSeq and NextSeq dataset for each MAF threshold. We chose this dataset because these two topologies were the most different in terms of interspecies relationships. The results of the SH tests were non-significant ($p > 0.05$; Appendix 3.3).

3.5 Discussion

3.5.1 *Reproducibility and compatibility of de novo*-aligned *ddRAD data*

As expected, the *de novo* NextSeq catalog recovered more loci than the *de novo* HiSeq catalog - albeit with an overall lower mean read depth per locus. This can be attributed to the complexity reduction step in the library prep of the HiSeq-generated data (Sonah *et al.* 2013). Despite differences in number of loci and read depth, SNP calling in the combined datasets appears to be robust. This is shown by overall high sequence similarities between the NextSeq- and HiSeq-generated sequence reads for each specimen. While percent sequence similarity for each individual in the combined HiSeq and NextSeq dataset ranged between 80% - 98% after filtering, even individuals with the lowest similarity values consistently clustered together in both phylogenetic and population genetic analyses (Fig. 3.3). In addition, sequence similarity across species decreased as phylogenetic distance increased (Fig. 3.1). These results are of

particular significance for demonstrating that *de novo*-aligned ddRAD data is not only reproducible across sequencing runs and platforms, but also bioinformatically compatible in the absence of reference-based tools for locus construction and SNP calling.

The low sequence similarity values and high amounts of missing data in the HiSeq-generated *S. aphrodite 9634* and *S. callippe 9638* and the HiSeq- and NextSeq-generated *S. zerene 10417* sequences likely indicate fewer loci in common with the rest of the dataset. Despite this, all four sequences consistently clustered in their respective species clades in all analyses, indicating that these sequences retained enough informative SNPs for downstream phylogenetic inferences.

3.5.2 *Phylogenetic and population results*

Our results consistently show the same few trees for the combined datasets, despite some topological incongruence. Interestingly, low and high MAF thresholds were more likely to show alternate topologies, while a more intermediate MAF threshold of 5% resulted in the depiction of the dominant topology for every dataset except R_3. Filtering using MAF involves a trade-off between removing potential sequencing error and retaining uncommon, but informative, variants. This is most easily observed in the *S. hesperis* species cluster, which was rendered paraphyletic with respect to *S. atlantis* in most datasets that used a more stringent MAF threshold (Appendix 3.2). This change in topology indicates a close relationship between these two clades, but also a loss of informative SNPs that support clade distinction due to overly strict filtering parameters.

The Structure results and SH tests further support these observations. Structure only discriminated 3 clusters in the data rather than one for each of the 10 putative species (Fig. 3.3). These three clusters loosely correspond to larger species complexes in the group, but notably

reveal admixture in some species: both *S. mormonia* and *S. hesperis* share ancestry with the *S. hydaspe* clade. These results suggest that the phylogenetic inconsistencies of *S. mormonia* and *S. hesperis/S. atlantis* are at least partially explained by close evolutionary relationships. The SH tests similarly indicate that these topological differences were minor, as all tests were non-significant (Appendix 3.3).

Given these observations, it is unlikely that differences in read depth contributed directly to the observed topological incongruences. The combined HiSeq and NextSeq dataset, which depicted an alternate topology when filtered using minor allele frequencies of 3% and 10%, had an average read depth per locus of 109. This is far higher than the thresholds that have been shown to impact genotyping (Nielsen *et al.* 2011; Buerkle & Gompert, 2012, Andrews *et al.* 2016). Moreover, the NextSeq dataset had a mean read depth of 50, yet depicted the dominant topology for all three MAF thresholds. Therefore, the inclusion or exclusion of uncommon loci due to sampling bias and locus dropout may be a more likely explanation for why much of the topological variation occurred in phylogenetic trees with more extreme MAF thresholds and those inferred from combined datasets (Andrews *et al.* 2016).

3.5.3 *Phylogenetic considerations*

Phylogenetic discordance in *Speyeria* has been previously documented, and may be attributed in part to overall close relationships between species in this genus. A recent paper used molecular dating to estimate that the divergence of *Speyeria* from *Argynnis* likely occurred 6-5 MYA, and was followed by rapid diversification of *Speyeria* across North America (de Moya *et al.* 2017). It is likely that close relationships in this genus complicate the genetic and morphological diagnosability of species. For instance, Brittnacher *et al.* (1978) used allozymes to

infer species delimitation in *Speyeria* collected from California. Their results differ substantially from both our dominant and alternate topologies. In addition, Dunford (2007) depicted several topological inconsistencies between phylogenies from morphological and gene sequence data. Brittnacher *et al.* (1978) also noted lower levels of genetic divergence in *Speyeria* than other related lepidopteran species and, despite some morphological differences between species, they failed to detect any diagnostic allozyme loci for *S. zerene, S. atlantis, S. callippe,* and *S. egleis*. Recent divergences in this genus are further supported by Hammond *et al.* (2013), whose laboratory hybridization study demonstrated that most species produced viable hybrid offspring when mated by hand.

Likely due to these putatively close relationships, Dunford (2007) and McHugh *et al.* (2013) found at least some degree of paraphyly and polyphyly for many morphological species. For McHugh *et al.* (2013) this was particularly prevalent in *S. zerene, S. atlantis, S. callippe,* and *S. egleis*, while Dunford (2007) demonstrated this to a lesser degree. Incomplete lineage sorting and introgression between closely related lineages has been shown to result in incongruences between datasets (Sang & Zhong 2000; Pollard *et al.* 2006), and provide likely explanations for the observed topological discordance presented in this study.

Our results indicate consistent monophyly for species represented by more than one individual, with the only exception limited to the already discussed paraphyly of *S. hesperis* in some analyses. In addition, all species represented in this dataset by a single individual were clearly distinguished from species clades containing multiple individuals. Thus, despite the topological incongruence of our competing phylogenies, this fine-scale genome-wide approach has characterized what are likely to be recent divergence events within *Speyeria*.

Recent divergence, as well as the effect of uncommon alleles in combined datasets, provides a plausible explanation for the observed phylogenetic discordance between and within datasets. Testing this method in a system with clearer species boundaries may help tease apart the effects of these two factors. Other explanations for more broad scale phylogenetic incongruences in *Speyeria* such as the effects of missing data and incomplete taxon sampling (see Nabhan & Sarkar, 2012, Roure *et al.* 2012, and Huang & Knowles, 2016) should be explored, but are beyond the scope of this paper.

3.5.4 *Conclusions*

This study is, to our knowledge, the first to explore the potential compatibility of ddRAD and two-enzyme GBS datasets using *de novo* locus assembly. We demonstrate a high degree of genotyping consistency across platforms, and high bioinformatic compatibility across a range of data combinations and filtering parameters. In addition, our study substantiates the utility of these methods for revealing fine-scale patterns of divergence and admixture in closely related, non-model species. Our results support the current species delimitation of several species of *Speyeria*, though interspecific relationships within this genus were less consistent, and warrant future focus.

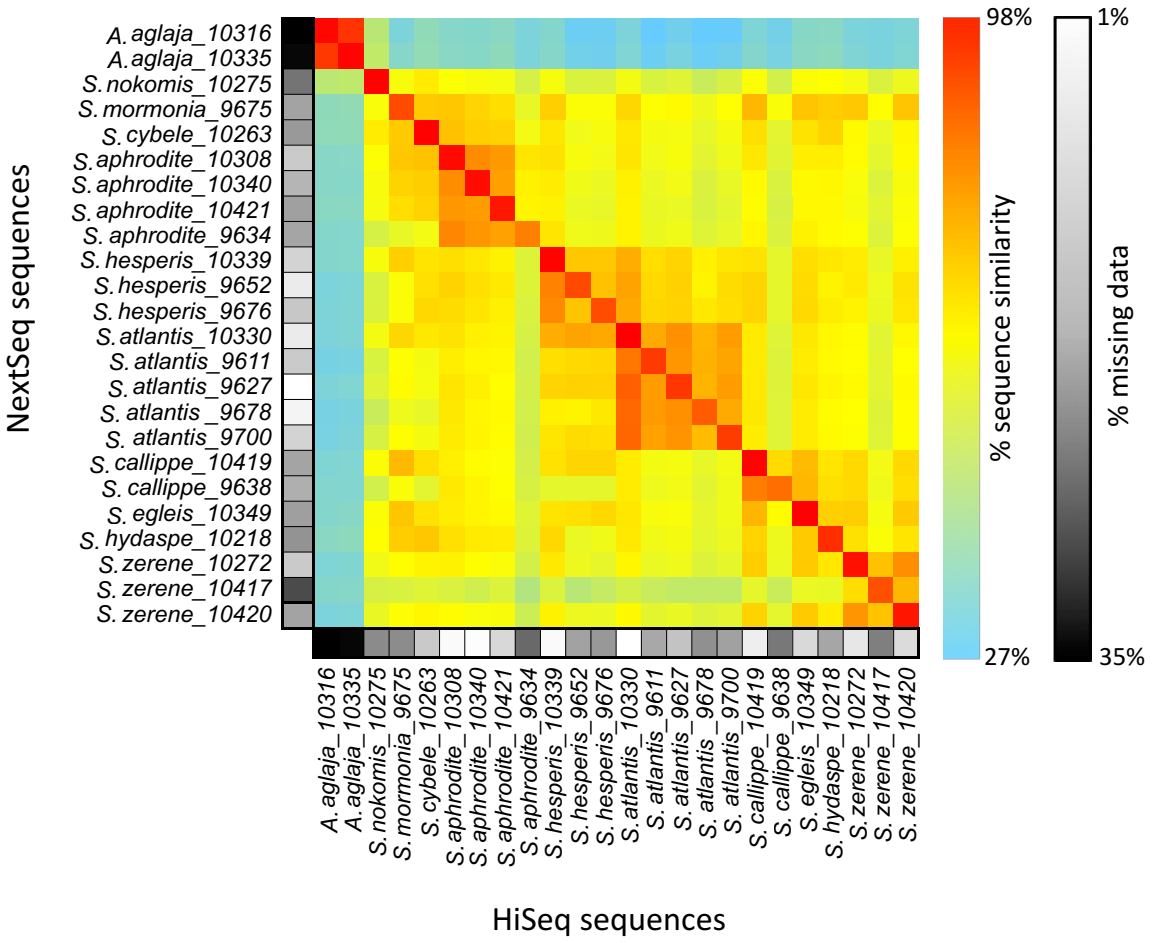


Figure 3.1 Pairwise percent sequence similarity and percent missing data for all sequences in the combined HiSeq and NextSeq dataset filtered with a MAF of 5%. HiSeq sequences, on the x-axis, were compared to NextSeq sequences along the y-axis. Similarity values range from 27% to 98%. The highest observed similarity values were between sequences of the same specimen, and range between 80% and 98%, with an average of 93%.

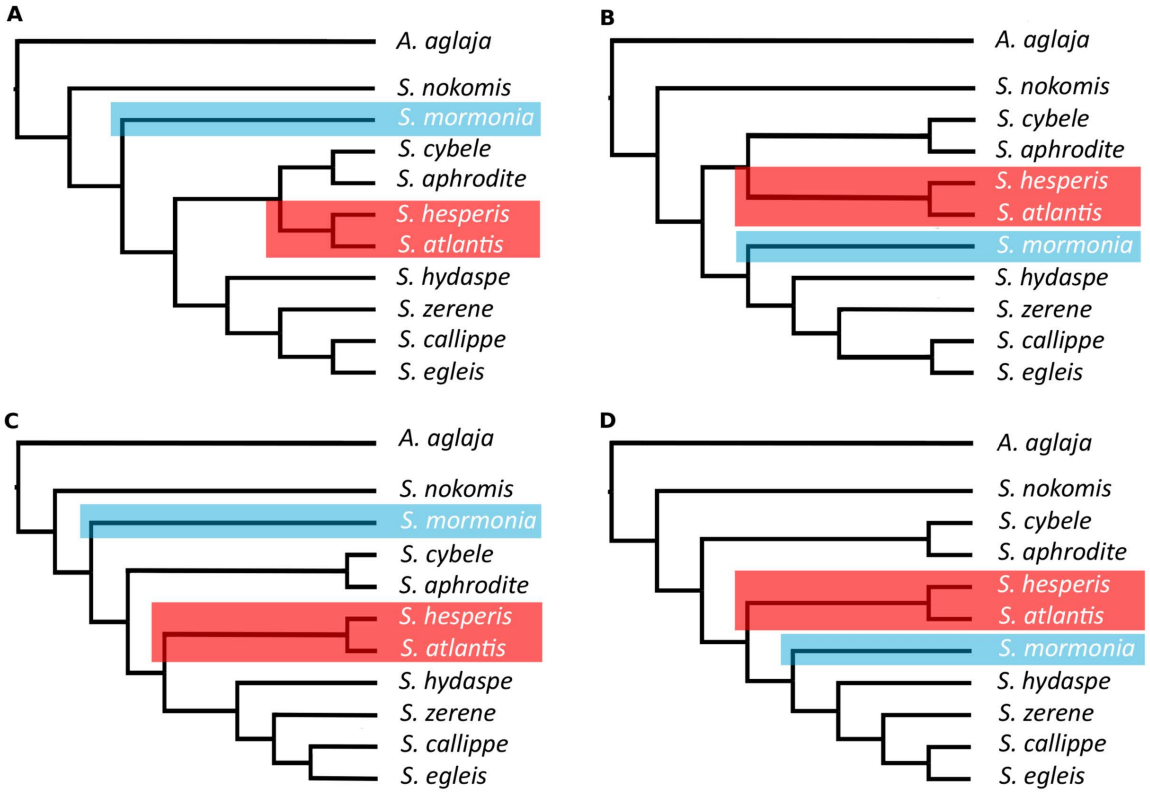


Figure 3.2 Cladograms for the dominant topology (a) and three alternate topologies (b-d) for ten *Speyeria* species and one outgroup species. The number of SNPs and loci in each dataset are summarized in Table 3.1. Maximum-likelihood phylogenies for each of the 21 analyses are in Appendix 3.2.

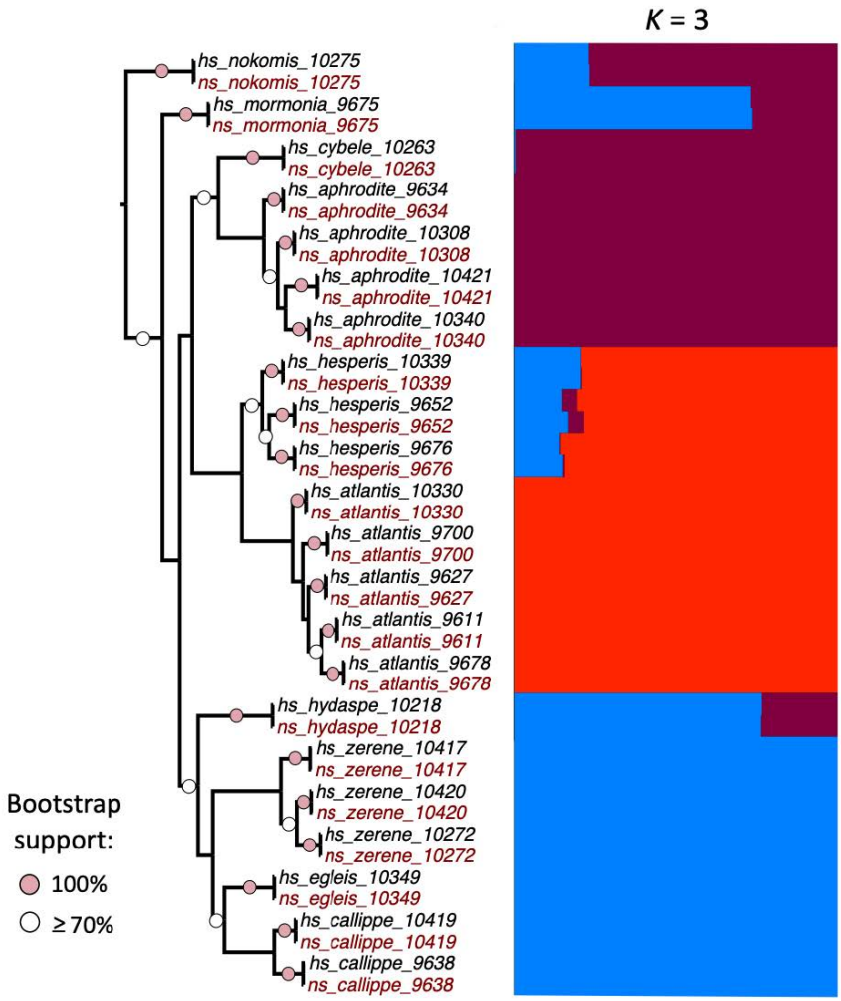


Figure 3.3 Maximum-likelihood tree (left) and Structure plot (right) for the combined HiSeq and NextSeq dataset with a MAF of 5%. Each of 22 ingroup specimens has two sets of concatenated loci, one sequenced on a HiSeq 2000 (hs) after a complexity reduction step and another, without the complexity reduction step, sequenced on a NextSeq 500 (ns). Terminals on the tree correspond to adjacent bars in the Structure plot.

Table 3.1 Summary information for each of seven datasets filtered for three minor allele frequency (MAF) thresholds

				3% MAF		5% MAF		10% MAF	
Dataset	No. Catalog loci	$\bar{x}$ Read depth	σ Read depth	No. loci	No. SNPs	No. loci	No. SNPs	No. loci	No. SNPs
HiSeq	69,753	171	149	1,169	2,485	967	1,679	570	749
NextSeq	105,503	50	31	1,847	4,334	1,556	2,919	974	1,305
All	133,330	109	85	1,052	2,364	887	1,573	531	706
R_1	93,273	117	95	928	2,037	770	1,363	447	580
R_2	94,998	125	106	1,109	2,480	928	1,670	546	724
R_3	88,331	124	101	1,053	2,276	878	1,539	507	664
R_4	94,716	119	98	969	2,174	816	1,470	487	642

Table 3.2 Dataset descriptions and summary of phylogenetic relationships

		Resulting topology		
Dataset	Composition	MAF 3%	MAF 5%	MAF 10%
HiSeq	n=24, all HiSeq	A	A	A
NextSeq	n=24, all NextSeq	A	A	A
All	24 HiSeq, 24 NextSeq	D	A	D
R_1	15 HiSeq, 9 NextSeq	A	A	A
R_2	13 HiSeq, 11 NextSeq	B	A	B
R_3	13 HiSeq, 11 NextSeq	C	C	A
R_4	14 HiSeq, 10 NextSeq	C	A	A

3.6 Literature cited

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016) Harnessing the power of RADseq for ecological and evolutionary genetics. *Nature Reviews Genetics*, **17**, 81- 92.

Avise, J. C., Lansman, R. A., & Shade, R. O. (1979) The use of restriction endonucleases to measure mitochondrial DNA sequence relatedness in natural populations. I. Population structure and evolution in the genus *Peromyscus*. *Genetics*, **92**, 279-295.

Breed, G. A., Stichter, S., & Crone, E. E. (2012) Climate-driven changes in northeastern US butterfly communities. *Nature Climate Change*, **3**, 142-145.

Brittnacher, J. G., Sims, S. R., & Ayala, F. J. (1978). Genetic differentiation between species of the genus *Speyeria* (Lepidoptera: Nymphalidae). *Evolution*, **32**, 199-210.

Buerkle, C. A. & Gompert, Z. (2013) Population genomics based on low coverage sequencing: how low should we go? *Molecular Ecology*, **22**, 3028-3035.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**:421. DOI: 10.1186/1471-2105-10-421

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011) Stacks: building and genotyping loci de novo from short-read sequences. *G3*, **1**, 171-182.

DaCosta, J. M. & Sorenson, M. D. (2015) ddRAD-seq phylogenetics based on nucleotide, indel, and presence-absence polymorphisms: Analyses of two avian genera with contrasting histories. *Molecular Phylogenetics and Evolution*, **94**, 122-135.

DaCosta, J. M. & Sorenson, M.D. (2014) Amplification biases and consistent recovery of loci in a double-digest RAD-seq protocol. *PLoS One*, **9**, e106713.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E, DePristo, M. A., … 1000 Genomes Analysis Group (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156-2158.

Davey, J. W., Cezard, T., Fuentes-Utrilla, P., Eland, C., Gharbi, K., & Blaxter, M. (2013) Special features of RAD sequencing data: implications for genotyping. *Molecular Ecology*, **22**, 3151-3164.

de Moya, R. S., Savage, W. K., Tenney, C. Bao, X., Whalberg, N., & Hill, R. I. (2017) Interrelationships and diversification of *Argynnis* Fabricius and *Speyeria* Scudder butterflies. *Systematic Entomology*. DOI: 10.1111/syen.12236

Dent, E. A. & vonHoldt, B. M. (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **2**, 359-361.

dos Passos, C. R. & Grey, L. P. (1947) Systematic catalogue of *Speyeria* (Lepidoptera, Nymphalidae) with designations of types and fixations of type localities. *American Museum Novitates*, **1370**, 1-30.

Dunford, J. C. (2007) The genus *Speyeria* and the *Speyeria atlantis/Speyeria hesperis* complex: species and subspecies accounts, systematics, and biogeography (Lepidoptera: Nymphalidae). Dissertation, University of Florida.

Evanno, G., Regnaut, S., & Goudet, J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611-2620.

Guindon, S., Dufayard, J., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2009) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, **59**, 307-321.

Hammond, P. C. (1981) The Colonization of violets and *Speyeria* butterflies on the ash-pumice fields deposited by Cascadian volcanoes. *Journal of Research on the Lepidoptera*, **20**, 179-191.

Hammond, P. C., McCorkle, D. V. and Bergman, W. (2013) Hybridization studies of genomic compatibility and phenotypic expression in the greater fritillary butterflies (Nymphalidae: Argynnini). *Journal of the Lepidopterists' Society,* **67**, 263-273.

Hong, H., Xu, L., Lio, J., Jones, W. D., Su, Z., Ning, B., …Shi, L. (2012) Technical reproducibility of genotyping SNP arrays used in genome-wide association studies. *PLoS One*, **7**, e44483.

Huang, H. & Knowles, L. L. (2016) Unforeseen consequences of excluding missing data from next-generation sequences: simulation study of RAD sequences. *Systematic Biology*, **65**, 357-365.

Jakobsson, M, & Rosenberg, N. A. (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801-1806.

Jones, C. J., Edwards, K. J., Castaglione, S., Winfield, M. O., Sala, F., van de Wiel, C., … Karp, A. (1997) Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Molecular Breeding*, **3**, 381-390.

Lischer, H. E. L. & Excoffier, L. (2012) PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics,* **28**, 298-299.

Lynch, M. (2009) Estimation of allele frequencies from high-coverage genome-sequencing projects. *Genetics Society of America*, **182**, 295-301.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, **17**, 10-12.

McHugh, A., Bierzychudek, P., Greever, C., Marzulla, T., Van Buskirk, R., & Binford, G. (2013) A molecular phylogenetic analysis of *Speyeria* and its implications for the management of the threatened *Speyeria zerene Hippolyta*. *Journal of Insect Conservation*, **17**, 1237-1253.

Minh, B. Q., Nguyen, M. A. T., & von Haeseler, A. (2013) Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution,* **30**, 1188-1195.

Nabhan, A. R. & Sarkar, I. N. (2011) The impact of taxon sampling on phylogenetic inference: a review of two decades of controversy. *Briefings in Bioinformatics*, **13**, 122-134.

Nielsen, R., Korneliussen, T., Albrechtsen, A., Li, Y., & Wang, J. (2012) SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data, *PLoS One*, **7**, e37558.

Nielsen, R., Paul, J. S., Albrechtsen, A., & Song, Y. S. (2011) Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics*, **12**, 443-451.

Penner, G. A., Bush, A., Wise, R., Kim, W., Domier, L., Kasha, K., … Fedak, G. (1993) Reproducibility of Random Amplified Polymorphic DNA (RAPD) analysis among laboratories. *Genome Research*, **2**, 341-345.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012) Double Digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS One*, **7**, e37135.

Poland, J. A., Brown, P. J., Sorrells, M. E., & Jannink, J. (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS One*, **7**, e32253.

Pollard, D. A., Iyer, V. N., Moses, A. M., & Eisen, M. B. (2006) Widespread discordance of gene trees with species trees in *Drosophila*: evidence for incomplete lineage sorting. *PLoS Genetics*, **2** e173.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics,* **155,** 945-959.

Roure, B., Baurain, D., & Philippe, H. (2012) Impact of missing data on phylogenies inferred from empirical phylogenomic data sets. *Molecular Biology and Evolution*, **30**, 197-214.

Sang, T., & Zhong, Y. (2000) Testing hybridization hypotheses based on incongruent gene trees. *Systematic Biology*, **49**, 422-434.

Shimodaira, H. & Hasegawa, M. (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution*, **16**, 1114-1116.

Simonsen, T. J., Wahlberg, N., Brower, A. V. Z., & de Jong, R. (2006) Morphology, molecules and fritillaries: approaching a stable phylogeny for Argynnini (Lepidoptera: Nymphalidae). *Insect Systematics and Evolution*, **37**, 405-418.

Sonah, H., Bastien, M., Iquira, E., Tardivel, A., Légaré, G., Boyle, B., … Belzile, F. (2013) An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. *PLoS One*, **8**, e54603.

Stothard, P. (2000) The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniqu*es, **28**, 1102-1104.

Swofford, D. L. (2002) PAUP*: Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts

Tabangin, M. E., Woo, J. G., & Martin, L. J. (2009) The effect of minor allele frequency on the likelihood of obtaining false positives. *BMC Proceedings*, **3**, S41.

Trifinopoulos, J., Nguyen, L. T., von Haeseler, A., & Minh, B. Q. (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucl. Acids. Res.,* **44**, W232-W235.

Warren, A. D., Davis, K. J., Grishin, N. V., Pelham, J. P., & Stangeland, E. M. (2012) Interactive listing of American butterflies. [accessed Nov 15 2016]. http://www.butterfliesofamerica.com

Williams, J. G. K., Kubelik, A. R., LIvak, K. J., Rafalski, J. A., & Tingey, S. V. (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*, **16**, 6531-6535.

Chapter 4

Single nucleotide polymorphism-based species phylogeny of greater fritillary butterflies (Lepidoptera: Nymphalidae: *Speyeria*) demonstrates widespread mito-nuclear discordance

A version of this chapter has been published as Campbell EO, Gage EV, Gage RV, and Sperling FAH (2019) "Single nucleotide polymorphism-based species phylogeny of greater fritillary butterflies (Lepidoptera: Nymphalidae: Speyeria) demonstrates widespread mito-nuclear discordance", Systematic Entomology, DOI:10.1111/syen.12393.

4.1 Summary

The systematics of *Speyeria* Scudder, 1872 butterflies has historically been complicated by intraspecific variability that has challenged efforts to delimit species and reconstruct phylogenies. Our study presents a phylogenetic comparison of genomic single nucleotide polymorphisms (SNPs) and mitochondrial *COI* gene sequences, with comprehensive taxon sampling that includes 15 species and 46 subspecies. Increased sampling of genetic markers and taxa improved the match between genetic clusters, obtained with both phylogenetic and cluster-based analyses, and species previously detected using morphology, as well as showing two species delimitations that may need revision. We also recovered extensive mito-nuclear discordance between genomic SNPs and the *COI* gene, confirming that mitochondrial DNA does not reliably identify several species at broad geographic scales. Resolution of the relationships of *Speyeria* species demonstrates the importance of sampling variation across the whole genome, and provides an essential foundation for understanding the evolution of this charismatic clade of North American butterflies.

4.2 Introduction

Speyeria Scudder, 1872 is a charismatic clade of greater fritillary butterflies endemic to North America that is notorious for its confusing intraspecific morphological variation (Moeck 1975; Hammond *et al.* 2013). Variously ranked as a genus (Hammond 1981; de Moya *et al.* 2017; Hill *et al.* 2018) or a subgenus (Simonsen *et al.* 2006), sixteen *Speyeria* species are recognised in Pelham's (2008) widely used catalogue of North American butterflies. Morphological species identification has typically relied on wing pattern and colour, as well as differences in size and sexual dimorphism (Moeck 1975; Dunford 2009). Much of the variation within and between *Speyeria* species is geographically correlated (Hammond 1978) and shows a wide range of wing colour patterns within some species (e.g. *S. hesperis* (Edwards, 1864)*, S. callippe* (Boisduval, 1852)) while different species can be morphologically similar in regions of sympatry and parapatry. It has therefore historically been difficult to delimit species of *Speyeria* and to characterise their phylogenetic relationships, leading to a comment that "the literature is replete with erroneous determinations and many more are of doubtful validity" (dos Passos & Grey 1947). Today, despite the integration of DNA sequence data, species relationships of *Speyeria* remain largely unresolved (Dunford 2007; McHugh *et al.* 2013; de Moya *et al.* 2017), yet a stable phylogeny would support growing conservation concern for several species (Hammond 1995; Breed *et al.* 2012; McHugh *et al.* 2013; Wells & Tonkin 2014; Sims 2017; Hill *et al.* 2018).

Numerous lines of evidence suggest close relationships within *Speyeria*. These species are the only North American representatives of the Argynnini*,* a large tribe of fritillaries with a Holarctic distribution (Simonsen *et al.* 2006). *Speyeria* likely colonised and began to radiate across North America approximately 6 million years ago (de Moya *et al.* 2017), with several

species diversifying after the last glacial maximum 20,000 years ago. Larvae of all species feed on *Viola* Linnaeus (Violaceae) with little or no apparent host-plant specialisation among species (Hammond 1981; Dunford 2009), and *S. mormonia* (Boisduval, 1869) may also feed on *Bistorta bistortoides* (Pursh) Small, 1906 (Polygonaceae) (Wolfe 2017). While differences in genitalic characters have been proposed as the basis for separating *Speyeria* into two major groups, *Semnopsyche* and *Callippe* (dos Passos & Grey 1945; Hammond 1978), genitalic characters are not useful for species identifications, and many species of *Speyeria* can be hybridised in laboratory conditions to produce viable offspring (Hammond *et al.* 2013).

Because of morphological ambiguity and apparently close relationships among many *Speyeria*, recent studies have used molecular characters to clarify relationships among species, typically using mitochondrial or a few nuclear genes (Dunford 2007; McHugh *et al.* 2013, de Moya *et al.* 2017). Their results were often discordant with one another and showed high levels of paraphyly and polyphyly among morphologically recognised species (Fig. 4.1). This may be explained by various combinations of: 1. incorrect morphology-based delimitation or identification of species; 2. unresolved molecular phylogenies due to the use of few markers with low phylogenetic signal (Dupuis *et al.* 2012); 3. limited taxon sampling that does not reflect species-level genetic diversity (Nabham & Sarkar 2012); or 4. biologically ambiguous relationships due to retained ancestral polymorphism and/or hybridisation events (Rosenberg 2003; Leaché & McGuire 2006). A recent exception to the findings of para/polyphyly among *Speyeria* species is a study published by Hill *et al.* (2018). They focused on the utility of the mitochondrial *COI* gene for identifying four sympatric *Speyeria* species that inhabit the San Francisco Bay area and recovered monophyletic mtDNA haplotypes for each of them. However, many of the systematic challenges in this clade occur at larger scales (i.e.: the species level),

creating a need for a more inclusive phylogeny that incorporates a greater proportion of the genetic diversity within *Speyeria.*

Only one study has used increased marker sampling via genome-wide SNPs (Campbell *et al.* 2017). This study was a methodological assessment of the analytical compatibility between different next-generation sequencing (NGS) datasets; taxon sampling was limited due to practical constraints. Genome-wide SNPs recovered monophyletic groupings of *a priori* species, with variable branch support (Fig. 4.1), and revealed probable discordance within the nuclear genome that would otherwise be missed using low numbers of markers. The next step toward a stable molecular phylogeny for *Speyeria* is thus to extend sampling to include more species and subspecies across a broader geographic range.

Here, we sample fifteen species and approximately 46 subspecies of *Speyeria* across the United States and Canada, representing the most comprehensive sampling effort for a molecular systematic analysis of *Speyeria* to date. We conduct a phylogenetic comparison between genome-wide SNPs and mitochondrial *COI* sequence to identify mito-nuclear discordance, and use cluster analyses to assess genetic structuring at the species level. With this approach, we aim to better identify likely sources of phylogenetic discordance and unstable species delimitation in *Speyeria*, and in doing so, provide a framework for future studies to further clarify the evolutionary history of this group.

4.3 Methods

4.3.1 *Specimen collection and identification*

Unless otherwise specified in Appendix 3.1, butterflies were collected using aerial nets at sites across Canada and the United States. Three *Speyeria diana* (Cramer, 1779) specimens were

reared and preserved for this study. When possible, samples were immediately frozen at -20°C, but where this was not feasible the samples were preserved in 70% or 95% ethanol until they were frozen.

Given the known difficulties inherent in morphological identification of *Speyeria* species and subspecies, we approached specimen identification iteratively. Species and subspecies identifications were initially determined using morphological field markings and species ranges as described in several major resources (Moeck 1975; Bird *et al.* 1995; Brock & Kaufman 2003; Dunford 2009; Warren *et al.* 2012). We additionally consulted the Bean Museum collection at Brigham Young University, as well as a personal reference collection (E. Gage). We then ran preliminary molecular phylogenetic analyses on these specimens using SNPs (following methodologies outlined below) to assess the consistency of our morphological identifications. From these analyses, we re-assessed the morphological identification of specimens that did not cluster genetically with other specimens with the same morphological *a priori* species identification. This approach indicated initial species misidentifications of three specimens of *S. aphrodite ethne* (Hemming, 1933) *as S. hesperis ratonensis* Scott, 1981*,* and four *S. zerene picta* (McDunnough, 1924) specimens were misidentified as *S. egleis* (Behr, 1862)*,* one as *S. callippe,* and four *S. zerene zerene* (Boisduval, 1852) specimens were misidentified as *S. hesperis.* In each case, the initial misidentification was due to morphological convergence between species inhabiting the same or geographically proximate areas. Misidentifications were corrected by comparing the original descriptions of putative morphological species against the genetic groups obtained from cluster-based and phylogenetic analyses (Boisduval 1852, 1869; Edwards 1864).

This iterative approach also allowed us to use the recovered phylogenetic clusters to inform additional sampling (which was conducted over several years) to maximise subspecific

diversity in our dataset. Subspecies identities were verified by consulting original descriptions (Edwards 1869, 1872, 1874, 1879, 1886; Comstock 1925), additional summaries of subspecies morphology (Bird *et al.* 1995; Dunford 2007), described subspecies ranges (Moeck 1975), and finally compared to specimen photos on the Butterflies of America website (Warren e*t al.* 2012) which contains images for all recognised specific and subspecific taxa, including numerous type specimens. The names of species and subspecies used in our study follow Pelham (2008).

In total, we analysed 155 *Speyeria* specimens representing fifteen of the sixteen recognised species (Pelham 2008). *Speyeria adiaste* (Edwards, 1864) was not included in the dataset due to poor DNA quality of the few specimens we had, and logistical difficulties arising from its small distribution restricted to a few areas in California (Zaman *et al.* 2015) that made re-sampling unfeasible, but we were able to access mitochondrial sequence on GenBank for this species. For the fifteen species sampled, we identified a total of 46 putative morphological subspecies (about one third of the currently recognised diversity) based on a combination of morphology, phylogeny, and collection locality information, as outlined above. When possible, a minimum of ten specimens per *a priori* species were included in the dataset, however a few species (*S. diana, S. idalia* (Drury, 1773), *S. carolae* (dos Passos & Grey, 1942)*,* and *S. edwardsii* (Reakirt, 1866)) were represented by numbers ranging from two to eight specimens. We also included three outgroup specimens comprising two species, *Argynnis aglaja* (Linnaeus, 1758) and *Argynnis paphia* (Linnaeus, 1758), which were collected in Europe.

4.3.2 *DNA extraction and sequencing*

DNA was isolated using a Qiagen DNeasy kit with RNAse A treatment, and each sample was prepared for Sanger sequencing of the mitochondrial *COI* barcoding region following

Hebert *et al.* (2003). For next-generation sequencing (NGS), we followed Campbell *et al.* (2017), with library preparation and single-end sequencing performed using either the ddRAD protocol of Peterson *et al.* (2012) on an Illumina Nextseq 500 or the two-enzyme GBS protocol of Poland *et al.* (2012) on an Illumina Hiseq 2000. Details of the methodology used for each specimen can be found in Appendix 4.1.

4.3.3 *Mitochondrial data processing*

For each specimen, the forward and reverse sequences of the barcoding region of the *COI* gene were aligned into a single consensus sequence and then quality checked by eye using Geneious v. 10.1.3 (www.geneious.com). A *COI* gene sequence for *S. adiaste,* published in de Moya *et al.* (2017), was downloaded from GenBank (accession number in Appendix 4.1) and included in the mitochondrial dataset to complete the species-level sampling for all sixteen recognised species (Pelham 2008). A multiple sequence alignment was then constructed in MAFFT online version 7 (Katoh *et al.* 2017) using default settings and then manually inspected and trimmed to 648 basepairs in Mesquite 3.51 (Maddison & Maddison 2018).

4.3.4 *SNP data processing and parameter testing*

ddRAD and two-enzyme GBS data were processed using the *de novo* Stacks 2.0b pipeline (Catchen *et al.* 2011) on the Cedar cluster hosted by Compute Canada. The ddRAD and two-enzyme GBS sequence data were different lengths, as they were sequenced on different Illumina platforms, and sequence read trimming and initial data processing followed Campbell *et al.* (2017). The final length of each sequence read input into Stacks for *de novo* locus construction was 62 basepairs.

Following Paris *et al.* (2017), we conducted testing to determine the optimal values for Stacks parameters *M* and *n*, which control the number of mismatches allowed between alleles in the same individual (*M*) and between individuals (*n*) during locus construction (Paris *et al.* 2017; Rochette & Catchen 2017). Low values for *M* and *n* tolerate fewer mismatches between sequence reads, while higher values of *M* and *n* tolerate a greater number of mismatches and should therefore result in fewer loci being built by Stacks, but an overall increase in per-locus polymorphism. In general, higher values are appropriate in scenarios where evolutionary divergence is relatively great and/or sampled populations are known to be highly polymorphic, while lower values are more appropriate for closely related or less polymorphic populations. Since our sequence reads were shorter than those in Paris *et al.* (2017), we tested a subset of the parameter values that represented the same approximate proportion of mismatches (*M* = 0-5, *n* = 0-6). We also adhered to the "r80" principle suggested in Paris *et al.* (2017), which only outputs loci that are found in 80% of the specimens for any given population (in our case, population = *a priori* species). We additionally specified a maximum per-locus heterozygosity value of 80%, and only required any given locus to be present in at least one population.

During parameter testing and final data processing in Stacks, we output all SNPs for each locus. Some studies have advocated the inclusion of invariant flanking sequence in phylogenetic analysis of SNPs (Wagner *et al.* 2013; Leaché *et al.* 2015), however given the evolutionary distance in this study we expected to retain fewer loci in our dataset relative to other studies at shallower evolutionary scales (Arnold *et al.* 2013; Lee *et al.* 2018), as well as to find an overall increase in per-locus polymorphism (Paris *et al.* 2017). While parameter testing with the *r80* rule reduces the number of spurious or paralogous loci being built in Stacks, it does not control removal of potentially erroneous SNPs *within* a locus; instead, those SNPs are typically filtered

by specifying a minimum genotype quality score and/or minor allele frequency. Therefore, if we included the invariant region of each locus in the dataset, after filtering we would have to remove entire loci in order to correct for a small number of potentially erroneous SNPs. In an effort to retain as much data as possible for phylogenetic analysis we opted to output SNPs without the invariant flanking region and removed only those individual SNPs that failed to meet our minimum quality thresholds.

Using vcftools 0.1.14 (Danecek *et al.* 2011), we filtered the global dataset at a minimum genotype quality score of 30, a minimum minor allele frequency of 3%, and total missing data per locus at a maximum of 50% for subsequent phylogenetic analyses. For population structure analyses, which are more sensitive to genomic linkage and missing data, we output only a single random SNP from each locus and reduced total missing data per locus to a maximum of 20%.

4.3.5 *Phylogenetic reconstruction*

We conducted phylogenetic analyses of the mitochondrial and SNP datasets using both maximum likelihood and Bayesian inference. Bayesian inference was conducted in MrBayes 3.2.6 (Ronquist *et al.* 2012) using four chains, a mixed model, and a gamma rate variation distribution. The mitochondrial dataset was additionally run with the invariant sites model (*invgamma*). MrBayes analyses of mitochondrial and SNP datasets were allowed to run until the standard deviation of split frequencies approached 0.01, the potential scale reduction factor approached 1, and ESS values were high (>200), indicating stationarity. We conducted maximum likelihood inference in IQ-TREE 1.3.10 **(**Nguyen *et al.* 2015). Model testing to infer the optimal substitution model for both datasets was conducted using this program (Kalyaanamoorthy *et al.* 2017) and analyses were run with 1000 replicates each of ultrafast

bootstrapping (Hoang *et al.* 2018), and SH-aLRT branch testing. We used the ASC+ flag in IQ-TREE to correct for possible ascertainment bias in the SNP dataset due to the aforementioned exclusion of invariant flanking sequence from each locus. MrBayes trees were summarised as 50% majority rule consensus trees, and posterior probability values were plotted on each branch. Maximum likelihood analyses were summarised as extended majority rule consensus trees, an approach that first builds a tree using nodes found in at least 50% of the retained trees and then iteratively adds less-supported nodes until the tree is fully resolved. Bootstrap values were plotted on each branch.

4.3.6 *Cluster-based analyses*

We ran Structure 2.3.4 (Pritchard *et al.* 2000) on the filtered SNP data for fifteen putative *Speyeria* species (n=155) for five million generations with a burn-in period of 10%, using the admixture model and without specifying *a priori* population assignments. We tested values of *K* between 1 and 17, with ten replicates per *K.* We then ran hierarchical analyses to assess substructure in unresolved clusters. Each of these substructure analyses was run for 1 million generations with a 10% burn-in period. For each analysis, the optimal *K* was inferred considering the methods of both Evanno *et al.* (2005) and Pritchard *et al.* (2000), and replicate runs were visualized and averaged together in CLUMPAK (Kopelman *et al.* 2015). Finally, we conducted hierarchical Principle Component Analyses (PCA) in R 3.4.3 (R Core Team 2017) using the package adegenet 2.1.1 (Jombart 2008). The resulting PCAs were plotted with ggplot2 2.2.1 (Wickham 2016).

4.4 Results

4.4.1 *SNP parameter testing and genetic data processing*

We assessed parameter optimality by graphing the number of monomorphic vs. polymorphic loci and counting the number of SNPs output for each parameter value, and then determined the value (or range of values) at which these numbers appeared to stabilise (Paris *et al.* 2017). Overall, variation in these numbers stabilised at moderate parameter values of 3 to 4 for both *M* and *n*. The parameter combination of *M*3*n*4 yielded the highest number of polymorphic loci (16,088) and the highest number of SNPs (64,714) prior to downstream filtering, and was thus considered optimal. We also observed high levels of polymorphism in the data, as expected. The number of SNPs per locus ranged from 1 to 27, although most loci had a moderate number of SNPs (approximately 56% of loci had three or fewer SNPs) and very few loci were extremely polymorphic. We also observed patterns of missing data that largely corresponded to species-specific locus dropouts. Further filtering of the data for phylogenetic analysis retained 2,458 SNPs across 825 unique loci (supplementary data, File S1), and more stringent filtering for population structure analysis yielded 208 SNPs (1 SNP for each of 208 unique loci). Mean read depth per site ranged from 11 to 253 with an average value of 56, and mean read depth per individual varied from 7 to 144 with an average of 57.

Visual inspection of the aligned *COI* data matrix (supplementary data, File S2) confirmed sequence homology across all 158 specimens, and missing data was minimal at approximately 0.09%. This dataset contained 496 constant sites, 29 singleton sites, and 123 parsimony-informative sites.

4.4.2 *Phylogenetic reconstruction*

Stationarity in the MrBayes analyses of the mitochondrial dataset was checked and confirmed after 1 million generations, and after 5 million generations for the SNP dataset. Phylogenetic reconstructions of the SNP dataset and the mitochondrial dataset were generally congruent across Bayesian and maximum likelihood methods, but not between SNP markers and mitochondrial gene sequence. We therefore limit the following discussion to trees derived from the Bayesian analyses only, as the majority rule consensus trees output by MrBayes better represent uncertainty in the data by collapsing poorly supported nodes. The maximum likelihood trees can be found in Appendix 4.2.

The SNP phylogeny was fully resolved and generally highly supported, with posterior probability support values > 0.9 for each species cluster, and each *a priori* species was recovered as monophyletic with the exception of *S. hesperis* and *S. coronis* (Behr, 1864), which were paraphyletic in relation to *S. atlantis* (Edwards, 1862) and *S. carolae,* respectively (Fig 2a). The mitochondrial phylogeny was also generally well supported, although less resolved; one well supported clade contained specimens of *S. callippe, S. egleis, S. edwardsii, S. zerene* (Boisduval, 1852)*,* and *S. coronis*, making each of these species polyphyletic with respect to mitochondrial *COI* (Fig. 4.2b). Two distinct clusters of *S. cybele* (Fabricius, 1775) were indicated by both SNP and mitochondrial analyses, with these clusters corresponding to populations that range west versus east of the Rocky Mountains. The same pattern of distinct populations to the southwest and northeast of the Rockies was recovered for *S. hesperis,* but only with the SNP phylogeny.

We observed substantial mito-nuclear discordance between the SNP and the *COI* phylogenies (Fig. 4.2), with extensive polyphyletic and paraphyletic groupings on the mitochondrial phylogeny, and different SNP and mitochondrial sister relationships between

several species. For instance, *S. idalia* was recovered as the most basal ingroup species on the SNP phylogeny, but the mitochondrial phylogeny recovered *S. nokomis* (Edwards, 1862) in this position. The SNP phylogeny also placed *S. aphrodite* (Fabricius, 1787) as the sister group to *S. cybele*, but the mitochondrial analysis instead depicted *S. hydaspe* (Boisduval, 1869) as the sister to *S. cybele. Speyeria hesperis* and *S. atlantis* were paraphyletic on the SNP phylogeny, but were recovered as more distantly related on the mitochondrial tree. *Speyeria hydaspe* was paraphyletic with a single specimen of *S. adiaste* on the mitochondrial phylogeny.

4.4.3 *Cluster-based analyses*

Structure analysis for SNPs of the fifteen *a priori* species indicated an optimal *K* of 13, although *K*=3 was also moderately supported by Δ*K* (Fig. 4.3a). At *K*=3, *S. coronis, S. egleis, S. callippe,* and *S. carolae* were recovered as a single cluster, however hierarchical analysis of this group resolved all but *S. carolae. K*=3 depicted all other species except for *S. nokomis* as being comprised of varying degrees of "pink", "green", or "grey" ancestry. *K*=13 resolved several of these species, including *S. hydaspe, S. aphrodite, S. atlantis, S. mormonia, S. edwardsii,* and *S. idalia*, however *S. diana, S. cybele,* and *S. hesperis* were not fully resolved with *K*=13. Hierarchical analysis of species that contained only "green" or "grey" ancestry with Q values (proportions of admixture) between 0.2-0.8 suggested *K*=6, but this analysis had similar results to *K*=13 and did not further resolve species clusters.

Structure analyses depicted two unresolved clusters for *S. cybele* that corresponded to the same eastern and western *S. cybele* groups present in the SNP and mitochondrial phylogenetic analyses; the western cluster of *S. cybele* was admixed with *S. hydaspe* for *K*=13 and the hierarchical analysis, which corresponds with the sister relationship between these two species

on the mitochondrial phylogeny. Structure also recovered two distinct groupings of *S. hesperis*, with one resolved cluster that corresponded to the *S. hesperis* clade southwest of the Rocky Mountains on the SNP phylogeny, and a second unresolved cluster that was admixed with *S. atlantis* and corresponded to the *S. hesperis* clade northeast of the Rocky Mountains in the SNP phylogeny.

We additionally observed multiple instances of individual admixed specimens in the Structure analysis, which were primarily limited to *S. zerene, S. hesperis*, *S. mormonia,* and to a lesser extent *S. aphrodite* (Fig. 4.3a). Many of these specimens also exhibited discordance in their species assignment on the SNP and mitochondrial phylogenies (Fig 2); for instance some specimens clustered with *S. zerene* on the SNP tree, but with *S. hesperis* on the mitochondrial tree.

The Principle Component Analysis of fifteen *Speyeria* species largely supported both *a priori* species identification and the species clusters recovered in the SNP phylogenetic analysis and the Structure analyses. *Speyeria nokomis* and *S. edwardsii* were the most genetically distinct species, and accounted for most of the variance along PC axis 1, at approximately 15.7%, and PC axis 2, at approximately 10.9% (Fig. 4.3b). After removing *S. nokomis* and *S. edwardsii,* the remaining thirteen species generally formed distinct clusters, although *S. callippe, S. carolae, S. coronis,* and *S. egleis* tended to cluster closely together (Fig. 4.3c). PC axis 3 separated both the eastern and western groups of *S. cybele* and the southwestern and northeastern groups of *S. hesperis* that were found in the phylogenetic and Structure analyses (Fig. 4.3c).

4.5 Discussion

4.5.1 *How many species in Speyeria?*

Phylogenetic analysis of SNPs indicates monophyletic genetic clusters for most of the fifteen *a priori* species considered, with the exception of *S. hesperis* and *S. coronis* (Fig. 4.2a). Since our SNP dataset did not contain representatives of *S. adiaste*, we do not know whether *S. hydaspe* would remain monophyletic with the inclusion of *S. adiaste*, or non-monophyletic as in the mitochondrial phylogeny (Fig. 4.2b). In addition, both *S. cybele* and *S. hesperis* exhibited sub-structuring on the SNP phylogeny that was also present in cluster-based analyses (Fig. 4.3), and the same applied to the mitochondrial phylogeny for *S. cybele*. This suggests that the Rocky Mountains form a barrier to gene flow within each of these species.

The genetic substructuring of *S. cybele* generally corresponds to differences in the morphology of eastern versus western *S. cybele* subspecies. The western populations tend to exhibit more sexual dimorphism, having bright orange males and pale yellow-white females with heavy, dark brown or black markings, while many eastern populations have less pronounced differences between the sexes (Dunford 2009). In regions where western and eastern subspecies co-occur, as they do in parts of Alberta (Bird *et al.* 1995), the extent of admixture between populations is unclear, and more intensive sampling is needed to assess whether the eastern and western forms of *S. cybele* constitute one or two species.

Speyeria hesperis was recovered in the SNP phylogeny as a paraphyletic clade that included a monophyletic *S. atlantis*, however the mitochondrial phylogeny depicted *S. hesperis* as marginally polyphyletic. The mitochondrial tree also showed a more distant relationship between these two species. The *S. atlantis-hesperis* complex has historically been difficult to classify, and several authors have used morphology, behaviour and range information to

alternately classify this complex as either a single morphologically variable species called *S. atlantis* (Grey 1951; Hammond *et al.* 2013) or a closely related complex of two species, *S. atlantis* and *S. hesperis* (Scott *et al.* 1998; Opler & Warren 2005; Dunford 2009). A recent study additionally treated the subspecies *S. atlantis hollandi* (Chermock & Chermock, 1940) as a distinct species, *S. hollandi* (McHugh *et al.* 2013). Our results support *S. atlantis* as a distinct genetic entity that does not appear to hybridise or otherwise mix with *S. hesperis* in regions where both taxa co-occur. Additionally, we did not find *S. a. hollandi* to be genetically distinct from the eastern subspecies, *S. atlantis canadensis* (dos Passos, 1935) (Fig. 4.2, Fig. 4.3). However, our Structure analyses provide some evidence that *S. hesperis* may occasionally hybridise with *S. zerene* and *S. aphrodite.* Also of note is the single specimen of *S. hesperis irene* (Boisduval, 1869) in our dataset, which appears to be intermediate and potentially a hybrid between northeastern *S. hesperis* and *S. coronis* (Fig. 4.3). There has been informal discussion among *Speyeria* enthusiasts about whether *S. hesperis irene* belongs in the *S. hesperis* species group, or if it is yet another example of an erroneous delimitation. Since a single specimen does not resolve this issue, we recommend additional sampling and genotyping to clarify the relationship of this enigmatic taxon to other species of *Speyeria*.

Our Structure analyses did not distinguish discrete clusters for *S. carolae* and *S. diana* (Fig. 4.3a), however these species were represented by only two and three specimens, respectively, limiting the ability of the program to accurately estimate the ancestral allele frequencies of these species (Lawson *et al.* 2018). Despite the lack of resolution in the Structure analysis, both species formed monophyletic clades on the SNP phylogeny (though *S. coronis* was paraphyletic with respect to *S. carolae*), and *S. diana* was also monophyletic in the mitochondrial phylogeny (Fig. 4.2). *Speyeria diana* formed a distinct cluster on the PCA, while *S. carolae*

clustered very closely with *S. coronis* (Fig. 4.3b-c). de Moya *et al.* (2017) estimated that *S. carolae* diverged only a few hundred thousand years ago, whereas *S. diana* likely diverged between 3-4 million years ago, making it a relatively old lineage within *Speyeria.* Laboratory hybridisation experiments using *S. diana* also indicated a high degree of reproductive isolation relative to other species of *Speyeria* (Hammond *et al.* 2013). Small sample sizes can impact the reliability of Structure analysis (Puechmaille 2016), and it is therefore likely that the unresolved *S. diana* and *S. carolae* clusters in our Structure plots at least partially reflect insufficient sampling. However, close relationships between *S. carolae, S. coronis* and *S. callippe* may also have impacted this analysis. Increased sampling should clarify the status of these species.

4.5.2 *Phylogenetic marker comparison reveals extensive mito-nuclear discordance*

Our marker comparison revealed discordant phylogenetic relationships among several species. Neither mitochondrial nor nuclear phylogenetic results fully recovered the proposed *Semnopsyche* or *Callippe* clades, which group species together primarily based on differences in genitalic morphology (dos Passos & Grey 1945; Hammond 1978). The *Semnopsyche* group was originally considered to contain species basal to the genus: *S. diana, S. cybele,* and *S. aphrodite,* while *S. idalia* and *S. nokomis,* which have intermediate genitalic forms but have been long considered to be similarly basal, have sometimes been included in this group (Hammond 1978). Our SNP phylogeny places all five *Semnopsyche* species of Hammond (1978) as a paraphyletic grade at the base of the tree, with all *Callippe* species monophyletic and sister to *S. cybele* + *S. aphrodite*. The five putative *Semnopsyche* species are further interdigitated with *Callippe* species in the mitochondrial tree. Our results are largely congruent with those of Dunford (2007) and de Moya *et al.* (2017).

Our mitochondrial tree recovered *S. nokomis* as the most basal ingroup taxon, while our SNP tree recovered *S. idalia* in this position, which is consistent with de Moya *et al.* (2017). The mitochondrial work of Dunford (2007) also placed *S. idalia* at the base of *Speyeria*, however when morphological data was combined with *COI* gene sequence in the same study, *S. nokomis* was instead recovered as the basal taxon. In our study, branch support for the position of *S. idalia* on the mitochondrial tree and position of *S. nokomis* on the SNP tree had lower posterior probability values than most other species-level clades, indicating some uncertainty in the placement of both species in our analyses (Fig. 4.2). Since *S. nokomis* and *S. idalia* are both morphologically distinct (Dunford 2009) and appear to be relatively genetically distinct (Fig. 4.2, Fig. 4.3), it is possible that long branch attraction (LBA) is confounding phylogenetic consistency in the SNP and/or mitochondrial phylogenies; LBA is more common when there is rate heterogeneity across sites, which is almost certainly the case when using genomic SNPs for phylogenetic reconstruction, although gene sequence data can be similarly susceptible (Bergsten 2005; Kolaczowski & Thornton 2009; Kück *et al.* 2012). Additionally, while we took care to reduce the incidence of paralagous loci in the SNP dataset, the use of SNP data at the level of species may introduce other artefacts that can impact phylogenetic reconstruction. For instance, *de novo* locus construction may erroneously build non-orthologous loci (Diaz-Arce & Rodriguez-Ezpeleta 2019), and the inclusion of only variable SNPs (see Methods), an approach we took to retain more loci in our dataset, can introduce ascertainment bias that may also contribute to topological inconsistencies or inaccuracies (Leaché *et al.* 2015). Allelic dropout, or the failure to sequence a restriction site due to the presence of a mutation in older or fast-evolving lineages, can also bias which loci are represented in the dataset (Leaché and Oaks 2017). Thus, SNP genotyping above the species level may have contributed to some

phylogenetic uncertainty. However the observed mito-nuclear discordance in our study is not easily accounted for by this phenomenon, since there was often tight clustering of specimens based on SNPs while single members of these clusters had divergent mitochondrial haplotypes.

Speyeria cybele and *S. hydaspe* were recovered as sisters in our mitochondrial gene tree analysis (Fig. 4.2b), and Structure analyses (Fig. 4.3a) similarly indicated a close relationship between *S. hydaspe* and western subspecies of *S. cybele*. In contrast, our phylogenetic analysis of nuclear SNPs indicated a distant relationship between these two species and a sister relationship between *S. cybele* and *S. aphrodite* (Fig. 4.2a). *Speyeria hydaspe* and *S. cybele* exhibit pronounced morphological differences in size, wing colour and patterning, and genitalic morphology (dos Passos & Grey 1945; Dunford 2009), and so the mitochondrial relationship between them is not intuitive. However, *S. hydaspe* and the western subspecies of *S. cybele* are both present in the western Rocky Mountains and the Great Basin region (Dunford 2009), so contemporary and/or historical contact between these groups is possible. In contrast, the eastern subspecies of *S. cybele* are outside the range of *S. hydaspe*, even in Alberta where both *S. cybele* clades occur. Morphological characteristics (including genitalic similarities) and extensive range overlap suggest a closer relationship between *S. cybele* and *S. aphrodite* (dos Passos & Grey 1945; Dunford 2009), and this relationship is shown on the SNP tree (Fig. 4.2a) which has these species as sisters, and in the Structure plot, which suggests shared ancestry between *S. aphrodite* and the eastern grouping of *S. cybele* (Fig. 4.3a).

The mitochondrial phylogeny shows *S. hesperis* and *S. atlantis* with relatively distantly related *COI* haplotypes, but these two species together form a monophyletic clade on the SNP tree (Fig. 4.2). *Speyeria atlantis* has a mostly eastern distribution, extending west as far as Alberta, while *S. hesperis* is primarily a western species with many subspecies that range in or

near the Rocky Mountains, as far south as Arizona and New Mexico, and north into the Canadian prairies, British Columbia, and the Northwest Territories (Layberry *et al.* 1998; Dunford 2009). Like *S. hydaspe* and *S. cybele,* while *S. hesperis* and *S. atlantis* overlap and are common in Alberta, they have different habitat preferences that may limit significant contact between species (Bird *et al.* 1995). The northeastern clade of *S. hesperis* appears to be more genetically similar to *S. atlantis* than to the southwestern clade of *S. hesperis* in the cluster-based analyses (Fig. 4.3), which is congruent with the relationship depicted in the SNP phylogeny and also the eastern distribution of both clades. Although *S. atlantis* forms a cluster distinct from *S. hesperis*, we cannot unambiguously state whether this supports the delimitation of *S. atlantis* as a species distinct from *S. hesperis* or, alternatively, a subspecies relationship. The alternate genetic relationships between *S. atlantis* and *S. hesperis* depicted on the *COI* gene tree, the SNP species tree, and in the clustering analyses indicate a complex evolutionary history that will require more extensive genomic and geographic sampling and focused analyses to elucidate.

Our Structure analysis also suggests that occasional hybridisation may occur between a few species (see *S. zerene, S. hesperis,* and *S. aphrodite* clusters on the Structure plot in Fig. 4.3a), although admixture between species in the regions we sampled did not suggest high levels of ongoing hybridisation. Nonetheless, one *S. zerene zerene* specimen from California showed substantial genetic similarity to southwestern *S. hesperis,* another *S. zerene zerene* from the same region appeared similar *to S. mormonia,* a *S. aphrodite ethne* specimen from Colorado had major genetic similarity to northeastern *S. hesperis*, and a southwestern *S. hesperis irene* specimen from California was largely similar to *S. coronis*. In addition, the polyphyly on our mitochondrial phylogeny does not clearly correspond to geography (Fig. 4.2b), which would be expected if

contemporary hybridisation was the primary source of mito-nuclear discordance in these species (Sonsthagen *et al.* 2016).

4.5.3 *COI sequence is an unreliable indicator of species relationships in Speyeria*

The pronounced non-monophyly of several species on our mitochondrial phylogeny corroborates previous studies (summarised in Fig. 4.1), and suggests that the *COI* mitochondrial gene has insufficient power to resolve species in *Speyeria,* particularly among closely related species or those with complex phylogeographic histories; this has also been shown in numerous other groups that have similarly close relationships (Dupuis *et al.* 2012). Previous studies have suggested that diversification in *Speyeria* mostly occurred during and after the last glacial retreat, and several species, particularly those that range in western montane habitats, such as *S. mormonia, S. zerene,* and *S. hesperis*, likely experienced recurrent periods of isolation in glacial refugia followed by recolonisation and secondary contact (Grey 1951; Dunford 2009; de Moya *et al.* 2017). Such stochastic events can at least partially explain the observed mito-nuclear discordance in our study (Hewitt 2008; Toews & Brelsford 2012), while close relationships resulting in incomplete lineage sorting between *S. callippe, S. coronis, S. carolae,* and *S. egleis* may be another explanation (Rosenberg 2003). A clear example of phylogenetic discordance in *Speyeria* is presented by McHugh *et al.* (2013), who produced multiple phylogenies using the same specimens but different molecular markers, with each phylogeny exhibiting alternate relationships between the species included in their study (Fig. 4.1).

4.5.4 *Conclusions*

Our study further substantiates the use of genome-wide SNPs for phylogenetic reconstruction of closely related insect species, and shows that the *COI* gene is not a reliable marker for *Speyeria* at broad geographic scales. The phylogenetic marker comparison and cluster-based analyses in our study of *Speyeria* butterflies revealed aspects of the evolutionary history of this group that has so far only received speculation, and opens many avenues for future study. We suggest that mito-nuclear discordance in *Speyeria* may result from a number of different scenarios, including retained ancestral polymorphism, hybridisation with introgression or methodological artefacts, and suggest relevant areas for future work on *Speyeria* to further clarify the sources of this discordance. In particular, more intensive sampling for phylogeographic analyses and assessments of historical introgression will clarify species boundaries and elucidate patterns of population structure and diversification in this clade.

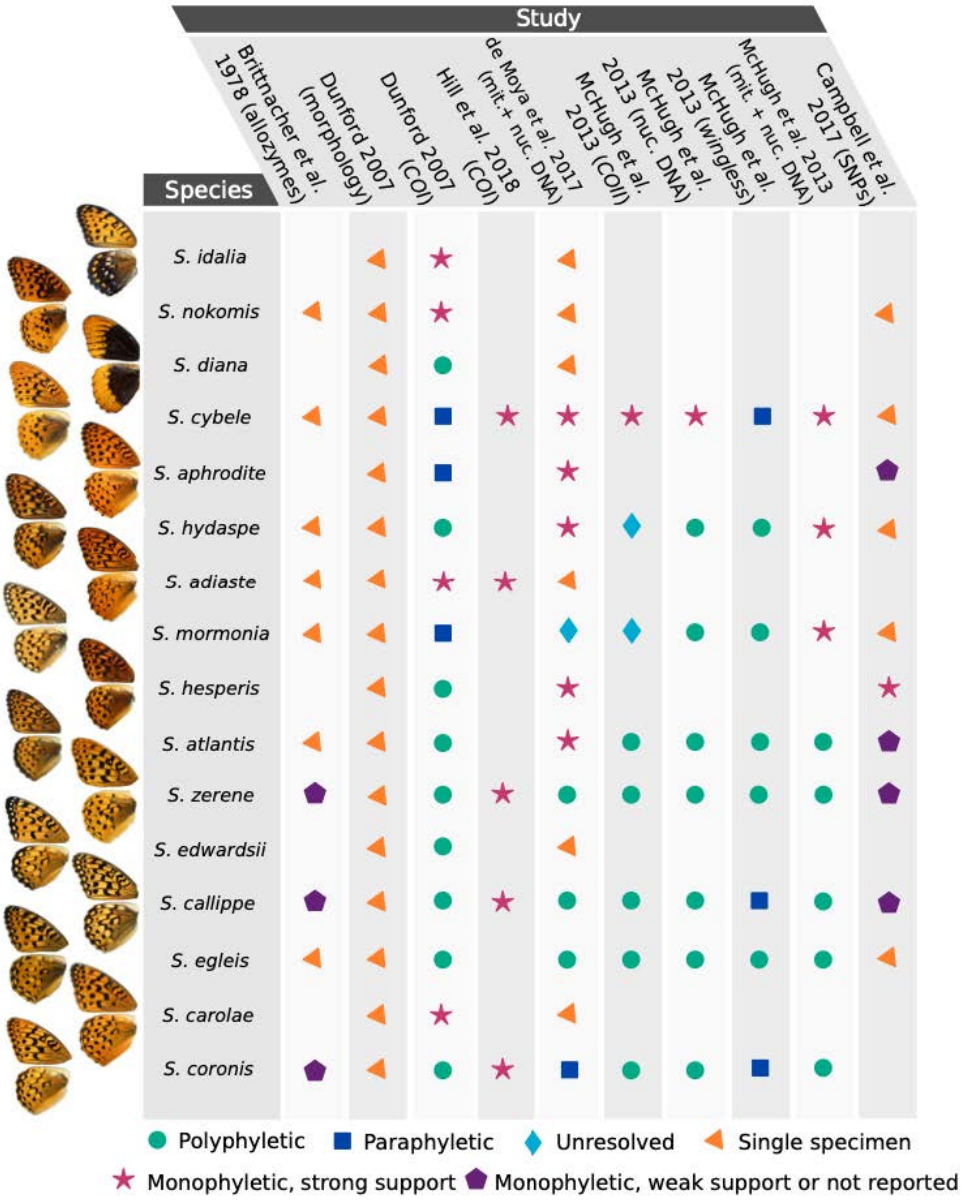


Figure 4.1 Summary of taxon sampling and clade topologies obtained in previous phylogenetic studies on *Speyeria*. McHugh *et al.* (2013) used different markers from the same specimens, obtaining varied phylogenetic results. We considered strong branch support for monophyletic clades to have bootstrap values > 70% or posterior probability values > 0.9.

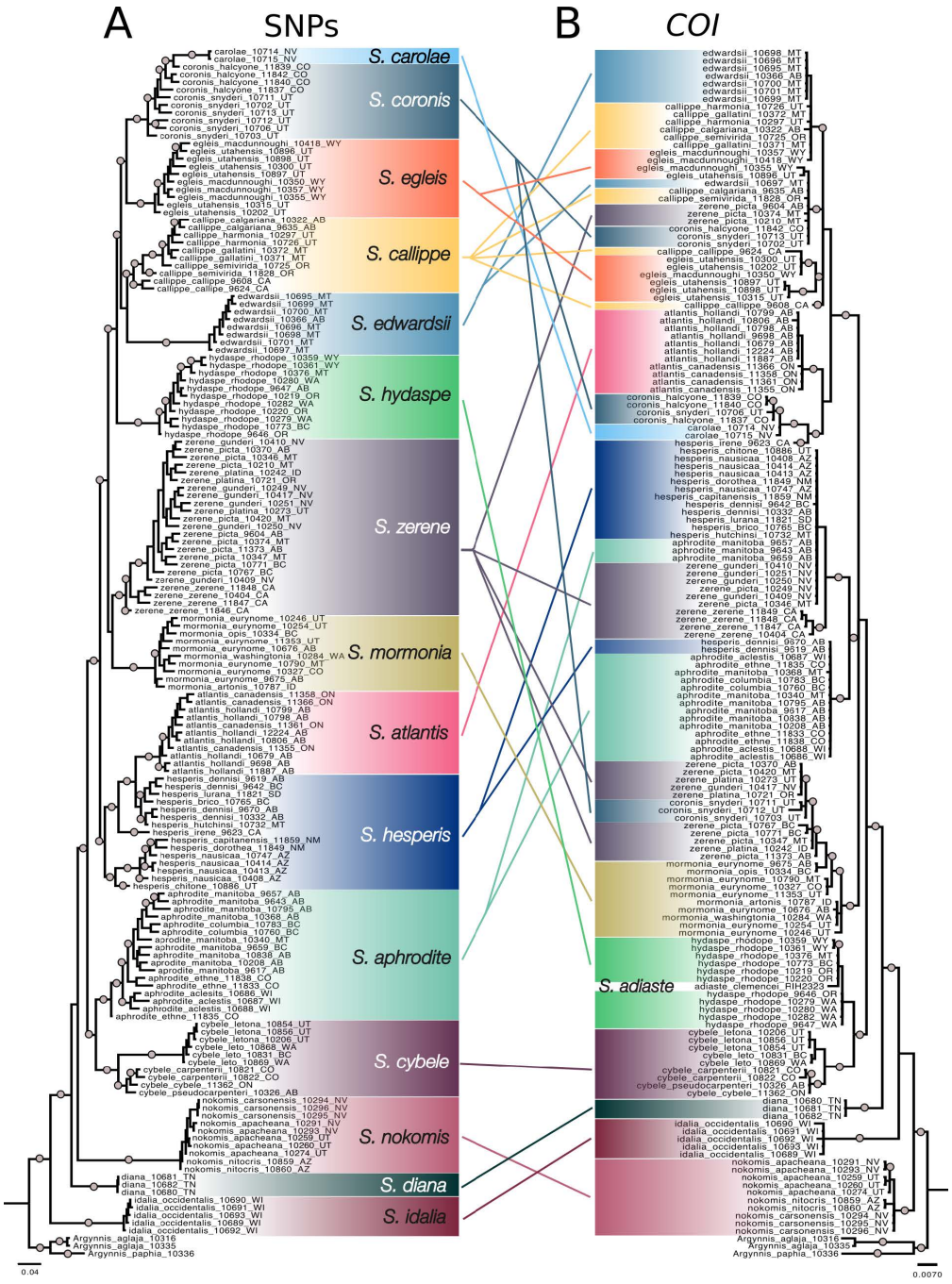


Figure 4.2 Bayesian phylogenetic comparison of A, genomic SNPs and B, *COI* sequences. Posterior probabilities are located on each branch, and specimens on the *COI* tree (B) are coloured according to their cluster identity on the SNP tree (A). Grey circles along branches indicate posterior probabilities > 0.9. Scale bars indicate the mean number of substitutions per site.

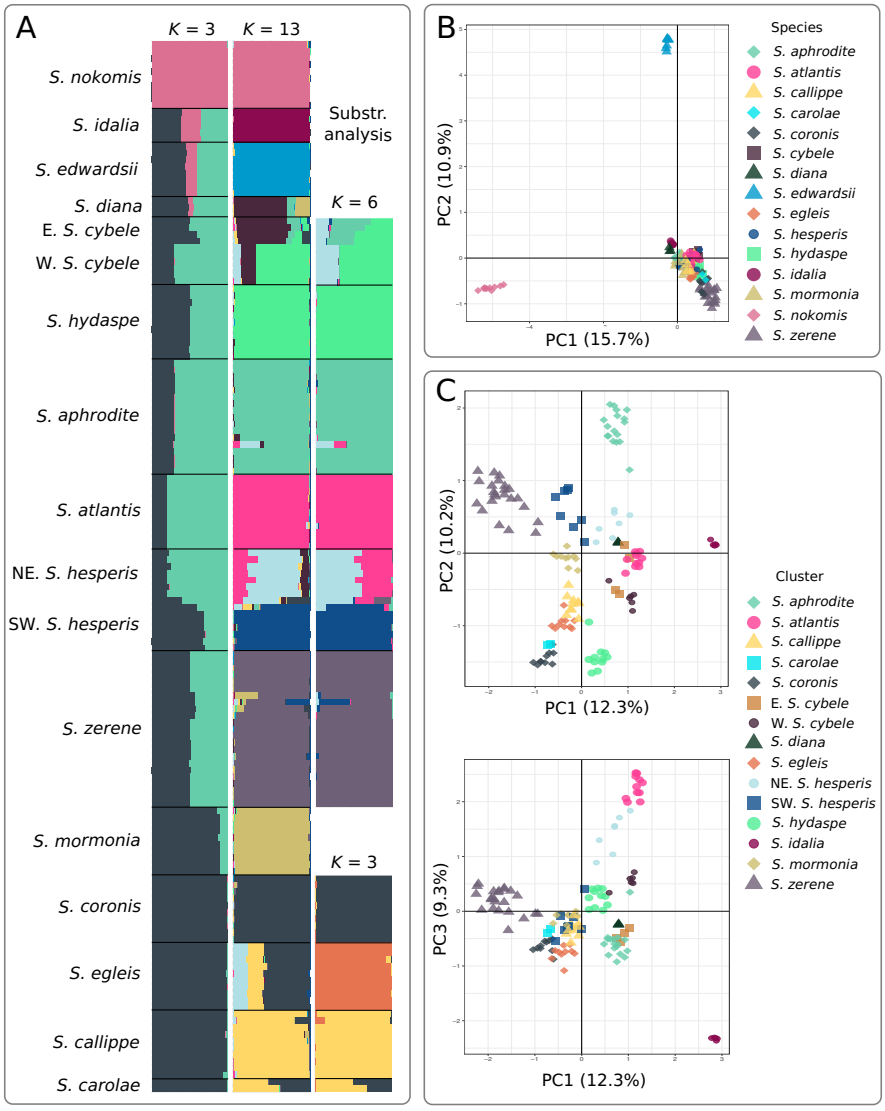


Figure 4.3 Structure and PCA analyses of SNP data. A, Structure analysis of 155 specimens from fifteen species of *Speyeria* recovered *K*=13 as optimal, which fully resolved 11 *a priori* species. The remaining species exhibited variable degrees of shared ancestry, although hierarchical analysis resolved an additional species, *S. egleis*. Sub-clustering in *S. cybele* and *S. hesperis* corresponds to populations sampled on opposite sides of the Rocky Mountains. B-C, (B) PCA of all sampled *Speyeria* indicates that *S. nokomis* and *S. edwardsii* are the most genetically distinctive species. (C) The PCA after removing *S. nokomis* and *S. edwardsii* shows sub-clustering of *S. cybele* and *S. hesperis* like that in the Structure analysis.

4.6 Literature cited

Arnold, B., Corbett-Detig, R. B., Hartl, D., and Bomblies, K. (2013) RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology*, **22**, 3179-3190

Bergsten, J. (2005) A review of long-branch attraction. *Cladistics*, **21**, 163-193.

Bird, C. D., Hilchie, G. J., Kondla, N. G., Pike, E. M., and Sperling, F. A. H. (1995) Alberta Butterflies, The Provincial Museum of Alberta, Edmonton, Alberta, Canada. Boisduval, J. A. 1869. Lépidoptères de la Californie. *Annales de la Société entomologique de Belgique*, **12**, 7-94.

Boisduval, J. A. (1852) Lépidoptères de la Californie. *Annales de la Société entomologique de France,* **10**, 275-324.

Breed, G. A., Stichter, S., and Crone, E. E. (2012) Climate-driven changes in northeastern US butterfly communities. *Nature Climate Change*, **3**, 142-145.

Brittnacher, J. G., Sims, S. R., and Ayala, F. J. (1978) Genetic differentiation between species of the genus *Speyeria* (Lepidoptera: Nymphalidae). *Evolution*, **32**, 199-210.

Brock, J. P. and Kaufman, K. (2003) *Kaufman Field Guide to Butterflies of North America*, New York:Houghton Mifflin Co.

Campbell, E. O., Davis, C. S., Dupuis, J. R., Muirhead, K., and Sperling, F. A. H. (2017) Cross-platform compatibility of *de novo*-aligned SNPs in a nonmodel butterfly genus. *Molecular Ecology Resources*, **6**, e84-e93.

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., and Postlethwait, J. H. (2011) Stacks: building and genotyping loci de novo from short-read sequences. *G3*, **1**, 171-182.

Comstock, J. A. (1925) Eight new races, forms or aberrations of California butterflies. *Bulletin of the Southern California Academy of Sciences*, **24**, 62-68.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … 1000 Genomes Analysis Group. (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156-2158.

de Moya, R. S., Savage, W. K., Tenney, C., Bao, X., Wahlberg, N., and Hill, R. I. (2017) Interrelationships and diversification of *Argynnis* Fabricius and *Speyeria* Scudder butterflies. *Systematic Entomology*, **42**, 635-649.

Diaz-Arce, M., and Rodriguez-Ezpeleta, N. (2019) Selecting RAD-Seq data analysis parameters for population genetics: the more the better? *Frontiers in Genetics,* **10**, 1-10.

dos Passos, C. F. and Grey, L. P. (1945) A genitalic survey of Argynninae (Lepidoptera, Nymphalidae). *American Museum Novitates*, **1296**.

dos Passos, C. F. and Grey, L. P. (1947) Systematic catalogue of *Speyeria* (Lepidoptera, Nymphalidae) with designations of types and fixations of type localities. *American Museum Novitates*, **1370**.

Dunford, J. C. (2007) The genus *Speyeria* and the *Speyeria atlants/hesperis* complex: Species and subspecies accounts, systematics, and biogeography (Lepidoptera, Nymphalidae). Unpublished Dissertation, University of Florida.

Dunford, J.C. (2009) Taxonomic overview of the greater fritillary genus *Speyeria* Scudder and the atlantis-hesperis species complexes, with species accounts, type images, and relevant literature (Lepidoptera: Nymphalidae). *Insecta Mundi.* 0090: 1-74.

Dupuis, J.R., Roe, A.D., and Sperling, F.A.H. (2012) Multi-locus species delimitation in closely related animals and fungi: one marker is not enough. *Molecular Ecology.* 21: 4422-4436.

Edwards, W. H. (1864) Description of certain species of diurnal Lepidoptera found within the limits of the United States and British North America. *Proceedings of the Entomological Society of Philadelphia*, **2**, 501-507.

Edwards, W. H. (1869) Descriptions of new species of diurnal Lepidoptera found within the United States. *Transactions of the American Entomological Society,* **2**, 369-382.

Edwards, W. H. (1872) Descriptions of new species of diurnal Lepidoptera found within the United States. *Transactions of the American Entomological Society*, **4**, 61-70.

Edwards, W. H. (1874) Descriptions of new species of diurnal Lepidoptera found in North America. *Transactions of the American Entomological Society*, **5**, 103-111.

Edwards, W. H. (1879) Descriptions of new species of North American butterflies; also, notes upon certain species. *The Canadian Entomologist*, **11**, 81-89.

Edwards, W. H. (1886) Descriptions of new species of butterflies found in the United States. *The Canadian Entomologist*, **18**, 61-65.

Evanno, G., Regnaut, S., and Goudet, J. (2005) Detecting the number of clusters of individuals using the software Structure: A simulation study. *Molecular Ecology*, **14**, 2611-2620.

Grey, L. P. (1951) The subspeciation of *Speyeria atlantis*. *The Lepidopterist's News*, **5**, 31-35.

Hammond, P. C. (1978) Geographic variation and speciation in the Nymphalid butterfly genus *Speyeria*. Ph.D. Dissertation, Virginia Polytechnic Institute and State University, Blacksburg.

Hammond, P. C. (1981) The Colonization of Violets and *Speyeria* butterflies on the ash-pumice fields deposited by Cascadian volcanoes. *Journal of Research on the Lepidoptera*, **20**, 179-91.

Hammond, P. C. (1995) Conservation of the biodiversity in native prairie communities in the United States. *Journal of the Kansas Entomological Society*, **68**, 1-6.

Hammond, P. C., McCorkle, D., and Bergman, W. (2013) Hybridization studies of genomic compatibility and phenotypic expression in the greater fritillary butterflies (Nymphalidae: Argynnini). *Journal of the Lepidopterists' Society*, **67**, 263-273.

Hebert, P.D.N., Cywinska, A., Ball, S.L., and deWaard, J.R. (2003) Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B.* 270: 313-321

Hewitt, G. M. (2008) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247-276.

Hill, R. I., Ganeshan, M., Wourms, L., Kronforst, M. R., Mullen, S. P., and Savage, W. K. (2018) Effectiveness of DNA barcoding in *Speyeria* butterflies at small geographic scales. *Diversity*, **10**, 130.

Hoang, D. T., Chernomor, L., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018) UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, **35**, 518-522.

Jombart, T. (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403-1405.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermiin, L. S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, **14**, 587-589.

Katoh, K., Rozewicki, J., and Yamada, K. D. (2017) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, bbx108.

Kolaczkowski, B., and Thornton, J. W. (2009) Long-branch attraction bias and inconsistency in Bayesian phylogenetics. *PLoS ONE*, **4**, e7871

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, M. A., and Mayrose, I. (2015) Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, **15**, 1179-1191.

Kück, P., Mayer, C., Wägele, J. W., and Misof, B. (2012) Long branch effects distor maximum likelihood phylogenies in simulations despite selection of the correct model. *PLoS ONE*, **7,** e36593.

Lawson, D. J., van Dorp, L, and Falush, D. (2018) A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communications*, **9**, 3258.

Layberry, R. A., Hall, P. W., and Lafontaine, J. D. (1998) *The Butterflies of Canada*. University of Toronto Press.

Leaché, A. D., Banbury, B. L., Felsenstein, J., nieto-Montes de Oca, A., and Stamatakis, A. (2015) Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. *Systematic Biology*, **6**, 1032-1047.

Leaché, A. D., and McGuire, J. A. (2006) Phylogenetic relationships of horned lizards (*Phrynosoma*) based on nuclear and mitochondrial data: evidence for a misleading mitochondrial gene tree. *Molecular Phylogenetics and Evolution*, **39**, 628-644.

Leaché, A. D., and Oaks, J. R. (2017) The Utility of single nucleotide polymorphism (SNP) data in phylogenetics, *Annual Review of Ecology, Evolution, and Systematics*, **48**, 69-84.

Lee, K. M., Kivelä, S. M., Ivanov, V., Hausmann, A., Kaila, L, Wahlberg, N., and Mutanen, M. (2018) Information dropout patterns in restriction site associated DNA phylogenomic and

a comparison with multilocus Sanger data in a species-rich moth genus. *Systematic Biology*, **67**, 925-939.

Maddison, W. P., and Maddison, D. R. (2018) Mesquite: a modular system for evolutionary analysis. Version 3.51. http://www.mesquiteproject.org

Martin, M. 2011. Cutadapt removed adapter sequences from high-throughput sequencing reads. *EMBnet Journal*, **17**, 10-12.

McHugh, A., Greever, C., and Van Buskirk, R. (2013) A molecular phylogenetic analysis of *Speyeria* and its implications for the management of the threatened *Speyeria zerene hippolyta*. *Journal of Insect Conservation*, **17**, 1237-1253.

Moeck, A. H. (1975) Geographic variability in *Speyeria*: comments, records, and description of a new subspecies. Los Angeles: Entomological Reprint Specialists.

Nabhan, A. R., and Sarkar, I. N. (2012) the impact of taxon sampling on phylogenetic inference: a review of two decades of controversy. *Briefings in Bioinformatics*, **13**, 122-134.

Nguyen, L. -T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Molecular Biology and Evolution*, **32**:268-274.

Opler, P. A. and Warren, A. D. (2005) Lepidoptera of North America. 4. Scientific names list for butterfly species of North America, North of Mexico. Contributions of the C. P. Gillette Museum of Arthropod Diversity, Colorado State University, Fort Collins, Colorado, USA

Paris, J. P., Stevens, J. R., and Catchen, J. M. (2017) Lost in parameter space: a road map for STACKS. *Methods in Ecology and Evolution*, **8**, 1360-1373.

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. *The Lepidoptera*

Research Foundation/Journal of Research on the Lepidoptera, Volume 40, Beverly Hills, CA, USA.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., and Hoekstra, H. E. (2012) Double digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS One* 7: e37135. doi:10.1371/journal.pone.0037135.

Poland, J. A., Brown, P. J., Sorrells, M. E., & Jannink, J. (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE*, **7**, e32253.

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945-959.

Puechmaille, S. J. (2016) The program STRUCTURE does not reliably recover the correct population structure when sampling is uneven: sub-sampling and new estimator alleviate the problem. *Molecular Ecology Resources*, **16**, 608-627.

R Core Team. (2017) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rochette, N. C., and Catchen J. M. (2017) Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols*, **12**, 2640-2659.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., and Huelsenbeck, J. P. (2012) MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**, 539-542.

Rosenberg, N. A. (2003) the shapes of neutral gene genealogies in two species: probabilities of monophyly, paraphyly, and polyphyly in a coalescent model. *Evolution*, **57**, 1465-1477.

Scott, J. A., Kondla, N. G., and Spomer, S. M. (1998) *Speyeria hesperis* and *Speyeria atlantis* are distinct species. *Papilio,* **8**, 1-31.

Simonsen, T. J., Wahlberg, N., Brower, A. V. Z, and de Jong, R. (2006) Morphology, molecules and fritillaries: approaching a stable phylogeny for Argynnini (Lepidoptera: Nymphalidae). *Insect Systematics and Evolution*, **37**, 405-418.

Sims, S. R. (2017) *Speyeria* (Lepidoptera: Nymphalidae) Conservation. *Insects*, **8**, 45.

Sonsthagen, S. A., Wilson, R. E., Chesser, R. T., Pons, J. M., Crochet, P. A., Driskell, A., and Dove, C. (2016) Recurrent hybridization and recent origin obscure phylogenetic relationships within the 'white-headed' gull (*Larus* sp.) complex. *Molecular Phylogenetics and Evolution*, **103**, 41-54.

Toews, D. P. L. and Brelsford, A. (2012) The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, **21**, 3907-3930.

Wagner, C. E., Keller, I., Wittwer, S., Selz, O. M., Mwaiko, S., Greuter, L., Sivasundar, A., and Seehausen, O. (2013) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*, **22**, 787-798.

Warren, A. D., Davis, K. J., Grishing, N. V., Pelham, J. P., and Stangeland, E. M. (2012) Interactive listing of American butterflies <www.butterfliesofamerica.com>

Wells, C. N., and Tonkin, D. W. (2014) Range collapse in the Diana fritillary, *Speyeria diana* (Nymphalidae). *Insect Conservation and Diversity*, **7**, 365-380.

Wolfe, K. (2017) Natural non-Viola hostplant of a Sierran *Speyeria* (Nymphalidae, Heliconiinae) and an associated parasitoid (Diptera, Tachinidae). *Journal of the Lepidopterists' Society*, **71**, 129-131.

Zaman, K., Tenney, C., Rush, C. E., and Hill, R. I. (2015) Population ecology of a California endemic: *Speyeria adiaste clemencei*, *Journal of Insect Conservation*, **19**, 753-763.

Chapter 5

Do not adjust your (tree) set:

ambiguity in the *Speyeria atlantis-hesperis* butterfly species complex is real

5.1 Summary

Speyeria atlantis (Edwards, 1862) and *S. hesperis* (Edwards, 1864) (Lepidoptera: Nymphalidae) form a species complex that is notorious for its species delimitations being complicated by morphological variation and mito-nuclear discordance. But more accurate species delimitations are urgently needed, in part because recent evidence has suggested that these species may have a history of hybridization with a more distantly related species, *S. zerene* (Boisduval, 1852), that has several populations currently experiencing large declines. Using genomic SNPs, we re-assessed the species boundaries of *S. atlantis* and *S. hesperis*, and inferred how their phylogeographic history may have impacted relationships both within this complex and with *S. zerene*. Our results uphold the current delimitation of *S. atlantis* as a distinct species, and demonstrate major genomic divergences within *S. hesperis* that may have resulted from a combination of past introgression with *S. zerene* and rapid, post-glacial range expansion. We recommend taxonomic recognition of *S. hesperis* populations as two distinct species, with those in the northeastern half of its range retaining the name *S. hesperis*, and populations in the southwest being elevated to species status. For the populations we genotyped, the oldest name is *S. nausicaa*; however other subspecies in the southwest region remain to be sampled.

5.2 Introduction

Species delimitation has biological, social, economic, and political consequences that are becoming more pronounced amidst rapid habitat and biodiversity loss (Coates *et al.* 2018; Stanton *et al.* 2019). However, assessing species limits is often operationally problematic, in large part because speciation can proceed over long periods of time and is influenced by varied processes that impact how we identify discrete units (Sites & Marshall 2003; de Queiroz 2007). Assessment of young lineages is particularly difficult, since species indicators such as discrete morphology or genetic divergence are often partial or contradictory at these early stages (de Queiroz 1998, 2007). For instance, cryptic or highly variable morphology can conflict with ecological and genetic assessments of species (Razkin *et al.* 2016; MacGuigan *et al.* 2017; Freitas *et al.* 2018), and mitochondrial and nuclear markers may be differentially impacted by hybridization, incomplete lineage sorting, population bottlenecks, and the transmission of endosymbionts, all of which can contribute to discordant genetic patterns (Ebel *et al.* 2015; Papakostas *et al.* 2016; Weigand *et al.* 2017).

While these factors present a practical challenge for species delimitation, data discordance also provides an opportunity to identify factors that contribute to early lineage diversification and its characterization has been facilitated by recent advances in molecular techniques (Wagner *et al.* 2012; Ivanov *et al.* 2018; Hinojosa *et al.* 2019). Development of RADseq (Baird *et al.* 2008) and related methods has allowed genotyping of hundreds to thousands of SNPs, greatly extending the capacity to detect fine-scale genomic divergences in non-model organisms and identify the processes that lead to them (Hohenlohe *et al.* 2012; Wagner *et al.* 2012; Escudero *et al.* 2014; Vargas *et al.* 2017; Abdelkrim *et al.* 2018; Campbell

et al. 2018; Hinojosa *et al.* 2019; Hundsdoerfer *et al.* 2019). Genomic data has become so widespread that it is now commonly used in integrative approaches to species delimitation (Carstens *et al.* 2013; Stanton *et al.* 2019), and new analytical techniques like the multispecies coalescent, which detects the genomic boundary between population-level and species-level processes (Degnan & Rosenberg 2009). Therefore, such genomic data are valuable for re-assessing species limits in taxa that have been historically difficult to characterize using morphological or ecological data.

The butterfly genus *Speyeria* Scudder, 1872 (Lepidoptera: Nymphalidae) is well known for its phenotypic variability and the ambiguous evolutionary relationships of its component lineages, which have historically made it difficult to delimit and identify its species (dos Passos & Grey 1947; Moeck 1975; Dunford 2009). Reliance on one or a few genetic markers has consistently failed to recover stable genetic clusters that match morphological species delimitations in *Speyeria* (McHugh *et al.* 2013; de Moya *et al.* 2017; Campbell *et al.* 2019). Surveys of genome-wide variation have recently improved the correspondence between morphological and genetic species delimitations of *Speyeria* (Campbell *et al.* 2017; Thompson *et al.* 2019), but have also demonstrated extensive mito-nuclear discordance (Campbell *et al.* 2019).

A total of 16 species are currently recognized in *Speyeria*, as well as over 110 morphologically variable subspecies and several species complexes with poorly understood evolutionary relationships (dos Passos & Grey 1947; Scott *et al.* 1998; Dunford 2009; Pelham 2019). Among these, *S. hesperis* (Edwards, 1864) and *S. atlantis* (Edwards, 1862) form a large complex containing 26 subspecies (Pelham 2019). These include five subspecies in *S. atlantis* that are broadly distributed in conifer woodlands across North America from the Rocky Mountains to Newfoundland. More variation is taxonomically recognized in *S. hesperis*, which

has 21 subspecies in drier meadows and open forests throughout western North America and east to South Dakota and southeastern Manitoba (Pelham 2019). The two species contact each other in mixed forest areas from Manitoba to British Columbia and south along the Rocky Mountains to Colorado, exhibiting substantial similarity between species in some areas as well as geographic variation within each species (dos Passos & Grey 1947; Moeck 1975; Dunford 2009). Some taxonomic treatments have considered *S. hesperis* to be a subspecies of *S. atlantis,* based on their overall morphological similarity (Grey 1951; Miller and Brown 1981; Hammond *et al.* 2013). Current practice generally recognizes these taxa as two distinct species, based on assessments of morphological differences and an apparent lack of hybridization between sympatric populations, but the taxa remain difficult to identify reliably (Bird *et al.* 1995; Scott *et al.* 1998; Opler & Warren 2005). While recent genetic evidence indicating that *S. atlantis* is distinct from *S. hesperis* has lent support to the separate species status of these taxa (Campbell *et al.* 2017; de Moya *et al.* 2017; Campbell *et al.* 2019; Thompson *et al.* 2019), their relationships remain unclear. Notably, Campbell *et al.* (2019) has also recently shown, based on a limited number of specimens, that there are substantial SNP differences between populations of *S. hesperis* that occur northeast of the Rocky Mountains (i.e.: in Canada, Montana, and South Dakota) versus the southwestern United States.

In addition to ambiguities within the *S. atlantis-hesperis* species complex, further complexity is provided by interactions with species that have not traditionally been considered part of this complex, including intermediates between *S. hesperis* and *S. zerene* (Boisduval, 1852) in some regions Campbell *et al.* (2019). This brings conservation issues into play, since *S. zerene* has several subspecies that are currently being managed to reverse significant population declines (McHugh *et al.* 2013; Sims 2017). Therefore, re-assessing genomic relationships within

the *S. atlantis-hesperis* species complex and its relationship to other *Speyeria* should improve our understanding of the evolution of these species and provide a foundation for continued conservation studies on the genus (Coates *et al.* 2018; Stanton *et al.* 2019).

Our objective is thus to provide a focused phylogenomic and population genomic reassessment of species limits in the *S. atlantis-hesperis* complex. We use *de novo* SNPs to recover distinct genetic clusters that maintain their genomic integrity in regions of contact (Sperling 2003), testing alternate hypotheses on species delimitation and phylogeographic or ecological factors that may have contributed to the genetic patterns we recovered. Our results support current recognition of *S. atlantis* as a distinct species, and suggest that rapid range expansion and past introgression with *S. zerene* have contributed to major nuclear genomic divergence within *S. hesperis* that is not reflected in mitochondrial DNA. Based on these results, we recommend that *S. hesperis* should be taxonomically recognized as two distinct species.

5.3 Methods

5.3.1 *Specimen collection and identification*

Specimens were collected by net and were either preserved in ethanol or frozen at -20°C until DNA was extracted from each sample. Following Campbell *et al.* (2019), morphological identifications to subspecies were made using multiple sources, including range information (Moeck 1975), field markings (Bird *et al.* 1995; Brock & Kaufman 2003; Dunford 2009; Warren *et al.* 2012), and comparison to specimens in the Bean Museum collection at Brigham Young University or the personal reference collection of E. Gage. In addition to specimens of *S. atlantis* and *S. hesperis*, we included several specimens of *S. aphrodite* (Fabricius, 1787) and *S. zerene* to test for both ancient and contemporary admixture, since SNP-based Structure analysis of

Campbell *et al.* (2019) indicated limited hybridization between these species and *S. hesperis*, and phylogenetic analyses additionally showed mito-nuclear discordance in relationships among *S. hesperis, S. atlantis, S. zerene,* and *S. aphrodite*. A single specimen tentatively identified as *S. hesperis irene* (Boisduval, 1869) was included in Campbell *et al.* (2019) but was not included here since its identity was uncertain and may represent an intermediate with *S. coronis* (Behr, 1864). The total dataset was comprised of 111 specimens in 12 subspecies of *S. hesperis,* 18 specimens in three subspecies of *S. atlantis,* 16 specimens in four subspecies of *S. zerene*, six specimens in three subspecies of *S. aphrodite*, and two outgroup specimens of *S. cybele cybele* (Fabricius, 1775) (Appendix 5.1).

5.3.2 *Molecular data generation and processing*

DNA extraction, mitochondrial *COI* gene amplification, ddRAD/two-enzyme GBS library preparation and sequencing, and initial mitochondrial and SNP data processing follows Campbell *et al.* (2017; 2019), and so are not described in detail here. SNPs were genotyped *de novo* using the Stacks 2.3 pipeline (Catchen *et al.* 2011; Rochette *et al.* 2019) on the Béluga cluster of Compute Canada, with default parameter settings except for the following: the *n* parameter, which controls the number of mismatches tolerated per locus during catalog construction in *cstacks*, was set to 2 instead of 1; we only retained loci that were found in 80% of any single population (the "r80" principle of Paris *et al.* (2017)); and a single, random SNP from each locus was output during final processing in the *populations* program of Stacks to reduce genomic linkage. We conducted additional SNP filtering in vcftools 0.1.14 (Danecek *et al.* 2011) to retain only loci with a minimum minor allele frequency of 3% and to further reduce the global

missing data per locus to a maximum of either 20% for phylogenetic and population genetic analyses or 10% for species delimitation and phylogeographic analyses.

5.3.3 *Phylogenetic and population genetic analyses*

Phylogenetic analyses for both the *COI* gene and the filtered genomic SNPs were conducted in IQ-TREE 1.3.10 (Nguyen *et al.* 2015). Model testing, SH-aLRT branch testing, and 1000 replicates of ultrafast bootstrapping (Hoang *et al.* 2018) were conducted in the program.

Mitochondrial *COI* gene data was used to build a minimum spanning haplotype network (Bandelt *et al.* 1999) in the program PopART (Leigh & Bryant 2015), which outputs a visual representation of the population genetic relationships between *COI* haplotypes. We used Structure version 2.3.4 (Pritchard *et al.* 2000) and TESS version 2.3.1 (Chen *et al.* 2007) to infer population structure in the SNP dataset. While both programs take a similar Bayesian approach to population clustering based on changes in allele frequency, TESS differs from Structure by additionally incorporating a spatial component for inferring genetically disparate populations that may result from geographic discontinuities. This is particularly useful when genetic structure is correlated with isolation by distance (IBD), which can contribute to population over-splitting in non-spatial programs (Chen *et al.* 2007). Campbell *et al.* (2019) showed strong genetic sub-structuring within *S. hesperis* that corresponded to populations sampled southwest and northeast of the Rocky Mountains; TESS and Structure were compared to clarify the extent that geography influenced these results.

Structure analyses were run using the admixture model without assigning individuals to *a priori* populations. We tested *K* values 1-10 with a burn-in period of 250,000 generations, 1 million MCMC chains, and 10 replicate runs for each *K* value. We then ran Structure separately

on the northeast cluster of *S. hesperis* (identified in the maximum likelihood phylogeny presented in this study and in the Structure analyses of Campbell *et al.* (2019)) using the same parameters, but testing K = 1-5. We used CLUMPAK (Kopelman *et al.* 2005) to determine the optimal K considering both the Evanno method (Evanno *et al.* 2005) and lnPr(K) (Pritchard & Wen 2004). To assess the influence of spatial structure on genetic clustering, we used TESS to infer K values 2-10 using the latitude and longitude coordinates of the sampling location for each specimen. We ran this analysis using the CAR admixture model (Durand *et al.* 2009) for 20 replicates per K, with a burn-in period of 10,000 and 50,000 sweeps (analogous to "generations" in Structure), and sampled the spatial interaction parameter and variance during the MCMC runs. Following program recommendations we identified the best 10 runs for each value of K using the Deviance Information Criterion (DIC) score, and these scores were then averaged and plotted. The optimal K was identified as the lowest value at which the DIC scores stabilized. We also calculated expected and observed heterozygosity and pairwise F_{st} for each major SNP cluster of *S. hesperis* and *S. atlantis* identified by Structure and TESS.

5.3.4 *Species delimitation and introgression analyses*

We conducted species delimitation testing using BFD* (Leaché *et al.* 2014) implemented in the SNAPP plug-in (Bryant *et al.* 2012) for BEAST 2 (Bouckaert *et al.* 2014). SNAPP uses the multispecies coalescent (MSC) to estimate trees, effective population sizes, and divergence times from SNPs by inferring probabilities of allele frequency change, and then outputs a posterior distribution that represents different estimations of the species tree (Bryant *et al.* 2012). BFD* is a program built specifically for species delimitation using SNPs that uses the MSC approach implemented in SNAPP to consider each of these independent genealogies and test

alternate species delimitation models specified by the user (Leaché *et al.* 2014). In BFD*, the user runs alternate species scenarios, and then uses the marginal likelihood (MLE) output by each analysis to calculate Bayes Factors (Grummer *et al.* 2014), which are used to determine the best supported model (Leaché *et al.* 2014). Because this program is computationally demanding, we used the dataset with 433 SNPs and further reduced the total number of individuals to 17 to speed up the analysis. Using the *K*=9 Structure analysis as a guide (described below), we included either two or three specimens from major genetic clusters as follows: three each from *S. zerene* (sampled from AB, UT, and CA)*, S. atlantis* (sampled from AB, ON, and CO)*,* and the northeast cluster of *S. hesperis* (sampled from AB, BC, and SD)*,* and two specimens each from the northern UT, southern UT, northern NM, and southern NM *S. hesperis* clusters. We did not sample from the Arizonan population of *S. hesperis* as these specimens did not form a distinct cluster, and treated *S. zerene* as a single cluster, despite admixture in some specimens (discussed below). We note that the Californian *S. zerene* specimen we used for this analysis was not the apparent hybrid mix between *S. zerene/S. hesperis* (see results below), and that we chose individuals for the BFD* analysis with little or no genomic admixture in the Structure and TESS analyses to ensure that these analyses weren't biased by contemporary hybridization. We also excluded *S. cybele* (Fabricius, 1775) and *S. aphrodite* from this analysis as the SNP phylogeny indicated a basal relationship of these taxa to the *S. hesperis* and *S. zerene* clade; the inclusion of outgroups or similarly basal taxa can bias the estimation of marginal likelihoods (Leaché *et al.* 2014).

Using the genetic clusters recovered from the maximum likelihood phylogeny and the Structure/TESS results, we tested seven species delimitation scenarios focused on the relationship between *S. atlantis, S. hesperis,* and *S. zerene*: (i) the "*a priori*" scenario, which

follows the current species delimitation for *S. zerene, S. hesperis,* and *S. atlantis*; (ii) the "*K*=2" scenario, which represents the results of the Structure analysis indicating an optimal *K* of 2; (iii) the "*K*=9" scenario generally following the *K*=9 Structure results, which indicated strong geographic sub-structuring of the southwestern *S. hesperis* cluster; (iv) the "*K*=6" scenario that follows the TESS results; and (v-vii) three additional "clade" scenarios based on alternate species groupings from the recovered maximum likelihood phylogeny. Details about the groupings tested for each scenario can be found in Table 5.1.

Following the program recommendations, we set the mutation parameters *u* and *v* to 1, and allowed the coalescence rate to be sampled via MCMC to reflect probable differences in population size between lineages. We also included non-polymorphic sites in the analysis because our dataset contained some missing data. BFD* implements a birth-only Yule tree prior, which we set to have a gamma distribution with a single parameter, λ, governing speciation rate (Leaché *et al.* 2014). We calculated λ from the maximum likelihood consensus SNP tree output from IQ-TREE using the package phytools 0.6-99 (Revell 2012) implemented in R 3.6.1 (R Core Team 2017), and used it to determine the β scale parameter with an α shape parameter of 2. This gave us a gamma distribution of λ=65.96, α=2, and β=32.9. In order to avoid potential sampling bias by the program due to the narrow parameter distribution for λ and β indicated by the data, we further relaxed our λ to 200 and our β to 100. Finally, following program recommendations, we set our rate priors to also follow a gamma distribution, with λ=10, α=1, and β=250 (Leaché and Bouckaert 2018), and ran each scenario with 500,000 MCMC chains, 100,000 burn-in replicates, and 24 path sampling steps. Convergence was assessed using the program Tracer (Rambaut *et al.* 2018), and TreeAnnotator 2.4.7 (Drummond & Rambaut 2007) was used to generate the maximum clade credibility tree. We additionally used DensiTree (Bouckaert 2010)

to visualize topological discordance in the posterior distribution of trees recovered by this analysis.

Finally, we used the program TreeMix (Pickrell & Pritchard 2012) to calculate *f*3 statistics, which estimate admixture between specified parental and admixed populations. These tests were conducted on three populations at a time (for example: [A;B,C] tests for admixture in population A resulting from introgression between two parental populations, B and C) (Reich *et al.* 2009; Pickrell & Pritchard 2012) for all possible combinations using the genetic clusters identified in the *K*=9 Structure results and including *S. aphrodite*.

5.4 Results

5.4.1 *SNP and COI dataset construction*

We produced two SNP datasets, one filtered for phylogenetic and population genetic analyses that allowed up to 20% missing data (728 SNPs retained, 1 SNP per locus, min. locus depth: 16, max. locus depth: 234, mean locus depth: 72.6), and the other only allowing up to 10% missing data (433 SNPs retained, 1 SNP per locus, min. locus depth: 25, max. locus depth: 234, mean locus depth: 80.07) to be used for species delimitation and introgression analyses. The *COI* dataset consisted of 579 invariant and 69 variant sites, 62 of which were phylogenetically informative.

5.4.2 *Phylogenetic reconstruction*

The SNP species tree shows *S. cybele, S. aphrodite* and *S. zerene* as monophyletic clades, and *S. aphrodite* as sister to a clade containing *S. atlantis, S. hesperis,* and *S. zerene* (Fig. 5.1a). The recovered topology indicates a polyphyletic relationship for northeastern and southwestern

S. hesperis populations. The northeastern cluster was additionally paraphyletic with *S. atlantis,* and contained subspecies *S. h. beani* (Barnes & Benjamin, 1926) and *S. h. dennisi* dos Passos & Grey, 1945 from Alberta, *S. h. hutchinsi* dos Passos & Grey, 1947 from Montana*, S. h. brico* Kondla, Scott & Spomer, 1998 and *S. h. beani* from British Columbia*, S. h. lurana* dos Passos & Grey, 1945 from South Dakota, and *S. h. ratonensis* Scott, 1981 from Colorado, but did not exhibit consistent geographic sub-clustering. The southwestern *S. hesperis* cluster was paraphyletic with *S. zerene*, and itself contained two major geographic clusters: one with *S. hesperis* from Utah and southeastern Colorado, and the other containing *S. hesperis* from New Mexico and Arizona, as well as two specimens collected from southeast Utah near the Utah-Colorado border. Three specimens of *S. hesperis viola* dos Passos & Grey 1945 sampled from Idaho were divided between the two southwestern clusters. Within the Arizona-New Mexico cluster, we recovered fine-scale population structuring that broadly separated a southern New Mexico population of *S. h. capitanensis* Holland, 1988 sampled in the Sacramento Mountains from the more northern population of *S. h. dorothea* Moeck, 1947 sampled in the Sandia Mountains, and *S. h. nausicaa* (Edwards, 1874) sampled in Arizona. Similarly, the Utah cluster of *S. hesperis* had two major groupings, one of *S. h. tetonia*, dos Passos & Grey, 1945 sampled in northern Utah, and another consisting of *S. h. chitone* (Edwards, 1879) sampled from southeastern Utah and southwestern Colorado. Branch support for both the southwestern *S. hesperis/S. zerene* and the northeastern *S. atlantis/S. hesperis* clades was robust, with bootstrap values at 97% and 99%, respectively, indicating a high degree of topological congruence among bootstrap replicates.

Only *S. cybele* and *S. atlantis* were monophyletic on the *COI* gene tree (Fig. 5.1b); all other previously recognized species were either para- or polyphyletic. The geographic sub-

clustering of *S. hesperis* on the SNP phylogeny was absent in the *COI* phylogeny, which instead depicted a single haplotype for all *S. hesperis* specimens, interdigitated with *S. zerene* from California and Nevada, as well as a single specimen of *S. aphrodite manitoba* (Chermock & Chermock, 1940) collected in Alberta. In the *COI* tree, *S. atlantis* had a sister relationship with *S. zerene* sampled in Alberta, Utah, Idaho, and Montana, and *S. aphrodite* was sister to the mixed *S. hesperis* clade. The major branches of this gene tree were well supported, with bootstrap values over 70% (Fig. 5.1b).

5.4.3 *Haplotype network, Structure, and TESS*

Structure analysis of SNP data suggested two optimal values of *K*. The Evanno method of Δ*K* (Evanno *et al.* 2005) supported *K*=2, which grouped northeast *S. hesperis* with *S. atlantis,* and the southwestern cluster of *S. hesperis* with *S. zerene. Speyeria aphrodite* was unresolved and admixed between these broadly "northern" and "southern" clusters (Fig. 5.2a). Alternately, the lnPr(*K*) method (Pritchard & Wen 2004) indicated an optimal *K* of 9. This resolved *S. aphrodite* and *S. atlantis,* and additionally separated *S. hesperis* into five geographically-defined clusters: 1. northeast (British Columbia, Alberta, Montana, South Dakota, and southeastern Colorado specimens), 2. central (northern Utah, Idaho, and southwestern Colorado specimens), 3. southern Utah, 4. northern New Mexico, and 5. southern New Mexico. Two of the northern New Mexico specimens and all the Arizonan specimens appeared to be mixtures between northern and southern New Mexico populations. Similarly, the Colorado *S. hesperis* specimens were intermediate between the northeast and central populations, and the Idaho specimens were intermediate between the central and southern Utah populations (Fig. 5.2b). The *K*=9 results also indicate unique genetic differences in *S. zerene* from California that were not shared by any *S.*

zerene specimens sampled from Nevada, Utah, Idaho, Montana and Alberta. Hierarchical Structure analysis of the northeast *S. hesperis* cluster did not indicate additional geographic substructuring; both lnPr(*K*) and Δ*K* indicated an optimal *K* of 2, however this did not produce any meaningful sub-structure in the data that corresponded to sampling locality, and given that Δ*K* cannot estimate *K*=1, we suggest that *K*=1 is a more meaningful result (results not shown).

TESS suggested an optimal *K* of 6. This analysis was largely congruent with the *K*=9 Structure results, except that it united the central and southern Utah populations into a single cluster and the New Mexico and Arizonan populations into a single cluster (Fig. 5.2a). Both TESS and Structure indicated a few likely hybrids: two specimens of *S. hesperis*, sampled from Alberta and South Dakota, shared ancestry with *S. aphrodite*; one *S. hesperis* sampled from southern Utah shared ancestry with the northern New Mexico *S. hesperis* population, and one *S. zerene* from California shared ancestry with the southern Utah *S. hesperis* population.

The minimum spanning haplotype network depicted distinct *S. atlantis* and *S. hesperis* clusters, however there was very little haplotype variation within either *a priori* species (Fig. 5.2c). In all cases there were only one or two nucleotide differences between the "distinct" specimens and the major haplotype group for each species. For *S. atlantis,* this haplotype variation largely correlated to sampling location - eastern *S. atlantis canadensis* (dos Passos, 1935) specimens sampled in Manitoba, Ontario, and Quebec and the *S. atlantis sorocko* Kondla & Spomer, 1998 specimen from Colorado were marginally distinct from the *S. atlantis hollandi* (Chermock & Chermock, 1940) specimens sampled in Alberta; only one *S. atlantis hollandi* from Alberta contained a single nucleotide difference from the major "*hollandi*" haplotype group. This geographic pattern was not observed in *S. hesperis,* as almost all the specimens sampled had identical haplotypes regardless of sampling location. Comparatively, *S. zerene* had

much more haplotype diversity in the minimum spanning network, with three distinct haplogroups. One group consisted of the *S. zerene* specimens sampled from Alberta, Idaho, Montana, and Utah and was the most distinct from *S. hesperis* and *S. atlantis*, and a second group containing specimens from California was intermediate between the *S. hesperis* and *S. atlantis* haplogroups. Interestingly, the third *S. zerene* haplotype that was found in specimens sampled from Nevada was identical to the major *S. hesperis* haplotype (Fig. 5.2c), but these individuals did not appear admixed with *S. hesperis* in the Structure or TESS analyses of SNPs; the single likely *S. zerene* hybrid was instead from California, and had a haplotype consistent with the rest of the Californian *S. zerene* specimens.

Expected heterozygosity (H_e) was higher than observed heterozygosity (H_o) in all populations, but the difference between these values was most pronounced in the southern New Mexico/Arizonan population ($H_e = 0.29$, $H_o = 0.1$, Table 5.2). The northeast *S. hesperis* clade had the lowest observed heterozygosity at 0.07, followed by the northern New Mexico *S. hesperis* population and *S. atlantis*. F_{st} values were generally quite high for each pairwise comparison (between 0.13-0.64, Table 5.2), however they indicated that the northeast cluster of *S. hesperis* was more genetically similar to *S. atlantis* than to any other *S. hesperis* population, and that geographically proximate populations of *S. hesperis* (excluding the northeastern cluster) were more similar to each other than to geographically distant populations (Table 5.2).

5.4.4 *Introgression and species delimitation analyses*

Given the discordant relationship between *S. zerene* and *S. atlantis/S. hesperis* in our phylogenetic analyses (Fig. 5.1) we used the three-population *f*3 statistic to test for admixture in the dataset that might indicate past introgression events between these species. We calculated *f*3

statistics for every three-population combination using the major genetic clusters recovered in the Structure *K*=9 results (Fig. 5.2a); a significantly negative *f*3 statistic supports the hypothesis of an introgression event between two parental populations and a putatively admixed population (Reich *et al.* 2009). Of the 168 tests we computed, ten had negative *f*3 statistics, and only one was significant with a p-value < 0.01 (Appendix 5.2). This test indicated likely admixture from *S. zerene* and the southern Utah population of *S. hesperis* into the central *S. hesperis* population.

Of the seven species scenarios tested in BFD*, the *clades1* scenario, which tested four species (*S. zerene, S. atlantis,* northeast *S. hesperis*, and southwest *S. hesperis*) had the highest marginal likelihood estimate (MLE) and the most strongly negative Bayes Factor (BF), indicating that this model was the best supported by the data (Table 5.1). The *K=6* scenario reflecting the TESS results (Fig. 5.2a) was the second best supported model, and the *a priori* model had the lowest support of all seven scenarios, followed closely by the *K=2* model. The maximum clade credibility (MCC) tree of the *clades1* model indicated strong support for the northeastern *S. hesperis* and *S. atlantis* clades (posterior probability of 1), but the clade containing *S. zerene* and southwestern *S. hesperis* was much more poorly supported, with a posterior probability of 0.69 (Fig. 5.3). The MCC tree additionally indicated that the southwestern *S. hesperis/S. zerene* clade diverged slightly earlier than the clade containing *S. atlantis* and northeastern *S. hesperis*, and that the southwestern *S. hesperis* lineage had the largest effective population size, indicated by a larger theta value (Fig. 5.3). *Speyeria atlantis* had the lowest theta value, indicating the smallest estimate for effective population size.

The DensiTree visualization of the *clades1* analysis indicated discordance in the relationship between *S. zerene* and the remaining three species on the tree. The major recovered topology, which accounted for 69% of the sampled trees (shown in blue in Fig. 5.3), was the

same as the topology presented in the MCC tree and indicated a sister relationship between *S. zerene* and southwestern *S. hesperis*, however the second most common topology (representing 20.2% of the sampled trees, shown in purple) depicted *S. zerene* as the sister taxon to the *S. atlantis/*northeastern *S. hesperis* clade. A third topology accounting for the remaining 10.9% of the sampled trees (shown in orange) depicted *S. zerene* as basal to the other three species groupings on the tree. The *S. atlantis*/northeastern *S. hesperis* sister relationship was consistently recovered in each sampled topology.

5.5 Discussion

5.5.1 *Historical introgression and mito-nuclear discordance*

While Structure and TESS indicated little contemporary hybridization between the sampled populations of *S. atlantis, S. hesperis* and *S. zerene*, multiple analyses suggest a complex, shared evolutionary history between them. The three topologies depicted in the DensiTree visualization of the BFD* analysis in SNAPP unambiguously supported *S. atlantis* and the northeastern lineage of *S. hesperis* as sister taxa, but the relationship of *S. zerene* to the other clades varied (Fig. 5.3). Previous work on *Speyeria* species relationships using mtDNA sequence or low numbers of nuclear genes showed a lack of monophyly for both *S. hesperis* and *S. zerene* (McHugh *et al.* 2013; de Moya *et al.* 2017), however greater numbers of nuclear loci have recovered these species as monophyletic and consistently support a non-sister relationship between them (Campbell *et al.* 2017, Thompson *et al.* 2019, but see Campbell *et al.* (2019), which recovered *S. hesperis* as paraphyletic with *S. atlantis*).

Mito-nuclear discordance can be expected to be caused both by incomplete lineage sorting (ILS) and past hybridization events (Linnen & Farrell 2007; Li *et al.* 2016). SNAPP

analysis does not distinguish between these two processes (Bryant *et al.* 2012). However, *f*3 statistical tests (Appendix 5.2) and the lack of sequence variation in the mitochondrial haplotype shared between *S. zerene* and *S. hesperis* (Fig. 5.2c) lends support to the hypothesis that at least one lineage of *S. hesperis* has introgressed into *S. zerene*. The distribution of this haplotype throughout *S. hesperis* but only in the Nevadan part of the range of *S. zerene* suggests that it originated in *S. hesperis*. In contrast to mtDNA, nuclear SNPs across the range of *S. hesperis* do not show an obvious reduction in genetic variability, suggesting that a strong selective sweep leading to a severe bottleneck event has recently caused the loss of other variable mitochondrial haplotypes (Sonsthagen *et al.* 2017; Hurst & Jiggins 2005).

A candidate for facilitating such a process is *Wolbachia* Hertig & Wolbach, 1924 bacteria (Werren *et al.* 2008; Kodandaramaiah *et al.* 2013; Ahmed *et al.* 2015). *Wolbachia* infections have been reported in *S. diana* (Cramer, 1777) and *S. idalia* (Drury, 1773) (Hamm *et al.* 2014) and a few specimens of *S. zerene* (McHugh *et al.* 2013), however *Wolbachia* testing of a few *S. atlantis* specimens did not indicate any infection (McHugh *et al.* 2013). *Wolbachia* infection offers a plausible and testable hypothesis that could explain the observed haplotype sharing between Nevadan *S. zerene* and *S. hesperis* in the absence of contemporary nuclear admixture between the sampled specimens in the Structure and TESS analyses, which may further clarify the historical relationship between these taxa.

5.5.2 *Phylogeography suggests a recent, northward expansion of S. hesperis and S. atlantis*

Structuring of nuclear SNPs in southwestern *S. hesperis* was most pronounced in the north New Mexico/Arizonan population, where H_e was almost three times higher than H_o (Table 5. 2). Such high genetic diversity with low heterozygosity indicates strong population isolation

(Keller & Waller 2002; Sattler *et al.* 2017), and suggests that the heterogenous landscape in and around the southern Rocky Mountains is a significant barrier to gene flow.

In contrast, the lack of population structure (Fig. 5.2a-b) and low H_o (Table 5.2) within each of the northeastern *S. hesperis* and *S. atlantis* populations suggests greater gene flow, whether due to fewer current barriers to dispersal and/or more recent postglacial range expansion of these taxa northward and into the Great Plains (Milá *et al.* 2007; Excoffier *et al.* 2009). This is supported by the BFD*/SNAPP analysis in Fig. 5.3, which indicates a slightly more recent divergence estimate of northeastern *S. hesperis* and *S. atlantis*. This contrasts with the traditional hypothesis that *S. hesperis* is an evolutionary offshoot of *S. atlantis* (dos Passos & Grey 1947), and suggests that some southern populations of *S. hesperis* may pre-date the divergence of *S. atlantis*. These results also corroborate other studies that suggest several western *Speyeria* species inhabited regions in or near the American Rocky Mountains and experienced cycles of isolation in glacial refugia and secondary contact during the Pleistocene, before recently diverging and radiating north (Dunford 2009; de Moya *et al.* 2017; Thompson *et al.* 2019).

5.5.3 *Species delimitations of S. hesperis* and *S. atlantis*

Our results consistently show that the northeastern *S. hesperis* clade is more closely related to *S. atlantis* than to any of the southwestern *S. hesperis* clusters (Figs. 5.1 and 5.3, Table 5.2), but we did not recover any evidence of admixture between these taxa (Fig. 5.2). Other work has noted differences in micro-habitat preference between *S. atlantis* and *S. hesperis* in regions where they co-occur (Bird *et al.* 1995; Guppy & Shepard 2001; Dunford 2009; Riva *et al.* in press), and no genetic admixture was reported between these taxa in other genetic surveys (Thompson *et al.* 2019; Campbell *et al.* 2019). Thus, *S. atlantis* is an independently evolving

lineage distinct from *S. hesperis*, and recognition of the species status of these two taxa should be maintained.

In contrast to the delimitation of *S. atlantis*, our re-assessment of the species boundary of *S. hesperis* gave variable results, with strong genetic differentiation of SNPs in the southwest (Fig. 5.2a-b). This contrasts with the haplotype network (Fig. 5.2c) and *COI* gene tree (Fig. 5.1b) that had one major, invariant haplotype of *S. hesperis* and only a few specimens containing one or two mutations. The fact that this major haplotype was also found in some specimens of *S. zerene* means that *COI* is only partially informative for clarifying species boundaries of *S. hesperis*. Structure and TESS analyses of SNPs indicated that the northeastern and southwestern lineages of *S. hesperis* were consistently genetically distinct and that the central population containing specimens from northern Utah, Idaho, and western Colorado may represent a region of introgression between these lineages. Both analyses also recovered strong substructure in the southwestern lineage of *S. hesperis* that is likely to have been a result of population isolation in the Rocky Mountains, as discussed above.

BFD* species delimitation using SNAPP indicated clear support for the *clades1* model that split *S. hesperis* into two species - one comprised of the southwestern lineages and the other representing the northeastern lineage (Fig. 5.3). In contrast to programs like TESS and Structure, which estimate the probability of assignment based on allele frequencies, coalescent methods estimate the probability of allele frequency change through time to determine the shift from population-level processes to those of species (Bryant *et al.* 2012; Leaché *et al.* 2014). Proponents of this approach have suggested that the Bayesian nature of MSC-based species delimitation algorithms should be robust to large intraspecific divergences that result from population structure and thereby reduce oversplitting of lineages (Yang & Rannala 2017, but see

Sukumaran & Knowles 2017), hence the strong support for the *clades1* model over the more speciose scenarios supported by Structure and TESS (Fig. 5.2, Table 5.1)

Pairwise F_{st} values also indicated greater genetic similarity between northeastern *S. hesperis* and *S. atlantis* than between northeastern and southwestern *S. hesperis* lineages (Table 5.2), and there are noted morphological differences between broadly southern and northern populations of *S. hesperis*. Northern and central subspecies are variably silvered or unsilvered on the hind wing underside, and have wingspans typically between 40-55 mm (e.g. *S. h. dennisi, S. h. beani, S. h. lurana, S. h. brico, S. h. tetonia, S. h. chitone*), while southern populations, particularly in New Mexico and Arizona, are always silvered on the underwing and have wingspans between 60-79 mm (e.g. *S. h. capitanensis, S. h. dorothea, S. h. nausicaa, S. h. schellbachi* Garth, 1949) (Dunford 2009).

Together, these results unambiguously support a recent evolutionary divergence between northeastern and southwestern lineages of *S. hesperis*, but also indicate that these lineages currently exist as distinct entities shaped by interactions with other species that nonetheless maintain their genomic integrity when in contact. Thus, we recommend that the two lineages should be recognized as distinct species. Following taxonomic priority (Pelham 2019), the northeastern lineage should continue to be referred to as *S. hesperis*, and the oldest southwestern subspecies name sampled in this study, *S. h nausicaa*, should be elevated to species level, *S. nausicaa*, to represent the whole southwestern clade. This taxonomic revision remains a hypothesis that requires further sampling and genomic assessment, particularly in California, where even older, pre-Pleistocene lineages may persist. Several *S. hesperis* subspecies have been recognized in California, and one of these, *S. h. irene*, has taxonomic priority over *S. nausicaa*. But it is currently not clear whether the Californian populations will be more closely related to

the northeastern or the southwestern populations of *S. hesperis*, or will represent a third distinct lineage.

5.5.4 *Conclusions*

Our results provide strong genomic evidence for recent speciation and range expansion events in the *S. atlantis-hesperis* species complex, despite persistent morphological and genetic ambiguities. Using SNPs, we clarify the evolutionary relationship between *S. atlantis* and *S. hesperis* and suggest phylogeographic hypotheses for the diversification of these species. Our results support recognition of the southwestern lineage of *S. hesperis* as a distinct species, *S. nausicaa*. We also detected introgression between *S. zerene* and *S. hesperis/S. nausicaa* that may be a source of nuclear gene tree-species tree discordance. This should be pursued due to the conservation status of several *S. zerene* subspecies. Whole genome sequencing will be especially useful for assessing the extent of genomic introgression between *S. zerene* and other *Speyeria*, and will have ramifications for informing ongoing conservation and management strategies.

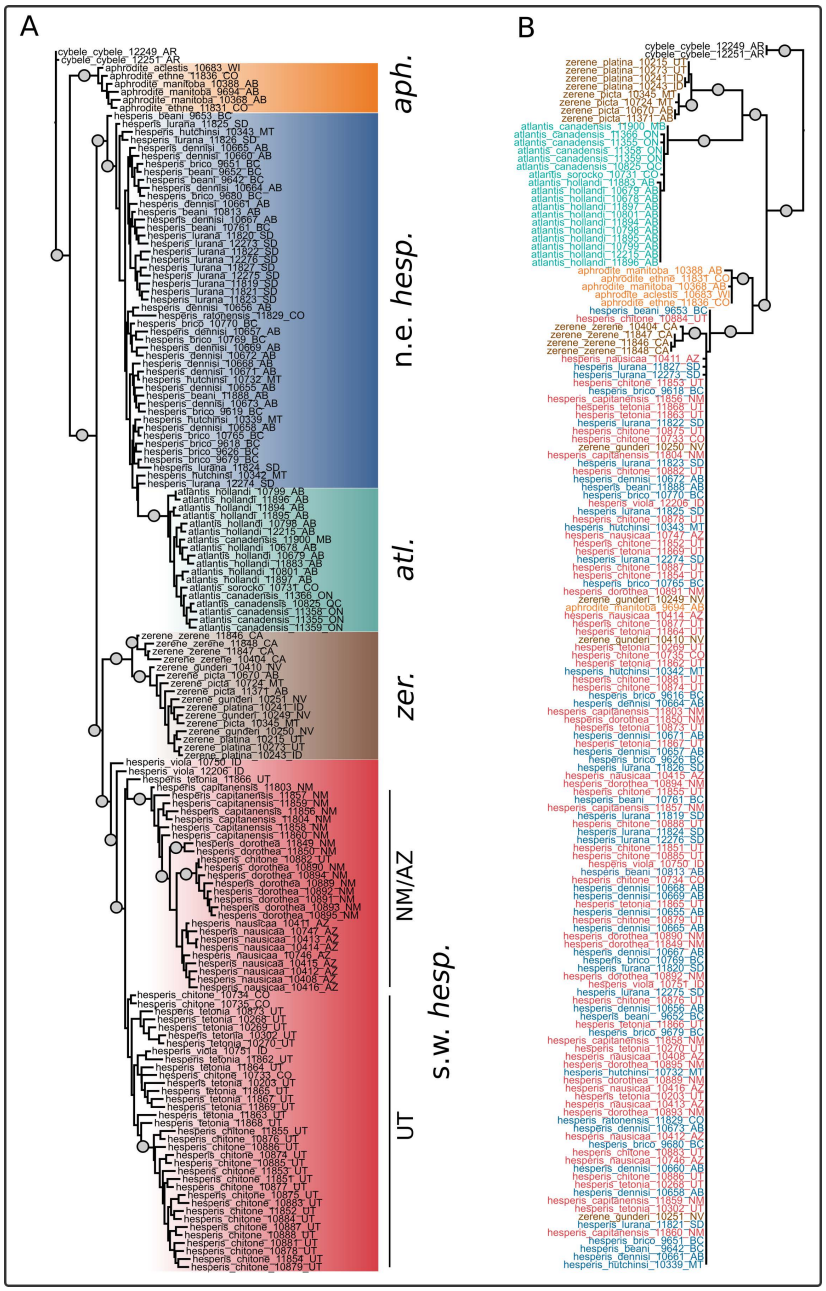


Figure 5.1 Maximum likelihood phylogenies based on (A) nuclear SNPs and (B) mitochondrial *COI* sequence. Specimens in (B) are coloured according to their group membership in (A). The UT and NM/AZ *S. hesperis* designations on the SNP tree (A) represent broad regional groupings that also include a minority of specimens sampled from CO and ID, and do not form monophyletic clades based on locality. Grey circles on nodes indicate bootstrap values >70%.

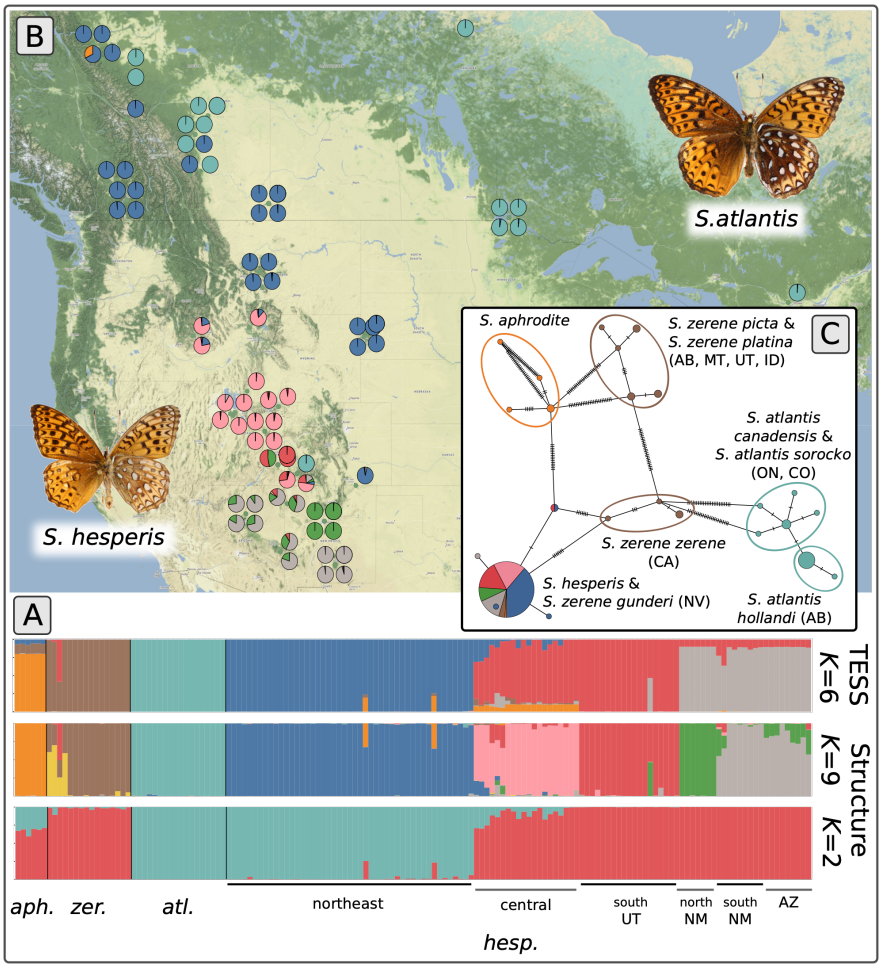


Figure 5.2 Geographic assessment of population genetic clustering of *S. atlantis* and *S. hesperis* using SNPs. Structure and TESS results (A) consistently indicate major genetic divergence between northeastern and southwestern *S. hesperis* lineages. Populations of *S. hesperis* and *S. atlantis* identified in the *K*=9 Structure analysis are plotted as pie charts on the map in (B) to show correspondence between genetic and geographic structure. The *COI* gene haplotype network (C) depicts a lack of geographic structure in *S. hesperis* and haplotype sharing between *S. hesperis* and Nevadan *S. zerene*. Hatches along the branches in (C) indicate the number of nucleotide differences between sequences. Top right image: *S. atlantis hollandi*; lower left image: *S. hesperis beani*. The dorsal wing surface is shown on the left side of each specimen, and the ventral surface is depicted on the right.

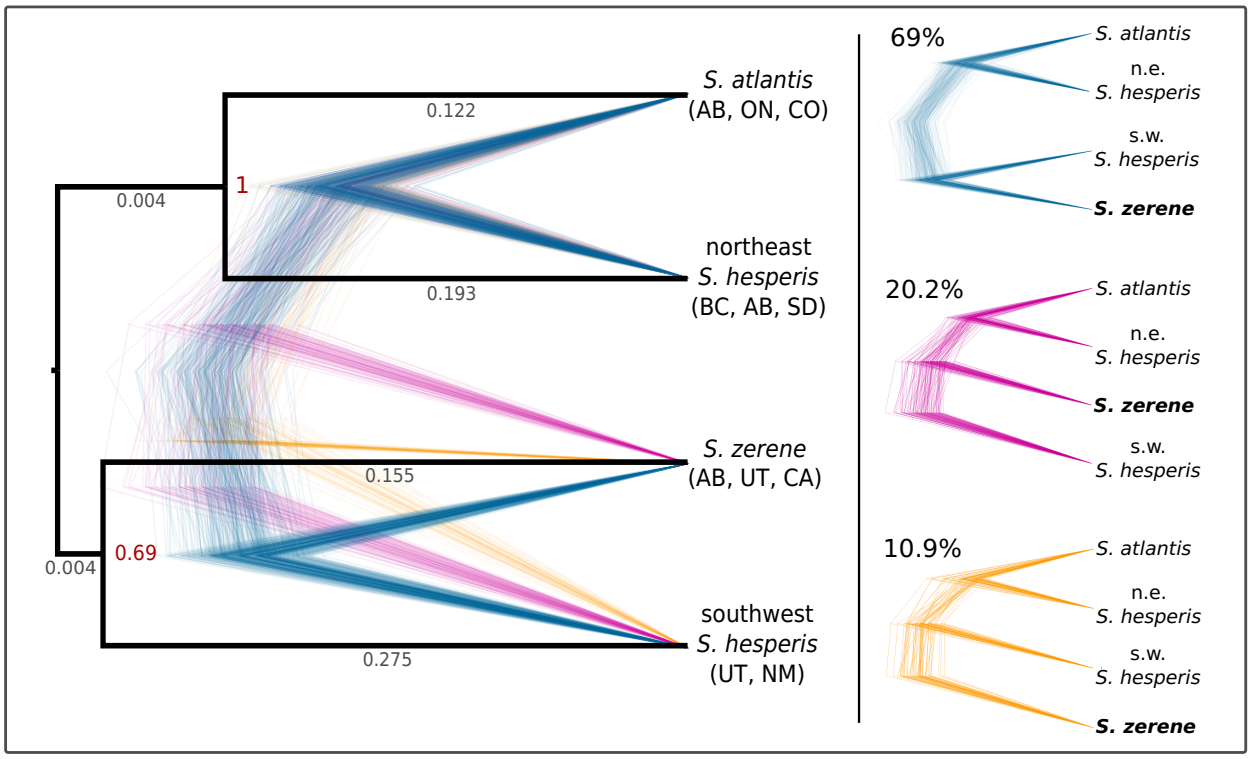


Figure 5.3 Left panel: Results of the BFD* species delimitation analysis of the *clades1* scenario visualized as both a maximum clade credibility (MCC) tree (overlaid in black), and as a "tree cloud" output by DensiTree (in colour) supporting northeastern and southwestern *S. hesperis* as distinct species. Posterior probability values are indicated in red at both nodes on the MCC tree, and grey values along the branches indicate the estimates of theta (=N_e) for each lineage. Right panel: DensiTree visualizes genealogical discordance between SNPs as the source of poor branch support for the *S. zerene*/southwest *S. hesperis* clade on the MCC tree in the left panel, and indicates three alternate relationships between *S. zerene* and *S. atlantis/S. hesperis*. Percentages above each Densitree represent the proportion of trees in the posterior distribution depicting that topology.

Table 5.1 BFD* model selection results. The *clades* scenarios are derived from alternate groupings of clusters on the SNP phylogeny in Fig. 1a. The highest marginal likelihood estimation (MLE) and most negative Bayes factor (BF) values indicate strongest model support, therefore the *clades1* scenario was the most optimal model.

Model	No. of species	Species tested	MLE	BF
a priori	3	current delimitation: *S. atlantis, S. hesperis, S. zerene*	-3765.21	--
K2	2	*atlantis* + n.e. *hesperis*, s.w. *hesperis* + *zerene*	-3729.03	-72.36
K6	5	*atlantis*, n.e. *hesperis*, New Mexico *hesperis*, Utah *hesperis*, *zerene*	-3642.78	-244.9
K9	7	*atlantis*, n.e. *hesperis*, central *hesperis*, s. UT *hesperis*, n. NM *hesperis*, s. NM *hesperis*, *zerene*	-3638.61	-253.2
clades1	4	*atlantis*, n.e. *hesperis*, s.w. *hesperis*, *zerene*	-3630.2	-270
clades2	4	*atlantis* + n.e. *hesperis*, NM *hesperis*, UT *hesperis*, *zerene*	-3637.64	-255.1
clades3	3	*atlantis* + n.e. *hesperis*, s.w. *hesperis*, *zerene*	-3643.58	-243.3

Table 5.2 Heterozygosity and F_{st} estimates for genetic clusters of *S. atlantis* and *S. hesperis*

		Pairwise F_{st} estimates						
N		*atlantis*	n.e. *hesperis*	central *hesperis*	s. UT *hesperis*	n. NM *hesperis*	H_o	H_e
18	*atlantis*	--					0.087	0.123
47	n.e. *hesperis*	0.28	--				0.07	0.129
20	central *hesperis*	0.436	0.303	--			0.122	0.17
19	s. UT *hesperis*	0.509	0.399	0.128	--		0.134	0.153
9	n. NM *hesperis*	0.638	0.517	0.271	0.307	--	0.085	0.127
16	s. NM/AZ *hesperis*	0.57	0.465	0.201	0.253	0.18	0.101	0.289

5.6 Literature cited

Abdelkrim, J., Aznar-Cormando, L, Buge, B., Fedosov, A., Kantor, Y., Zaharias, P, and Puillandre, N. (2018) Delimiting species of marine gastropods (Turridae, Conoidea) using RAD sequencing in an integrative taxonomy framework. *Molecular Ecology*, **27**, 4591-4611.

Ahmed, M. Z., Araujo-Jnr, E. V., Welch, J. J. and Kawahara, A. Y. (2015) *Wolbachia* in butterflies and moths: geographic structure in infection frequency. *Frontiers in Zoology*, **12**, 1-9.

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

Bandelt, H., Forster, P., and Röhl, A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37-48.

Bouckaert, R. (2010) DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics*, **26**, 1372-1373.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C., Xie, D., Suchard, M. A., Rambaut, A., and Drummond, A. J. (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, **10**, e1003537.

Brock, J. P. and Kaufman, K. (2003) *Kaufman Field Guide to Butterflies of North America*, New York: Houghton Mifflin Co.

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., and Choudhury, A. R. (2012) Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, **29**, 1917-1932.

Campbell, E. O., Davis, C. S., Dupuis, J. R., Muirhead, K., and Sperling, F. A. H. (2017) Cross-platform compatibility of *de novo*-aligned SNPs in a nonmodel butterfly genus. *Molecular Ecology Resources*, **6**, e84-e93.

Campbell, E. O., Gage, E., V., Gage, R. V., and Sperling, F. A. H. (2019) SNP-based species phylogeny of greater fritillary butterflies (Lepidoptera: Nymphalidae: *Speyeria*) demonstrates widespread mito-nuclear discordance. *Systematic Entomology*, DOI:10.1111/syent.12393

Carstens, B. C., Pelletier, T. A., Reid, N. M., and Satler, J. D. (2013) How to fail at species delimitation. *Molecular Ecology*, **22**, 4369-4383.

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., and Postlethwait, J. H. (2011) Stacks: building and genotyping loci de novo from short-read sequences. *G3*, **1**, 171-182.

Chen, C., Durand, E., Forbes, F., and François, O. (2007) Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Molecular Ecology Notes*, **7**, 747-756.

Coates, D. J., Byrne, M., and Moritz, C. (2018) Genetic diversity and conservation units: dealing with the species-population continuum in the age of genomics. *Frontiers in Ecology and Evolution*, **6**, 165.

Cornuet, J. M., Pudlo, P., Veyssier, J., Dehne-Garcia, A. Gautier, M., Leblois, R., Marin, J. M., and Estoup, A. (2014) DIYABC v2.0: a software to make approximate Bayesian computataion inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics*, **30**, 1187-1189.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … 1000 Genomes Analysis Group. (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156-2158.

Degnan, J. H., and Rosenberg, N. A. (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology and Evolution*, **24**, 332-340.

de Queiroz, K. (1998) The general lineage concept of species, species criteria, and the process of speciation. IN Howard, D. J., Berlocher, S. H. (Eds.) *Endless forms: species and speciation*. Oxford University Press, New York.

de Queiroz, K. (2007) Species Concepts and Species Delimitation. *Systematic Biology*, **56**, 879-886.

dos Passos, C. F., and Grey, L. P. (1947) Systematic catalogue of *Speyeria* (Lepidoptera, Nymphalidae) with designations of types and fixations of type localities. *American Museum Novitates,* **1370**, 1-30.

Drummond, A. J., and Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**: 214.

Dunford, J. C. (2009) Taxonomic overview of the greater fritillary genus *Speyeria* Scudder and the atlantis-hesperis species complexes, with species accounts, type images, and relevant literature (Lepidoptera: Nymphalidae). *Insecta Mundi*, **90**, 1-74.

Durand, E., Jay, F., Gaggiotti, O. E., and François, O. (2009) Spatial inference of admixture proportions and secondary contact zones. *Molecular Biology and Evolution*, **14**, 2611-2620.

Ebel, E. R., DaCosta, J. M., Sorenson, M. D., Hill, R. I., Briscoe, A. D., Willmott, K. R., and Mullen, S. P. (2015) Rapid diversification associated with ecological specialization in Neotropical *Adelpha* butterflies. *Molecular Ecology*, **24**, 2392-2405.

Escudero, M., Eaton, D. A. R., Hahn, M., Hipp, A. L. (2014) Genotyping-by-sequencing as a tool to infer phylogeny and ancestral hybridization: a case study in *Carex* (Cyperaceae). *Molecular Phylogenetics and Evolution*, **79**, 359-367.

Evanno, G., Regnaut, S., and Goudet, J. (2005) Detecting the number of clusters of individuals using the software Structure: A simulation study. *Molecular Ecology*, **14**, 2611-2620.

Excoffier, L. Foll, M., and Petit, R. J. (2009) Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 481-501.

Freitas, L., Mello, F., and Schrago, C. G. (2018) Multispecies coalescent analysis confirms standing phylogenetic instability in Hexapoda. *Journal of Evolutionary Biology*, **31**, 1623-1631.

Grey, L. P. (1951) The subspeciation of *Speyeria atlantis*. *The Lepidopterist's News*, **5**, 31-35.

Grummer, J. A., Bryson, R. W. Jr., and Reeder, T. W. (2014) Species delimitation using Bayes factors: simulations and applications to the *Sceloporus scalarsis* species group (Squamata: Phrynosomatidae). *Systematic Biology*, **63**, 119-133.

Grummer, J. A., Morando, M. M., Avila, L. J., Sites, J. W., Jr, and Leaché, A. D. (2018) Phylogenomic evidence for a recent and rapid radiation of lizards in the Patagonian *Liolaemus fitzingerii* species group. *Molecular Phylogenetics and Evolution*, **125**, 243-254.

Guppy, C. S., and Shepard, J. H. (2001) Butterflies of British Columbia. UBC Press; Vancouver, British Columbia. 414p.

Hamm, C. A., Handley, C. A., Pike, A., Forister, M. L., Fordyce, J. A., and Nice, C. C. (2014) *Wolbachia* infection and Lepidoptera of conservation concern. *Journal of Insect Science*, **14**:6.

Hammond, P. C., McCorkle, D., and Bergman, W. (2013) Hybridization studies of genomic compatibility and phenotypic expression in the greater fritillary butterflies (Nymphalidae: Argynnini). *Journal of the Lepidopterists' Society*, **67**, 263-273.

Hebert, P. D. N., Penton, E. H., Burns, J. M., Janzen, D. H., and Hallwachs, W. (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 14812-14817.

Hinojosa, J. C., Koubínová, D., Szenteczki, M. A., Pitteloud, C., Dincă, V., Alvarez, N., and Vila, R. (2019) A mirage of cryptic species: genomics uncover striking mitonuclear discordance in the butterfly *Thymelicus sylvestris*. *Molecular Ecology*. DOI: https://doi.org/10.1111/mec.15153

Hoang, D. T., Chernomor, O., von Haeseler, A., Ming, B. Q., and Vinh, L. S. (2018) UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution,* **35**, 518-522.

Hohenlohe, P. A., Day, M. D., Amish, S. J., Miller, M. R., Kamps-Hughes, N., Boyer, M. C., Muhlfeld, C. C., Allendorf, F. W., Johnson, E. A., and Luikart, G. (2013) Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. *Molecular Ecology*, **22**, 3002-3013.

Hundsdoerfer, A., K., Lee, K. M., Kitching, I. J., and Mutanen, M. (2019) Genome-wide SNP data reveal an overestimation of species diversity in a group of hawkmoths. *Genome Biology and Evolution*, **11**, 2136-2150.

Hurst, G. D. D. and Jiggins, F. M. (2005) Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society B*, **272**, 1525-1534.

Ivanov, V., Lee, K. M., and Mutanen, M. (2018) Mitonuclear discordance in wolf spiders: genomic evidence for species integrity and introgression. *Molecular Ecology*, **27**, 1681-1695.

Keller, L. F., and Waller, D. M. (2002) Inbreeding effects in wild populations. *Trends in Ecology and Evolution*, **17**, 230-241.

Kodandaramaiah, U., Simonsen, T. J., Bromilow, S., Wahlberg, N., and Sperling, F. A. H. (2013) Deceptive single-locus taxonomy and phylogeography: *Wolbachia*-associated divergence in mitochondrial DNA is not reflected in morphology and nuclear markers in a butterfly species. *Ecology and Evolution*, **3**, 5167-5176.

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., and Mayrose, I. (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, **15**, 179-1191.

Layberry, R. A., Hall, P. W., and Lafontaine, J. D. (1998) *The Butterflies of Canada*. University of Toronto Press

Leaché, A. D. and Bouckaert, R. R. (2018) Species trees and species delimitation with SNAPP: a tutorial and worked example. *Workshop on population and speciation genomics,* Český Krumlov

Leaché, A. D., Fujita, M. K., Minin, V. N., and Bouchaert, R. R. (2014) Species delimitation using genome-wide SNP data. *Systematic Biology*, **63**, 534-542.

Leigh, J. W., and Bryant, D. (2015) PopART: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, **6**, 1110-1116.

Li, G., Davis, B. W., Eizirik, E., and Murphy, W. J. (2016) Phylogenomic evidence for ancient hybridization in the genomes of living cats (Felidae). *Genome Resources*, **26**, 1-11.

Linnen, C. R., and Farrell, B. D. (2007) Rampant mitochondrial introgression in *Neodiprion* (Hymenoptera: Diprionidae) sawflies. *Evolution*, **61**, 1417-1438.

MacGuigan, D. J., Geneva, A. J., and Glor, R. E. (2017) A genomic assessment of species boundaries and hybridization in a group of highly polymorphic anoles (*distichus* species complex). *Ecology and Evolution*, **7**, 3657-3671.

McHugh, A. Bierzychudek, P., Greever, C., Marzulla, T., Van Buskirk, R., and Binford, G., (2013) A molecular phylogenetic analysis of *Speyeria* and its implications for the management of the threatened *S. zerene hippolyta*. *Journal of Insect Conservation*, **17**, 1237-1253.

Milá, B., McCormack, j. E., Castañeda, G., Wayne, R. K., and Smith, T. B. (2007) Recent postglacial range expansion drives the rapid diversification of a songbird lineage in the genus *Junco*. *Proceedings of the Royal Society B*, **274**, 2653-2660.

Miller, L. D., and Brown, F. M. (1981) Memoir No. 2: A catalogue/ checklist of the butterflies of America north of Mexico. Pp. 143. The Lepidopterists' Society.

Moeck, A. H. (1975) Geographic variability in *Speyeria*: comments, records, and description of a new subspecies. Los Angeles: Entomological Reprint Specialists.

Nguyen, L. -T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Molecular Biology and Evolution*, **32**:268-274.

Opler, P. A. and Warren, A. D. (2005) Lepidoptera of North America. 4. Scientific names list for butterfly species of North America, North of Mexico. Contributions of the C. P. Gillette Museum of Arthropod Diversity, Colorado State University, Fort Collins, Colorado, USA

Papakostas, S., Michaloudi, E., Proios, K., Brehm, M., Verhage, L., Rota, J., Peña, C., Stamou, G. Prithcard, V. L., Fontaneto, D., and Declerck, S. A. J. (2016) Integrative taxonomy recognizes evolutionary units despite widespread mitonuclear discordance: evidence from a Rotifer cryptic species complex. *Systematic Biology*, **65**, 508-524.

Pelham, J. P. (2019) A catalogue of the butterflies of the United States and Canada; with a complete bibliography of the descriptive and systematic literature. *The Lepidoptera Research Foundation/Journal of Research on the Lepidoptera,* Volume 40, Beverly Hills, CA, USA.

Pickrell, J. K., and Pritchard, J. K. (2012) Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, **8**, e1002967-17.

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945-959.

Pritchard, J. K., and Wen, X. (2004) *Documentation for structure software: Version 2*. University of Chicago Press, Chicago.

R Core Team. (2017) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rambaut, A. Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018) Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, **67**, 901-904.

Razkin, O., Gómez-Moliner, B. J., Vardinoyannis, K., Martínez-Ortí, A., and Madeira, M. J. (2016) Species delimitation for cryptic species complexes: case study of *Pyramidula* (Gastropoda: Pulmonata). *Zoological Scripta*, **46**, 55-72.

Reich, D. Thangaraj, K., Patterson, N., Price, A. L., and Singh, L. (2009) Reconstructing Indian population history. *Nature*, **461**, 489-494.

Revell, L. J. (2012) phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, **3**, 217-223.

Riva, F., Campbell, E. O., Carroll, F., and Acorn, J. H. (in press) Identification "by eye": integrative character assessment informs regional field identification of greater fritillary butterflies (Nymphalidae: *Speyeria*). *Journal of Insect Conservation*.

Rochette, N. C., Rivera-Colón, A. G., and Catchen, J. M. (2019) Stacks 2: analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, DOI:10.1111/MEC.15253

Rosenberg, N. A. (2002) The probability of topological concordance of gene trees and species trees. *Theoretical Population Biology*, **61**, 225-247.

Sattler, R. L., Willoughby, J. R., and Swanson, B. J. (2017) Decline of heterozygosity in a large but isolated population: a 45-year examination of mosse gnetic diversity on Isle Royale. *PeerJ*, **5**, e3584.

Scott, J. A., Kondla, N. G., and Spomer, S. M. (1998) *Speyeria hesperis* and *Speyeria atlantis* are distinct species. *Papilio New Series*, **8**, 1-32.

Sims, S. R. (2017) *Speyeria* (Lepidoptera: Nymphalidae) conservation. *Insects*, **8**, 45.

Sites, J. W., Jr., and Marshall, J. C. (2003) Delimiting species: a Renaissance issue in systematic biology. *Trends in Ecology and Evolution*, **18**, 462-470.

Sonsthagen, S. A., Wilson, R. E., and underwood, J. G. (2017) Genetic implications of bottleneck effects of differing severities of genetic diversity in naturally recovering populations: An example from Hawaiian coot and Hawaiian gallinule. *Ecology and Evolution*, **7**, 9925-9934.

Sperling, F. A. H. (2003) Butterfly molecular systematics: from species definitions to higher level phylogenies. Pp. 431-458. IN Boggs, C., Erlich, P. and Watt, W. (Eds). *Ecology and Evolution Taking Flight: Butterflies as model study systems*. University of Chicago Press, Chicago.

Stanton, D. W. G., Frandsen, P., Waples, R. K., Russo, I-R. M., Orozco-terWengel, P. A., Tingskov Pedersen, C-E., Siegismund H. R., and Bruford, M. W. (2019) More grist for the mill? Species delimitation in the genomic era and its implications for conservation. *Conservation Genetics*, **20**, 101-113.

Sukumaran, J., and Knowles, L. L. (2017) Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences of the United States of America*, **114**, 1607-1612.

Thompson, E., Baumsteiger, J., and Hill, R. I. (2019) Phylogenomic analyses clarify true species within the butterfly genus *Speyeria* despite evidence of a recent adaptive radiation. *Insects*, **10**, 209-224.

Vargas, O. M., Ortiz, E. M., Simpson, B. B. (2017) Conflicting phylogenomic signals reveal a pattern of reticulate evolution in a recent high-Andean diversification (Asteraceae: Astereae: *Diplostephium*). *New Phytologist*, **214**, 1736-1750.

Wagner, C. E., Keller, I., Wittwer, S., Selz, O. M., Mwaiko, S., Greuter, L., Sivasundar, A., and Seehausen, O. (2013) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*, **22**, 787-798.

Warren, A. D., Davis, K. J., Grishin, N. V., Pelham, J. P., and Stangeland, E. M. (2012) Interactive listing of American butterflies <www.butterfliesofamerica.com>

Weigand, H., Weiss, M., Cai, H., Li, Y., Yu, L., Zhang, C., and Leese, F. (2017) Deciphering the origin of mito-nuclear discordance in two sibling caddisfly species. *Molecular Ecology*, **26**, 5705-5715.

Werren, J. H., Baldo, L., and Clark, M. E. (2008) *Wolbachia*: master manipulators of invertebrate biology. *Nature Reviews Microbiology*, **6**, 741-751.

Yang, Z., and Rannala, B. (2017) Bayesian species identification under the multispecies coalescent provides significant improvements to DNA barcoding analyses. *Molecular Ecology*, **26**, 3028-3036.

Chapter 6

General Conclusions

6.1 Thesis Overview

Systematic research is currently experiencing a resurgence facilitated in large part by the development of methods for producing genomic data in non-model organisms (Sites & Marshall 2003; Andrews *et al.* 2016). These advances have yielded unprecedented insight into the processes of evolution and speciation, and are likely to contribute to robust systematic assessments of taxa that have been otherwise difficult to resolve using more traditional systematic approaches (McCormack *et al.* 2013; Vargas *et al.* 2017; Hinojosa *et al.* 2019). However, the rapid proliferation of these methods has also created technical ambiguities that may have consequences for the reliability of systematic inference using these genomic techniques (Puritz *et al.* 2014; Li *et al.* 2016). In this thesis, I first clarify the relationships among several of these new methods and assess the consistency of locus recovery and genotyping in two of the most commonly used methods in molecular systematics. Using these methods, I then conduct systematic and population genomic analyses to characterize the evolution and phylogenetic relationships of *Speyeria* Scudder, 1872 butterflies, which, due to recent divergence and morphological inconsistencies, have otherwise been hard to detect (Scott *et al* 1998; Dunford 2009; de Moya *et al.* 2017; Thompson *et al.* 2019).

Periods of conceptual advance are often facilitated by technical innovations in science. The development of PCR, for instance, revolutionized the field of molecular biology (Innis *et al.* 1995), and the development of DNA barcoding (Hebert *et al.* 2003) had a similar impact on the field of systematics. However, the development of new techniques is typically followed by a period of "proof of concept" testing, which is necessary for understanding the research

applications and shortcomings of new methods, and for ensuring they are used appropriately. In systematics, for instance, it has become clear that DNA barcoding is an often-unreliable indicator of species, particularly when they are closely related (Dupuis *et al.* 2012), and that barcoding may be more useful for specimen identification than for species delimitation (Collins & Cruickshank 2012).

As an alternative, multi-locus genomic approaches have since become widely used in systematics and population genomics. While there have been several reviews that have discussed the technical differences and applications of common methods (Davey *et al.* 2011; Puritz *et al.* 2014; McCormack *et al.* 2013; Andrews *et al.* 2016) none of these has been adequately comprehensive, and so many of the differences between methods remain poorly characterized. In **Chapter 2**, we conducted a comprehensive review of the technical and etymological relationships of 36 RRS techniques, which currently are the most commonly used NGS methods in systematics and population genomics. We found that many methods are only subtly different from the "parent" techniques they were derived from, and that these differences often represented optimizations for specific taxa rather than large methodological innovations. In addition, we found that several derived techniques were methodologically convergent, meaning that there are several duplicate, yet distinctly named, techniques in the literature. We also conducted a literature review of papers that cited at least one of the five most commonly used RRS methods, and evaluated the rates of inconsistent name usage and improper attribution among them, which we defined as the alternative use of a name to refer to a specific technique. We found that these techniques were commonly misattributed in the literature, which suggests there is confusion among researchers regarding the differences between these methods. We

propose that greater restraint in naming new techniques will improve clarity and proper attribution.

While the production of RRS data for systematic research has continued to increase (McCormack *et al.* 2013), there are still many unknowns about whether this data will remain useful on medium- to long-term time scales amidst rapid sequencing advances, particularly for non-model organisms. *de novo* SNP calling is highly influenced by the sequence data input into bioinformatic pipelines, and so initial differences in this data caused by the use of different lab protocols may have impacts on the consistency of data filtering and downstream analyses. Data consistency is essential for robust analyses, long-term studies, and has implications for continued dataset augmentation that is facilitated by online data repositories. In **Chapter 3**, we thus conducted the first empirical assessment of the cross-platform compatibility of two-enzyme GBS (Poland *et al.* 2012) and ddRAD (Peterson *et al.* 2012), two of the most widely used RRS methods in molecular systematics. We generated SNP data for 22 *Speyeria* butterflies using both RRS techniques, and then assessed their compatibility post-processing. There were large initial differences in read depth and genomic coverage between methods, however bioinformatic data processing largely equalized these differences, and demonstrated the consistency of these filtered data in downstream phylogenomic and population genomic analyses. This study also represented the first use of SNPs to infer relationships among *Speyeria*: despite limited taxon sampling, SNPs improved the monophyly of species clusters and also revealed discordant nuclear gene histories that may reflect recent divergence and/or hybridization between species.

Most *Speyeria* species exhibit pronounced morphological inconsistencies, particularly at broad geographic scales, which has contributed to the approximately 120 subspecies delimitations in this genus (Dunford 2007; Pelham 2019). Yet, systematic studies of *Speyeria*

have most often been regionally focused (McHugh *et al.* 2013; Hill *et al.* 2018), or include only one or two subspecies to represent broadly distributed taxa (de Moya *et al.* 2017; Thompson *et al.* 2019). In **Chapter 4**, we present the most comprehensive sampling effort of *Speyeria* to date, and compare genomic SNPs to the mitochondrial *COI* barcoding gene to clarify the species-level phylogenetic relationships in this genus. Our results indicate extensive mito-nuclear discordance between the SNP species phylogeny and the *COI* gene phylogeny: most *a priori* morphological species were non-monophyletic on the *COI* phylogeny, however genomic SNPs recovered monophyletic species clusters in most cases, and two species, *S. cybele* (Fabricius, 1775) and *S. hesperis* (Edwards, 1864), exhibited pronounced genomic sub-structuring that had not previously been detected by other studies. The observed mito-nuclear discordance in *Speyeria* likely has multiple sources, including introgressive hybridization and incomplete lineage sorting, which should be investigated in future studies. We also detected a few hybrids, notably of *S. hesperis* with *S. zerene* (Boisduval, 1852) and *S. aphrodite* (Fabricius, 1787), indicating that occasional, natural hybridization between non-sister species occurs in *Speyeria*. Importantly, our results demonstrate the limitations of DNA barcoding for identifying *Speyeria*, and further validate the effectiveness of SNPs for phylogenetic reconstruction in this genus. In addition, the sub-clustering of *S. cybele* and *S. hesperis* emphasizes the importance of geographic sampling for species-level assessments. Given the genomic distinctiveness of the *S. cybele* and *S. hesperis* sub-clusters, we further suggest that systematic revisions of both species will be necessary to clarify their species delimitation.

Chapter 5 presents a focused molecular, population-level assessment of the species limits and phylogeography of the *S. atlantis-hesperis* species complex, which has been historically hard to detect given the morphological variability among all 26 subspecies, and the

degree of sympatry between *S. atlantis* (Edwards, 1862) and *S. hesperis* (Moeck 1975; Scott *et al.* 1998; Dunford 2009). As a follow-up to Chapter 4, we also included specimens of *S. zerene* and *S. aphrodite* to better characterize the extent of contemporary and ancient admixture with *S. hesperis*, which may be of particular significance given the urgent need to define ESUs for *S. zerene* (McHugh *et al.* 2013; Sims 2017). We used SNPs and the *COI* gene to infer genetic structure and admixture of populations of *S. hesperis* and *S. atlantis* throughout much of their range, and used the multispecies coalescent to re-assess the species delimitation of *S. hesperis,* given the pronounced genomic sub-structuring recovered in Chapter 4, and to infer the historical relationship between *S. hesperis* and *S. zerene.*

Under the genomic integrity species concept (Sperling 2003), our results upheld the current delimitation of *S. atlantis* as a species distinct from *S. hesperis,* and additionally supported the recognition of *S. hesperis* as two distinct species - one that is northeast of the Rocky Mountains in distribution, and another southwest of the Rockies. However, formal elevation of the southwestern lineage of *S. hesperis* will be contingent upon additional sampling in California and the Great Basin, which contains older, pre-Pleistocene lineages that were not sampled in this study. We also recovered substantial mito-nuclear discordance in *S. hesperis* that likely reflects a strong, recent selective sweep and subsequent mitochondrial bottleneck event that is not reflected in nuclear SNPs. Multispecies coalescent analyses indicated likely ancient introgression between *S. zerene* and *S. hesperis*, which was corroborated by regional mitochondrial introgression between these species. Finally, our results indicated distinct geographic structuring in the southwestern populations of *S. hesperis,* but not in *S. atlantis* or the northeastern *S. hesperis* lineage. This suggests a recent and rapid radiation northward, likely amidst the Pleistocene glacial retreat.

These results bring much-needed clarity to the evolution of the *S. atlantis-hesperis* species complex, and demonstrate the power of genomic data for characterizing evolutionary and demographic events that can lead to data discordance and unstable taxonomies.

6.2 Future Research

This thesis assesses the utility and consistency of RRS for clarifying long-standing systematic problems in *Speyeria*. However, there are many areas of this work that will benefit from additional follow-up studies, some of which are underway. From a methodological perspective, the use of NGS methods in molecular ecology has recently begun to shift toward whole-genome sequencing (WGS), which is becoming more feasible as sequencing costs continue to decrease. The integration of WGS to systematics will increase our ability to detect and characterize early lineage divergence and introgression, and is tractable even with dried, museum specimens (Condamine *et al.* 2018). For instance, WGS can be used to make inferences about genomic architecture, or structural genomic differences, between species. Recent work has both theorized and shown that changes in genomic architecture are often facilitators of early speciation, and interspecific genomic comparisons can help identify so-called "speciation genes", or genomic regions that appear to be disproportionately affected by both divergent selection and hybridization (Mallet 2005; Feder *et al.* 2012; Kronforst *et al.* 2013; Rogers *et al.* 2015; Campbell *et al.* 2018). This stands to greatly improve our current understanding of speciation with gene flow, as well as the taxonomic recognition of species, subspecies, and ESUs for conservation. WGS data should also be compatible with existing RRS data, which would facilitate long-term NGS studies, even as NGS methods continue to advance. The assembly and

annotation of a reference genome for *Speyeria* will greatly facilitate these types of studies in the future.

As well, Chapters 4 and 5 indicate that additional taxon sampling will further improve our understanding of *Speyeria* evolution. Greater sampling of *S. hesperis*, particularly in California and the Great Basin region, will be necessary to validate our recognition of two distinct species within *S. hesperis*, and increased sampling of *S. zerene* in regions that are sympatric with *S. hesperis* will clarify the extent to which introgression between these species has occurred in the past. With the help of collaborators, I have also been accumulating range-wide sampling of *S. cybele* to better clarify the genomic sub-clustering recovered in Chapter 4, and will soon begin analyses that will re-assess the species delimitation of *S. cybele* and its historical interactions with other *Speyeria.*

While this thesis has been focused on assessing genomic signals of speciation and introgression, morphological analyses should also be integrated into future work on *Speyeria*. I recently co-authored a study that used genetic clusters to assess the consistency of several commonly used field markers for the regional identification of three sympatric *Speyeria* species (Riva *et al.* in press). Such approaches could be applied on an even greater taxonomic scale to identify which wing features are the most reliable for species identification, or used to "train" machine learning software to search for morphological characters that correlate to genetic clusters (Wäldchen & Mäder 2018). Geometric morphometrics (Zelditch *et al.* 2004) and the quantification of wing colour patterning (Van Belleghem *et al.* 2018) will bring additional clarity to what we, perhaps superficially, perceive to be morphological ambiguity or variability. Morphological re-assessments of *Speyeria* will not only improve field identifications, which is important for conservation and biodiversity assessments, but will also link genomic patterns of

speciation to morphological variation, and contribute to a holistic understanding of how biological, ecological, and environmental factors interact to produce wing pattern variation in *Speyeria.*

In summary, I hope that this work provides some affirmation to those lepidopterists who have found themselves consistently challenged and befuddled by *Speyeria*. Genomic analysis has demonstrated that the systematic challenges in this genus are not only wing-deep, but reflective of the complexities that underlie speciation and evolution. These butterflies contain enough mystery to warrant several additional theses, and with continued work that incorporates new methodologies and analytical tools, we may yet resolve many of them.

6.3 Literature Cited

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., and Hohenlohe, P. A. (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, **17**, 81-92.

Campbell, C. R., Poelstra, J. W., and Yoder, A. D. (2018) What is speciation genomics? The roles of ecology, gene flow, and genomic architecture in the formation of species. *Biological Journal of the Linnean Society*, **124**, 561-583.

Collins, R. A., and Cruickshank, R. H. (2013) The seven deadly since of DNA barcoding. *Molecular Ecology Resources,* **13**, 969-975.

Condamine, F. L., Nabholz, B., Clamens, A-L., Dupuis, J. R., and Sperling, F. A. H. (2018) Mitochondrial phylogenomics, the origin of swallowtail butterflies, and the impact of the number of clocks in Bayesian molecular dating. *Systematic Entomology*, **43**, 460-480.

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., and Blaxter M. L. (2011) Genome-wide marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, **12**, 499-510.

de Moya, R. S., Savage, W. K., Tenney, C., Bao, X., Wahlberg, N., and Hill, R. I. (2017) Interrelationships and diversification of *Argynnis* Fabricius and *Speyeria* Scudder butterflies. *Systematic Entomology*, **42**, 635-649.

Dunford, J. C. (2007) *The genus Speyeria and the Speyeria atlantis/hesperis complex: species and subspecies accounts, systematics, and biogeography (Lepidoptera: Nymphalidae)*. Unpublished Dissertation, University of Florida.

Dunford, J. C. (2009) Taxonomic overview of the greater fritillary genus *Speyeria* Scudder and the atlantis-hesperis species complexes, with species accounts, type images, and relevant literature (Lepidoptera: Nymphalidae). *Insecta Mundi*, **90**, 1-74

Dupuis, J. R., Roe, A. D., and Sperling, F. A. H. (2012) Multi-locus species delimitation in closely related animals and fungi: one marker is not enough. *Molecular Ecology*, **21**, 4422-4436.

Feder, J. L., Egan, S. P., and Nosil, P. (2012) The genomics of speciation-with-gene-flow. *Trends in Genetics*, **28**, 342-350.

Hebert, P. D. N., Cywinska, A., Ball, S. L., and deWaard, J. R. (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B.* **270**, 313-321.

Hill, R. I., Ganeshan, M., Wourms., L., Kronforst, M. R., Mullen, S. P., and Savage, W. K. (2018) Effectiveness of DNA barcoding in *Speyeria* at small geographic scales. *Diversity*, **10**, 30.

Hinojosa, J. C., Koubínová, D., Szenteczki, M. A., Pitteloud, C., Dincă, V., Alvarez, N., and Vila, R. (2019) A mirage of cryptic species: genomics uncover striking mitonuclear discordance in the butterfly *Thymelicus sylvestris*. *Molecular Ecology*. DOI: https://doi.org/10.1111/mec.15153

Innis, M. A., Gelfand, D. H., and Sninsky, J. J. (Eds.) (1995) *PCR Strategies*, Academic Press, San Diego, California.

Kronforst, M. R., Hansen, M. E. B., Crawford, N. G., Gallant, J. R., Zhang, W., Kulathinal, R. J., Kapan, D. D., and Mullen, S. P. (2013) Hybridization reveals the evolving genomic architecture of speciation. *Cell Reports*, **5**, 666-677.

Li, G., Davis, B. W., Eizirik, E., and Murphy, W. J. (2016) Phylogenomic evidence for ancient hybridization in the genomes of living cats (Felidae). *Genome Resources*, **26**, 1-11.

Mallet, J. (2005) Hybridization as an invasion of the genome. *TRENDS in Ecology and Evolution*, **20**, 229-237.

McCormack, J.E., Hird, S. M., Zellmer, A. J., Carstens, B. C., and Brumfield, R. T. (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, **66**, 526-538.

McHugh, A. Greever, C., and Van Buskirk, R. (2013) A molecular phylogenetic analysis of *Speyeria* and its implications for the management of the threatened *Speyeria zerene hippolyta. Journal of Insect Conservation,* **17**, 1237-1253.

Moeck, A. H. (1975) Geographic variability in *Speyeria*: comments, records, and description of a new subspecies. Los Angeles: Entomological Reprint Specialists.

Pelham, J. P. (2019) A catalogue of the butterflies of the United States and Canada; with a complete bibliography of the descriptive and systematic literature. *The Lepidoptera Research Foundation/Journal of Research on the Lepidoptera,* Volume 40, Beverly Hills, CA, USA.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., and Hoekstra, H. E. (2012) Double digest RADseq: an inexpensive method for *de novo* discovery and genotyping in model and non-model species. *PLoS ONE*, **7**, e37135.

Poland, J. A., Brown, P. J., Sorrells, M. E., and Jannink, J. (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE*, **7**, e32253.

Puritz, J. B., Matz, M. V., Toonen, R. J., Weber, J. N., Bolnick, D. I., and Bird, C. E. (2014) Demystifying the RAD fad. *Molecular Ecology*, **23**, 5937-5942.

Riva, F., Campbell, E. O., Carroll, F., and Acorn, J. H. (in press) Identification "by eye": integrative character assessment informs regional field identification of greater fritillary butterflies (Nymphalidae: *Speyeria*). *Journal of Insect Conservation*.

Rogers, S. M., Mee, J. A., and Bowles, E. (2015) The consequences of genomic architecture on ecological speciation. *Current Zoology*, **59**, 53-71.

Scott, J. A., Kondla, N. G., and Spomer, S. M. (1998) *Speyeria hesperis* and *Speyeria atlantis* are distinct species. *Papilio New Series*, **8**, 1-32.

Sims, S. R. (2017) *Speyeria* (Lepidoptera: Nymphalidae) conservation. *Insects*, **8**, 45.

Sites, J. W., Jr., and Marshall, J. C. (2003) Delimiting species: a Renaissance issue in systematic biology. *TRENDS in Ecology and Evolution*, **18**, 462-470.

Sperling, F. A. H. (2003) Butterfly molecular systematics: from species definitions to higher level phylogenies. Pp. 431-458. IN Boggs, C., Erlich, P. and Watt, W. (Eds). *Ecology and Evolution Taking Flight: Butterflies as model study systems*. University of Chicago Press, Chicago.

Thompson, E., Baumsteiger, J., and Hill, R. I. (2019) Phylogenomic analyses clarify true species within the butterfly genus *Speyeria* despite evidence of a recent adaptive radiation. *Insects*, **10**, 209-224.

Van Belleghem, S. M., Papa, R., Ortiz-Zuazaga, J., Hendrickx, F., Jiggins, C. D., McMillan, W. O., and Counterman, B. A. (2018) Patternize: an R package for quantifying colour pattern variation. *Methods in Ecology and Evolution*, **9**, 390-398.

Vargas, O. M., Ortiz, E. M., Simpson, B. B. (2017) Conflicting phylogenomic signals reveal a pattern of reticulate evolution in a recent high-Andean diversification (Asteraceae: Astereae: *Diplostephium*). *New Phytologist*, **214**, 1736-1750.

Wäldchen, J. and Mäder, P. (2018) Machine learning for image based species identification. *Methods in Ecology and Evolution*, **9**, 2216-2225.

Will, K., Mishler, B. D., and Wheeler, Q. D. (2005) The perils of DNA barcoding and the need for integrative taxonomy. *Systematic Biology*, **54**, 844-851.

Zelditch, M. L., Swiderski, D. L., Sheets, H. D., and Fink, W. L. (2004) *Geometric Morphometrics for Biologists: A Primer*, Elsevier/Academic Press, Amsterdam.

Biography

I was born on August 1, 1984, in Grande Prairie, Alberta, to my parents Barb and Dave Campbell. I was the first of three children, and much of my childhood was spent with my two younger brothers, either fishing for scuds and snails off the dock at the reservoir across the street from our house, or building forts in the adjacent woods. As a young child, I was fascinated by "bugs" and often toted around pasta jars or margarine containers full of spiders and caterpillars, that I was firmly told were "not allowed in the house". They sometimes found their way inside anyway, and I remember chasing down escapees more than once, so that I would not be found out.

As I got older I became an avid reader, and this would fuel my decision to move to Edmonton and pursue a degree in English literature at the University of Alberta, where I became particularly interested in modern and post-modern poetry. After graduation, I worked for a few years in marketing at a local paper but I eventually decided to search for something more fulfilling, and that led me back to my early love of biology and entomology. I enrolled again at the University of Alberta for an after-degree in biology, and things began to click. I got a summer job at Nature Alberta, where I ran the bird conservation program, and eventually began to work as a summer student in Felix Sperling's lab doing DNA extractions on spruce budworm. This grew into a student research project on *Speyeria*…which then grew into a Master's project…and then into this Ph.D thesis. I have since been converted to a true *Speyeria*-phile, and look forward to unravelling more of their mysteries.

Bibliography

Acquadro, A., Barchi, L., Portis, E., Carrasquilla-Garcia, N., Tirone, M., Lanteri, S., & Comino, C. (2016) RAD2seq: an efficient protocol for plant genotyping by sequencing. Acta Horticulturae, **1147**, 1-8.

Ahmed, M. Z., Araujo-Jnr, E. V., Welch, J. J. and Kawahara, A. Y. (2015) *Wolbachia* in butterflies and moths: geographic structure in infection frequency. *Frontiers in Zoology*, **12**, 1-9.

Ali, O. A., O'Rourke, S. M., Amish, S. J., Meek, M. H., Luikart, G., Jeffres, C., & Miller, M. R. (2016) RAD Capture (Rapture): Flexible and Efficient sequence-based genotyping. Genetics, **202**, 389-400.

Altshuler, D., Pollara, V. J., Cowles, C. R., Van Etten, W. J., Baldwin, J., Linton, L., and Lander, E. S. (2000) A SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature*, **407**, 513-516.

Andolfatto, P., Davison, D., Erezyilmaz, D., Hu, T. T., Mast, J., Sunayama-Morita, T., & Stern, D. L. (2011) Multiplexed shotgun genotyping for rapid and efficient genetic mapping. Genome Research, **21**, 610-617.

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., and Hohenlohe, P. A. (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, **17**, 81-92.

Andrews, K. R., and Luikart, G. (2014) Recent novel approaches for population genomics data analysis. *Molecular Ecology*, **23**, 1661-1667.

Armstrong, K. F., and Ball, S. L. (2005) DNA barcodes for biosecurity: invasive species identification. *Philosophical Transactions of the Royal Society B*, **10**, **S10.**

Arnold, B., Corbett-Detig, R. B., and Bomblies, K. (2013) RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology*, **22**, 3179-3190.

Avise, J. C., Lansman, R. A., & Shade, R. O. (1979) The use of restriction endonucleases to measure mitochondrial DNA sequence relatedness in natural populations. I. Population structure and evolution in the genus *Peromyscus*. *Genetics*, **92**, 279-295.

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

Bandelt, H., Forster, P., and Röhl, A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37-48.

Bayerl, H., Kraus, R.H.S., Nowak, C., Foerster, D.W., Fickel, J., & Kuehn, R. (2017) Fast and cost-effective single nucleotide polymorphism (SNP) detection in the absence of a reference genome using semideep next-generation Random Amplicon Sequencing (RAMseq). Molecular Ecology Resources, **18**, 107-117.

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Wheeler, D. L. (2007) Genbank. *Nucleic Acids Research*, **35**, D21-D25.

Bergsten, J. (2005) A review of long-branch attraction. *Cladistics*, **21**, 163-193.

Bird, C. D., Hilchie, G. J., Kondla, N. G., Pike, E. M., and Sperling, F. A. H. (1995) *Alberta Butterflies*, The Provincial Museum of Alberta, Edmonton, Alberta, Canada.

Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., Ingram, K. K., and Das, I. (2007) Cryptic species as a window on diversity and conservation. *TRENDS in Ecology and Evolution*, **22**, 148-155.

Boisduval, J. A. (1852) Lépidoptères de la Californie. *Annales de la Société entomologique de France,* **10**, 275-324.

Bouckaert, R. (2010) DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics*, **26**, 1372-1373.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C., Xie, D., Suchard, M. A., Rambaut, A., and Drummond, A. J. (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, **10**, e1003537.

Bradley, R. D., and Baker, R. J. (2001) A test of the genetic species concept: cytochrome-*b* sequences and mammals. *Journal of mammology*, **82**, 960-973.

Breed, G. A., Stichter, S., and Crone, E. E. (2012) Climate-driven changes in northeastern US butterfly communities. *Nature Climate Change*, **3**, 142-145.

Brittnacher, J. G., Sims, S. R., and Ayala, F. J. (1978) Genetic Differentiation Between Species of the Genus Speyeria (Lepidoptera: Nymphalidae). *Evolution.* 32: 199-210.

Brock, J. P. and Kaufman, K. (2003) *Kaufman Field Guide to Butterflies of North America*. ftrHoughton Mifflin Co., New York.

Brower, A. V. Z. (2006) Problems with DNA barcodes for species delimitation: 'ten species' of *Astraptes fulgerator* reassessed (Lepidoptera: Hesperiidae). *Systematics and Biodiversity*, **4**, 127-132.

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., and Choudhury, A. R. (2012) Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, **29**, 1917-1932.

Buerkle, C. A. & Gompert, Z. (2013) Population genomics based on low coverage sequencing: how low should we go? *Molecular Ecology*, **22**, 3028-3035.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**:421. DOI: 10.1186/1471-2105-10-421.

Campbell, E. O., Brunet, B. T. M., Dupuis, J. R., and Sperling, F. A. H. (2018) Would an RRS by any other name sound as RAD? *Methods in Ecology and Evolution*, **9**, 1920-1927.

Campbell, E. O., Davis, C. S., Dupuis, J. R., Muirhead, K., and Sperling, F. A. H. (2017) Cross-platform compatibility of *de novo*-aligned SNPs in a nonmodel butterfly genus. *Molecular Ecology Resources*, **6**, e84-e93.

Campbell, E. O., Gage, E., V., Gage, R. V., and Sperling, F. A. H. (2019) SNP-based species phylogeny of greater fritillary butterflies (Lepidoptera: Nymphalidae: *Speyeria*) demonstrates widespread mito-nuclear discordance. *Systematic Entomology*. DOI:10.1111/syent.12393

Campbell, C. R., Poelstra, J. W., and Yoder, A. D. (2018) What is speciation genomics? The roles of ecology, gene flow, and genomic architecture in the formation of species. *Biological Journal of the Linnean Society*, **124**, 561-583.

Carstens, B. C., Pelletier, T. A., Reid, N. M., and Satler, J. D. (2013) How to fail at species delimitation. *Molecular Ecology*, **22**, 4369-4383.

Catchen, J. M., Hohenlohe, P. A., Bernatchez, L., Funk, W. C., Andrews, K. R., & Allendorf, F. W. (2017) Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. Molecular Ecology Resources, **17**, 362-365.

Caterino, M. S., Cho, S., and Sperling, F. A. H. (2000) The current state of insect molecular systematics: a thriving tower of Babel. *Annual Review of Entomology*, **45**, 1.54.

Chen, Q., Ma, Y., Yang, Y., Chen, Z., Liao, R., Xie, X., … Pan, Y. (2013) Genotyping by Genome Reducing and Sequencing for Outbred Animals. PLoS One **8**, e67500.

Coates, D. J., Byrne, M., and Moritz, C. (2018) Genetic diversity and conservation units: dealing with the species-population continuum in the age of genomics. *Frontiers in Ecology and Evolution*, **6**, 165.

Collins, R. A., and Cruickshank, R. H. (2013) The seven deadly since of DNA barcoding. *Molecular Ecology Resources,* **13**, 969-975.

Comstock, J. A. (1925) Eight new races, forms or aberrations of California butterflies. *Bulletin of the Sothern California Academy of Sciences*, **24**, 62-68.

Condamine, F. L., Nabholz, B., Clamens, A-L., Dupuis, J. R., and Sperling, F. A. H. (2018) Mitochondrial phylogenomics, the origin of swallowtail butterflies, and the impact of the number of clocks in Bayesian molecular dating. *Systematic Entomology*, **43**, 460-480.

Cornuet, J. M., Pudlo, P., Veyssier, J., Dehne-Garcia, A. Gautier, M., Leblois, R., Marin, J. M., and Estoup, A. (2014) DIYABC v2.0: a software to make approximate Bayesian computataion inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics*, **30**, 1187-1189.

DaCosta, J. M. & Sorenson, M. D. (2015) ddRAD-seq phylogenetics based on nucleotide, indel, and presence-absence polymorphisms: Analyses of two avian genera with contrasting histories. *Molecular Phylogenetics and Evolution*, **94**, 122-135.

DaCosta, J. M. & Sorenson, M.D. (2014) Amplification biases and consistent recovery of loci in a double-digest RAD-seq protocol. *PLoS One*, **9**, e106713.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E, DePristo, M. A., … 1000 Genomes Analysis Group (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156-2158.

Dasmahapatra, K. K. Elias, M., Hill, R. I., Hoffman, J I., and Mallet, J. (2010) Mitochondrial DNA barcoding detects some species that are real, and some that are not. *Molecular Ecology Resources*, **10**, 264-273.

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., and Blaxter M. L. (2011) Genome-wide marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, **12**, 499-510.

Degnan, J. H., and Rosenberg, N. A. (2006) Discordance of species trees with their most likely gene trees. *PLoS Genetics*, **2**, e68.

Degnan, J. H., and Rosenberg, N. A. (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *TRENDS in Ecology and Evolution*, **24**, 332-340.

de Moya, R. S., Savage, W. K., Tenney, C., Bao, X., Wahlberg, N., and Hill, R. I. (2017) Interrelationships and diversification of *Argynnis* Fabricius and *Speyeria* Scudder butterflies. *Systematic Entomology*, **42**, 635-649.

de Queiroz, K. (1998) The general lineage concept of species, species criteria, and the process of speciation. Pp. 57-75. IN Howard, D. J., Berlocher, S. H. (Eds.) *Endless forms: species and speciation*. Oxford University Press, New York.

de Queiroz, K. (2005) Ernst Mayr and the modern concept of species. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 6660-6607.

de Queiroz, K. (2007) Species Concepts and Species Delimitation. *Systematic Biology*, **56**, 879-886.

Dent, E. A. & vonHoldt, B. M. (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **2**, 359-361.

Devitt, T. J., Baird, S. J. E., and Moritz, C. (2011) Asymmetric reproductive isolation between terminal forms of the salamander ring species *Ensatina eschscholtzii* revealed by fine-scale genetic analysis of a hybrid zone. *MBC Evolutionary Biology*, **11**, 245.

Diaz-Arce, M., and Rodriguez-Ezpeleta, N. (2019) Selecting RAD-Seq data analysis parameters for population genetics: the more the better? *Frontiers in Genetics,* **10**, 1-10.

dos Passos, C. F. and Grey, L. P. (1945) A genitalic survey of Argynninae (Lepidoptera, Nymphalidae). *American Museum Novitates*, **1296**.

dos Passos, C. F. and Grey, L. P. (1947) Systematic catalogue of *Speyeria* (Lepidoptera, Nymphalidae) with designations of types and fixations of type localities. *American Museum Novitates*, **1370**, 1-30.

Drummond, A. J., and Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**: 214.

Dunford, J. C. (2007) *The genus Speyeria and the Speyeria atlantis/hesperis complex: species and subspecies accounts, systematics, and biogeography (Lepidoptera: Nymphalidae)*. Unpublished Dissertation, University of Florida.

Dunford, J. C. (2009) Taxonomic overview of the greater fritillary genus *Speyeria* Scudder and the atlantis-hesperis species complexes, with species accounts, type images, and relevant literature (Lepidoptera: Nymphalidae). *Insecta Mundi*, **90**, 1-74

Dupuis, J. R., Roe, A. D., and Sperling, F. A. H. (2012) Multi-locus species delimitation in closely related animals and fungi: one marker is not enough. *Molecular Ecology*, **21**, 4422-4436.

Durand, E., Jay, F., Gaggiotti, O. E., and François, O. (2009) Spatial inference of admixture proportions and secondary contact zones. *Molecular Biology and Evolution*, **14**, 2611-2620.

Eaton, D. A. R. (2014) PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics, **30**, 1844-1849.

Edwards, S. V. (2009) Is a new and general theory of molecular systematics emerging? *Evolution*, **63**, 1-19.

Edwards, W. H. (1864) Description of certain species of diurnal Lepidoptera found within the limits of the United States and British North America. *Proceedings of the Entomological Society of Philadelphia*, **2**, 501-507.

Edwards, W. H. (1869) Descriptions of new species of diurnal Lepidoptera found within the United States. *Transactions of the American Entomological Society,* **2**, 369-382.

Edwards, W. H. (1872) Descriptions of new species of diurnal Lepidoptera found within the United States. *Transactions of the American Entomological Society*, **4**, 61-70.

Edwards, W. H. (1874) Descriptions of new species of diurnal Lepidoptera found in North America. *Transactions of the American Entomological Society*, **5**, 103-111.

Edwards, W. H. (1879) Descriptions of new species of North American butterflies; also, notes upon certain species. *The Canadian Entomologist*, **11**, 81-89.

Edwards, W. H. (1886) Descriptions of new species of butterflies found in the United States. *The Canadian Entomologist*, **18**, 61-65.

Ebel, E. R., DaCosta, J. M., Sorenson, M. D., Hill, R. I., Briscoe, A. D., Willmott, K. R., and Mullen, S. P. (2015) Rapid diversification associated with ecological specialization in Neotropical *Adelpha* butterflies. *Molecular Ecology*, **24**, 2392-2405.

Elshire, R.J., Glaubitz, J. C., 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.

Elwes, H.J. (1889) Revision of *Argynnis*. Transactions of the Royal Entomological Society of London. 37: 535-575.

Escudero, M., Eaton, D. A. R., Hahn, M., Hipp, A. L. (2014) Genotyping-by-sequencing as a tool to infer phylogeny and ancestral hybridization: a case study in *Carex* (Cyperaceae). *Molecular Phylogenetics and Evolution*, **79**, 359-367.

Evanno, G., Regnaut, S., & Goudet, J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611-2620.

Excoffier, L. Foll, M., and Petit, R. J. (2009) Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 481-501.

Feder, J. L., Egan, S. P., and Nosil, P. (2012) The genomics of speciation-with-gene-flow. *Trends in Genetics*, **28**, 342-350.

Fišer, C., Robinson, C. T., and Malard, F. (2018) Cryptic species as a window into the paradigm shift of the species concept. *Molecular Ecology*, **27**, 613-635.

Flaherty, M. (2017) Resignations at 'Third World Quarterly'. Inside Higher Ed. Retrieved from https://www.insidehighered.com/news/2017/09/20/much-third-world-quarterlys-editorial-board-resigns-saying-controversial-article

Flanagan, S. P. and Jones, A. G. (2018) Substantial differences in bias between single-digest and double-digest RAD-seq libraries: a case study. *Molecular Ecology Resources*, **18**, 264-280.

Franchini, P. Parera, D. M., Kautt, A. F., & Meyer, A. (2017) quaddRAD: a new high-multiplexing and PCR duplicate removal ddRAD protocol produces novel evolutionary insights in a nonradiating cichlid lineage. Molecular Ecology, **28**, 2783-2795.

Freitas, L., Mello, F., and Schrago, C. G. (2018) Multispecies coalescent analysis confirms standing phylogenetic instability in Hexapoda. *Journal of Evolutionary Biology*, **31**, 1623-1631.

Fu, Z., Epstein, B., Kelley, J. L., Zheng, Q., Bergland, A. O., Carrillo, C. I. C., … Snyder, W. E. (2017) Using NextRAD sequencing to infer movement of herbivores among host plants. PLoS One, **12**, e0177742.

Gallien, M., & Roelofs, P. (2017) What the success of a journal article supporting colonialism says about academia in a clickbait era. Scroll.in. Retrieved from https://scroll.in/article/851215/after-the-media-the-lure-of-clickbait-has-taken-over-academia

Galtier, N. (2018) Delineating species in the speciation continuum: a proposal. *Evolutionary Applications*, **12**, 657-663.

Gast, R. J., Ledee, D. R., Fuerst, P. A., and Byers, T. J. (1996) Sybgenus systematics of *Acanthamoeba:* four nuclear 18S rDNA sequence types. *Eukaryotic Microbiology*, **43**, 498-504.

Ghiselin, M. T. (1974) A radical solution to the species problem. *Systematic Zoology*, **23**, 536-544.

Goodwin, S., McPherson J. D., & McCombi, W. R. (2016) Coming of age: ten years of next-generation sequencing technologies. Nature Reviews Genetics**, 27**, 333-351.

Graham, C. F., Glenn, T. C., McArthur, A. G., Boreham, D. R., Kieran, T., Lance, S., … Somers, C. M. (2015) Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADseq), Molecular Ecology Resources, **15**, 1304-1315.

Grey, L. P. (1951) The subspeciation of *Speyeria atlantis*. *The Lepidopterist's News*, **5**, 31-35.

Guindon, S., Dufayard, J., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2009) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, **59**, 307-321.

Guo, Y., Fang, D., Song, L., Liu, Y., Liu, Y., Wu, L., … Zhang, H. (2014) An improved 2b-RAD approach (I2B-RAD) offering genotyping tested by a rice (Oryza sativa L.) F2 population. BMC genomics, **15**, 956.

Grummer, J. A., Bryson, R. W. Jr., and Reeder, T. W. (2014) Species delimitation using Bayes factors: simulations and applications to the *Sceloporus scalarsis* species group (Squamata: Phrynosomatidae). *Systematic Biology*, **63**, 119-133.

Grummer, J. A., Morando, M. M., Avila, L. J., Sites, J. W., Jr, and Leaché, A. D. (2018) Phylogenomic evidence for a recent and rapid radiation of lizards in the Patagonian *Liolaemus fitzingerii* species group. *Molecular Phylogenetics and Evolution*, **125**, 243-254.

Guppy, C. S., and Shepard, J. H. (2001) *Butterflies of British Columbia*, UBC Press, Vancouver, British Columbia.

Hadfield, J. and Retief, J. (2018) A profusion of confusion in NGS methods naming. *Nature Methods*, **15**, 7-9.

Hamm, C. A., Handley, C. A., Pike, A., Forister, M. L., Fordyce, J. A., and Nice, C. C. (2014) *Wolbachia* infection and Lepidoptera of conservation concern. *Journal of Insect Science*, **14**:6.

Hammond, P. C. (1978) *Geographic variation and speciation in the Nymphalid butterfly genus Speyeria.* PhD Dissertation, Virginia Polytechnic Institute and State University, Blacksburg.

Hammond, P. C. (1981) The colonization of violets and *Speyeria* butterflies on the ash-pumice fields deposited by Cascadian volcanoes. *Journal of the Research on Lepidoptera*, **20**, 179-191.

Hammond, P. C. (1990) Patterns of geographic variation and evolution in polytypic butterflies. *Journal of Research on the Lepidoptera*, **29**, 54-76.

Hammond, P. C. (1995) Conservation of the biodiversity in native prairie communities in the United States. *Journal of the Kansas Entomological Society*, **68**, 1-6.

Hammond, P. C., McCorkle, D., and Bergman, W. (2013) Hybridization studies of genomic compatibility and phenotypic expression in the greater fritillary butterflies (Nymphalidae: Argynnini). *Journal of the Lepidopterists' Society*, **67**, 263-273.

Hebert, P. D. N., Cywinska, A., Ball, S. L., and deWaard, J. R. (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B.* **270**, 313-321.

Heled, J. and Drummond, A. J. (2010) Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, **27**, 570-580.

Hewitt, G. M. (2008) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247-276.

Hilario, E., Barron, L., Deng, C. H., Datson, P. M., De Silva, N., Davy, M. W., & Storey, R. D. (2015). Random tagging genotyping by sequencing (rtGBS), an unbiased approach to locate restriction enzyme sites across the target genome. PLoS One, **10**, e0143193.

Hill, R. I., Ganeshan, M., Wourms, L., Kronforst, M. R., Mullen, S. P., and Savage, W. K. (2018) Effectiveness of DNA barcoding in *Speyeria* butterflies at small geographic scales. *Diversity*, **10**, 130.

Hillis, D. M., Moritz, C., and Mabel, B. K. (Eds.) (1996) *Molecular Systematics* (2nd edition), Sinauer, Sunderland, Massachusetts.

Hinojosa, J. C., Koubínová, D., Szenteczki, M. A., Pitteloud, C., Dincă, V., Alvarez, N., and Vila, R. (2019) A mirage of cryptic species: genomics uncover striking mitonuclear discordance in the butterfly *Thymelicus sylvestris*. *Molecular Ecology*. DOI: https://doi.org/10.1111/mec.15153

Hoang, D. T., Chernomor, L., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018) UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, **35**, 518-522.

Hoffberg, S. L., Kieran, T. J., Catchen, J. M., Devault, A. Faircloth, B. C., Mauricio, R., & Glenn, T. C. (2016) RADcap: sequence capture of dual-digest RADseq libraries with identifiable duplicates and reduced missing data. Molecular Ecology Resources, **16**, 1264-1278.

Hohenlohe, P. A., Day, M. D., Amish, S. J., Miller, M. R., Kamps-Hughes, N., Boyer, M. C., Muhlfeld, C. C., Allendorf, F. W., Johnson, E. A., and Luikart, G. (2013) Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. *Molecular Ecology*, **22**, 3002-3013.

Holland, W.J.,(1898) *The Butterfly Book: A Popular Guide to a Knowledge of the Butterflies of North America*. Doubleday, Page, and Company, New York. Reprinted 1907.

Holland, W. J. (1931) Notes on some American butterflies mainly relating to classification and nomenclature. Part 3. *Annals of the Carnegie Museum,* **20**, 255-265.

Hong, H., Xu, L., Lio, J., Jones, W. D., Su, Z., Ning, B., …Shi, L. (2012) Technical reproducibility of genotyping SNP arrays used in genome-wide association studies. *PLoS One*, **7**, e44483.

Huang, H. & Knowles, L. L. (2016) Unforeseen consequences of excluding missing data from next-generation sequences: simulation study of RAD sequences. *Systematic Biology*, **65**, 357-365.

Hundsdoerfer, A., K., Lee, K. M., Kitching, I. J., and Mutanen, M. (2019) Genome-wide SNP data reveal an overestimation of species diversity in a group of hawkmoths. *Genome Biology and Evolution*, **11**, 2136-2150.

Hurst, G. D. D. and Jiggins, F. M. (2005) Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society B*, **272**, 1525-1534.

Innis, M. A., Gelfand, D. H., and Sninsky, J. J. (Eds.) (1995) *PCR Strategies*, Academic Press, San Diego, California.

Irwin, D. E., Irwin, J. H., and Price, T. D. (2001) Ring species as bridges between microevolution and speciation. *Genetica*, **112-113**, 223-243.

Ivanov, V., Lee, K. M., and Mutanen, M. (2018) Mitonuclear discordance in wolf spiders: genomic evidence for species integrity and introgression. *Molecular Ecology*, **27**, 1681-1695.

Jakobsson, M, & Rosenberg, N. A. (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801-1806.

James, D. G., and Nunnallee, D. (2011) *Life Histories of Cascadia Butterflies*, Oregon State University press, Corvallis, Oregon.

Jiang, Z., Wang, H., Michal, J. J., Zhou, X., Liu, B., Solberg Woods, L. C., & Fuchs, R. A. (2016) Genome wide sampling sequencing for SNP genotyping: methods, challenges, and future development. Int. J. Biol. Sci., **12**, 100-108.

Jombart, T. (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403-1405.

Jones, C. J., Edwards, K. J., Castaglione, S., Winfield, M. O., Sala, F., van de Wiel, C., … Karp, A. (1997) Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Molecular Breeding*, **3**, 381-390.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermiin, L. S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, **14**, 587-589.

Katoh, K., Rozewicki, J., and Yamada, K. D. (2017) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, bbx108.

Keller, L. F., and Waller, D. M. (2002) Inbreeding effects in wild populations. *Trends in Ecology and Evolution*, **17**, 230-241.

KeyGene (2016). KeyGene's SBG patent upheld by the USPTO after ex parte reexamination. [online] Available at: http://www.keygene.com/wp-content/uploads/2016/03/Press-

release-KeyGenes-SBG-patent-upheld-by-the-USPTO-after-ex-parte-reexamination.pdf [Accessed 20 Nov. 2017].

Knowles, L. L., and Carstens, B. C. (2007) Delimiting species without monophyletic gene trees. *Systematic Biology*, **56**, 887-895.

Kodandaramaiah, U., Simonsen, T. J., Bromilow, S., Wahlberg, N., and Sperling, F. A. H. (2013) Deceptive single-locus taxonomy and phylogeography: *Wolbachia*-associated divergence in mitochondrial DNA is not reflected in morphology and nuclear markers in a butterfly species. *Ecology and Evolution*, **3**, 5167-5176.

Kolaczkowski, B., and Thornton, J. W. (2009) Long-branch attraction bias and inconsistency in Bayesian phylogenetics. *PLoS ONE*, **4**, e7871

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, M. A., and Mayrose, I. (2015) Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, **15**, 1179-1191.

Kronforst, M. R., Hansen, M. E. B., Crawford, N. G., Gallant, J. R., Zhang, W., Kulathinal, R. J., Kapan, D. D., and Mullen, S. P. (2013) Hybridization reveals the evolving genomic architecture of speciation. *Cell Reports*, **5**, 666-677.

Kück, P., Mayer, C., Wägele, J. W., and Misof, B. (2012) Long branch effects distor maximum likelihood phylogenies in simulations despite selection of the correct model. *PLoS ONE*, **7,** e36593.

Lavretsky, P., Hernández-Baños, B. E., and Peters, J. L. (2014) Rapid radiation and hybridization contribute to weak differentiation and hinder phylogenetic inference in the New World Mallard complex (*Anas* spp.). *BioOne Complete*, **131**, 524-538.

Lawson, D. J., van Dorp, L, and Falush, D. (2018) A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communications*, **9**, 3258.

Layberry, R. A., Hall, P. W., and Lafontaine, J. D. (1998) *The Butterflies of Canada*. University of Toronto Press, Toronto.

Leaché, A. D., Banbury, B. L., Felsentein, J., Nieto-Montes de Oca, A., and Stamatakis, A. (2015) Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. *Systematic Biology*, **64**, 1032-1047.

Leaché, A. D. and Bouckaert, R. R. (2018) Species trees and species delimitation with SNAPP: a tutorial and worked example. *Workshop on population and speciation genomics,* Český Krumlov

Leaché, A. D., Fujita, M. K., Minin, V. N., and Bouchaert, R. R. (2014) Species delimitation using genome-wide SNP data. *Systematic Biology*, **63**, 534-542.

Leaché, A. D., Harris, R. B., Rannala, B., and Yang, Z. (2014) The influence of gene flow on specie tree estimation: a simulation study. *Systematic Biology*, **63**, 17-30.

Leaché, A. D. and McGuire, J. A. (2006) Phylogenetic relationships of horned lizards (*Phrynosoma*) based on nuclear and mitochondrial data: evidence for a misleading mitochondrial gene tree. *Molecular Phylogenetics and Evolution*, **39**, 628-644.

Leaché, A. D., and Oaks, J. R. (2017) The Utility of single nucleotide polymorphism (SNP) data in phylogenetics, *Annual Review of Ecology, Evolution, and Systematics*, **48**, 69-84.

Lee, K. M., Kivelä, S. M., Ivanov, V., Hausmann, A., Kaila, L, Wahlberg, N., and Mutanen, M. (2018) Information dropout patterns in restriction site associated DNA phylogenomic and a comparison with multilocus Sanger data in a species-rich moth genus. *Systematic Biology*, **67**, 925-939.

Leigh, J. W., and Bryant, D. (2015) PopART: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, **6**, 1110-1116.

Li, G., Davis, B. W., Eizirik, E., and Murphy, W. J. (2016) Phylogenomic evidence for ancient hybridization in the genomes of living cats (Felidae). *Genome Resources*, **26**, 1-11.

Lischer, H. E. L. & Excoffier, L. (2012) PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics,* **28**, 298-299.

Linnen, C. R., and Farrell, B. D. (2007) Rampant mitochondrial introgression in *Neodiprion* (Hymenoptera: Diprionidae) sawflies. *Evolution*, **61**, 1417-1438.

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer, A. (2016) Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. Molecular Ecology Resources, **17**, 142-152.

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer, A. (2017) Responsible RAD: Striving for best practices in population genomic studies of adaptation. Molecular Ecology Resources, **17**, 366-369.

Lynch, M. (2009) Estimation of allele frequencies from high-coverage genome-sequencing projects. *Genetics Society of America*, **182**, 295-301.

MacGuigan, D. J., Geneva, A. J., and Glor, R. E. (2017) A genomic assessment of species boundaries and hybridization in a group of highly polymorphic anoles (*distichus* species complex). *Ecology and Evolution*, **7**, 3657-3671.

Maddison, W. P. (1997) Gene trees in species trees. *Systematic Biology*, **46**, 523-536.

Maddison, W. P. and Knowles, L. L. (2006) Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology*, **55** 21-30.

Maddison, W. P., and Maddison, D. R. (2018) Mesquite: a modular system for evolutionary analysis. Version 3.51. http://www.mesquiteproject.org

Mallet, J. (1995) A species definition for the modern synthesis. *TRENDS in Ecology and Evolution*, **10**, 294-299

Mallet, J. (2005) Hybridization as an invasion of the genome. *TRENDS in Ecology and Evolution*, **20**, 229-237.

Mallet, J. (2008) Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society B*, **363**, 2971-2986.

Marek, P. E., and Bond, J. E. (2009) A Müllerian mimicry ring in Appalachian millipedes. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 9755-9760.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, **17**, 10-12.

Mascher, M., Wu, S., St. Amand, P., Stein, N., & Poland, J. (2013) Application of genotyping-by-sequencing on semiconductor sequencing platforms: a comparison of genetic and reference-based marker ordering in barley. PLoS One, **8**, e76925.

Mayden, R. L. (1997) A hierarchy of species concepts: The denouement in the saga of the species problem. Pp. 381-424. IN Claridge, M.F., Dawah, H. A., and Wilson, M.R (Eds.) *Species: the units of biodiversity*, Chapman and hall, London.

Mayr, E. (1942) *Systematics and the origin of species*, Columbia University Press, New York.

McCormack, J.E., Hird, S. M., Zellmer, A. J., Carstens, B. C., and Brumfield, R. T. (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, **66**, 526-538.

McHugh, A. Greever, C., and Van Buskirk, R. (2013) A molecular phylogenetic analysis of *Speyeria* and its implications for the management of the threatened *Speyeria zerene hippolyta. Journal of Insect Conservation,* **17**, 1237-1253.

McKinney G. J., Larson, W. A., Seeb, L. W., & Seeb, J. E. (2017) RADseq provides unprecedented insights into molecular ecology and evolutionary insights: comment on Breaking RAD by Lowry et al. (2016). Molecular Ecology Resources, **17**, 356-361.

Milá, B., McCormack, j. E., Castañeda, G., Wayne, R. K., and Smith, T. B. (2007) Recent postglacial range expansion drives the rapid diversification of a songbird lineage in the genus *Junco*. *Proceedings of the Royal Society B*, **274**, 2653-2660.

Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A., & Johnston, E. A. (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Research, **17**, 240-248.

Minh, B. Q., Nguyen, M. A. T., & von Haeseler, A. (2013) Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution,* **30**, 1188-1195.

Moeck, A. H. (1975) Geographic variability in *Speyeria*: comments, records, and description of a new subspecies. Los Angeles: Entomological Reprint Specialists.

Moritz, C. (1994) Defining 'evolutionarily significant units' for conservation. *TRENDS in Ecology and Evolution*, **9**, 372-375.

Nabhan, A. R. & Sarkar, I. N. (2011) The impact of taxon sampling on phylogenetic inference: a review of two decades of controversy. *Briefings in Bioinformatics*, **13**, 122-134.

Narum, S. R., Buerkle, C. A., Davey, J. W., Miller, M. R., & Hohenlohe, P. A. (2013) Genotyping-by-sequencing in ecological and conservation genomics. Molecular Ecology, **22**, 2841-2847.

Nguyen, L. -T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Molecular Biology and Evolution*, **32**:268-274.

Nielsen, R., Korneliussen, T., Albrechtsen, A., Li, Y., & Wang, J. (2012) SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data, *PLoS One*, **7**, e37558.

Nielsen, R., Paul, J. S., Albrechtsen, A., & Song, Y. S. (2011) Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics*, **12**, 443-451.

NUAP (no unnecessary acronyms please). (2011) *Nature Methods*, **8**, 521.

Opler, P. A., and Warren, A. D. (2005) *Lepidoptera of North America. 4. Scientific names list for butterfly species of North America, north of Mexico.* Contributions of the C. P. Gillette Museum of Arthropod Diversity, Colorado State University, Fort Collins, CO.

Ott, A., Liu, S., Schnable, J.C., Yeh, C.-T., Wang, K.-S., & Schnable, P.S. (2017) tGBS® genotyping-by-sequencing enables reliable genotyping of heterozygous loci. Nucleic Acids Research, **45**, e178.

Papakostas, S., Michaloudi, E., Proios, K., Brehm, M., Verhage, L., Rota, J., Peña, C., Stamou, G. Prithcard, V. L., Fontaneto, D., and Declerck, S. A. J. (2016) Integrative taxonomy recognizes evolutionary units despite widespread mitonuclear discordance: evidence from a Rotifer cryptic species complex. *Systematic Biology*, **65**, 508-524.

Paris, J. P., Stevens, J. R., and Catchen, J. M. (2017) Lost in parameter space: a road map for STACKS. *Methods in Ecology and Evolution*, **8**, 1360-1373.

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. *The Lepidoptera Research Foundation/Journal of Research on the Lepidoptera*, Volume 40, Beverly Hills, CA, USA.

Pelham, J. P. (2019) A catalogue of the butterflies of the United States and Canada; with a complete bibliography of the descriptive and systematic literature. *The Lepidoptera Research Foundation/Journal of Research on the Lepidoptera,* Volume 40, Beverly Hills, CA, USA.

Penner, G. A., Bush, A., Wise, R., Kim, W., Domier, L., Kasha, K., … Fedak, G. (1993) Reproducibility of Random Amplified Polymorphic DNA (RAPD) analysis among laboratories. *Genome Research*, **2**, 341-345.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., and Hoekstra, H. E. (2012) Double digest RADseq: an inexpensive method for *de novo* discovery and genotyping in model and non-model species. *PLoS ONE*, **7**, e37135.

Pickrell, J. K., and Pritchard, J. K. (2012) Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, **8**, e1002967-17.

Powell, A. F. L. A., Busby, W. H., and Kindscher, K. (2006) Status of the regal fritillary (*Speyeria idalia*) and effects of fire management on its abundance in northeastern Kansas, USA. *Journal of Insect Conservation*, **11**, 299-308.

Poland, J. A., Brown, P. J., Sorrells, M. E., and Jannink, J. (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE*, **7**, e32253.

Pollard, D. A., Iyer, V. N., Moses, A. M., & Eisen, M. B. (2006) Widespread discordance of gene trees with species trees in *Drosophila*: evidence for incomplete lineage sorting. *PLoS Genetics*, **2** e173.

Pons, J., Barraclough, T. G., Gomez-Zurita, J., Cardoso, A., Duran, D. P., Hazell, S., Kamoun, S., Sumlin, W. D., and Vogler, A. P. (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, **55**, 595-609.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics,* **155,** 945-959.

Pritchard, J. K., and Wen, X. (2004) *Documentation for structure software: Version 2*. University of Chicago Press, Chicago.

Puechmaille, S. J. (2016) The program STRUCTURE does not reliably recover the correct population structure when sampling is uneven: sub-sampling and new estimator alleviate the problem. *Molecular Ecology Resources*, **16**, 608-627.

Puritz, J. B., Matz, M. V., Toonen, R. J., Weber, J. N., Bolnick, D. I., and Bird, C. E. (2014) Demystifying the RAD fad. *Molecular Ecology*, **23**, 5937-5942.

R Core Team (2017) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org

Rambaut, A. Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018) Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, **67**, 901-904.

Ratnasingham, S., and Hebert, P. D. (2007) BOLD: the barcode of life data system. *Molecular Ecology Notes*, **7**, 355-364.

Razkin, O., Gómez-Moliner, B. J., Vardinoyannis, K., Martínez-Ortí, A., and Madeira, M. J. (2016) Species delimitation for cryptic species complexes: case study of *Pyramidula* (Gastropoda: Pulmonata). *Zoological Scripta*, **46**, 55-72.

Recknagel, H. Jacobs, A., Herzyk., P., & Elmer, K. R. (2015) Double-digest RAD sequencing using ion proton semiconductor platform (ddRADseq-ion) with nonmodel organisms. Molecular Ecology Resources, **15**, 1316-1329.

Reich, D. Thangaraj, K., Patterson, N., Price, A. L., and Singh, L. (2009) Reconstructing Indian population history. *Nature*, **461**, 489-494.

Revell, L. J. (2012) phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, **3**, 217-223.

Rife, T. W., Wu, S., Bowden, R. L., & Poland, J. A. (2015) Spiked GBS: a unified, open platform for single marker genotyping and whole-genome profiling. BMC Genomics, **16**, 248.

Riva, F., Campbell, E. O., Carroll, F., and Acorn, J. H. (in press) Identification "by eye": integrative character assessment informs regional field identification of greater fritillary butterflies (Nymphalidae: *Speyeria*). *Journal of Insect Conservation*.

Rochette, N. C., and Catchen J. M. (2017) Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols*, **12**, 2640-2659.

Rogers, S. M., Mee, J. A., and Bowles, E. (2015) The consequences of genomic architecture on ecological speciation. *Current Zoology*, **59**, 53-71.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., and Huelsenbeck, J. P. (2012) MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**, 539-542.

Rosenberg, N. A. (2002) The probability of topological concordance of gene trees and species trees. *Theoretical Population Biology*, **61**, 225-247.

Rosenberg, N. A. (2003) The shapes of neutral gene genealogies in two species: probabilities of monophyly, paraphyly, and polyphyly in a coalescent model. *Evolution*, **57**, 1465-1477.

Roure, B., Baurain, D., & Philippe, H. (2012) Impact of missing data on phylogenies inferred from empirical phylogenomic data sets. *Molecular Biology and Evolution*, **30**, 197-214.

Ruse, M. (1969) Definitions of species in biology. *British Journal for the Philosophy of Science*, **20**, 97-119.

Ryder, O. A. (1986) Species conservation and systematics: the dilemma of subspecies. *TRENDS in Ecology and Evolution*, **1**, 9-10.

Sanford, M. P. (2011) Improving conservation and management of the imperiled Carson valley silverspot butterfly *Speyeria nokomis carsonensis* (Lepidoptera: Nymphalidae) based on rapid assessments of distribution, habitat, and threats. *Journal of Insect Conservation*, **15**, 715-725.

Sang, T., & Zhong, Y. (2000) Testing hybridization hypotheses based on incongruent gene trees. *Systematic Biology*, **49**, 422-434.

Sattler, R. L., Willoughby, J. R., and Swanson, B. J. (2017) Decline of heterozygosity in a large but isolated population: a 45-year examination of mosse gnetic diversity on Isle Royale. *PeerJ*, **5**, e3584.

Schield, D.R., Walsh, M.R., Card, D.C., Andrew, A.L., Adams, R.H., Castoe, T.A., & Bunce, M. (2016) EpiRADseq: scalable analysis of genomewide patterns of methylation using next-generation sequencing. Methods in Ecology and Evolution, **7**, 60-69.

Schmid, S., Genevest, R., Gobet, E., Suchan, T., Sperisen, C., Tinner, W., & Alvarez, N. (2017) hyRAD-X, a versatile method combining exome capture and RAD sequencing to extract genomic information from ancient DNA. Methods in Ecology and Evolution, **8**, 1374-1388.

Scott, J. A., Kondla, N. G., and Spomer, S. M. (1998) *Speyeria hesperis* and *Speyeria atlantis* are distinct species. *Papilio,* **8**, 1-31.

Shakespeare, W. (1591-94) "Romeo and Juliet" The complete works of William Shakespeare comprising his plays and poems, The World Syndicate Publishing Company, 1930, pp. 901.

Shakespeare, W. (1605-08) "Coriolanus" The complete works of William Shakespeare comprising his plays and poems, The World Syndicate Publishing Company, 1930, pp. 703.

Shimodaira, H. & Hasegawa, M. (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution*, **16**, 1114-1116.

Scott, J. A., Kondla, N. G., and Spomer, S. M. 1998. *Speyeria hesperis* and *Speyeria atlantis* are distinct species. *Papilio New Series*, **8**, 1-32.

Shepherd, S., and Debinski, D. M. (2005) Reintroduction of regal fritillary (*Speyeria idalia*) to a restored prairie. *Ecological Restoration*, **23**, 244 - 250.

Simonsen, T. J. (2006a) Fritillary phylogeny, classification, and larval host plants: reconstructed mainly on the basis of male and female genitalic morphology (Lepidoptera: Nymphalidae: Argynnini). *Biological Journal of the Linnaean Society*, **89**, 627-673.

Simonsen, T. J. (2006b) Glands, muscles and genitalia. Morphological and phylogenetic implications of histological characters in the male genitalia of Fritillary butterflies (Lepidoptera: Nymphalidae: Argynnini). *Zoologica Scripta*, **35**, 231-241.

Simonsen, T. J., Wahlberg, N., Brower, A. V. Z., and de Jong, R. (2006) Morphology, molecules and fritillaries: approaching a satble phylogeny for Argynnini (Lepidoptera: Nymphaliade). *Insect Systematics and Evolution*, **37**, 405-418.

Simpson, G. G. (1951) The species concept. *Evolution*, **5**, 285-298.

Sims, S. R. (2017) *Speyeria* (Lepidoptera: Nymphalidae) conservation. *Insects*, **8**, 45.

Sites, J. W., Jr., and Marshall, J. C. (2003) Delimiting species: a Renaissance issue in systematic biology. *TRENDS in Ecology and Evolution*, **18**, 462-470.

Snyder, A. J. (1900) The argynnids of North America. *Occasional Memoirs of the Chicago Entomological Society*, **1**, 27-38.

Sokal, R. R., and Crovello, T. J. (1970) The biological species concept: a critical evaluation. *The American Society of Naturalists*, **104**, 127-153.

Sonah, H., Bastien, M., Iquira, E., Tardivel, A., Légaré, G., Boyle, B., … Belzile, F. (2013) An improved Genotyping by Sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. PLoS One, **8**, e54603.

Sonsthagen, S. A., Wilson, R. E., Chesser, R. T., Pons, J. M., Crochet, P. A., Driskell, A., and Dove, C. (2016) Recurrent hybridization and recent origin obscure phylogenetic

relationships within the 'white-headed' gull (*Larus* sp.) complex. *Molecular Phylogenetics and Evolution*, **103**, 41-54.

Sousa, V., and Hey, J. (2013) Understanding the origin of species with genome-scale data: modelling gene flow. *Nature Reviews Genetics*, **14**, 404-414.

Sperling, F. A. H. (2003) Butterfly molecular systematics: from species definitions to higher level phylogenies. Pp. 431-458. IN Boggs, C., Erlich, P. and Watt, W. (Eds). *Ecology and Evolution Taking Flight: Butterflies as model study systems*. University of Chicago Press, Chicago.

Sperling, J. L., Silva-Brandao, K. L., Brandao, M. M., Lloyd, V. K., Dang, S., Davis, C. S., Sperling, F. A. H., and Magor, K. E. (2017) Comparison of bacterial 16S rRNA variable regions for microbiome surveys of ticks. *Ticks and Tick-borne Diseases*, **8**, 453-461.

Stanton, D. W.G., Frandsen, P., Waples, R. K., Heller, R., Russo, I-R. M., Orozco-terWengel, P. A., Pedersen, C-E. T., Siegismund, H. R., and Bruford, M. W. (2019) More grist for the mill? Species delimitation in the genomic era and its implcations for conservation. *Conservation Genetics*, **20**, 101-113.

Stolle, E., & Moritz, R. F. A. (2013) RESTseq - Efficient benchtop population genomics with RESTriction Fragment SEQuencing. PLoS One, **8**, e63960.

Stothard, P. (2000) The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniqu*es, **28**, 1102-1104.

Suchan, T., Pitteloud, C., Gerasimova, N. S., Kostikova, A., Schmid, S., Arrigo, N., … Alvarez, N. (2016) Hybridization capture using RAD probes (hyRAD), a new tool for performing genomic analyses on collection specimens. PLoS One, **11**, e0151651.

Sukumaran, J., and Knowles, L. L. (2017) Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences of the United States of America*, **114**, 1607-1612.

Sun, X., Liu, D., Zhang, X., Li, W., Liu, H., Hong, W., … Zheng, H. (2013) SLAF-seq: An efficient method of large-scale De Novo SNP discovery and genotyping using high-throughput sequencing. PLoS One, **8**, e58700.

Suyama, Y., & Matsuki, Y. (2015) MIG-seq: an effective PCR-based method for genome-wide single nucleotide polymorphism genotyping using the next-generation sequencing platform. Scientific Reports, **5**, 16963.

Swofford, D. L. (2002) PAUP*: Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Tabangin, M. E., Woo, J. G., & Martin, L. J. (2009) The effect of minor allele frequency on the likelihood of obtaining false positives. *BMC Proceedings*, **3**, S41.

Templeton, A. R. (1989) The meaning of species and speciation: A genetic perspective. Pp. 3-27. IN Howard, D. J., and Berlocher, S. H., (Eds.). *Endless forms: species and speciation*, Oxford University Press, New York.

Thompson, E., Baumsteiger, J., and Hill, R. I. (2019) Phylogenomic analyses clarify true species within the butterfly genus *Speyeria* despite evidence of a recent adaptive radation. *Insects*, **10**, 209-224.

Toews, D. P. L. and Brelsford, A. (2012) The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, **21**, 3907-3930.

Toonen, R. J., Puritz, J. B., Forsman, Z. H., Whitney, J. L., Fernandez-Silva, I., Andrews, K. R., & Bird, C. E. (2013) ezRAD: a simplified method for genomic genotyping in non-model organisms. PeerJ, **l**, e203.

Trifinopoulos, J., Nguyen, L. T., von Haeseler, A., & Minh, B. Q. (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucl. Acids. Res.,* **44**, W232-W235.

Trucchi, E., Mazzarella, A. B., Gilfillan, G. D., Lorenzo, M. T., Schönswetter, P., & Paun, O. (2016) BsRADseq: screening DNA methylation in natural populations of non-model species. Molecular Ecology, **25**, 1697-1713.

Truong, H. T., Ramos, A. M., Yalcin, F., de Ruiter, M., van der Poel H. J. A., Huvenaars, K. H. J., … van Eijk, J. T. (2012) Sequence-based genotyping for marker discovery and co-dominant scoring in germplasm and populations. PLoS One, **7**, e37565.

Turner, J. R. G. (1976) Adaptive radiation and convergence in subdivisions of butterfly genus *Heliconius* (Lepidoptera: Nymphalidae). *Zoological Journal of the Linnean Society,* **58**, 297-308.

Van Belleghem, S. M., Papa, R., Ortiz-Zuazaga, J., Hendrickx, F., Jiggins, C. D., McMillan, W. O., and Counterman, B. A. (2018) Patternize: an R package for quantifying colour pattern variation. *Methods in Ecology and Evolution*, **9**, 390-398.

van Dijk, E. L., Jaszczyszyn, Y., and Thermes, C. (2014) Library preparation methods for next-generation sequencing: tone down the bias. *Experimental Cell Research*, **32**, 12-20.

van Gurp, T. P., Wagemaker, N. C. A. M., Wouters, B., Vergeer, P., Ouborg, J. N. J., & Verhoeven, K. J. F. (2016) epiGBS: reference-free reduced representation bisulfite sequencing. Nature Methods, **13**, 322-328.

van Orsouw, N. J., Hogers, R. C. J., Janssen, A., Yalcin, F., Snoeihers, S., Verstege, E., … van Eijk, M. J. T. (2007) Complexity reduction of polymorphic sequences (CRoPs™): a novel approach for large-scale polymorphism discovery in complex genomes. PLoS One, **2**, e1172.

Van Tassell, C. P., Smith, T. P. L., Matukumalli, L. K., Taylor, J. F., Schnabel, R. D., Lawley, C. T., … Sonstegard, T. S. (2008) SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nature Methods, **5**, 247-252.

Van Valen, L. (1976) Ecological species, multispecies, and oaks. *Taxon*, **25**, 233-239.

Vargas, O. M., Ortiz, E. M., Simpson, B. B. (2017) Conflicting phylogenomic signals reveal a pattern of reticulate evolution in a recent high-Andean diversification (Asteraceae: Astereae: *Diplostephium*). *New Phytologist*, **214**, 1736-1750.

Wagner, C. E., Keller, I., Wittwer, S., Selz, O. M., Mwaiko, S., Greuter, L., Sivasundar, A., and Seehausen, O. (2013) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*, **22**, 787-798.

Wahlberg, N. and Wheat, C. W. (2008) Genomic outposts serve the phylogenomic pioneers: designing novel nuclear markers for genomic DNA extractions of Lepidoptera. *Systematic Biology*, **57**, 231-242.

Wäldchen, J. and Mäder, P. (2018) Machine learning for image based species identification. *Methods in Ecology and Evolution*, **9**, 2216-2225.

Wang, S., Meyer, E., McKay, J. K., & Matz, M. V. (2012) 2b-RAD: a simple and flexible method for genome-wide genotyping. Nature Methods, **9**, 808-812.

Warren, A. D., Davis, K. J., Grishin, N. V., Pelham, J. P., & Stangeland, E. M. (2012) Interactive listing of American butterflies. [accessed Nov 15 2016]. http://www.butterfliesofamerica.com

Weigand, H., Weiss, M., Cai, H., Li, Y., Yu, L., Zhang, C., and Leese, F. (2017) Deciphering the origin of mito-nuclear discordance in two sibling caddisfly species. *Molecular Ecology*, **26**, 5705-5715.

Weins, J. J. (2007) Species delimitation: new approaches for discovering diversity. *Systematic Biology*, **56**, 87-878.

Wells, C. N., Marko, P. B., and Tonkyn, D. W. (2015) The phylogeographic history of the threatened Diana fritillary, *Speyeria diana* (Lepidoptera: Nymphalidae): with implications for conservation. *Conservation Genetics*, **16**, 703-716.

Wells, C. N. and Tonkyn, D. W. (2014) Range collapse in the Diana fritillary, *Speyeria diana* (Nymphalidae). *Insect Conservation and Diversity*, **7**, 365-380.

Werren, J. H., Baldo, L., and Clark, M. E. (2008) *Wolbachia*: master manipulators of invertebrate biology. *Nature Reviews Microbiology*, **6**, 741-751.

Wickham, H. (2009) ggplot2: Elegant graphics for data analysis. Springer-Verlag, New York, United States of America. https:// ggplot2.org

Will, K., Mishler, B. D., and Wheeler, Q. D. (2005) The perils of DNA barcoding and the need for integrative taxonomy. *Systematic Biology*, **54**, 844-851.

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

Williams, B. L. (2001) Patterns of morphological variation in *Speyeria idalia* (Lepidoptera: Nymphalidae) with implications for taxonomy and conservation. *Annals of the Entomological Society of America*, **94**, 239-243.

Williams, B. L. (2002) Conservation genetics, extinction, and taxonomic status: a case history of the regal fritillary. *Conservation Biology*, **16**, 148-157.

Williams, J. G. K., Kubelik, A. R., LIvak, K. J., Rafalski, J. A., & Tingey, S. V. (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*, 16, 6531-6535.

Wolfe, K. (2017) Natural non-*Viola* hostplant of a Sierran *Speyeria* (Nymphalidae, Heliconiinae) and an associated parasite (Diptera, Tachinidae). *Journal of the Lepidopterists' Society*, **71**, 129-131.

Yang, G., Chen, Y., Wang, J., Guo, C., Zhao, L., Wang, X., … Guo, Z. (2016) Development of a universal and simplified ddRAD library preparation approach for SNP discovery and genotyping in angiosperm plants. Plant Methods, **12**, 39.

Yang, Z., and Rannala, B. (2017) Bayesian species identification under the multispecies coalescent provides significant improvements to DNA barcoding analyses. *Molecular Ecology*, **26**, 3028-3036.

Zaman, K., Tenney, C., Rush, C. E., and Hill, R. I. (2015) Population ecology of a California endemic: *Speyeria adiaste clemencei*, *Journal of Insect Conservation*, **19**, 753-763.

Zelditch, M. L., Swiderski, D. L., Sheets, H. D., and Fink, W. L. (2004) *Geometric Morphometrics for Biologists: A Primer*, Elsevier/Academic Press, Amsterdam.

Appendix 2.1 Features of RRS techniques used to inform the concept map (Fig. 2.1). Citation numbers include all citations on Web of Science, not just those from journal articles, and are current to March 13, 2018.

Row	Method	Publication year	No. citations	Parent RE protocol	Multiplexing capability	Mass of input DNA/cDNA	DNA shearing?
1	RRS	2000	438	n/a	No	100 -150 ng	No
2	CRoPS	2007	106	n/a	Yes	100-500 ng	No
3	RAD markers	2007	409	n/a	No	2000 ng	Yes
4	RADseq	2008	1143	RAD markers	Yes	100-1000 ng	Yes
5	RRL	2008	342	RRS	Yes	5500 ng	Yes
6	GBS	2011	1452	RADseq	Yes	100 ng	No
7	MSG	2011	178	RADseq	Yes	10-100 ng	No
8	two-enzyme GBS	2012	449	GBS	Yes	200 ng	No
9	2b-RAD	2012	121	RADseq	Yes	100-200 ng	No
10	SBG	2012	46	CRoPs	Yes	100-500 ng	No
11	ddRADseq	2012	554	RADseq	Yes	100 ng or less	No
12	SLAF-seq	2013	145	RRL	Yes	not specified	No
13	ezRAD	2013	55	RADseq	Yes	1500 ng	No
14	RESTseq	2013	21	n/a	Yes	2000 ng	No
15	GBS for semiconductor	2013	55	two-enzyme GBS	Yes	200 ng	No
16	GGRS	2013	24	RADseq and GBS	Yes	100 ng	No
17	Selective amp. GBS	2013	149	GBS	Yes	100 ng	No
18	l2b-RAD	2014	9	2b-RAD	Yes	200 ng	No
19	Spiked GBS	2015	6	GBS for semiconductor	Yes	150 ng	No
20	ddRADseq-ion	2015	11	ddRAD	Yes	1000 ng	No
21	rtGBS	2015	2	GBS	Yes	28 ng	No
22	RAD2Seq	2015	2	ddRAD	Yes	not specified	No
23	3RAD	2015	16	ddRAD	Yes	100 ng	No

Row	Method	Publication year	No. citations	Parent RE protocol	Multiplexing capability	Mass of input DNA/cDNA	DNA shearing?
24	MIG-seq	2015	9	n/a	Yes	2-100 ng	No
25	Rapture	2016	33	RADseq	Yes	50 ng	Yes
26	epiGBS	2016	11	GBS	Yes	400 ng	No
27	hyRAD	2016	22	ddRAD	Yes	100 ng or less	No
28	MiddRAD (Protocol B)	2016	0	ddRAD	Yes	200 ng	No
29	RADcap	2016	11	3RAD	Yes	100 ng	No
30	BsRADseq	2016	10	RADseq	Yes	not specified	Yes
31	EpiRADseq	2016	12	ddRAD	Yes	300 ng	No
32	quaddRAD	2017	2	ddRAD	Yes	0.1-100 ng	No
33	NextRAD	2017	2	Similar to RADseq	Yes	<50 ng	No
34	hyRAD-X	2017	1	hyRAD/RADseq	Yes	20 ng	No
35	RAMseq	2017	0	RRL	Yes	not specified	No
36	tGBS	2017	1	GBS	Yes	120 ng	No

Row	Tagmentation via transposomes?	PCR amplification or RE digestion?	No. of REs	Methylation-sensitive REs?	Barcodes added via PCR or ligation?	Selected size range	Seq. platform
1	No	RE	1	No	n/a	564 bp/400-650 bp	Sanger-based
2	No	RE	2	Optional	PCR	n/a	Roche 454
3	No	RE	1	No	n/a	n/a	n/a
4	No	RE	1	No	ligation	300-700 bp	Illumina
5	No	RE	1	No	ligation	70-130 bp	Illumina
6	No	RE	1	Yes	ligation	170-350 bp	Illumina
7	No	RE	1	No	ligation	250-300 bp	Illumina
8	No	RE	2	No	ligation	n/a	Illumina
9	No	IIB REs	1	No	ligation	76 bp / 96 bp	SOLiD/Illumina

Row	Tagmentation via transposomes?	PCR amplification or RE digestion?	No. of REs	Methylation-sensitive REs?	Barcodes added via PCR or ligation?	Selected size range	Seq. platform
10	No	RE	2	No	ligation	n/a	Illumina
11	No	RE	2	No	ligation	300 bp	Illumina
12	No	RE	2	No	PCR	450-500 bp	Illumina
13	No	RE	2	Yes	ligation	400-500 bp	Illumina
14	No	RE	7	No	ligation	70-105 bp	Ion Torrent
15	No	RE	2	No	ligation	n/a	Ion Torrent
16	No	RE	1	No	ligation	300-400 bp	Illumina
17	No	RE	1	Yes	ligation	170-350 bp	Illumina
18	No	IIB REs	1	No	ligation	150-200 bp	Illumina
19	No	PCR / RE	2	No	PCR / ligation	200-250 bp	Ion Torrent
20	No	RE	2	No	ligation	130-320 bp	Ion Torrent
21	No	PCR / RE	1	No	ligation	374-396 bp	Illumina
22	No	RE	2	No	ligation	n/a	Illumina
23	No	RE	3	No	ligation	600 bp	Illumina
24	No	PCR	n/a	No	PCR	300-800 bp	Illumina
25	No	RE	1	No	ligation	n/a	Illumina
26	No	RE	1	Yes	ligation	150-400 bp	Illumina
27	No	RE	2	No	ligation	270 bp	Illumina
28	No	RE	2	No	ligation	600-700 bp	Illumina
29	No	RE	3	No	ligation	550 bp	Illumina
30	No	RE	1	No	ligation	not specified	Illumina
31	No	RE	2	Yes	ligation	640-790 bp	Illumina
32	No	RE	2	No	ligation	620-680 bp	Illumina

Row	Tagmentation via transposomes?	PCR amplification or RE digestion?	No. of REs	Methylation-sensitive REs?	Barcodes added via PCR or ligation?	Selected size range	Seq. platform
33	Yes	n/a	n/a	No	ligation	n/a	Illumina
34	No	RE	1	No	ligation	270 bp	Illumina
35	No	PCR	n/a	No	not specified	n/a	Roche 454
36	No	RE	2	No	ligation	200-300 bp	Ion Torrent

Appendix 2.2 Links to online discussion about the difference between various RAD or GBS based methods (last accessed Dec. 5 2017)

Discussion of RAD vs. GBS derivative methods:

http://seqanswers.com/forums/showthread.php?t=62131

Difference between RADseq and GBS:

https://www.biostars.org/p/77927/
https://www.biostars.org/p/77927/
https://www.biostars.org/p/253474/#253491

SBG and ddRAD (from Floragenex website):

http://www.floragenex.com/sbg-ddrad-seq/

Appendix 2.3 Raw list with exact spelling of author keywords pertaining to GBS, SBG and RAD, and corresponding search string used for searching titles, abstracts and keywords of journal articles citing Baird et al. 2008, Elshire et al. 2011, Poland et al. 2012, Truong et al. 2012, and Peterson et al. 2012. Citation lists extracted from Web of Science® on February 6, 2018.

Keyword	Search String Used
GBS	GBS
GBS bioinformatics	GBS
GBS marker	GBS
GBS SNPs	GBS
GBS-SNP	GBS
GBS-SNP-CROP	GBS
genotyping by sequencing (GBS)	GBS
Genotyping by sequencing (GBS)	GBS
Genotyping-by sequencing (GBS)	GBS
Genotyping-by-sequencing (GBS)	GBS
Genotyping-by-sequencing(GBS)	GBS
methylation sensitive GBS	GBS
Genotype by sequencing	Genotype by sequencing
Genotype-by-sequencing	Genotype-by-sequencing
genotyping by sequencing	genotyping by sequencing
genotyping-by-sequencing	genotyping-by-sequencing
Genotyping-by-sequencing	Genotyping-by-sequencing
2b-RAD	RAD
2b-RAD genotyping	RAD
3RAD	RAD
bsRADseq	RAD
ddRAD	RAD
ddRAD seq	RAD
ddRAD sequencing	RAD

Keyword	Search String Used
ddRADseq	RAD
ddRAD-seq	RAD
double digest RADseq	RAD
Double digest restriction site associated DNA sequencing (ddRADseq)	RAD
double-digest RAD sequencing	RAD
double-digest RADSeq	RAD
double-digested RADseq	RAD
hyRAD	RAD
mbRADseq	RAD
MiddRAD	RAD
Next-generation sequencing ddRADSeq	RAD
NextRAD	RAD
phylogeny by RAD-seq	RAD
PredRAD	RAD
PyRAD	RAD
quaddRAD	RAD
RAD	RAD
RAD genotyping	RAD
RAD loci	RAD
RAD markers	RAD
RAD paired-end sequencing	RAD
RAD sequencing	RAD
RAD tag	RAD
RAD tag sequencing	RAD
RAD tags	RAD

Keyword	Search String Used
RADseg	RAD
RADseq	RAD
RAD-seq	RAD
RAD-seq-based SNP	RAD
RAD-sequencing	RAD
RAD-tag	RAD
RADtag sequencing	RAD
RAD-tag sequencing	RAD
RAD-Tag SNP	RAD
RAD-tag-based GBS	RAD
RADtags	RAD
RAD-tags	RAD
Restriction site associated DNA (RAD)	RAD
Restriction site associated DNA (RAD) markers	RAD
Restriction site associated DNA sequencing (RADseq)	RAD
Restriction site associated DNA sequencing (RAD-Seq)	RAD
Restriction site-associated DNA (RAD)	RAD
Restriction site-associated DNA (RAD) sequencing	RAD
restriction site-associated DNA (RAD) tags	RAD
restriction site-associated DNA sequencing (RADseq)	RAD
restriction site-associated DNA sequencing (RAD-seq)	RAD
restriction-site associated DNA (RAD)	RAD
Restriction-site associated DNA (RAD) sequencing	RAD
Restriction-site associated DNA (RAD) tags	RAD
restriction-site associated DNA (RADseq)	RAD

Keyword	Search String Used
Restriction-site associated DNA sequencing (RADseq)	RAD
Restriction-site associated DNA sequencing (RAD-Seq)	RAD
Restriction associated DNA genotyping by sequencing (RAD-GBS)	RAD/GBS
double digest restriction associated DNA sequencing	restriction associated
Restriction associated sequencing	Restriction associated
Restriction site associated DNA	Restriction site associated
restriction site associated DNA polymorphism	Restriction site associated
Restriction site associated DNA sequencing	Restriction site associated
Double digest restriction site associated DNA sequencing	restriction site-associated
Double digest restriction site-associated DNA sequencing	restriction site-associated
Restriction site-associated DNA	restriction site-associated
restriction site-associated DNA sequencing	restriction site-associated
restriction site-associated DNA tag	restriction site-associated
restriction site-associated sequencing	restriction site-associated
restriction-associated DNA	restriction-associated
restriction-associated DNA sequencing	restriction-associated
restriction-associated DNA tags	restriction-associated
restriction-site associate DNA	restriction-site associate
Restriction-site associated DNA	restriction-site associate
Restriction-site associated DNA sequencing	restriction-site associate
restriction-site associated DNA tags	restriction-site associate
Restriction-site-associated DNA	restriction-site-associated
Restriction-site-associated DNA sequencing	restriction-site-associated
SBG	SBG
sequence-based genotyping	sequence-based genotyping

Appendix 2.4 Methods used in literature review

Step	Description
1	Full citation lists for Baird *et al.* (2008), Elshire *et al.* (2011), Poland *et al.* (2012), Truong *et al.* (2012), and Peterson *et al.* (2012) were exported from Web of Science on February 6, 2018.[1]
2	Citation lists were copied into separate excel spreadsheets and merged where multiple files were necessary during export - Web of Science has an export limit of 500 articles.
3	A column was created to indicate citation source in each list (*eg.* Baird, Elshire, Poland, Peterson, Truong)
4	Unnecessary columns were deleted from the citation lists (*ie.* only publication type, author, title, author keywords, abstract, journal, and publication year were retained)
5	All five lists were merged into single new sheet.
6	Entries that were not journal articles were removed from the merged list
7	The publication type column was deleted
8	The merged list was copied into a new sheet and duplicate entries were removed ignoring the citation source column
9	A new column was created to run a series of countif commands to count the number of occurrences of each entry among the individual citation lists from step 2.
10	The entire list of author keywords from step 2 was copied into a new sheet
11	Keywords were parsed into separate columns using text-to-columns delimited by semi-colons
12	Keywords were aligned into a single column and sorted
13	Sorted keywords were mined for terms pertaining to GBS/SBG/RAD methods and these were copied to a separate column
14	A new column was created to define a search string to be used for automated discovery of each copied keyword from step 13 using excel formulas (see Supplementary Table 5)
15	The search string column was copied to a new column and duplicates were removed
16	If(isnumber(search())) formulas were run in each of three new columns of the sheet in step 9 to find the presence/absence of GBS/SBG/RAD search strings from step 15 in the title, abstract, and keyword columns of each entry, respectively[2]
17	A new column was created in the sheet from step 9 to count the number of times search strings from step 15 were found in each journal entry
18	Entries were sorted by the new column in step 17 (i.e. the number of times GBS/SBG/RAD were found) and all entries with counts greater than zero were copied to a new sheet.

Step	Description
19	A new column was created in the new sheet from step 18 to record the combined citation source (eg. Ba, El, Po, Tr, Pe, Ba/El, Ba/El/Po, Ba/El/Po/Tr, BA/El/Po/Tr/Pe, etc) for each entry
20	New columns were created to count the number of times each of GBS, SBG, or RAD was mentioned in each entrie's title, abstract or keywords.
21	New columns were created to repeat step 20 but for each of the five citation sources.
22	Data was summarized in tabular form

Appendix 2.5 The number of journal articles uniquely citing Baird et al. (2008), Elshire et al. (2011), Poland et al. (2012), Truong et al. (2012), or Peterson et al. (2012) as of February 6, 2018 (Web of Science®) and referencing either "GBS", "SBG", "RAD" or more than two (i.e. ">2") of these names within the title, abstract or keywords.

	Baird et al. 2008				Elshire et al. 2011			
Year	GBS	SBG	RAD	≥2	GBS	SBG	RAD	≥2
2010	0	0	4	0	0	0	0	0
2011	2	0	9	0	0	0	0	0
2012	0	0	14	2	7	0	0	0
2013	0	0	40	3	10	0	2	0
2014	3	0	40	2	26	0	2	2
2015	1	0	60	3	51	0	4	1
2016	3	0	59	4	97	0	1	3
2017	1	0	50	1	99	0	0	4
Total	10	0	276	15	290	0	9	10

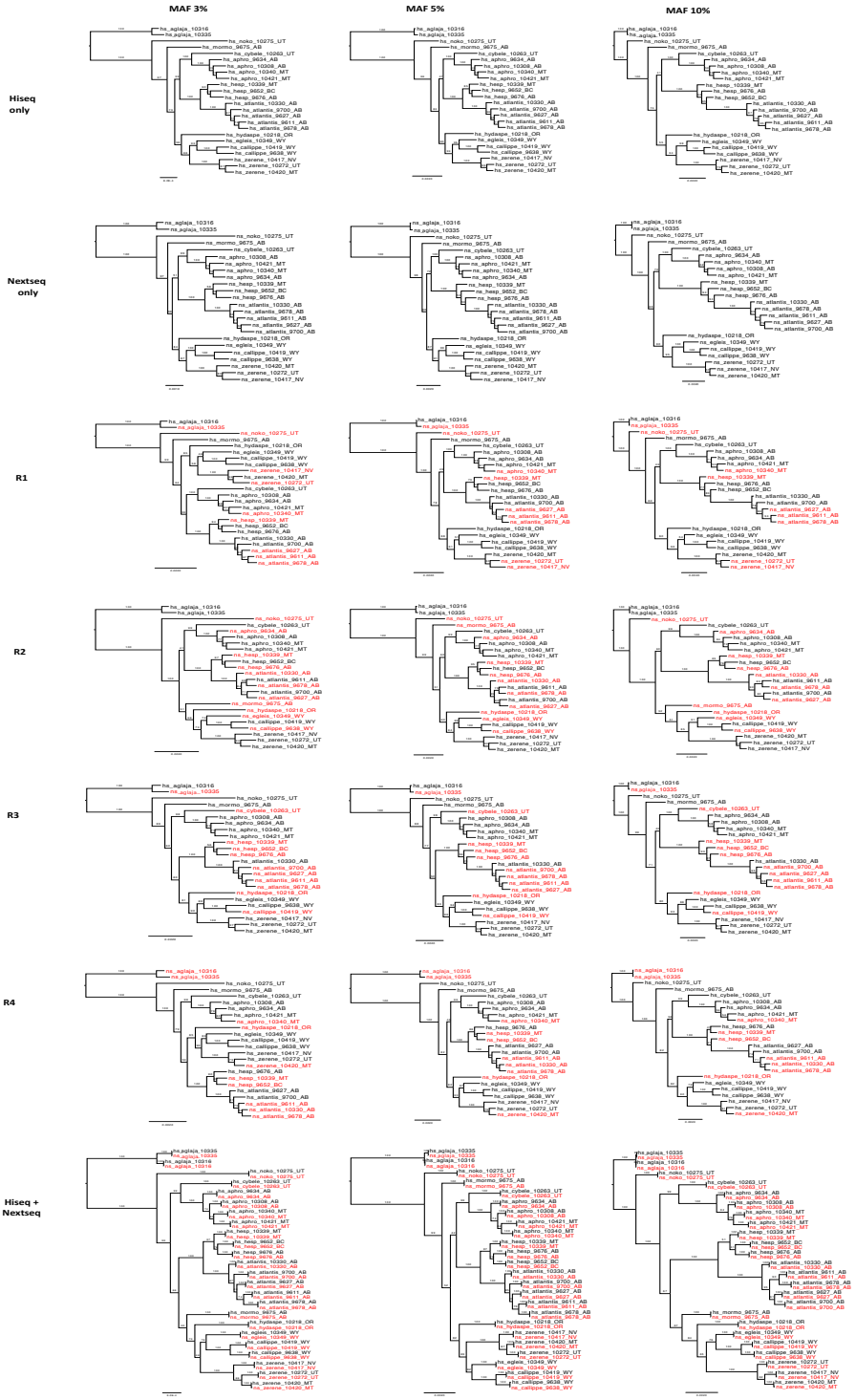
	Poland et al. 2012				Truong et al. 2012			
Year	GBS	SBG	RAD	≥2	GBS	SBG	RAD	≥2
2010	0	0	0	0	0	0	0	0
2011	0	0	0	0	0	0	0	0
2012	0	0	0	0	0	0	0	0
2013	0	0	0	0	0	0	0	0
2014	0	0	1	0	0	0	0	0
2015	6	1	1	0	0	0	1	0
2016	12	0	2	1	0	1	1	0
2017	13	0	3	0	0	0	0	0
Total	31	1	7	1	0	1	2	0

	Peterson et al. 2012				Total (all
Year	GBS	SBG	RAD	≥2	methods)
2010	0	0	0	0	4
2011	0	0	0	0	11
2012	0	0	0	0	23
2013	0	0	0	0	55
2014	0	0	5	1	82
2015	0	0	17	2	148
2016	0	0	38	0	222
2017	4	0	64	4	243
Total	4	0	124	7	788

Appendix 3.1 Collection information for the 24 specimens used in this study.

Row	ID	Genus	Species	Country	Province	Collection Locality	Date (D/M/Y)	Collector(s)
1	10316	*Argynnis*	*aglaja*	Spain	n/a	Banesqe	28-vi-2002	T. Simonsen
2	10335	*Argynnis*	*aglaja*	Finland	Aland	n/a	16-viii-2014	F. Sperling
3	10275	*Speyeria*	*nokomis*	USA	Utah	E. of Torrey	15-viii-2014	E. & R. Gage
4	9675	*Speyeria*	*mormonia*	CAN	Alberta	Pigeon Lake	13-vii-2001	F. Sperling
5	10263	*Speyeria*	*cybele*	USA	Utah	La Sal Mtns	16-viii-2014	E. & R. Gage
6	10308	*Speyeria*	*aphrodite*	CAN	Alberta	Highland Park	18-viii-2002	M. Hervieux
7	10340	*Speyeria*	*aphrodite*	USA	Montana	Little Belt Mtns	23-vii-2014	E. & R. Gage
8	10421	*Speyeria*	*aphrodite*	USA	Montana	Little Belt Mtns	23-vii-2014	E. & R. Gage
9	9634	*Speyeria*	*aphrodite*	CAN	Alberta	Red Deer River	16-viii-2013	J. Dupuis
10	10339	*Speyeria*	*hesperis*	USA	Montana	Little Belt Mtns	21-vii-2014	F. & T. Sperling; Ferguson, S.
11	9652	*Speyeria*	*hesperis*	CAN	British Columbia	S. of Chetwynd	10-viii-2013	J. Dupuis
12	9676	*Speyeria*	*hesperis*	CAN	Alberta	Pigeon Lake	2-vii-2006	F. Sperling
13	10330	*Speyeria*	*atlantis*	CAN	Alberta	Pigeon Lake	30-vii-2011	F. Sperling
14	9611	*Speyeria*	*atlantis*	CAN	Alberta	Pigeon Lake	1-viii-2009	FAH Sperling
15	9627	*Speyeria*	*atlantis*	CAN	Alberta	Schrader's Hill	14-vii-2011	J. Dupuis
16	9678	*Speyeria*	*atlantis*	CAN	Alberta	Bragg Creek	30-vii-2010	F. Sperling
17	9700	*Speyeria*	*atlantis*	CAN	Alberta	Pigeon Lake	2-vii-2001	FAH Sperling
18	10419	*Speyeria*	*callippe*	USA	Wyoming	Salt River Pass	20-vii-2014	F. Sperling
19	9638	*Speyeria*	*callippe*	CAN	Alberta	Wintering Hills	22-vii-2010	J. Dupuis
20	10349	*Speyeria*	*egleis*	USA	Wyoming	Salt River Pass	20-vii-2014	F. & T. Sperling; Ferguson, S.
21	10218	*Speyeria*	*hydaspe*	USA	Oregon	Le Grande	15-vii-2014	J. Dupuis
22	10272	*Speyeria*	*zerene*	USA	Utah	Middle Canyon	2-viii-2014	E. & R. Gage
23	10417	*Speyeria*	*zerene*	USA	Wyoming	Humbolt Natl. Forest	19-vii-1996	F. Sperling
24	10420	*Speyeria*	*zerene*	USA	Montana	Little Belt Mtns	23-vii-2014	F. Sperling

Appendix 3.2 Maximum likelihood phylogenies for 21 analyses (7 datasets filtered X 3 MAF thresholds). All bootstrap values are indicated on the branches. Sequences produced on the HiSeq 2000 are indicated by “hs” preceding the specimen names, and “ns” if the sequences were produced on the NextSeq 500. In mixed datasets, NextSeq-generated sequences are indicated in red.



Appendix 3.3 p-values for the SH test results. Each test was run using the combined HiSeq and NextSeq dataset, with minor allele frequency (MAF) thresholds of 3%, 5%, and 10%. All tests were non-significant (p-value > 0.05)

	3% MAF matrix		5% MAF matrix		10% MAF matrix	
MAF	5%	10%	5%	10%	5%	10%
3%	0.39	0.16	0.45	0.19	0.48	0.45
5%		0.27		0.22		0.53

Appendix 4.1 Specimen collection locality information

Row	ID #	Genus	Species	Subspecies	Collection date	Country	State/Prov.
1	10316	*Argynnis*	*aglaja*	n/a	28-vi-2002	Spain	n/a
2	10335	*Argynnis*	*aglaja*	n/a	16-viii-2014	Finland	Aland
3	10336	*Argynnis*	*paphia*	n/a	16-viii-2014	Finland	Aland
4	RIH2323	*Speyeria*	*adiaste*	*clemencei*	n/a	United States	California
5	9617	*Speyeria*	*aphrodite*	*manitoba*	28-viii-2013	Canada	Alberta
6	9643	*Speyeria*	*aphrodite*	*manitoba*	15-vii-2011	Canada	Alberta
7	9657	*Speyeria*	*aphrodite*	*manitoba*	30-vii-2010	Canada	Alberta
8	9659	*Speyeria*	*aphrodite*	*manitoba*	9-viii-2013	Canada	British Columbia
9	10208	*Speyeria*	*aphrodite*	*manitoba*	26-vii-2005	Canada	Alberta
10	10340	*Speyeria*	*aphrodite*	*manitoba*	23-vii-2014	United States	Montana
11	10368	*Speyeria*	*aphrodite*	*manitoba*	13-vii-2014	Canada	Alberta
12	10686	*Speyeria*	*aphrodite*	*aclestis*	21-vi-2015	United States	Wisconsin
13	10687	*Speyeria*	*aphrodite*	*aclestis*	21-vi-2015	United States	Wisconsin
14	10688	*Speyeria*	*aphrodite*	*aclestis*	21-vi-2015	United States	Wisconsin
15	10760	*Speyeria*	*aphrodite*	*columbia*	28-vii-2015	Canada	British Columbia
16	10783	*Speyeria*	*aphrodite*	*columbia*	23-vii-2015	Canada	British Columbia
17	10795	*Speyeria*	*aphrodite*	*manitoba*	9-viii-2015	Canada	Alberta
18	10838	*Speyeria*	*aphrodite*	*manitoba*	16-viii-2013	Canada	Alberta
19	11833	*Speyeria*	*aphrodite*	*ethne*	7-vii-2017	United States	Colorado
20	11835	*Speyeria*	*aphrodite*	*ethne*	7-vii-2017	United States	Colorado
21	11838	*Speyeria*	*aphrodite*	*ethne*	7-vii-2017	United States	Colorado
22	9698	*Speyeria*	*atlantis*	*hollandi*	2-vii-2002	Canada	Alberta
23	10679	*Speyeria*	*atlantis*	*hollandi*	2-vii-2015	Canada	Alberta

Row	ID #	Genus	Species	Subspecies	Collection date	Country	State/Prov.
24	10798	*Speyeria*	*atlantis*	*hollandi*	9-viii-2015	Canada	Alberta
25	10799	*Speyeria*	*atlantis*	*hollandi*	9-viii-2015	Canada	Alberta
26	10806	*Speyeria*	*atlantis*	*hollandi*	9-viii-2015	Canada	Alberta
27	11355	*Speyeria*	*atlantis*	*atlantis*	5-vii-2016	Canada	Ontario
28	11358	*Speyeria*	*atlantis*	*atlantis*	9-vii-2016	Canada	Ontario
29	11361	*Speyeria*	*atlantis*	*atlantis*	16-vii-2016	Canada	Ontario
30	11366	*Speyeria*	*atlantis*	*atlantis*	16-vii-2016	Canada	Ontario
31	11887	*Speyeria*	*atlantis*	*hollandi*	27-vii-2008	Canada	Alberta
32	12224	*Speyeria*	*atlantis*	*hollandi*	9-viii-2015	Canada	Alberta
33	9608	*Speyeria*	*callippe*	*callippe*	5-vi-1999	United States	California
34	9624	*Speyeria*	*callippe*	*callippe*	5-vi-1999	United States	California
35	9635	*Speyeria*	*callippe*	*calgariana*	15-vii-2011	Canada	Alberta
36	10297	*Speyeria*	*callippe*	*harmonia*	18-vii-2014	United States	Utah
37	10322	*Speyeria*	*callippe*	*calgariana*	13-vii-2012	Canada	Alberta
38	10371	*Speyeria*	*callippe*	*gallatini*	22-vii-2014	United States	Montana
39	10372	*Speyeria*	*callippe*	*gallatini*	22-vii-2014	United States	Montana
40	10725	*Speyeria*	*callippe*	*semivirida*	1-viiii-2015	United States	Oregon
41	10726	*Speyeria*	*callippe*	*harmonia*	16-vi-2015	United States	Utah
42	11828	*Speyeria*	*callippe*	*elaine*	15-vii-2017	United States	Oregon
43	10714	*Speyeria*	*carolae*	n/a	11-vii-2015	United States	Nevada
44	10715	*Speyeria*	*carolae*	n/a	11-vii-2015	United States	Nevada
45	10702	*Speyeria*	*coronis*	*snyderi*	16-vi-2015	United States	Utah
46	10703	*Speyeria*	*coronis*	*snyderi*	16-vi-2015	United States	Utah
47	10706	*Speyeria*	*coronis*	*snyderi*	18-vi-2015	United States	Utah
48	10711	*Speyeria*	*coronis*	*snyderi*	17-viiii-2015	United States	Utah

Row	ID #	Genus	Species	Subspecies	Collection date	Country	State/Prov.
49	10712	*Speyeria*	*coronis*	*snyderi*	17-viiii-2015	United States	Utah
50	10713	*Speyeria*	*coronis*	*snyderi*	17-viiii-2015	United States	Utah
51	11837	*Speyeria*	*coronis*	*halcyone*	6-vii-2017	United States	Colorado
52	11839	*Speyeria*	*coronis*	*halcyone*	6-vii-2017	United States	Colorado
53	11840	*Speyeria*	*coronis*	*halcyone*	6-vii-2017	United States	Colorado
54	11842	*Speyeria*	*coronis*	*halcyone*	6-vii-2017	United States	Colorado
55	10206	*Speyeria*	*cybele*	*letona*	18-vii-2014	United States	Utah
56	10326	*Speyeria*	*cybele*	*pseudocarpenteri*	6-vii-1998	Canada	Alberta
57	10821	*Speyeria*	*cybele*	*carpenterii*	5-viii-2015	United States	Colorado
58	10822	*Speyeria*	*cybele*	*carpenterii*	5-viii-2015	United States	Colorado
59	10831	*Speyeria*	*cybele*	*leto*	13-vii-2015	Canada	British Columbia
60	10854	*Speyeria*	*cybele*	*letona*	5-viii-2014	United States	Utah
61	10856	*Speyeria*	*cybele*	*letona*	5-viii-2014	United States	Utah
62	10868	*Speyeria*	*cybele*	*leto*	29-vii-2016	United States	Washington
63	10869	*Speyeria*	*cybele*	*leto*	29-vii-2016	United States	Washington
64	11362	*Speyeria*	*cybele*	*krautwurmi*	17-vii-2016	Canada	Ontario
65	10680	*Speyeria*	*diana*	n/a	2015	United States	Tennessee
66	10681	*Speyeria*	*diana*	n/a	2015	United States	Tennessee
67	10682	*Speyeria*	*diana*	n/a	2015	United States	Tennessee
68	10366	*Speyeria*	*edwardsii*	n/a	12-vii-2014	Canada	Alberta
69	10695	*Speyeria*	*edwardsii*	n/a	10-vii-2015	United States	Montana
70	10696	*Speyeria*	*edwardsii*	n/a	10-vii-2015	United States	Montana
71	10697	*Speyeria*	*edwardsii*	n/a	10-vii-2015	United States	Montana
72	10698	*Speyeria*	*edwardsii*	n/a	10-vii-2015	United States	Montana
73	10699	*Speyeria*	*edwardsii*	n/a	10-vii-2015	United States	Montana

Row	ID #	Genus	Species	Subspecies	Collection date	Country	State/Prov.
74	10700	*Speyeria*	*edwardsii*	n/a	10-vii-2015	United States	Montana
75	10701	*Speyeria*	*edwardsii*	n/a	10-vii-2015	United States	Montana
76	10202	*Speyeria*	*egleis*	*utahensis*	vii-2014	United States	Utah
77	10300	*Speyeria*	*egleis*	*utahensis*	20-vii-2014	United States	Utah
78	10315	*Speyeria*	*egleis*	*utahensis*	vii-2014	United States	Utah
79	10350	*Speyeria*	*egleis*	*macdunnoughi*	20-vii-2014	United States	Wyoming
80	10355	*Speyeria*	*egleis*	*macdunnoughi*	20-vii-2014	United States	Wyoming
81	10357	*Speyeria*	*egleis*	*macdunnoughi*	20-vii-2014	United States	Wyoming
82	10418	*Speyeria*	*egleis*	*macdunnoughi*	20-vii-2014	United States	Wyoming
83	10896	*Speyeria*	*egleis*	*utahensis*	12-vii-2016	United States	Utah
84	10897	*Speyeria*	*egleis*	*utahensis*	12-vii-2016	United States	Utah
85	10898	*Speyeria*	*egleis*	*utahensis*	12-vii-2016	United States	Utah
86	9619	*Speyeria*	*hesperis*	*dennisi*	3-viii-2013	Canada	Alberta
87	9623	*Speyeria*	*hesperis*	*irene*	9-viii-1998	United States	California
88	9642	*Speyeria*	*hesperis*	*dennisi*	9-viii-2013	Canada	British Columbia
89	9670	*Speyeria*	*hesperis*	*dennisi*	21-vii-2003	Canada	Alberta
90	10332	*Speyeria*	*hesperis*	*dennisi*	2-vii-2012	Canada	Alberta
91	10408	*Speyeria*	*hesperis*	*schellbachi*	20-vii-1993	United States	Arizona
92	10413	*Speyeria*	*hesperis*	*schellbachi*	20-vii-1993	United States	Arizona
93	10414	*Speyeria*	*hesperis*	*schellbachi*	20-vii-1993	United States	Arizona
94	10732	*Speyeria*	*hesperis*	*hutchinsi*	14-vii-2015	United States	Montana
95	10747	*Speyeria*	*hesperis*	*nausicaa*	30-vi-2015	United States	Arizona
96	10765	*Speyeria*	*hesperis*	*brico*	23-vii-2015	Canada	British Columbia
97	10886	*Speyeria*	*hesperis*	*chitone*	5-vii-2016	United States	Utah
98	11821	*Speyeria*	*hesperis*	*lurana*	5-viii-2015	United States	South Dakota

Row	ID #	Genus	Species	Subspecies	Collection date	Country	State/Prov.
99	11849	*Speyeria*	*hesperis*	*dorothea*	13-vii-2017	United States	New Mexico
100	11859	*Speyeria*	*hesperis*	*capitanensis*	7-viii-2017	United States	New Mexico
101	9646	*Speyeria*	*hydaspe*	*rhodope*	25-vii-2001	United States	Oregon
102	9647	*Speyeria*	*hydaspe*	*rhodope*	6-viii-2006	Canada	Alberta
103	10219	*Speyeria*	*hydaspe*	*rhodope*	15-vii-2014	United States	Oregon
104	10220	*Speyeria*	*hydaspe*	*rhodope*	15-vii-2014	United States	Oregon
105	10279	*Speyeria*	*hydaspe*	*rhodope*	22-viii-2014	United States	Washington
106	10280	*Speyeria*	*hydaspe*	*rhodope*	22-viii-2014	United States	Washington
107	10282	*Speyeria*	*hydaspe*	*rhodope*	22-viii-2014	United States	Washington
108	10359	*Speyeria*	*hydaspe*	*rhodope*	20-vii-2014	United States	Wyoming
109	10361	*Speyeria*	*hydaspe*	*rhodope*	20-vii-2014	United States	Wyoming
110	10376	*Speyeria*	*hydaspe*	*rhodope*	22-vii-2014	United States	Montana
111	10773	*Speyeria*	*hydaspe*	*rhodope*	23-vii-2015	Canada	British Columbia
112	10689	*Speyeria*	*idalia*	*occidentalis*	21-vi-2015	United States	Wisconsin
113	10690	*Speyeria*	*idalia*	*occidentalis*	27-vi-2015	United States	Wisconsin
114	10691	*Speyeria*	*idalia*	*occidentalis*	27-vi-2015	United States	Wisconsin
115	10692	*Speyeria*	*idalia*	*occidentalis*	27-vi-2015	United States	Wisconsin
116	10693	*Speyeria*	*idalia*	*occidentalis*	27-vi-2015	United States	Wisconsin
117	9675	*Speyeria*	*mormonia*	*eurynome*	13-vii-2001	Canada	Alberta
118	10246	*Speyeria*	*mormonia*	*mormonia*	6-viii-2014	United States	Utah
119	10254	*Speyeria*	*mormonia*	*mormonia*	11-viii-2014	United States	Utah
120	10284	*Speyeria*	*mormonia*	*washingtonia*	22-viii-2014	United States	Washington
121	10327	*Speyeria*	*mormonia*	*eurynome*	24-vii-2012	United States	Colorado
122	10334	*Speyeria*	*mormonia*	*opis*	4-vii-1998	Canada	British Columbia
123	10676	*Speyeria*	*mormonia*	*eurynome*	12-viii-2015	Canada	Alberta

Row	ID #	Genus	Species	Subspecies	Collection date	Country	State/Prov.
124	10787	*Speyeria*	*mormonia*	*artonis*	5-viii-2015	United States	Idaho
125	10790	*Speyeria*	*mormonia*	*eurynome*	14-vii-2015	United States	Montana
126	11353	*Speyeria*	*mormonia*	*mormonia*	12-vii-2016	United States	Utah
127	10259	*Speyeria*	*nokomis*	*nokomis*	10-viii-2014	United States	Utah
128	10260	*Speyeria*	*nokomis*	*nokomis*	10-viii-2014	United States	Utah
129	10274	*Speyeria*	*nokomis*	*nokomis*	15-viii-2014	United States	Utah
130	10291	*Speyeria*	*nokomis*	*apacheana*	19-viii-2014	United States	Nevada
131	10293	*Speyeria*	*nokomis*	*apacheana*	19-viii-2014	United States	Nevada
132	10294	*Speyeria*	*nokomis*	*carsonensis*	19-viii-2014	United States	Nevada
133	10295	*Speyeria*	*nokomis*	*carsonensis*	19-viii-2014	United States	Nevada
134	10296	*Speyeria*	*nokomis*	*carsonensis*	19-viii-2014	United States	Nevada
135	10859	*Speyeria*	*nokomis*	*nitocris*	10-viii-2016	United States	Arizona
136	10860	*Speyeria*	*nokomis*	*nitocris*	10-viii-2016	United States	Arizona
137	9604	*Speyeria*	*zerene*	*picta*	5-vii-2001	Canada	Alberta
138	10210	*Speyeria*	*zerene*	*picta*	13-vii-2014	United States	Montana
139	10242	*Speyeria*	*zerene*	*platina*	16-vii-2014	United States	Idaho
140	10249	*Speyeria*	*zerene*	*gunderi*	19-viii-2014	United States	Nevada
141	10250	*Speyeria*	*zerene*	*gunderi*	19-viii-2014	United States	Nevada
142	10251	*Speyeria*	*zerene*	*gunderi*	19-viii-2014	United States	Nevada
143	10273	*Speyeria*	*zerene*	*platina*	7-viii-2014	United States	Utah
144	10346	*Speyeria*	*zerene*	*picta*	23-vii-2014	United States	Montana
145	10347	*Speyeria*	*zerene*	*picta*	23-vii-2014	United States	Montana
146	10370	*Speyeria*	*zerene*	*picta*	13-vii-2014	Canada	Alberta
147	10374	*Speyeria*	*zerene*	*picta*	22-vii-2014	United States	Montana
148	10404	*Speyeria*	*zerene*	*zerene*	26-vii-1996	United States	California

Row	ID #	Genus	Species	Subspecies	Collection date	Country	State/Prov.
149	10409	*Speyeria*	*zerene*	*gunderi*	19-vii-1996	United States	Nevada
150	10410	*Speyeria*	*zerene*	*gunderi*	19-vii-1996	United States	Nevada
151	10417	*Speyeria*	*zerene*	*gunderi*	20-vii-1996	United States	Nevada
152	10420	*Speyeria*	*zerene*	*picta*	23-vii-2014	United States	Montana
153	10721	*Speyeria*	*zerene*	*platina*	1-viiii-2015	United States	Oregon
154	10767	*Speyeria*	*zerene*	*picta*	23-vii-2015	Canada	British Columbia
155	10771	*Speyeria*	*zerene*	*picta*	23-vii-2015	Canada	British Columbia
156	11373	*Speyeria*	*zerene*	*picta*	21-vii-2016	Canada	Alberta
157	11846	*Speyeria*	*zerene*	*zerene*	13-vii-2017	United States	California
158	11847	*Speyeria*	*zerene*	*zerene*	13-vii-2017	United States	California
159	11848	*Speyeria*	*zerene*	*zerene*	13-vii-2017	United States	California

Row	ID #	Latitude	Longitude	Locality
1	10316	42.6	0.53	Banesque
2	10335	60.21	25.89	n/a
3	10336	60.21	25.89	n/a
4	RIH2323	n/a	n/a	Monterey County
5	9617	54.75	-112	RR 141 km 20-48, W side of Lac la Biche lake
6	9643	50.76	-114.57	Mesa Butte
7	9657	53.052	-114.74	Buck Mtn, near Pigeon Lake and N. of Buck Lake
8	9659	55.13	-120.99	S. of Tumbler Ridge, Murray River FSR
9	10208	56.23	-117.3	12' Davis
10	10340	46.84	-110.7	Little Belt Mountains along Hwy 89 mile 53
11	10368	49.64	-110.1	Cypress Hills

Row	ID #	Latitude	Longitude	Locality
12	10686	44.48	-89.5	Buena Vista Grassland
13	10687	44.48	-89.5	Buena Vista Grassland
14	10688	44.48	-89.5	Buena Vista Grassland
15	10760	50.61	-120.12	Pendelton Creek Rec Area
16	10783	50.61	-120.12	Pendelton Creek Rec Area
17	10795	52.37	-114.92	Rocky Mtn House
18	10838	52.306	-113.076	Red Deer R @ Hwy 21
19	11833	36.996	-104.366	Colorado State Wildlife Management Area
20	11835	36.996	-104.366	Colorado State Wildlife Management Area
21	11838	38.95	-104.893	Blodgett peak open space, Colorado Springs
22	9698	53.07	-114.08	Pigeon Lake
23	10679	54.84	-119.41	Two Lakes Road, 50 km S of Grande Prairie
24	10798	52.37	-114.92	Rocky Mtn House
25	10799	52.37	-114.92	Rocky Mtn House
26	10806	52.37	-114.92	Rocky Mtn House
27	11355	49.1	-94.315	Morson
28	11358	49.1	-94.315	Morson
29	11361	49.1	-94.315	Morson
30	11366	49.1	-94.315	Morson
31	11887	53.07	-114.08	Pigeon Lake, Itaska, Audobon Reserve Bog
32	12224	53.052	-114.74	Buck Mtn Road near Pigeon Lake
33	9608	36.38	-121.57	Hastings Natural History Reserve
34	9624	36.38	-121.57	Hastings Natural History Reserve
35	9635	50.76	-114.57	Mesa Butte
36	10297	40.38	-112.45	South willow canyon, 6 mi. SW of Grantsville

Row	ID #	Latitude	Longitude	Locality
37	10322	50.95	-114.57	Bragg Creek, Sperling Farm
38	10371	45.18	-110.63	Gallatin Natl. Forest, along Bear Creek Road
39	10372	45.18	-110.63	Gallatin Natl. Forest, along Bear Creek Road
40	10725	45.335	-117.22	Lake Wallowa
41	10726	40.204	-110.969	Deep Creek canyon
42	11828	43.796	-123.884	Umpqua River at Dog Creek, Hwy 138, Mile 46.5
43	10714	36.21	-115.71	Spring Mts: Mt. Charleston
44	10715	36.21	-115.71	Spring Mts: Mt. Charleston
45	10702	40.204	-110.969	Deep Creek canyon
46	10703	40.204	-110.969	Deep Creek canyon
47	10706	40.58	-112.43	South Willow Canyon
48	10711	40.204	-110.969	Deep Creek canyon
49	10712	40.204	-110.969	Deep Creek canyon
50	10713	40.204	-110.969	Deep Creek canyon
51	11837	38.95	-104.893	Blodgett peak open space, Colorado Springs
52	11839	38.95	-104.893	Blodgett peak open space, Colorado Springs
53	11840	38.95	-104.893	Blodgett peak open space, Colorado Springs
54	11842	38.95	-104.893	Blodgett peak open space, Colorado Springs
55	10206	41.53	-111.51	Wasatch Natl. Forest
56	10326	55.25	-118.54	Kleskun Hills
57	10821	37.34	-108.6	Taylor Creek, 0.5-2.4 mi N. of SR 145
58	10822	37.34	-108.6	Taylor Creek, 0.5-2.4 mi N. of SR 145
59	10831	50.11	-120.79	16 km E of Merritt, start of Kane Rd.
60	10854	39.509	-111.86	Deep Canyon, 6.4 miles South of Levan
61	10856	39.509	-111.86	Deep Canyon, 6.4 miles South of Levan

Row	ID #	Latitude	Longitude	Locality
62	10868	48.453	-117.932	Jct. of Cole creek and Hollar creek
63	10869	48.453	-117.932	Jct. of Cole creek and Hollar creek
64	11362	49.1	-94.315	Morson
65	10680	44.144	-103.624	reared; lat and long reflect collection site of parents
66	10681	44.144	-103.624	reared; lat and long reflect collection site of parents
67	10682	44.144	-103.624	reared; lat and long reflect collection site of parents
68	10366	49.64	-110.1	Cypress Hills
69	10695	46.021	-110.134	Crazy Mountains, Big Timber Canyon
70	10696	46.021	-110.134	Crazy Mountains, Big Timber Canyon
71	10697	46.021	-110.134	Crazy Mountains, Big Timber Canyon
72	10698	46.021	-110.134	Crazy Mountains, Big Timber Canyon
73	10699	46.021	-110.134	Crazy Mountains, Big Timber Canyon
74	10700	46.021	-110.134	Crazy Mountains, Big Timber Canyon
75	10701	46.021	-110.134	Crazy Mountains, Big Timber Canyon
76	10202	40.6	-111.28	Francis Park
77	10300	40.38	-112.45	South willow canyon, 6 mi. SW of Grantsville
78	10315	40.6	-111.27	Francis Park
79	10350	42.51	-110.91	Salt River Pass, Bridger Natl. Forest
80	10355	42.51	-110.91	Salt River Pass, Bridger Natl. Forest
81	10357	42.51	-110.91	Salt River Pass, Bridger Natl. Forest
82	10418	42.51	-110.91	Salt River Pass, Bridger Natl. Forest
83	10896	39.618	-111.311	Hwy 31 and Skyline drive Jct.
84	10897	39.618	-111.311	Hwy 31 and Skyline drive Jct.
85	10898	39.618	-111.311	Hwy 31 and Skyline drive Jct.
86	9619	52.0	-113.205	Elnora Rock Ranch

Row	ID #	Latitude	Longitude	Locality
87	9623	38.54	-119.9	Hermit Valley Campground
88	9642	55.13	-120.99	S. of Tumbler Ridge
89	9670	55.24	-118.53	Kleskun Hills
90	10332	53.07	-114.08	Pigeon Lake, Itaska municipal road
91	10408	35.58	-111.6	20 min N of Flagstaff
92	10413	35.58	-111.6	20 min N of Flagstaff
93	10414	35.58	-111.6	20 min N of Flagstaff
94	10732	46.167	-110.501	Crazy Mountains, Big Timber Canyon
95	10747	33.642	-109.329	Hannigan Meadows South of Alpine
96	10765	50.610	-120.12	Pendelton Creek Rec Area
97	10886	38.017	-109.488	Abajo Mts. Forest Rd 0079 off of Hwy 101 (N. creek rd)
98	11821	44.144	-103.624	6mi. West of Merritt, Black Hills National Forest
99	11849	36.083	-108.882	Southside of hwy 134/32, E. of Norbana Pass Picnic Area
100	11859	32.958	-105.742	Lincoln Nat. Forest, 1.4 mi. N. of Cloudcroft
101	9646	44.4	-121.85	Santiam Pass, Hwy 20, Cascades
102	9647	49.637	-114.502	Coleman
103	10219	45.33	-118.29	17 mi W of Le Grande, Emily Summit Rd
104	10220	45.33	-118.29	17 mi W of Le Grande, Emily Summit Rd
105	10279	47.020	-121.33	Raven Roost
106	10280	47.020	-121.33	Raven Roost
107	10282	47.020	-121.33	Raven Roost
108	10359	42.51	-110.91	Salt River Pass, Bridger Natl. Forest
109	10361	42.51	-110.91	Salt River Pass, Bridger Natl. Forest
110	10376	45.18	-110.63	Gallatin Natl. Forest, along Bear Creek Road
111	10773	50.61	-120.12	Pendelton Creek Rec Area

Row	ID #	Latitude	Longitude	Locality
112	10689	44.48	-89.5	Buena Vista Grassland
113	10690	44.48	-89.5	Buena Vista Grassland
114	10691	44.48	-89.5	Buena Vista Grassland
115	10692	44.48	-89.5	Buena Vista Grassland
116	10693	44.48	-89.5	Buena Vista Grassland
117	9675	53.07	-114.08	Pigeon Lake, Itaska
118	10246	37.64	-112.85	1 mile East of Cedar Breaks National Monument
119	10254	38.1	-111.5	Boulder Mountain, Dixie N.F., 25 mi. West on Rd. 178
120	10284	47.020	-121.33	Raven Roost
121	10327	39.664	-105.879	Loveland Pass
122	10334	57.07	-122.87	Pink Mt
123	10676	49.67	-110.1	Crowsnest area: Gould Dome Mtn.
124	10787	42.316	-113.659	Mt. Harrison
125	10790	46.017	-110.2	Crazy Mountains, Big Timber Canyon
126	11353	39.618	-111.311	Hwy 31 and Skyline drive Jct.
127	10259	38.51	-111.88	Koosharem
128	10260	38.51	-111.88	Koosharem
129	10274	38.26	-111.38	Morrill Ranch, 5.5 miles East of Torrey
130	10291	40.4	-115.44	Ruby Marsh South of Wells, Nevada
131	10293	40.4	-115.44	Ruby Marsh South of Wells, Nevada
132	10294	38.89	-119.82	Scossa Ranch, 3.6 miles South of Jct. 207 & 206
133	10295	38.89	-119.82	Scossa Ranch, 3.6 miles South of Jct. 207 & 206
134	10296	38.89	-119.82	Scossa Ranch, 3.6 miles South of Jct. 207 & 206
135	10859	33.831	-109.094	White Mountains, Luna lake
136	10860	33.831	-109.094	White Mountains, Luna lake

Row	ID #	Latitude	Longitude	Locality
137	9604	49.66	-110.3	Beaver Cr. Rec Area, Porcupine Hills
138	10210	46.54	-112.24	Helena Natl. Forest
139	10242	43.35	-114.84	W of Fairfield, Cat Creek Rd
140	10249	41.03	-115.09	Angel Lake South of Wells NV
141	10250	41.03	-115.09	Angel Lake South of Wells NV
142	10251	41.03	-115.09	Angel Lake South of Wells NV
143	10273	40.53	-112.32	Middle canyon, Tooele
144	10346	46.84	-110.7	Little Belt Mountains along Hwy 89 mile 42
145	10347	46.84	-110.7	Little Belt Mountains along Hwy 89 mile 51
146	10370	49.64	-110.1	Cypress Hills
147	10374	45.18	-110.63	Gallatin Natl. Forest, along Bear Creek Road
148	10404	39.43	-120.26	Sagehen Creek Cpgd.
149	10409	40.41	-115.49	NE of Ely, Timber Cr. Cpgd, Humbolt Nat. For
150	10410	40.41	-115.49	NE of Ely, Timber Cr. Cpgd, Humbolt Nat. For
151	10417	40.41	-115.49	NE of Ely, Timber Cr. Cpgd, Humbolt Nat. For
152	10420	46.84	-110.72	Little Belt Mtns, Hwy 89, Mi. 42
153	10721	45.57	-117.53	Promise Road N. of LaGrande
154	10767	50.61	-120.12	Pendelton Creek Rec Area
155	10771	50.61	-120.12	Pendelton Creek Rec Area
156	11373	49.083	-113.917	Waterton lookout
157	11846	39.65	-122.739	Alder Springs Rd. & Dept. of Corrections camp
158	11847	39.65	-122.739	Alder Springs Rd. & Dept. of Corrections camp
159	11848	39.65	-122.739	Alder Springs Rd. & Dept. of Corrections camp

Row	ID #	Collector(s)	NGS protocol	Sequencing platform
1	10316	T. Simonsen	Poland *et al.* 2012	HiSeq 2000
2	10335	F. Sperling	Poland *et al.* 2012	HiSeq 2000
3	10336	F. Sperling	Poland *et al.* 2012	HiSeq 2000
4	RIH2323	Genbank accession: RIH2323, from de Moya et. al. (2017)		
5	9617	C. McDonald	Poland *et al.* 2012	HiSeq 2000
6	9643	J. Dupuis	Poland *et al.* 2012	HiSeq 2000
7	9657	J. Dupuis	Poland *et al.* 2012	HiSeq 2000
8	9659	J. Dupuis	Poland *et al.* 2012	HiSeq 2000
9	10208	S. Bromilow	Poland *et al.* 2012	HiSeq 2000
10	10340	F. Sperling& T.; Ferguson, S.	Poland *et al.* 2012	HiSeq 2000
11	10368	F. & T. Sperling; Ferguson, S.	Poland *et al.* 2012	HiSeq 2000
12	10686	W. Anderson	Peterson *et al.* 2012	NextSeq 500
13	10687	W. Anderson	Peterson *et al.* 2012	NextSeq 500
14	10688	W. Anderson	Peterson *et al.* 2012	NextSeq 500
15	10760	J. Lee	Peterson *et al.* 2012	NextSeq 500
16	10783	J. Lee	Peterson *et al.* 2012	NextSeq 500
17	10795	J. Dupuis	Peterson *et al.* 2012	NextSeq 500
18	10838	J. Dupuis & B. Mori	Peterson *et al.* 2012	NextSeq 500
19	11833	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
20	11835	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
21	11838	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
22	9698	F. Sperling	Poland *et al.* 2012	HiSeq 2000
23	10679	J. Dupuis	Peterson *et al.* 2012	NextSeq 500
24	10798	E. Campbell	Peterson *et al.* 2012	NextSeq 500
25	10799	E. Campbell	Peterson *et al.* 2012	NextSeq 500

Row	ID #	Collector(s)	NGS protocol	Sequencing platform
26	10806	E. Campbell	Peterson *et al.* 2012	NextSeq 500
27	11355	Z. MacDonald	Peterson *et al.* 2012	NextSeq 500
28	11358	Z. MacDonald	Peterson *et al.* 2012	NextSeq 500
29	11361	Z. MacDonald	Peterson *et al.* 2012	NextSeq 500
30	11366	Z. MacDonald	Peterson *et al.* 2012	NextSeq 500
31	11887	F. Sperling	Peterson *et al.* 2012	NextSeq 500
32	12224	E. Campbell	Peterson *et al.* 2012	NextSeq 500
33	9608	F. Sperling	Poland *et al.* 2012	HiSeq 2000
34	9624	F. Sperling	Poland *et al.* 2012	HiSeq 2000
35	9635	J. Dupuis	Poland *et al.* 2012	HiSeq 2000
36	10297	E. & R. Gage	Poland *et al.* 2012	HiSeq 2000
37	10322	F. Sperling	Poland *et al.* 2012	HiSeq 2000
38	10371	F. & T. Sperling; S. Ferguson	Poland *et al.* 2012	HiSeq 2000
39	10372	F. & T. Sperling; S. Ferguson	Poland *et al.* 2012	HiSeq 2000
40	10725	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
41	10726	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
42	11828	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
43	10714	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
44	10715	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
45	10702	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
46	10703	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
47	10706	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
48	10711	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
49	10712	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
50	10713	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500

Row	ID #	Collector(s)	NGS protocol	Sequencing platform
51	11837	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
52	11839	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
53	11840	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
54	11842	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
55	10206	E. Campbell	Poland *et al.* 2012	HiSeq 2000
56	10326	F. Sperling	Poland *et al.* 2012	HiSeq 2000
57	10821	M. Fisher	Peterson *et al.* 2012	NextSeq 500
58	10822	M. Fisher	Peterson *et al.* 2012	NextSeq 500
59	10831	F. Sperling	Peterson *et al.* 2012	NextSeq 500
60	10854	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
61	10856	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
62	10868	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
63	10869	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
64	11362	Z. MacDonald	Peterson *et al.* 2012	NextSeq 500
65	10680	D. McCorkle	Peterson *et al.* 2012	NextSeq 500
66	10681	D. McCorkle	Peterson *et al.* 2012	NextSeq 500
67	10682	D. McCorkle	Peterson *et al.* 2012	NextSeq 500
68	10366	F. & T. Sperling; S. Ferguson	Poland *et al.* 2012	HiSeq 2000
69	10695	S. Kohler	Peterson *et al.* 2012	NextSeq 500
70	10696	S. Kohler	Peterson *et al.* 2012	NextSeq 500
71	10697	S. Kohler	Peterson *et al.* 2012	NextSeq 500
72	10698	S. Kohler	Peterson *et al.* 2012	NextSeq 500
73	10699	S. Kohler	Peterson *et al.* 2012	NextSeq 500
74	10700	S. Kohler	Peterson *et al.* 2012	NextSeq 500
75	10701	S. Kohler	Peterson *et al.* 2012	NextSeq 500

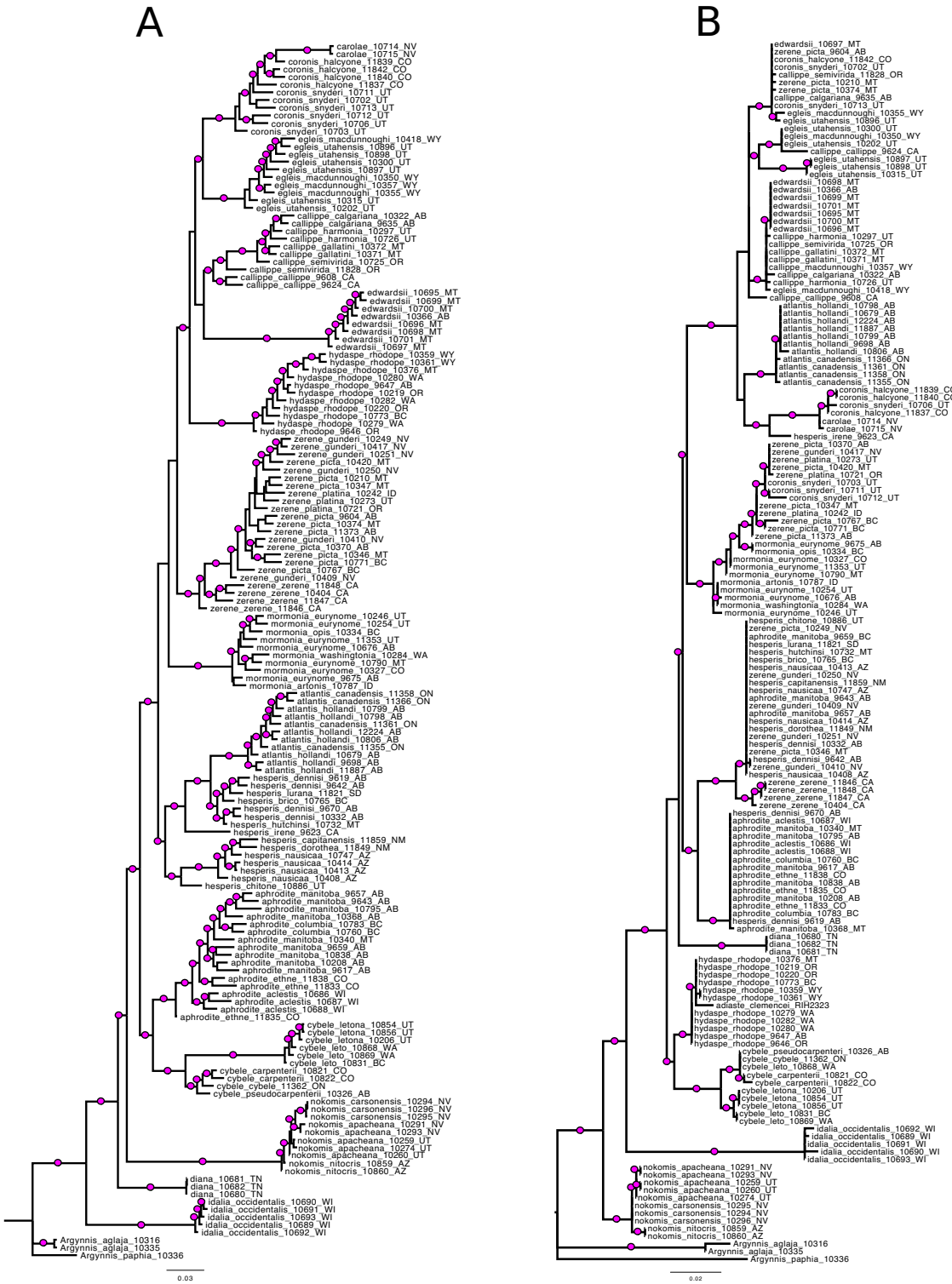
Row	ID #	Collector(s)	NGS protocol	Sequencing platform
76	10202	J. Pippen	Poland *et al.* 2012	HiSeq 2000
77	10300	E. & R. Gage	Poland *et al.* 2012	HiSeq 2000
78	10315	J. Pippen	Poland *et al.* 2012	HiSeq 2000
79	10350	F. & T. Sperling; S. Ferguson	Poland *et al.* 2012	HiSeq 2000
80	10355	F. & T. Sperling; S. Ferguson	Poland *et al.* 2012	HiSeq 2000
81	10357	F. & T. Sperling; S. Ferguson	Poland *et al.* 2012	HiSeq 2000
82	10418	F. & T. Sperling; S. Ferguson	Poland *et al.* 2012	HiSeq 2000
83	10896	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
84	10897	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
85	10898	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
86	9619	F. Sperling	Poland *et al.* 2012	HiSeq 2000
87	9623	F. Sperling	Poland *et al.* 2012	HiSeq 2000
88	9642	J. Dupuis	Poland *et al.* 2012	HiSeq 2000
89	9670	F. Sperling	Poland *et al.* 2012	HiSeq 2000
90	10332	F. Sperling	Poland *et al.* 2012	HiSeq 2000
91	10408	F. Sperling	Poland *et al.* 2012	HiSeq 2000
92	10413	F. Sperling	Poland *et al.* 2012	HiSeq 2000
93	10414	F. Sperling	Poland *et al.* 2012	HiSeq 2000
94	10732	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
95	10747	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
96	10765	E. Campbell	Peterson *et al.* 2012	NextSeq 500
97	10886	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
98	11821	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
99	11849	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
100	11859	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500

Row	ID #	Collector(s)	NGS protocol	Sequencing platform
101	9646	F. Sperling	Poland *et al.* 2012	HiSeq 2000
102	9647	F. Sperling	Poland *et al.* 2012	HiSeq 2000
103	10219	J. Dupuis	Poland *et al.* 2012	HiSeq 2000
104	10220	J. Dupuis	Poland *et al.* 2012	HiSeq 2000
105	10279	E. & R. Gage	Poland *et al.* 2012	HiSeq 2000
106	10280	E. & R. Gage	Poland *et al.* 2012	HiSeq 2000
107	10282	E. & R. Gage	Poland *et al.* 2012	HiSeq 2000
108	10359	F. & T. Sperling; S. Ferguson	Poland *et al.* 2012	HiSeq 2000
109	10361	F. & T. Sperling; S. Ferguson	Poland *et al.* 2012	HiSeq 2000
110	10376	F. & T. Sperling; S. Ferguson	Poland *et al.* 2012	HiSeq 2000
111	10773	J. Lee	Peterson *et al.* 2012	NextSeq 500
112	10689	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
113	10690	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
114	10691	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
115	10692	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
116	10693	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
117	9675	F. Sperling	Poland *et al.* 2012	HiSeq 2000
118	10246	E. & R. Gage	Poland *et al.* 2012	HiSeq 2000
119	10254	E. & R. Gage	Poland *et al.* 2012	HiSeq 2000
120	10284	E. & R. Gage	Poland *et al.* 2012	HiSeq 2000
121	10327	F. Sperling	Poland *et al.* 2012	HiSeq 2000
122	10334	F. Sperling	Poland *et al.* 2012	HiSeq 2000
123	10676	C. Whitehouse	Peterson *et al.* 2012	NextSeq 500
124	10787	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
125	10790	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500

Row	ID #	Collector(s)	NGS protocol	Sequencing platform
126	11353	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
127	10259	E. & R. Gage	Poland *et al.* 2012	HiSeq 2000
128	10260	E. & R. Gage	Poland *et al.* 2012	HiSeq 2000
129	10274	E. & R. Gage	Poland *et al.* 2012	HiSeq 2000
130	10291	E. & R. Gage	Poland *et al.* 2012	HiSeq 2000
131	10293	E. & R. Gage	Poland *et al.* 2012	HiSeq 2000
132	10294	E. & R. Gage	Poland *et al.* 2012	HiSeq 2000
133	10295	E. & R. Gage	Poland *et al.* 2012	HiSeq 2000
134	10296	E. & R. Gage	Poland *et al.* 2012	HiSeq 2000
135	10859	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
136	10860	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
137	9604	F. Sperling	Poland *et al.* 2012	HiSeq 2000
138	10210	E. Campbell	Poland *et al.* 2012	HiSeq 2000
139	10242	Dupuis, J.	Poland *et al.* 2012	HiSeq 2000
140	10249	E. & R. Gage	Poland *et al.* 2012	HiSeq 2000
141	10250	E. & R. Gage	Poland *et al.* 2012	HiSeq 2000
142	10251	E. & R. Gage	Poland *et al.* 2012	HiSeq 2000
143	10273	E. & R. Gage	Poland *et al.* 2012	HiSeq 2000
144	10346	F. & T. Sperling; S. Ferguson	Poland *et al.* 2012	HiSeq 2000
145	10347	F. & T. Sperling; S. Ferguson	Poland *et al.* 2012	HiSeq 2000
146	10370	F. & T. Sperling; S. Ferguson	Poland *et al.* 2012	HiSeq 2000
147	10374	F. & T. Sperling; S. Ferguson	Poland *et al.* 2012	HiSeq 2000
148	10404	F. Sperling	Poland *et al.* 2012	HiSeq 2000
149	10409	F. Sperling	Poland *et al.* 2012	HiSeq 2000
150	10410	F. Sperling	Poland *et al.* 2012	HiSeq 2000

Row	ID #	Collector(s)	NGS protocol	Sequencing platform
151	10417	F. Sperling	Poland *et al.* 2012	HiSeq 2000
152	10420	F. Sperling	Poland *et al.* 2012	HiSeq 2000
153	10721	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
154	10767	E. Campbell	Peterson *et al.* 2012	NextSeq 500
155	10771	E. Campbell	Peterson *et al.* 2012	NextSeq 500
156	11373	E. Campbell	Peterson *et al.* 2012	NextSeq 500
157	11846	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
158	11847	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500
159	11848	E. & R. Gage	Peterson *et al.* 2012	NextSeq 500

Appendix 4.2 Maximum likelihood SNP (A) and *COI* (B) phylogenies. Purple circles along branches indicate bootstrap support > 70%.



Appendix 5.1 Specimen collection locality information

Row	ID	Genus	Species	Subspecies	Country	State/Province	Longitude	Latitude
1	10683	*Speyeria*	*aphrodite*	*aclestis*	United States	Wisconsin	-89.5	48
2	11831	*Speyeria*	*aphrodite*	*ethne*	United States	Colorado	-104.37	37
3	11836	*Speyeria*	*aphrodite*	*ethne*	United States	Colorado	-104.37	37
4	9694	*Speyeria*	*aphrodite*	*manitoba*	Canada	Alberta	-114.08	53.07
5	10368	*Speyeria*	*aphrodite*	*manitoba*	Canada	Alberta	-110.1	49.64
6	10388	*Speyeria*	*aphrodite*	*manitoba*	Canada	Alberta	-112.93	53.8
7	10825	*Speyeria*	*atlantis*	*canadensis*	Canada	Quebec	-75.99	45.52
8	11355	*Speyeria*	*atlantis*	*canadensis*	Canada	Ontario	-94.32	49.1
9	11358	*Speyeria*	*atlantis*	*canadensis*	Canada	Ontario	-94.32	49.1
10	11359	*Speyeria*	*atlantis*	*canadensis*	Canada	Ontario	-94.32	49.1
11	11366	*Speyeria*	*atlantis*	*canadensis*	Canada	Ontario	-94.32	49.1
12	11900	*Speyeria*	*atlantis*	*canadensis*	Canada	Manitoba	-97.78	55.87
13	10678	*Speyeria*	*atlantis*	*hollandi*	Canada	Alberta	-119.41	54.84
14	10679	*Speyeria*	*atlantis*	*hollandi*	Canada	Alberta	-119.41	54.84
15	10798	*Speyeria*	*atlantis*	*hollandi*	Canada	Alberta	-114.92	52.37
16	10799	*Speyeria*	*atlantis*	*hollandi*	Canada	Alberta	-114.92	52.37
17	10801	*Speyeria*	*atlantis*	*hollandi*	Canada	Alberta	-114.92	52.37
18	11883	*Speyeria*	*atlantis*	*hollandi*	Canada	Alberta	-114.08	53.07
19	11894	*Speyeria*	*atlantis*	*hollandi*	Canada	Alberta	-114.08	53.07
20	11895	*Speyeria*	*atlantis*	*hollandi*	Canada	Alberta	-114.08	53.07
21	11896	*Speyeria*	*atlantis*	*hollandi*	Canada	Alberta	-114.08	53.07
22	11897	*Speyeria*	*atlantis*	*hollandi*	Canada	Alberta	-114.08	53.07
23	12215	*Speyeria*	*atlantis*	*hollandi*	Canada	Alberta	-114.57	50.95

Row	ID	Genus	Species	Subspecies	Country	State/Province	Longitude	Latitude
24	10731	*Speyeria*	*atlantis*	*sorocko*	United States	Colorado	-108.23	37.59
25	12249	*Speyeria*	*cybele*	*cybele*	United States	Arkansas	35.31	-93.91
26	12251	*Speyeria*	*cybele*	*cybele*	United States	Arkansas	35.31	-93.91
27	10813	*Speyeria*	*hesperis*	*beani*	Canada	Alberta	-114.92	52.37
28	11888	*Speyeria*	*hesperis*	*beani*	Canada	Alberta	-115.27	50.81
29	9616	*Speyeria*	*hesperis*	*brico*	Canada	Britsh Columbia	-119.41	49.8
30	9618	*Speyeria*	*hesperis*	*brico*	Canada	Britsh Columbia	-119.41	49.8
31	9626	*Speyeria*	*hesperis*	*brico*	Canada	Britsh Columbia	-121.63	55.67
32	9679	*Speyeria*	*hesperis*	*brico*	Canada	Britsh Columbia	-119.41	49.8
33	9680	*Speyeria*	*hesperis*	*brico*	Canada	Britsh Columbia	-119.41	49.8
34	10761	*Speyeria*	*hesperis*	*brico*	Canada	Britsh Columbia	-119.43	52.96
35	10765	*Speyeria*	*hesperis*	*brico*	Canada	Britsh Columbia	-120.12	50.61
36	10769	*Speyeria*	*hesperis*	*brico*	Canada	Britsh Columbia	-120.12	50.61
37	10770	*Speyeria*	*hesperis*	*brico*	Canada	Britsh Columbia	-120.12	50.61
38	11803	*Speyeria*	*hesperis*	*capitanensis*	United States	New Mexico	-105.74	32.96
39	11804	*Speyeria*	*hesperis*	*capitanensis*	United States	New Mexico	-105.74	32.96
40	11856	*Speyeria*	*hesperis*	*capitanensis*	United States	New Mexico	-105.74	32.96
41	11857	*Speyeria*	*hesperis*	*capitanensis*	United States	New Mexico	-105.74	32.96
42	11858	*Speyeria*	*hesperis*	*capitanensis*	United States	New Mexico	-105.74	32.96
43	11859	*Speyeria*	*hesperis*	*capitanensis*	United States	New Mexico	-105.74	32.96
44	11860	*Speyeria*	*hesperis*	*capitanensis*	United States	New Mexico	-105.74	32.96
45	10874	*Speyeria*	*hesperis*	*chitone*	United States	Utah	-109.49	38.02
46	10875	*Speyeria*	*hesperis*	*chitone*	United States	Utah	-109.49	38.02
47	10876	*Speyeria*	*hesperis*	*chitone*	United States	Utah	-109.49	38.02
48	10877	*Speyeria*	*hesperis*	*chitone*	United States	Utah	-109.49	38.02

Row	ID	Genus	Species	Subspecies	Country	State/Province	Longitude	Latitude
49	10878	*Speyeria*	*hesperis*	*chitone*	United States	Utah	-109.49	38.02
50	10879	*Speyeria*	*hesperis*	*chitone*	United States	Utah	-109.49	38.02
51	10881	*Speyeria*	*hesperis*	*chitone*	United States	Utah	-109.49	38.02
52	10882	*Speyeria*	*hesperis*	*chitone*	United States	Utah	-109.49	38.02
53	10883	*Speyeria*	*hesperis*	*chitone*	United States	Utah	-109.49	38.02
54	10884	*Speyeria*	*hesperis*	*chitone*	United States	Utah	-109.49	38.02
55	10885	*Speyeria*	*hesperis*	*chitone*	United States	Utah	-109.49	38.02
56	10886	*Speyeria*	*hesperis*	*chitone*	United States	Utah	-109.49	38.02
57	10887	*Speyeria*	*hesperis*	*chitone*	United States	Utah	-109.49	38.02
58	10888	*Speyeria*	*hesperis*	*chitone*	United States	Utah	-109.49	38.02
59	11851	*Speyeria*	*hesperis*	*chitone*	United States	Utah	-109.49	38.02
60	11852	*Speyeria*	*hesperis*	*chitone*	United States	Utah	-109.49	38.02
61	11853	*Speyeria*	*hesperis*	*chitone*	United States	Utah	-109.49	38.02
62	11854	*Speyeria*	*hesperis*	*chitone*	United States	Utah	-109.49	38.02
63	11855	*Speyeria*	*hesperis*	*chitone*	United States	Utah	-109.49	38.02
64	10733	*Speyeria*	*hesperis*	*chitone*	United States	Colorado	-108.23	37.59
65	10734	*Speyeria*	*hesperis*	*chitone*	United States	Colorado	-108.23	37.59
66	10735	*Speyeria*	*hesperis*	*chitone*	United States	Colorado	-108.23	37.59
67	9642	*Speyeria*	*hesperis*	*dennisi*	Canada	Britsh Columbia	-120.99	55.13
68	9651	*Speyeria*	*hesperis*	*dennisi*	Canada	Britsh Columbia	-121.63	55.67
69	9652	*Speyeria*	*hesperis*	*dennisi*	Canada	Britsh Columbia	-121.63	55.67
70	9653	*Speyeria*	*hesperis*	*dennisi*	Canada	Britsh Columbia	-121.63	55.67
71	10655	*Speyeria*	*hesperis*	*dennisi*	Canada	Alberta	-110.1	49.67
72	10656	*Speyeria*	*hesperis*	*dennisi*	Canada	Alberta	-110.1	49.67
73	10657	*Speyeria*	*hesperis*	*dennisi*	Canada	Alberta	-110.1	49.67

Row	ID	Genus	Species	Subspecies	Country	State/Province	Longitude	Latitude
74	10658	*Speyeria*	*hesperis*	*dennisi*	Canada	Alberta	-110.1	49.67
75	10660	*Speyeria*	*hesperis*	*dennisi*	Canada	Alberta	-110.1	49.67
76	10661	*Speyeria*	*hesperis*	*dennisi*	Canada	Alberta	-110.1	49.67
77	10664	*Speyeria*	*hesperis*	*dennisi*	Canada	Alberta	-110.1	49.67
78	10665	*Speyeria*	*hesperis*	*dennisi*	Canada	Alberta	-110.1	49.67
79	10667	*Speyeria*	*hesperis*	*dennisi*	Canada	Alberta	-110.1	49.67
80	10668	*Speyeria*	*hesperis*	*dennisi*	Canada	Alberta	-110.1	49.67
81	10669	*Speyeria*	*hesperis*	*dennisi*	Canada	Alberta	-110.1	49.67
82	10671	*Speyeria*	*hesperis*	*dennisi*	Canada	Alberta	-110.1	49.67
83	10672	*Speyeria*	*hesperis*	*dennisi*	Canada	Alberta	-110.1	49.67
84	10673	*Speyeria*	*hesperis*	*dennisi*	Canada	Alberta	-110.1	49.67
85	10889	*Speyeria*	*hesperis*	*dorothea*	United States	New Mexico	-106.44	35.2
86	10890	*Speyeria*	*hesperis*	*dorothea*	United States	New Mexico	-106.44	35.2
87	10891	*Speyeria*	*hesperis*	*dorothea*	United States	New Mexico	-106.44	35.2
88	10892	*Speyeria*	*hesperis*	*dorothea*	United States	New Mexico	-106.44	35.2
89	10893	*Speyeria*	*hesperis*	*dorothea*	United States	New Mexico	-106.44	35.2
90	10894	*Speyeria*	*hesperis*	*dorothea*	United States	New Mexico	-106.44	35.2
91	10895	*Speyeria*	*hesperis*	*dorothea*	United States	New Mexico	-106.44	35.2
92	11849	*Speyeria*	*hesperis*	*dorothea*	United States	New Mexico	-108.88	36.08
93	11850	*Speyeria*	*hesperis*	*dorothea*	United States	New Mexico	-108.88	36.08
94	10732	*Speyeria*	*hesperis*	*hutchinsi*	United States	Montana	-110.5	46.17
95	10339	*Speyeria*	*hesperis*	*hutchinsi*	United States	Montana	-110.7	46.84
96	10342	*Speyeria*	*hesperis*	*hutchinsi*	United States	Montana	-110.7	46.84
97	10343	*Speyeria*	*hesperis*	*hutchinsi*	United States	Montana	-110.7	46.84
98	11819	*Speyeria*	*hesperis*	*lurana*	United States	South Dakota	-103.62	44.14

Row	ID	Genus	Species	Subspecies	Country	State/Province	Longitude	Latitude
99	11820	*Speyeria*	*hesperis*	*lurana*	United States	South Dakota	-103.62	44.14
100	11821	*Speyeria*	*hesperis*	*lurana*	United States	South Dakota	-103.62	44.14
101	11822	*Speyeria*	*hesperis*	*lurana*	United States	South Dakota	-103.62	44.14
102	11823	*Speyeria*	*hesperis*	*lurana*	United States	South Dakota	-103.62	44.14
103	11824	*Speyeria*	*hesperis*	*lurana*	United States	South Dakota	-103.62	44.14
104	11825	*Speyeria*	*hesperis*	*lurana*	United States	South Dakota	-103.62	44.14
105	11826	*Speyeria*	*hesperis*	*lurana*	United States	South Dakota	-103.62	44.14
106	11827	*Speyeria*	*hesperis*	*lurana*	United States	South Dakota	-103.62	44.14
107	12273	*Speyeria*	*hesperis*	*lurana*	United States	South Dakota	-103.84	44.01
108	12274	*Speyeria*	*hesperis*	*lurana*	United States	South Dakota	-103.84	44.01
109	12275	*Speyeria*	*hesperis*	*lurana*	United States	South Dakota	-103.84	44.01
110	12276	*Speyeria*	*hesperis*	*lurana*	United States	South Dakota	-103.84	44.01
111	10746	*Speyeria*	*hesperis*	*nausicaa*	United States	Arizona	-109.32	33.64
112	10747	*Speyeria*	*hesperis*	*nausicaa*	United States	Arizona	-109.32	33.64
113	10408	*Speyeria*	*hesperis*	*nausicaa*	United States	Arizona	-111.6	35.58
114	10411	*Speyeria*	*hesperis*	*nausicaa*	United States	Arizona	-111.6	35.58
115	10412	*Speyeria*	*hesperis*	*nausicaa*	United States	Arizona	-111.6	35.58
116	10413	*Speyeria*	*hesperis*	*nausicaa*	United States	Arizona	-111.6	35.58
117	10414	*Speyeria*	*hesperis*	*nausicaa*	United States	Arizona	-111.6	35.58
118	10415	*Speyeria*	*hesperis*	*nausicaa*	United States	Arizona	-111.6	35.58
119	10416	*Speyeria*	*hesperis*	*nausicaa*	United States	Arizona	-111.6	35.58
120	11829	*Speyeria*	*hesperis*	*ratonensis*	United States	Colorado	-104.37	37
121	10203	*Speyeria*	*hesperis*	*tetonia*	United States	Utah	-111.51	41.53
122	10268	*Speyeria*	*hesperis*	*tetonia*	United States	Utah	-112.32	40.53
123	10269	*Speyeria*	*hesperis*	*tetonia*	United States	Utah	-112.32	40.53

Row	ID	Genus	Species	Subspecies	Country	State/Province	Longitude	Latitude
124	10270	*Speyeria*	*hesperis*	*tetonia*	United States	Utah	-112.32	40.53
125	10302	*Speyeria*	*hesperis*	*tetonia*	United States	Utah	-112.45	40.38
126	10873	*Speyeria*	*hesperis*	*tetonia*	United States	Utah	-112.62	40.48
127	11862	*Speyeria*	*hesperis*	*tetonia*	United States	Utah	-110.35	39.64
128	11863	*Speyeria*	*hesperis*	*tetonia*	United States	Utah	-110.35	39.64
129	11864	*Speyeria*	*hesperis*	*tetonia*	United States	Utah	-110.35	39.64
130	11865	*Speyeria*	*hesperis*	*tetonia*	United States	Utah	-110.35	39.64
131	11866	*Speyeria*	*hesperis*	*tetonia*	United States	Utah	-110.35	39.64
132	11867	*Speyeria*	*hesperis*	*tetonia*	United States	Utah	-110.35	39.64
133	11868	*Speyeria*	*hesperis*	*tetonia*	United States	Utah	-109.45	40.81
134	11869	*Speyeria*	*hesperis*	*tetonia*	United States	Utah	-109.45	40.81
135	10750	*Speyeria*	*hesperis*	*viola*	United States	Idaho	-115.08	44.06
136	10751	*Speyeria*	*hesperis*	*viola*	United States	Idaho	-115.08	44.06
137	12206	*Speyeria*	*hesperis*	*viola*	United States	Idaho	-111.37	44.42
138	10249	*Speyeria*	*zerene*	*gunderi*	United States	Nevada	-115.09	41.03
139	10250	*Speyeria*	*zerene*	*gunderi*	United States	Nevada	-115.09	41.03
140	10251	*Speyeria*	*zerene*	*gunderi*	United States	Nevada	-115.09	41.03
141	10410	*Speyeria*	*zerene*	*gunderi*	United States	Nevada	-115.49	40.41
142	10345	*Speyeria*	*zerene*	*picta*	United States	Montana	-110.7	46.84
143	10670	*Speyeria*	*zerene*	*picta*	Canada	Alberta	-110.1	49.67
144	10724	*Speyeria*	*zerene*	*picta*	United States	Montana	-110.2	46.02
145	11371	*Speyeria*	*zerene*	*picta*	Canada	Alberta	-113.92	49.08
146	10215	*Speyeria*	*zerene*	*platina*	United States	Utah	-111.51	41.53
147	10241	*Speyeria*	*zerene*	*platina*	United States	Idaho	-114.84	43.35
148	10243	*Speyeria*	*zerene*	*platina*	United States	Idaho	-114.84	43.35

Row	ID	Genus	Species	Subspecies	Country	State/Province	Longitude	Latitude
149	10273	*Speyeria*	*zerene*	*platina*	United States	Utah	-112.32	40.53
150	10404	*Speyeria*	*zerene*	*zerene*	United States	California	-120.26	39.43
151	11846	*Speyeria*	*zerene*	*zerene*	United States	California	-122.74	39.65
152	11847	*Speyeria*	*zerene*	*zerene*	United States	California	-122.74	39.65
153	11848	*Speyeria*	*zerene*	*zerene*	United States	California	-122.74	39.65

Row	ID	Locality	Collection date (D/M/Y)	Collector(s)
1	10683	Buena Vista Grassland	21-vi-2015	Anderson, W.
2	11831	Colorado State Wildlife Mgmt Area	7-vii-2017	Gage, E. & R.
3	11836	Colorado State Wildlife Mgmt Area	7-vii-2017	Gage, E. & R.
4	9694	Pigeon Lake	2-vii-2001	Sperling, F.
5	10368	Cypress Hills	13-vii-2014	Sperling, F. & T., Ferguson, S.
6	10388	Bruderheim	12-viii-2014	Wingert, B.
7	10825	Luskville	5-vii-2015	Nazari, V.
8	11355	Morson	5-vii-2016	MacDonald, Z.
9	11358	Morson	5-vii-2016	MacDonald, Z.
10	11359	Morson	5-vii-2016	MacDonald, Z.
11	11366	Morson	5-vii-2016	MacDonald, Z.
12	11900	Mystery Mountain	11-vii-2013	Dupuis, J.
13	10678	Two Lakes Road, S of Grande Prairie	2-vii-2015	Dupuis, J.
14	10679	Two Lakes Road, S of Grande Prairie	2-vii-2015	Dupuis, J.
15	10798	Rocky Mtn House	9-viii-2013	Campbell, E.
16	10799	Rocky Mtn House	9-viii-2013	Campbell, E.
17	10801	Rocky Mtn House	9-viii-2013	Campbell, E.
18	11883	Pigeon Lake	19-vii-2008	Sperling, F.

Row	ID	Locality	Collection date (D/M/Y)	Collector(s)
19	11894	Pigeon Lake	11-viii-2012	Sperling, F.
20	11895	Pigeon Lake	4-viii-2012	Sperling, F.
21	11896	Pigeon Lake	12-vii-2009	Sperling, F.
22	11897	Pigeon Lake	1-vii-2015	Sperling, F.
23	12215	Bragg Creek, Sperling Farm	21-viii-2016	Sperling, F.
24	10731	Taylor creek, 2.4-5 miles N of Hwy 145	5-viii-2015	Fisher, M.
25	12249	Logan Co.	1-viii-2018	Gage, E. & R.
26	12251	Logan Co.	1-viii-2018	Gage, E. & R.
27	10813	Rocky Mtn House	9-viii-2013	Campbell, E.
28	11888	Chester Lake	19-viii-2016	Nelson, T.
29	9616	Kettle Valley Trestle #1, Myra Bellevue Pk	14-vii-2013	Sperling, F.
30	9618	Kettle Valley Trestle #1, Myra Bellevue Pk	14-vii-2013	Sperling, F.
31	9626	S. of Chetwynd	10-viii-2013	Dupuis, J.
32	9679	Kettle Valley Trestle #1, Myra Bellevue Pk	14-vii-2013	Sperling, F.
33	9680	Kettle Valley Trestle #1, Myra Bellevue Pk	14-vii-2013	Sperling, F.
34	10761	Tete Jaune Cache	30-vii-2015	Campbell, E. & Lee, J.
35	10765	Pendelton Creek Rec Area	23-vii-2015	Campbell, E. & Lee, J.
36	10769	Pendelton Creek Rec Area	23-vii-2015	Campbell, E. & Lee, J.
37	10770	Pendelton Creek Rec Area	23-vii-2015	Campbell, E. & Lee, J.
38	11803	Lincoln Nat. Forest, 1.4 mi. N. of Cloudcroft	7-viii-2017	Gage, E. & R.
39	11804	Lincoln Nat. Forest, 1.4 mi. N. of Cloudcroft	7-viii-2017	Gage, E. & R.
40	11856	Lincoln Nat. Forest, 1.4 mi. N. of Cloudcroft	7-viii-2017	Gage, E. & R.
41	11857	Lincoln Nat. Forest, 1.4 mi. N. of Cloudcroft	7-viii-2017	Gage, E. & R.
42	11858	Lincoln Nat. Forest, 1.4 mi. N. of Cloudcroft	7-viii-2017	Gage, E. & R.
43	11859	Lincoln Nat. Forest, 1.4 mi. N. of Cloudcroft	7-viii-2017	Gage, E. & R.

Row	ID	Locality	Collection date (D/M/Y)	Collector(s)
44	11860	Lincoln Nat. Forest, 1.4 mi. N. of Cloudcroft	7-viii-2017	Gage, E. & R.
45	10874	Abajo Mts. Forest Rd 0079 off of Hwy 101	5-vii-2016	Gage, E. & R.
46	10875	Abajo Mts. Forest Rd 0079 off of Hwy 101	5-vii-2016	Gage, E. & R.
47	10876	Abajo Mts. Forest Rd 0079 off of Hwy 101	5-vii-2016	Gage, E. & R.
48	10877	Abajo Mts. Forest Rd 0079 off of Hwy 101	5-vii-2016	Gage, E. & R.
49	10878	Abajo Mts. Forest Rd 0079 off of Hwy 101	5-vii-2016	Gage, E. & R.
50	10879	Abajo Mts. Forest Rd 0079 off of Hwy 101	5-vii-2016	Gage, E. & R.
51	10881	Abajo Mts. Forest Rd 0079 off of Hwy 101	5-vii-2016	Gage, E. & R.
52	10882	Abajo Mts. Forest Rd 0079 off of Hwy 101	5-vii-2016	Gage, E. & R.
53	10883	Abajo Mts. Forest Rd 0079 off of Hwy 101	5-vii-2016	Gage, E. & R.
54	10884	Abajo Mts. Forest Rd 0079 off of Hwy 101	5-vii-2016	Gage, E. & R.
55	10885	Abajo Mts. Forest Rd 0079 off of Hwy 101	5-vii-2016	Gage, E. & R.
56	10886	Abajo Mts. Forest Rd 0079 off of Hwy 101	5-vii-2016	Gage, E. & R.
57	10887	Abajo Mts. Forest Rd 0079 off of Hwy 101	5-vii-2016	Gage, E. & R.
58	10888	Abajo Mts. Forest Rd 0079 off of Hwy 101	5-vii-2016	Gage, E. & R.
59	11851	Abajo Mts. Forest Rd 0079 off of Hwy 101	5-vii-2016	Gage, E. & R.
60	11852	Abajo Mts. Forest Rd 0079 off of Hwy 101	5-vii-2016	Gage, E. & R.
61	11853	Abajo Mts. Forest Rd 0079 off of Hwy 101	5-vii-2016	Gage, E. & R.
62	11854	Abajo Mts. Forest Rd 0079 off of Hwy 101	5-vii-2016	Gage, E. & R.
63	11855	Abajo Mts. Forest Rd 0079 off of Hwy 101	5-vii-2016	Gage, E. & R.
64	10733	Taylor Creek, 2.5-5 mi N. of Hwy 145	11-viii-2015	Fisher, M.
65	10734	Taylor Creek, 2.5-5 mi N. of Hwy 145	11-viii-2015	Fisher, M.
66	10735	Taylor Creek, 2.5-5 mi N. of Hwy 145	11-viii-2015	Fisher, M.
67	9642	S. of Tumbler Ridge	9-viii-2013	Dupuis, J.
68	9651	S. of Chetwynd	10-viii-2013	Dupuis, J.
69	9652	S. of Chetwynd	10-viii-2013	Dupuis, J.

Row	ID	Locality	Collection date (D/M/Y)	Collector(s)
70	9653	S. of Chetwynd	10-viii-2013	Dupuis, J.
71	10655	Reesor Lake Rd., Cypress Hills PP	4-vii-2015	Campbell, E. & Valgardson, K.
72	10656	Reesor Lake Rd., Cypress Hills PP	4-vii-2015	Campbell, E. & Valgardson, K.
73	10657	Reesor Lake Rd., Cypress Hills PP	4-vii-2015	Campbell, E. & Valgardson, K.
74	10658	Reesor Lake Rd., Cypress Hills PP	4-vii-2015	Campbell, E. & Valgardson, K.
75	10660	Reesor Lake Rd., Cypress Hills PP	4-vii-2015	Campbell, E. & Valgardson, K.
76	10661	Reesor Lake Rd., Cypress Hills PP	4-vii-2015	Campbell, E. & Valgardson, K.
77	10664	Reesor Lake Rd., Cypress Hills PP	4-vii-2015	Campbell, E. & Valgardson, K.
78	10665	Reesor Lake Rd., Cypress Hills PP	3-vii-2015	Campbell, E. & Valgardson, K.
79	10667	Reesor Lake Rd., Cypress Hills PP	3-vii-2015	Campbell, E. & Valgardson, K.
80	10668	Reesor Lake Rd., Cypress Hills PP	3-vii-2015	Campbell, E. & Valgardson, K.
81	10669	Reesor Lake Rd., Cypress Hills PP	3-vii-2015	Campbell, E. & Valgardson, K.
82	10671	Reesor Lake Rd., Cypress Hills PP	3-vii-2015	Campbell, E. & Valgardson, K.
83	10672	Reesor Lake Rd., Cypress Hills PP	3-vii-2015	Campbell, E. & Valgardson, K.
84	10673	Reesor Lake Rd., Cypress Hills PP	3-vii-2015	Campbell, E. & Valgardson, K.
85	10889	Sandia Mountains, East of Albuquerque	27-vii-2016	Gage, E. & R.
86	10890	Sandia Mountains, East of Albuquerque	27-vii-2016	Gage, E. & R.
87	10891	Sandia Mountains, East of Albuquerque	27-vii-2016	Gage, E. & R.
88	10892	Sandia Mountains, East of Albuquerque	27-vii-2016	Gage, E. & R.
89	10893	Sandia Mountains, East of Albuquerque	27-vii-2016	Gage, E. & R.
90	10894	Sandia Mountains, East of Albuquerque	27-vii-2016	Gage, E. & R.
91	10895	Sandia Mountains, East of Albuquerque	27-vii-2016	Gage, E. & R.
92	11849	Sandia Mountains, East of Albuquerque	7-viiii-2017	Gage, E. & R.
93	11850	Sandia Mountains, East of Albuquerque	7-viiii-2017	Gage, E. & R.
94	10732	Crazy Mountains	14-vii-2014	Gage, E. & R.
95	10339	Little Belt Mountains along Hwy 89 mile 51	23-vii-2014	Sperling, F. & T., Ferguson, S.

Row	ID	Locality	Collection date (D/M/Y)	Collector(s)
96	10342	Little Belt Mountains along Hwy 89 mile 53	23-vii-2014	Sperling, F. & T., Ferguson, S.
97	10343	Little Belt Mountains along Hwy 89 mile 51	23-vii-2014	Sperling, F. & T., Ferguson, S.
98	11819	West of merritt, Black Hills Natl. Forest	26-vi-2017	Gage, E. & R.
99	11820	West of merritt, Black Hills Natl. Forest	26-vi-2017	Gage, E. & R.
100	11821	West of merritt, Black Hills Natl. Forest	26-vi-2017	Gage, E. & R.
101	11822	West of merritt, Black Hills Natl. Forest	26-vi-2017	Gage, E. & R.
102	11823	West of merritt, Black Hills Natl. Forest	26-vi-2017	Gage, E. & R.
103	11824	West of merritt, Black Hills Natl. Forest	26-vi-2017	Gage, E. & R.
104	11825	West of merritt, Black Hills Natl. Forest	26-vi-2017	Gage, E. & R.
105	11826	West of merritt, Black Hills Natl. Forest	26-vi-2017	Gage, E. & R.
106	11827	West of merritt, Black Hills Natl. Forest	26-vi-2017	Gage, E. & R.
107	12273	Deerfield Reservoir	16-vii-2018	Gage, E. & R.
108	12274	Deerfield Reservoir	16-vii-2018	Gage, E. & R.
109	12275	Deerfield Reservoir	16-vii-2018	Gage, E. & R.
110	12276	Deerfield Reservoir	16-vii-2018	Gage, E. & R.
111	10746	Hannigan Meadows, S. of Alpine	30-vi-2015	Gage, E. & R.
112	10747	Hannigan Meadows, S. of Alpine	30-vi-2015	Gage, E. & R.
113	10408	20 min N. of Flagstaff	20-vii-1993	Sperling, F.
114	10411	20 min N. of Flagstaff	20-vii-1993	Sperling, F.
115	10412	20 min N. of Flagstaff	20-vii-1993	Sperling, F.
116	10413	20 min N. of Flagstaff	20-vii-1993	Sperling, F.
117	10414	20 min N. of Flagstaff	20-vii-1993	Sperling, F.
118	10415	20 min N. of Flagstaff	20-vii-1993	Sperling, F.
119	10416	20 min N. of Flagstaff	20-vii-1993	Sperling, F.
120	11829	Colorado State Wildlife Mgmt. Area	7-vii-2017	Gage, E. & R.
121	10203	Wasatch Natl. Forest	18-vii-2014	Campbell, E. & Lee, J.

Row	ID	Locality	Collection date (D/M/Y)	Collector(s)
122	10268	Middle Canyon	2-viii-2014	Gage, E. & R.
123	10269	Middle Canyon	2-viii-2014	Gage, E. & R.
124	10270	Middle Canyon	2-viii-2014	Gage, E. & R.
125	10302	South Willow Canyon, 6 m. SW of Grantsville	20-vii-2014	Gage, E. & R.
126	10873	South Willow Canyon	3-vii-2016	Gage, E. & R.
127	11862	Water Canyon Rd., Book Cliff Mtn. Range	29-vii-2016	Gage, E. & R.
128	11863	Water Canyon Rd., Book Cliff Mtn. Range	29-vii-2016	Gage, E. & R.
129	11864	Water Canyon Rd., Book Cliff Mtn. Range	29-vii-2016	Gage, E. & R.
130	11865	Water Canyon Rd., Book Cliff Mtn. Range	29-vii-2016	Gage, E. & R.
131	11866	Water Canyon Rd., Book Cliff Mtn. Range	29-vii-2016	Gage, E. & R.
132	11867	Water Canyon Rd., Book Cliff Mtn. Range	29-vii-2016	Gage, E. & R.
133	11868	Little brush creek, Kane hollow, Uinta Mts	29-vii-2016	Gage, E. & R.
134	11869	Little brush creek, Kane hollow, Uinta Mts	29-vii-2016	Gage, E. & R.
135	10750	Meadow Lake	11-viii-2015	Gage, E. & R.
136	10751	Meadow Lake	11-viii-2015	Gage, E. & R.
137	12206	Island Park	24-vii-2016	Campbell, E. & B.
138	10249	Angel Lake, South of Wells NV	19-viii-2014	Gage, E. & R.
139	10250	Angel Lake, South of Wells NV	19-viii-2014	Gage, E. & R.
140	10251	Angel Lake, South of Wells NV	19-viii-2014	Gage, E. & R.
141	10410	Timber Cr. Cpgd, Humbolt Nat. For	19-vii-1996	Sperling, F.
142	10345	Little Belt Mountains along Hwy 89 mile 51	23-vii-2014	Sperling, F. & T., Ferguson, S.
143	10670	Reesor Lake Rd., Cypress Hills PP	3-vii-2015	Campbell, E. & Valgardson, K.
144	10724	Crazy Mts. Big Timber Canyon	14-vii-2015	Gage, E. & R.
145	11371	Waterton lookout	21-vii-2016	Campbell, E.
146	10215	Wasatch Natl. Forest	18-vii-2014	Campbell, E. & Lee, J.
147	10241	W of Fairfield, Cat Creek Rd	16-vii-2014	Dupuis, J.

Row	ID	Locality	Collection date (D/M/Y)	Collector(s)
148	10243	W of Fairfield, Cat Creek Rd	16-vii-2014	Dupuis, J.
149	10273	Middle canyon, Tooele	7-viii-2014	Gage, E. & R.
150	10404	Sagehen Creek Cpgd.	26-vii-1996	Sperling, F.
151	11846	Alder Springs Rd. & Dept. of Corr. camp turnoff	13-vii-2017	Gage, E. & R.
152	11847	Alder Springs Rd. & Dept. of Corr. camp turnoff	13-vii-2017	Gage, E. & R.
153	11848	Alder Springs Rd. & Dept. of Corr. camp turnoff	13-vii-2017	Gage, E. & R.

Appendix 5.2 results of threepop tests for admixture, and respective p-values. Test results with a p-value of <0.01 were considered significant.

Admixed population	Ancestral populations	*f*3	Z-score	p-value
central hesperis	zerene,S.UT hesperis	0.00	-3.23	0.00
central hesperis	N.NM hesperis,NE hesperis	-0.01	-2.04	0.02
central hesperis	S.NM AZ hesperis,NE hesperis	0.00	-1.84	0.03
central hesperis	N.NM hesperis,atlantis	-0.01	-1.68	0.05
central hesperis	N.NM hesperis,zerene	0.00	-1.18	0.12
central hesperis	atlantis,S.UT hesperis	0.00	-1.06	0.14
central hesperis	atlantis,S.NM AZ hesperis	0.00	-1.06	0.14
NE hesperis	atlantis,S.UT hesperis	0.00	-0.43	0.34
central hesperis	NE hesperis,S.UT hesperis	0.00	-0.38	0.35
central hesperis	S.NM AZ hesperis,zerene	0.00	-0.29	0.39
central hesperis	aphrodite,S.UT hesperis	0.00	0.37	0.65
NE hesperis	N.NM hesperis,atlantis	0.00	0.39	0.65
NE hesperis	atlantis,S.NM AZ hesperis	0.00	0.49	0.69
S.NM AZ hesperis	N.NM hesperis,zerene	0.00	0.51	0.70
S.NM AZ hesperis	N.NM hesperis,atlantis	0.00	0.52	0.70
S.NM AZ hesperis	aphrodite,N.NM hesperis	0.00	0.55	0.71
NE hesperis	atlantis,central hesperis	0.00	0.59	0.72
S.NM AZ hesperis	N.NM hesperis,NE hesperis	0.00	0.70	0.76
S.UT hesperis	N.NM hesperis,NE hesperis	0.01	1.09	0.86
NE hesperis	atlantis,zerene	0.00	1.17	0.88
NE hesperis	aphrodite,atlantis	0.01	1.22	0.89
S.UT hesperis	N.NM hesperis,atlantis	0.01	1.43	0.92
central hesperis	aphrodite,N.NM hesperis	0.01	1.53	0.94
S.NM AZ hesperis	N.NM hesperis,S.UT hesperis	0.01	1.76	0.96
S.UT hesperis	atlantis,S.NM AZ hesperis	0.01	1.87	0.97
S.UT hesperis	S.NM AZ hesperis,NE hesperis	0.01	2.21	0.99
S.UT hesperis	N.NM hesperis,zerene	0.01	2.21	0.99
central hesperis	aphrodite,S.NM AZ hesperis	0.01	2.22	0.99
S.NM AZ hesperis	N.NM hesperis,central hesperis	0.00	2.32	0.99
S.UT hesperis	S.NM AZ hesperis,zerene	0.02	2.58	1.00
central hesperis	S.NM AZ hesperis,S.UT hesperis	0.01	3.05	1.00
central hesperis	atlantis,zerene	0.02	3.23	1.00
zerene	N.NM hesperis,atlantis	0.04	3.28	1.00
central hesperis	N.NM hesperis,S.UT hesperis	0.01	3.37	1.00
S.NM AZ hesperis	aphrodite,S.UT hesperis	0.03	3.51	1.00
central hesperis	NE hesperis,zerene	0.02	3.56	1.00

Admixed population	Ancestral populations	*f*3	Z-score	p-value
central hesperis	aphrodite,atlantis	0.03	3.64	1.00
central hesperis	aphrodite,NE hesperis	0.02	3.65	1.00
S.UT hesperis	aphrodite,N.NM hesperis	0.02	3.67	1.00
zerene	atlantis,S.NM AZ hesperis	0.04	3.76	1.00
aphrodite	N.NM hesperis,zerene	0.07	3.85	1.00
S.UT hesperis	N.NM hesperis,central hesperis	0.01	3.92	1.00
S.UT hesperis	aphrodite,S.NM AZ hesperis	0.02	3.94	1.00
aphrodite	N.NM hesperis,atlantis	0.05	4.25	1.00
aphrodite	atlantis,S.NM AZ hesperis	0.05	4.35	1.00
aphrodite	atlantis,zerene	0.07	4.39	1.00
aphrodite	N.NM hesperis,NE hesperis	0.06	4.44	1.00
zerene	atlantis,central hesperis	0.05	4.48	1.00
atlantis	aphrodite,NE hesperis	0.03	4.49	1.00
aphrodite	atlantis,S.UT hesperis	0.06	4.53	1.00
zerene	N.NM hesperis,NE hesperis	0.05	4.53	1.00
aphrodite	S.NM AZ hesperis,zerene	0.07	4.65	1.00
aphrodite	zerene,S.UT hesperis	0.07	4.71	1.00
S.UT hesperis	S.NM AZ hesperis,central hesperis	0.01	4.87	1.00
aphrodite	S.NM AZ hesperis,NE hesperis	0.06	4.94	1.00
zerene	atlantis,S.UT hesperis	0.05	4.98	1.00
S.NM AZ hesperis	S.UT hesperis,central hesperis	0.03	5.01	1.00
central hesperis	aphrodite,zerene	0.02	5.12	1.00
N.NM hesperis	S.NM AZ hesperis,NE hesperis	0.03	5.13	1.00
S.NM AZ hesperis	zerene,S.UT hesperis	0.03	5.13	1.00
zerene	aphrodite,N.NM hesperis	0.06	5.15	1.00
aphrodite	zerene,central hesperis	0.08	5.21	1.00
zerene	S.NM AZ hesperis,NE hesperis	0.05	5.24	1.00
N.NM hesperis	S.NM AZ hesperis,zerene	0.03	5.26	1.00
zerene	aphrodite,atlantis	0.06	5.29	1.00
S.UT hesperis	aphrodite,zerene	0.04	5.39	1.00
S.UT hesperis	aphrodite,central hesperis	0.02	5.45	1.00
aphrodite	atlantis,central hesperis	0.07	5.48	1.00
N.NM hesperis	S.NM AZ hesperis,central hesperis	0.03	5.48	1.00
atlantis	NE hesperis,zerene	0.04	5.48	1.00
aphrodite	NE hesperis,zerene	0.07	5.50	1.00
zerene	NE hesperis,central hesperis	0.05	5.54	1.00
S.NM AZ hesperis	NE hesperis,S.UT hesperis	0.04	5.58	1.00
aphrodite	N.NM hesperis,central hesperis	0.09	5.59	1.00
N.NM hesperis	atlantis,S.NM AZ hesperis	0.03	5.67	1.00

Admixed population	Ancestral populations	*f*3	Z-score	p-value
N.NM hesperis	S.NM AZ hesperis,S.UT hesperis	0.02	5.72	1.00
zerene	aphrodite,central hesperis	0.05	5.74	1.00
S.NM AZ hesperis	aphrodite,atlantis	0.06	5.82	1.00
atlantis	aphrodite,S.UT hesperis	0.08	5.88	1.00
aphrodite	S.NM AZ hesperis,central hesperis	0.09	5.91	1.00
S.UT hesperis	N.NM hesperis,S.NM AZ hesperis	0.04	5.94	1.00
aphrodite	S.NM AZ hesperis,S.UT hesperis	0.09	5.98	1.00
NE hesperis	aphrodite,zerene	0.04	6.00	1.00
zerene	aphrodite,NE hesperis	0.05	6.03	1.00
zerene	aphrodite,S.UT hesperis	0.05	6.04	1.00
aphrodite	NE hesperis,S.UT hesperis	0.07	6.07	1.00
zerene	aphrodite,S.NM AZ hesperis	0.06	6.23	1.00
zerene	NE hesperis,S.UT hesperis	0.05	6.26	1.00
atlantis	NE hesperis,S.UT hesperis	0.04	6.27	1.00
S.UT hesperis	aphrodite,atlantis	0.05	6.31	1.00
atlantis	N.NM hesperis,NE hesperis	0.04	6.32	1.00
atlantis	aphrodite,central hesperis	0.07	6.33	1.00
aphrodite	N.NM hesperis,S.UT hesperis	0.10	6.33	1.00
S.NM AZ hesperis	atlantis,S.UT hesperis	0.03	6.54	1.00
atlantis	S.NM AZ hesperis,NE hesperis	0.04	6.59	1.00
aphrodite	NE hesperis,central hesperis	0.07	6.67	1.00
S.NM AZ hesperis	aphrodite,central hesperis	0.03	6.74	1.00
S.UT hesperis	aphrodite,NE hesperis	0.04	6.80	1.00
atlantis	aphrodite,N.NM hesperis	0.08	6.82	1.00
central hesperis	N.NM hesperis,S.NM AZ hesperis	0.03	6.83	1.00
S.NM AZ hesperis	aphrodite,NE hesperis	0.06	6.86	1.00
atlantis	NE hesperis,central hesperis	0.04	6.88	1.00
S.UT hesperis	atlantis,central hesperis	0.02	6.98	1.00
NE hesperis	aphrodite,N.NM hesperis	0.05	7.03	1.00
S.NM AZ hesperis	NE hesperis,central hesperis	0.04	7.07	1.00
S.NM AZ hesperis	atlantis,zerene	0.06	7.13	1.00
central hesperis	atlantis,NE hesperis	0.06	7.16	1.00
NE hesperis	N.NM hesperis,zerene	0.05	7.19	1.00
aphrodite	S.UT hesperis,central hesperis	0.09	7.22	1.00
S.NM AZ hesperis	NE hesperis,zerene	0.06	7.27	1.00
NE hesperis	S.NM AZ hesperis,central hesperis	0.07	7.33	1.00
S.NM AZ hesperis	zerene,central hesperis	0.04	7.34	1.00
aphrodite	atlantis,NE hesperis	0.10	7.43	1.00
NE hesperis	S.NM AZ hesperis,zerene	0.05	7.59	1.00

Admixed population	Ancestral populations	*f*3	Z-score	p-value
S.NM AZ hesperis	aphrodite,zerene	0.05	7.61	1.00
atlantis	N.NM hesperis,zerene	0.08	7.61	1.00
S.UT hesperis	atlantis,zerene	0.05	7.66	1.00
atlantis	S.NM AZ hesperis,central hesperis	0.10	7.82	1.00
NE hesperis	aphrodite,S.UT hesperis	0.04	7.87	1.00
aphrodite	N.NM hesperis,S.NM AZ hesperis	0.12	7.93	1.00
NE hesperis	N.NM hesperis,central hesperis	0.07	8.05	1.00
atlantis	N.NM hesperis,central hesperis	0.11	8.07	1.00
atlantis	aphrodite,zerene	0.07	8.07	1.00
S.UT hesperis	zerene,central hesperis	0.03	8.26	1.00
N.NM hesperis	aphrodite,S.UT hesperis	0.05	8.30	1.00
N.NM hesperis	S.UT hesperis,central hesperis	0.05	8.30	1.00
atlantis	aphrodite,S.NM AZ hesperis	0.08	8.41	1.00
N.NM hesperis	zerene,S.UT hesperis	0.05	8.48	1.00
zerene	N.NM hesperis,central hesperis	0.07	8.63	1.00
zerene	N.NM hesperis,S.NM AZ hesperis	0.11	8.69	1.00
N.NM hesperis	aphrodite,S.NM AZ hesperis	0.03	8.69	1.00
NE hesperis	zerene,central hesperis	0.04	8.73	1.00
N.NM hesperis	NE hesperis,S.UT hesperis	0.06	8.78	1.00
zerene	N.NM hesperis,S.UT hesperis	0.09	8.81	1.00
S.UT hesperis	NE hesperis,central hesperis	0.02	8.98	1.00
NE hesperis	aphrodite,central hesperis	0.04	9.09	1.00
N.NM hesperis	NE hesperis,central hesperis	0.07	9.16	1.00
NE hesperis	S.NM AZ hesperis,S.UT hesperis	0.07	9.21	1.00
S.UT hesperis	NE hesperis,zerene	0.05	9.24	1.00
atlantis	S.NM AZ hesperis,S.UT hesperis	0.11	9.42	1.00
atlantis	S.NM AZ hesperis,zerene	0.08	9.44	1.00
zerene	atlantis,NE hesperis	0.09	9.51	1.00
S.NM AZ hesperis	atlantis,central hesperis	0.04	9.53	1.00
zerene	S.NM AZ hesperis,central hesperis	0.07	9.64	1.00
NE hesperis	S.UT hesperis,central hesperis	0.06	9.73	1.00
NE hesperis	N.NM hesperis,S.NM AZ hesperis	0.11	9.81	1.00
N.NM hesperis	zerene,central hesperis	0.06	9.82	1.00
atlantis	S.UT hesperis,central hesperis	0.10	9.93	1.00
zerene	S.UT hesperis,central hesperis	0.07	10.03	1.00
atlantis	N.NM hesperis,S.UT hesperis	0.12	10.09	1.00
N.NM hesperis	NE hesperis,zerene	0.09	10.18	1.00
NE hesperis	zerene,S.UT hesperis	0.04	10.20	1.00
N.NM hesperis	aphrodite,atlantis	0.09	10.32	1.00

Admixed population	Ancestral populations	*f*3	Z-score	p-value
S.NM AZ hesperis	atlantis,NE hesperis	0.11	10.35	1.00
N.NM hesperis	atlantis,S.UT hesperis	0.06	10.67	1.00
atlantis	N.NM hesperis,S.NM AZ hesperis	0.15	10.72	1.00
atlantis	zerene,S.UT hesperis	0.08	10.75	1.00
NE hesperis	N.NM hesperis,S.UT hesperis	0.08	10.91	1.00
NE hesperis	aphrodite,S.NM AZ hesperis	0.05	11.15	1.00
atlantis	zerene,central hesperis	0.08	11.32	1.00
N.NM hesperis	aphrodite,NE hesperis	0.09	12.04	1.00
zerene	S.NM AZ hesperis,S.UT hesperis	0.08	12.38	1.00
N.NM hesperis	atlantis,central hesperis	0.07	13.23	1.00
N.NM hesperis	atlantis,zerene	0.09	13.34	1.00
N.NM hesperis	atlantis,NE hesperis	0.14	13.47	1.00
N.NM hesperis	aphrodite,central hesperis	0.05	14.17	1.00
S.UT hesperis	atlantis,NE hesperis	0.09	14.75	1.00
N.NM hesperis	aphrodite, zerene	0.08	15.16	1.00