11-8-2018 2:30 PM

Spatial and temporal patterns of neutral and adaptive genetic variation in the alpine butterfly, Parnassius smintheus

Maryam Jangjoo, *The University of Western Ontario*

Supervisor: Keyghobadi, Nusha, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Biology

Abstract

Understanding how much genetic diversity exists in populations, and the processes that maintain that diversity, has been a central focus of population genetics. The evolutionary processes that determine patterns of genetic diversity depend on underlying ecological processes such as dispersal and changes in population size. In this thesis, I examine the influence of dispersal and population dynamics on neutral and adaptive genetic variation in a naturally occurring network of populations of the alpine butterfly, *Parnassius smintheus*.

My first objective was to determine the combined consequences of demographic bottlenecks and dispersal on neutral genetic variation within and among populations. Using microsatellite markers, I genotyped samples collected from across the network of populations over multiple years and tracked changes in genetic diversity and differentiation of populations across two documented bottlenecks. I also drew on long-term mark-recapture data characterizing population size and movement. I demonstrated that connectivity among populations rescues genetic diversity that is lost as a result of demographic bottlenecks. I also showed that levels and spatial patterns of genetic differentiation in the network change cyclically due to continual shifts in the relative dominance of genetic drift versus gene flow as populations fluctuate in size.

My second objective was to examine relationships between adaptive genetic variation and dispersal among populations. Using RNA sequencing, I compared gene expression patterns among individuals with differing dispersal histories. Individuals that had moved between patches (dispersers) upregulated genes involved in energy

metabolism, muscle development and stress responses compared to individuals that remained in the same patch (non-dispersers). I also examined whether variation at a candidate locus, the gene encoding the metabolic enzyme phosphoglucose isomerase (PGI), is associated with dispersal and movement. I found that individuals possessing the rare allele at each of two non-synonymous *Pgi* single nucleotide polymorphisms were either more likely to disperse or dispersed longer distances.

My work demonstrates how population size fluctuations, dispersal, and landscape structure interact to shape levels and patterns of genetic diversity. My work also provides insight into how two key global change factors, habitat fragmentation and climate change, may work synergistically to erode genetic diversity in natural populations.

Keywords

Alpine butterfly, Connectivity, Demographic bottleneck, Dispersal, Flight, Gene expression, Gene flow, Genetic differentiation, Genetic diversity, Isolation by distance, Molecular ecology, *Parnassius smintheus*, *Pgi*, Population dynamics, RNA sequencing, Single-nucleotide polymorphism, Spatial genetic structure, Thoracic temperature

Co-Authorship Statement

All chapters will be published with Nusha Keyghobadi as a co-author. For all chapters, Dr. Keyghobadi contributed to study conception and design, supervised the collection of genetic data, and helped with data interpretation and writing of the manuscripts.

Chapters 2, 3 and 5 are co-authored with Stephen F. Matter and Jens Roland, in addition to Nusha Keyghobadi. Dr. Matter and Dr. Roland contributed to study design, oversaw collection of mark-recapture data and samples, and analyzed the mark-recapture data. In addition to contributing to study design myself, I collected genetic data, conducted statistical analyses and drafted the manuscripts, with input from all other authors. Chapter 2 was published in Proceedings of the National Academy of Sciences of the United States of America. Chapter 3 is in revision at Heredity. Chapter 5 is being prepared for publication.

Chapter 4 will be co-authored with Joshua B. Benoit, Stephen F. Matter, Brent J. Sinclair, and Nusha Keyghobadi. I collected field data in collaboration with Dr. Matter. Dr. Benoit helped with de novo transcriptome assembly and gene expression analyses, and Dr. Sinclair helped with study design and interpretation of the results. I performed all laboratory work and analyses, and drafted the manuscript. Chapter 4 is being prepared for publication.

Acknowledgments

I express my special appreciation to my supervisor Dr. Nusha Keyghobadi for her invaluable support and guidance throughout my degree. You have been a tremendous mentor for me and you helped me to achieve more than I expected I could. I am truly grateful for your efforts every step of the way and for allowing me to grow as a research scientist.

I also express my gratitude to my committee members, Dr. Brent Sinclair and Dr. Elizabeth MacDougall-Shackleton for their support, guidance, and assessment of my work throughout my time at Western. I thank Dr. Stephen Matter and Dr. Jens Roland for providing a large part of the sampling and field data throughout my degree. Their academic advice and constructive feedback has helped me further develop my ability to interpret ecological and genetic data that improved the thesis.

I thank my former and current lab members: Jenna Siu, Benoit Talbot, Melissa Lucas, Helen Chen and Andrew Chaulk for their support. Special thanks to Helen, Melissa and Andrew for their continuous support, friendship, and humor that helped create an unforgettable experience and pushed me to move forward.

I thank the Biogeoscience Institute of the University of Calgary as well as the numerous field work teams for running a large mark-recapture study across the network of butterfly populations and providing samples. Financial support for this project was provided by the National Science and Engineering Research Council of Canada (NSERC), Canada Research Chairs and Western University.

A special thanks to my parents, Gholamreza and Afsaneh, and my sisters, Sara and Sarina. Words cannot express how grateful I am to my father, my mother, and my

sisters for their unconditional love and support which mean the world to me. Your support and prayer for me were what sustained me thus far. Finally, I would like to thank my wonderful husband and my best friend, Ehsan, who helps me in every step of the way. I am truly grateful for always understanding and for the love you have given me throughout this journey.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Symbols and Abbreviations

ΔAICc	Disparity in corrected Akaike information criterion
β	Estimated parameter coefficient
AIC	Akaike information criterion
AICc	Corrected Akaike information criterion
A_R	Allelic richness
ATP	Adenosine triphosphate
BLAST	Basic local alignment search tool
bp	Base pair
BP	Biological process
cDNA	Complementary DNA synthesized from an mRNA template
CI	Confidence interval
CS1	Common Sequence 1
CS2	Common Sequence 2
DEG	Differentially expressed gene
df	Degrees of freedom
DGAT	Diacylglycerol O-acyltransferase
DHAP	Dihydroxyacetone phosphate
DNA	Deoxyribonucleic acid
e-value	Expectation value
F6P	Fructose-6-phosphate
FDR	False discovery rate
for	Foraging gene
F_{ST}	Differentiation index
G3P	Glycerol-3-phosphate
G6P	Glucose-6-phosphate
GBS	Genotyping by sequencing
GLM	Generalized linear models
GO	Gene ontology
GYS	Glycogen synthase
H_e	Expected heterozygosity
hsp	Heat-shock protein
IAM	Infinite allele model
IDH1/IDH2	Isocitrate dehydrogenase1,2
IQR	Interquartile range
LD	Linkage disequilibrium
MAPK	Mitogen-activated protein kinases
ML	Maximum likelihood
MLPE	Maximum likelihood population effects
mRNA	Messenger ribonucleic acid
N	Nucleotide
NADH	Nicotinamide adenine dinucleotide
N_E	Effective population size
nsSNP	Non-synonymous single nucleotide polymorphism
PCR	Polymerase chain reaction

PDAT	Phospholipid: diacylglycerol acyltransferase
PDO	Pacific decadal oscillation
Pgi	Phosphoglucose isomerase
Phe	Phenylalanine
RADSeq	Restriction Site Associated DNA Sequencing
REML	Restricted log-likelihood
RNA	Ribonucleic acid
RNA-Seq	RNA sequencing or whole transcriptome shotgun sequencing
ROS	Reactive oxygen species
SD	Standard deviation
SE	Standard error
SMM	Stepwise mutational model
SNP	Single nucleotide polymorphism
ssSNP	Synonymous single nucleotide polymorphism
Taq	Thermophilus aquaticus
TCA cycle	Tricarboxylic acid cycle
Thr	Threonine
TMM	Trimmed Mean of M-values
TPM	Two-phase model
UQCRC2	Ubiquinol-cytochrome c reductase core protein 2
VEGF	Vascular endothelial growth factor
VMM	Virtual migration model
WSR	Wilcoxon signed rank
XDH	Xanthine dehydrogenase

List of Software Packages

DnaSP5	http://www.ub.edu/dnasp/
lme 4 1.1-10	http://CRAN.R-project.org/package=lme4
Ape 5.1	http://ape-package.ird.fr/
BCFtools 1.3.1	http://samtools.github.io/bcftools/
BOTTLENECK 1.2.02	http://www1.montpellier.inra.fr/CBGP/software/Bottleneck/bottleneck.html
Bowtie2 2.3.0	http://bowtie-bio.sourceforge.net/bowtie2
BUSCO	http://busco.ezlab.org
CLC Genomics Workbench 8.5	https://www.qiagenbioinformatics.com/
ClustVis	http://biit.cs.ut.ee/clustvis/
corMLPE 0.0.2	https://github.com/nspope/corMLPE
FASTAQC 0.11.5	http://www.bioinformatics.babraham.ac.uk/projects/fastqc
FreeNA	http://www1.montpellier.inra.fr/CBGP/software/FreeNA/
GeneMapper 4.0	Applied Biosystems (Forest City, CA)
Genepop 4.2	http://genepop.curtin.edu.au
GOseq 1.22.0	http://bioinf.wehi.edu.au/software/goseq/
HMMER 3.1	http://hmmer.org/
HP-Rare 1.0	http://www.montana.edu/kalinowski/
ImageJ	http://rsb.info.nih.gov/ij/
KEGG	http://www.genome.jp/kegg/
MEGA 6.06	http://www.megasoftware.net
Oases 0.2.08	https://www.ebi.ac.uk/~zerbino/oases/
edgeR 3.5	http://bioconductor.org/packages/edgeR/
Primer3 0.4.0	http://primer3.sourceforge.net/
R v.3.4.4	R Development Core Team, http://www.r-project.org/
RSEM 1.2.25	http://deweylab.biostat.wisc.edu/rsem
SAMtools 1.4.1	http://samtools.sourceforge.net/
SIFT	http://sift.jcvi.org/
Stats 3.6.0	https://github.com/arunsrinivasan/cran.stats
TRIMMOMATIC 0.36	http://www.usadellab.org/cms/index.php?page=trimmomatic
TRINITY 2.5.0	https://github.com/trinityrnaseq/trinityrnaseq/wiki
TRINOTATE 3.0.2	http://trinotate.github.io
VCFtools 0.1.13	https://vcftools.github.io/index.html
vegan 2.0-10	https://CRAN.R-project.org/package=vegan

Chapter 1

1 General Introduction

Biodiversity is the variation among all living things on earth (Rao and Hodgkin 2002; Benton 2016). Genetic diversity is that component of biodiversity represented by heritable variation among individuals and populations within a species (Rao and Hodgkin 2002). Genetic diversity represents the most fundamental level of biological diversity. Genetic diversity in populations is important for persistence (Saccheri et al. 1998) and the ability to adapt to environmental change (Lande and Shannon 1996); therefore, genetic diversity is arguably the foundation on which higher levels of biodiversity, namely species and ecosystem diversity, depend.

1.1 Dynamics of genetic diversity

The question of how much genetic diversity is contained in natural populations, and how that variation arises and is maintained over time, has been a central question of population genetics since the inception of the field (Charlesworth and Charlesworth 2017). There are four fundamental evolutionary processes that affect allele frequencies, and therefore levels of genetic diversity: mutation, gene flow, genetic drift, and selection (Hartl and Clark 1989).

Mutation is the ultimate source of genetic variation and leads to new alleles that can be acted upon by the other evolutionary forces (Fox and Wolf 2006; Frankham et al. 2010). Gene flow is the movement of alleles between populations; it occurs by movement of individuals or propagules followed by reproduction or establishment, and affects all parts of the genome. By introducing novel alleles into populations, gene flow increases genetic diversity within populations and also homogenizes allele frequencies between populations (Bohonak 1999; Keyghobadi et al. 2005; Fox and Wolf 2006).

Genetic drift is the change in allele frequencies due to random sampling of gametes from one generation to the next (Hartl and Clark 1989). Like gene flow, genetic drift affects all parts of the genome. In contrast to gene flow however, genetic drift reduces genetic variation within populations and increases, on average, differentiation among populations (Masel 2011). A variety of factors, including founder effects and demographic bottlenecks, determine the strength of genetic drift by influencing the effective number of breeding individuals in a population (Fox and Wolf 2006).

Finally, selection alters allele frequencies via the differential survival and reproduction of individuals with different genotypes. Selection acts on specific loci or regions of the genome, with nearby physically-linked regions potentially also being affected through ‘hitchhiking’ (Chevin et al. 2008). Selection can act to increase or decrease genetic diversity in populations depending on the exact nature of fitness differences among individuals with different genotypes. Balancing selection, for example, can maintain polymorphisms within populations, while purifying or directional selection can reduce variation at a given locus.

The fundamental evolutionary forces of gene flow, drift and selection are determined, in turn, by underlying ecological processes. Thus, dispersal and movement underlie gene flow (Ronce 2007). Population dynamics and mating structure within populations have a strong influence on levels of genetic drift (Kliman et al. 2008). Selection operates through differential ability of individuals to survive and reproduce within a given environment and ecological context (Pianka 2000).

With respect to genetic variation, a distinction is often made between neutral and adaptive variation (Holderegger et al. 2006; Frankham et al. 2010). In the case of neutral genetic variation, different possible alleles at a particular gene or locus do not have any direct effect on individual fitness. Synonymous substitutions in DNA sequences, which do not lead to differences at the amino acid level, are an example of potentially neutral variation. In the case of adaptive genetic variation, in contrast, different possible alleles at a gene or locus lead to differences in individual fitness. Adaptive variation is arguably most relevant to a population's persistence and growth, as it determines the population's direct response to environmental conditions (Fox and Wolf 2006; Forester et al. 2016).

All genetic variation is influenced by mutation, gene flow, and genetic drift, but only adaptive variation is also influenced by selection (Holderegger et al. 2006). The study of neutral genetic variation therefore allows us to understand the interplay between mutation, gene flow, and genetic drift that provides the backdrop against which selection can then act (Holderegger et al. 2006; Pélabon et al. 2010). Furthermore, the combined examination of both neutral and adaptive variation is needed to provide a complete picture of the dynamics of genetic variation in populations.

From the 1940s through the 1960s, there was considerable debate among theoretical population geneticists about both the expected levels of genetic diversity in populations and the processes that most strongly influence that variation (Crow 2010; Charlesworth and Charlesworth 2017). A 'classical' view of genetic variation suggested that, through the action of directional selection, most genes in populations would be represented by a single, favourable allele with alternative, deleterious alleles present only at very low frequencies (Muller 1950). In contrast, the 'balanced' view suggested that there might be many genes with alternative alleles that occurred at intermediate frequencies, as a result of balancing selection (Dobzhansky 1955).

Similarly, there was disagreement about the relative importance of selection in general relative to genetic drift, represented most famously by the arguments of the early theoreticians, Sir Ronald Fisher and Sewall Wright, respectively (Crow 2010). This debate continued in the 1960s and early 1970s with the introduction of the 'neutral model', which suggested that high levels of genetic variation within and between species could be maintained by genetic drift acting on neutral allelic variants (Kimura and Crow 1964; Clarke 1970).

One factor that made it difficult to resolve these debates was limited empirical data on levels and patterns of genetic variation in natural populations. Through the 1940s, 1950s and much of the 1960s, most available data on genetic variation was in the form of quantitative trait variation and visible chromosomal variation, along with a few, model cases of discrete morphological variation, such as banding patterns in the snail, *Cepaea nemoralis* (Lamotte 1959; Cain et al. 1960), or the wing-spot patterns of butterflies (Ford 1975). In many of these cases, the number of genes involved in determining the measured

traits and whether those genes had weak or strong effects was not known, making it difficult to infer levels of underlying genetic variability and examine the processes responsible for maintaining that diversity.

1.2 A brief history of molecular genetic variation studies

It was not until the mid-1960s that the first molecular-level data on genetic variation in natural populations became available. Over the next 50 years, the introduction of increasingly sophisticated yet affordable molecular techniques and tools for studying genetic variation led to an explosion of empirical data. In turn, there has been rapid progress in understanding levels of natural genetic variation, both within and among populations, and the underlying driving processes (Freeland et al. 2011). Technological innovations that have repeatedly revolutionized genetic data collection, and integration with theoretical studies and powerful data analysis methods, have allowed researchers to develop an increasingly accurate and complex understanding of genetic variability, even in non-model organisms (Manel et al. 2010; Rowan et al. 2011; Charlesworth and Charlesworth 2017).

The first studies to quantify population variability at a molecular level were by Lewontin and Hubby (1966) and Harris (1966). These studies used starch gel electrophoresis to differentiate enzyme variants of different charges, in fruit flies (*Drosophila pseudoobscura*) and humans, respectively. Although not measuring variation directly at the genetic level, such studies nonetheless provided a much clearer view of

natural genetic variation than previous studies based on morphological or chromosomal traits. These landmark studies opened the door to the use of protein electrophoresis to assay variation in a large potential set of markers, and researchers began studying levels of allozyme variability and characterizing genetic variation in a range of species. Allozyme surveys of wild populations increased rapidly through the 1970s and provided preliminary estimates of genetic variation in diverse taxa (Hamrick and Allard 1972; Harris and Hopkinson 1978; Allendorf 2017). They also provided a window into the processes, such as drift and balancing selection, that maintain genetic variation in nature (Watt 1997).

By the late 1970s, mapping of restriction enzyme sites was applied for detecting variation directly at the DNA level in natural populations (Loenen et al. 2014). Studies by Avise et al. (1979a, b) on restriction site variation in mitochondrial DNA of the pocket gopher, *Geomys pinetis*, represent the first studies that interpreted variation at the DNA level in the context of ecological and historical factors. These classic papers also represent the birth of the field of phylogeography. The first DNA sequencing study that characterized variation in a natural population was performed shortly after by Kreitman (1983), and revealed a large amount of previously hidden polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. This study revealed abundant diversity, particularly at synonymous sites, as well as small DNA insertions and deletions.

The development of microsatellite DNA loci, tandemly repeated short DNA sequences, as genetic markers in the 1990s represented another major step in studies of genetic variation. Because of high mutation rates that occur through a process of slipped strand mutation (Li et al. 2002), microsatellites are very variable. As a result, these

genetic markers had the resolution to uncover previously undetected genetic structure within and among natural populations, as well as the power to uniquely identify individuals and estimate relatedness among even close relatives (Blouin et al. 1996; Wagner et al. 2006; Lowe et al. 2010).

In the 1990s and 2000s, researchers also became increasingly interested in describing patterns of variation not just at a select few markers, but more widely across genomes. A large number of studies characterized genome-wide variation with hundreds to a few thousand markers using amplified fragment length polymorphism (AFLP), and related methods (Meudt and Clarke 2007; Bensch et al. 2008). In the mid-2000s, with the advent of DNA microarray and next generation sequencing technology, it became possible to characterize variation in non-model organisms at many thousand, genome-wide single nucleotide polymorphisms (SNPs; Davey and Blaxter 2010; Narum and Hess 2011; Schmitt et al. 2012). Today, whole genome sequencing is becoming more accessible to researchers working on natural populations, while next-generation sequencing of reduced representation libraries, through methods such as Restriction Site Associated DNA Sequencing (RADSeq; Davey and Blaxter 2010) and Genotyping by Sequencing (GBS: Elshire et al. 2011) are allowing genome-wide surveys of variability using hundreds to thousands of SNP markers.

1.3 The ecological context of genetic variation

As much as the ability to describe genetic variation in natural populations has been important, the ability to interpret the observed patterns within an ecological context has also been critical for improving our understanding of how those patterns arise and are maintained. We have, since the first characterizations of molecular-level variation in the 1960s (Harris 1966; Lewontin and Hubby 1966), developed a much stronger understanding of the balance of evolutionary forces that determine genetic variation, including the importance of drift (Crow 2010; Charlesworth and Charlesworth 2017). Studies in ecological genetics (Ford 1975; Conner and Hartl 2004), in turn, have revealed the detailed ecological basis of those evolutionary processes. Many studies have explored the complex links between dispersal and gene flow (Bohonak 1999; Keyghobadi et al. 2005; Derycke et al. 2013), population dynamics and genetic drift (Caplins et al. 2014) and individual survival and reproduction and selection (Wheat et al. 2006; Orsini et al. 2009).

Studies of the ecological basis of genetic variation patterns have themselves been spurred on over the past few decades by important technological advances in other areas. In particular, advances in remote sensing of the environment, and the analysis of spatial and geographic data, have allowed the use of spatial landscape data to flourish in diverse areas of ecology (Turner 1990). In the context of understanding genetic variation, these advances have allowed for in-depth analysis of the links among landscape structure, movement, and patterns of genetic variation. Thus, the field of 'landscape genetics' (Manel et al. 2003) has seen dramatic growth since the mid-2000s (Manel and Holderegger 2013; DiLeo and Wagner 2016).

Today, researchers are able to investigate genetic variation within natural populations using a variety of different tools and approaches. Researchers can target specific candidate genes that are known to code for a trait of interest (Mahamdallie and Ready 2012; Du et al. 2016) or whose dynamics are well understood in other systems (Wheat 2010). Researchers can also specifically target neutral loci, such as microsatellites (Keyghobadi et al. 1999; Lowe et al. 2010), or chose to study both neutral and adaptive variation in genome-wide studies (Whitehead and Crawford 2006; Candy et al. 2015). Researchers are also increasingly able to link the observed patterns of genetic variation to underlying ecological processes (Hughes et al. 2008).

1.4 Global change factors affecting genetic diversity

Biodiversity, including genetic diversity, is currently threatened by various global change factors. These include habitat loss and fragmentation, overexploitation of species, spread of invasive species, pollution, and climate change (Sala et al. 2000; Debinski and Holt 2000; Bax et al. 2003; Crow 2010). With respect to genetic diversity, the effects of habitat loss and fragmentation have probably been the most extensively studied (Takami et al. 2004; Keyghobadi 2007; Ortego et al. 2015). Researchers are also urgently attempting to understand the effects of climate change on genetic diversity (Pauls et al. 2013; Schierenbeck 2017).

1.4.1 Effects of habitat loss and fragmentation on genetic diversity

Urban, industrial, and agricultural expansion result in loss of natural habitats and decreases in the size of habitat patches (i.e., habitat loss), as well as greater isolation of habitat patches by unfavorable intervening land covers (i.e., habitat fragmentation) (Fahrig et al. 2011). Because of reduced availability of resources and smaller patch sizes, habitat loss leads to lower local effective population sizes, which enhances genetic drift (Young et al. 1996; Keyghobadi 2007). Loss and fragmentation of habitat both lead to a loss of connectivity among populations (Keyghobadi et al. 2005).

Connectivity, in a general sense, is the extent to which energy and material can move among populations, communities and ecosystems (Bishop et al. 2017). Connectivity depends on landscape structure, which is the relative abundance (‘composition’) and spatial arrangement (‘configuration’’) of different types of land cover and other geographic features (Turner 1989). However, connectivity among populations of organisms is ultimately a function of the interaction of those structural elements of the landscape (i.e., structural connectivity) with the movement behavior of individual species (i.e., functional connectivity) (Hanski 1994; Tischendorf and Fahrig 2000). Connectivity can be defined and studied at the level of the entire landscape (i.e., ‘landscape connectivity’) or individual habitat patches (‘patch connectivity’).

Connectivity among populations is necessary for gene flow (Keyghobadi 2007). Because gene flow introduces potentially novel alleles into populations, it is a process that tends to increase genetic variation within populations and counters the loss of diversity due to drift. Loss of connectivity is therefore predicted to be accompanied by reduced genetic diversity within populations and greater differentiation among

populations (Witkowski et al. 1997). These processes will be further accelerated if populations are experiencing greater levels of drift as a result of smaller population sizes. Therefore, habitat loss and fragmentation are expected to result in an erosion of genetic diversity from many populations (Keyghobadi 2007).

1.4.2 Effects of climate change on genetic diversity

Climate change as a result of increased atmospheric concentrations of greenhouse gases is driving increases in mean global temperature, and is also creating more variable local weather patterns (precipitation and temperature) worldwide (IPCC 2014). Climate change and associated climatic instability may result in unstable population dynamics and more frequent and severe fluctuations in population size, including demographic bottlenecks (i.e. severe but temporary reductions in population size) (Parmesan et al. 2000; Roland and Matter 2013). These unstable population dynamics could arise in response to greater variability in availability of resources or, particularly in ectotherms, direct effects of extreme weather conditions on individual survival (Roland and Matter 2016).

Population size is a key determinant of the rate of genetic drift, with smaller populations experiencing higher levels of drift (Slatkin 1987; Gauffre et al. 2008). In populations that fluctuate in size, the lowest population sizes experienced have the strongest influence on drift (Rich et al. 1979; Bouzat et al. 1998). As a result, when populations experience demographic bottlenecks, a considerable amount of genetic diversity can be lost due to drift (Bouzat et al. 1998; Spielman et al. 2004). The increased variability in population size that may accompany climate change is therefore predicted to

lead to lower genetic diversity in populations. At the same time, strong selective forces associated with a changing climate may arise and lead to directional selection on adaptive variation for traits such as body size (Gardner et al. 2011), dispersal (Thomas et al. 2001; Hill et al. 2011), and reproductive timing (Franks et al. 2007). These selective forces will undoubtedly affect patterns of adaptive genetic variation, although we still have a very limited understanding of the actual genes likely to be important for climate change adaptation (Franks and Hoffmann 2012).

1.4.3 Synergistic effects of global change factors

In general, most studies examining effects of global change factors on biodiversity have focused on the effects of a single factor, such as habitat loss or climate change. Potentially important interactions and synergies between different global change factors have been recognized however. Brook et al. (2008) suggested that extinction risk for many species has been underestimated because of failure to account for such synergies. They emphasized that better understanding of potential interactions between climate change and the other global change factors was needed, and that conservation actions focused on a single factor would be insufficient to prevent biodiversity loss (Brook et al. 2008; Metcalf et al. 2016; Davidson et al. 2018).

In recent years, more effort has been devoted to understanding how interactions and synergies among different causes of population decline affect biodiversity. Using an example of historical biodiversity loss, Metcalf et al. (2016) demonstrated that the combination of climate warming and human presence, rather than one of those factors

alone, drove the extinction of South American megafauna during the late Pleistocene. More recently, ecologists have explored combined effects of climate change and nutrient loading on terrestrial and aquatic ecosystems (Davidson et al. 2018). Despite the potential importance of such synergies, their effects on genetic diversity have not been studied as extensively as the effects of single factors acting in isolation.

Both habitat fragmentation and climate change may cause losses of genetic diversity in natural populations. However, the process by which this will occur is different in each case. The effects of habitat fragmentation are likely to be driven primarily by loss of connectivity and gene flow among populations, while the effects of climate change are more likely to be driven by lower effective population sizes and increased drift. Given that these two factors will affect genetic diversity through different mechanisms, there is a strong potential for synergies that will accelerate loss of diversity. That is, lack of connectivity due to fragmentation could exacerbate the effects of climate change, and vice versa.

In this thesis I examine neutral and adaptive genetic variation in a network of interconnected populations (a ‘metapopulation’; Levins 1969). I integrate data on neutral and putatively adaptive variation with data on dispersal and population dynamics. In Chapters 2 and 3 I focus explicitly on the effects of connectivity among habitat patches, the effects of fluctuations in population size (specifically, demographic bottlenecks), and their interaction, on genetic diversity. In doing so I address the first objective of my thesis, which is to examine potential synergies between habitat fragmentation (which reduces connectivity) and climate change (which increases demographic variability) on genetic diversity.

1.5 Genetic variation and dispersal

Dispersal, through its effect on gene flow, plays an important role in determining levels of genetic diversity within and among populations (Keyghobadi et al. 2005). However, the converse relationship, the effect of genetic variation on dispersal ability or tendency, is also of considerable interest to researchers (Niitepõld 2010; Edelsparre et al. 2014).

In many organisms, individuals display variation in dispersal ability or tendency. In some cases, differences in dispersal ability among individuals are a result of obvious, external morphological differences, such as winged and wingless forms (Roff 1986; Schwander and Leimar 2011). In other cases, the source of dispersal variation is more subtle and may include less obvious morphological differences (e.g., size of underlying muscle or relative wing size), behavior, or physiology (Bonte et al. 2012). There has been considerable interest in understanding the basis of variation in dispersal ability, including the underlying genetics (Saastamoinen et al. 2018).

Dispersal traits (i.e. morphological, physiological and behavioral aspects of dispersal) are often quantitative traits showing continuous variation within populations and potentially under polygenic control (Saastamoinen et al. 2018). The heritability of dispersal and associated traits (i.e., the proportion of phenotypic variation that is due to genetic variation) has been estimated in a variety of taxa using quantitative genetic approaches (reviewed in Zera and Brisson 2012, and Saastamoinen et al. 2018). These studies demonstrated a significant heritability of dispersal-related traits across many species, with a moderate average value of 0.35 in insects (Saastamoinen et al. 2018). Although such studies point to an important role of additive genetic variation in

determining dispersal, they also highlight that environmental variation and epigenetic effects, which modify gene expression patterns, can also explain much of the phenotypic variation in dispersal (Saastamoinen et al. 2018).

Some studies have incorporated genetic markers and mapping to assess the genetic architecture of dispersal (i.e., number of genes, their location, and relative effect), in particular using quantitative trait loci (QTL) approaches. This kind of genetic mapping approach has revealed that the genetic architecture underlying dispersal varies considerably among species, such that dispersal variation can be controlled by a single gene, a few genes (each of large effect), or by additive and interactive effects of many genes (Roff and Fairbairn 2001, 2007; Caillaud et al. 2002).

More recently, researchers have used transcriptome profiling, which reveals gene expression patterns at the RNA level, to evaluate gene expression differences associated with dispersal-related behaviours (e.g., long-distance flight) and have thereby identified large numbers of genes potentially important for dispersal (Margotta et al. 2012; Somervuo et al. 2014; Jones et al. 2015; Kvist et al. 2015). For example, in the cotton bollworm (*Helicoverpa armigera*), flight performance was linked to the differential expression of a suite of candidate genes involved in flight muscle structure and lipid metabolism (Jones et al. 2015). In the Glanville fritillary butterfly, *Melitaea cinxia*, long distance flight resulted in differential expression of over 1500 genes, including genes involved in energy metabolism, ribosome biogenesis and RNA processing, stress responses, and immunity (Kvist et al. 2015).

Finally, in a relatively small number of cases, individual genes and nucleotide variants with a large effect on dispersal have been identified. These include the *forager*

gene with rover and sitter alleles in *Drosophila melanogaster* (Sokolowski 1980; Edelsparre et al. 2014), the G-protein coupled receptor gene neuropeptide receptor-1 *(npr-1*) in *Caenorhabditis elegans* (de Bono and Bargmann 1998), and the phosphoglucose isomerase gene (*Pgi*) in the Glanville fritillary butterfly, *M. cinxia* (Saastamoinen et al. 2018). Allelic variation at the *for* gene in *D. melanogaster* influences larval foraging behaviour, as well as adult dispersal propensity and the probability of long-distance dispersal (Edelsparre et al. 2014). Likewise, variation at the G-protein coupled receptor *npr-1* leads to a behavioural polymorphism in *C. elegans* that is in analogous to the rover and sitter phenotypes associated with the *Drosophila for* gene (de Bono and Bargmann 1998). In the Glanville fritillary, *Pgi* allelic variation has been linked directly to movement, dispersal and flight metabolic rate in the field (Haag et al. 2005; Niitepõld et al. 2009).

Dispersal is critically important to the ecology and evolution of spatially structured populations and communities (Clobert et al. 2012; Travis et al. 2013). The second objective of my thesis is to explore the genetic basis of dispersal using transcriptomics (Chapter 4) and using *Pgi* as a candidate locus (Chapter 5). By developing a better understanding of not just how dispersal affects genetic variation (Chapters 2 and 3), but also how genetic variation influences dispersal (Chapters 4 and 5), we can develop a more complete and rich appreciation of the dynamics of genetic variation in spatially structured populations.

1.6 *Pgi* as a candidate gene in ecology and evolution

The gene *Pgi* encodes the enzyme phosphoglucose isomerase (PGI), which catalyzes the second step in glycolysis, converting glucose-6-phosphate (G6P) into fructose-6-phosphate (F6P) (Berg et al. 2002). Because this reaction is reversible, and since G6P can enter alternative pathways of the pentose phosphate shunt or glycogen biosynthesis, PGI is considered a branch-point enzyme (Wheat and Hill 2014), However, F6P is normally rapidly consumed in the next step of glycolysis so that it is unlikely to undergo the reverse reaction in most circumstances (Berg et al. 2002). Through the process of glycolysis, the high-energy compounds adenosine triphosphate (ATP) and reduced nicotinamide adenine dinucleotide (NADH) are produced, providing energy to sustain cellular activities.

The gene *Pgi* has been well-studied in ecological and evolutionary contexts (Wheat 2010). Many studies have demonstrated polymorphism in *Pgi* sequences (Haag et al. 2005; Hoffman 1981; Filatov and Charlesworth 1999; Wheat 2010), evidence for selection on *Pgi* (Watt 2003; Orsini et al. 2009), and effects of *Pgi* alleles on diverse aspects of performance and fitness (Watt 1983; Watt et al. 1983; Haag et al. 2005). In addition to effects on flight and dispersal, *Pgi* variants have also been shown to influence mating success, oviposition, running speed, thermal performance, and even population growth rate, across a wide range of taxa (Filatov and Charlesworth 1999; Dahlhoff and Rank 2000; Orsini et al. 2009; Wheat et al. 2010). Because of extensive evidence for selection on *Pgi*, as well as documented effects on ecologically relevant performance traits, this gene is considered by many to be a candidate adaptive locus for ecologically important traits, including movement and dispersal, particularly in arthropods (Wheat and Hill 2014). Although very different in structure, function and evolutionary history, the role of *Pgi* as a

candidate locus for arthropods is somewhat analogous to the role of the major histocompatibility (*MHC*) genes in vertebrates (Wheat 2010).

1.7 Introduction to study species and a model system

The Rocky Mountain apollo butterfly, *Parnassius smintheus*, occupies high-altitude alpine meadows (above *ca.* 2100 m) separated to varying degrees by montane forests in the North American Rocky Mountains (Fownes and Roland 2002; Ross et al. 2005). *Parnassius smintheus* has one generation per year. Adults have a single annual flight in July–August, during which they mate, and the females lay eggs. Individuals overwinter as a pharate larva inside the egg. The aposomatic larvae hatch the following year in May and feed on the main host plant, lanceleaf stonecrop (*Sedum lanceolatum*), and to a lesser degree, ledge stonecrop (*Rhodiola integrifolia*) (Fownes and Roland 2002; Matter et al. 2011). Adults do not disperse more than a few kilometers during their adult lifespan (Matter et al. 2011), and movement is hindered through forested areas (Roland et al. 2000). *Parnassius smintheus* is an ideal subject to study spatially-structured populations, dispersal, and gene flow because of its patchy distribution in mountain landscapes.

My thesis research was conducted on a network of populations of *P. smintheus* that occupy patches of alpine meadow habitat along Jumpingpound Ridge, and the adjacent Cox Hill and Lusk Ridge, in the Kananaskis region of Alberta, Canada (50° 57'N, 114° 54'W; Figure 1.1). The meadow patches are located above treeline, approximately 2100 m above sea level, and range in area from 0.2 ha to 22.7 ha. The

meadow patches are separated to varying degrees by intervening forest (Roland and Matter 2007). The total distance between pairs of patches, measured along ridge-tops and between the centroids of butterfly captures within each patch, range from ~ 150 m to ~11.18 km. This particular network of populations of *P. smintheus* has been studied continuously since 1995 through yearly mark-recapture studies and long-term collection of tissue samples for genetic analysis. The effects of both landscape structure and climate variation on population dynamics and dispersal in this system have been studied in detail (Roland and Matter 2007).

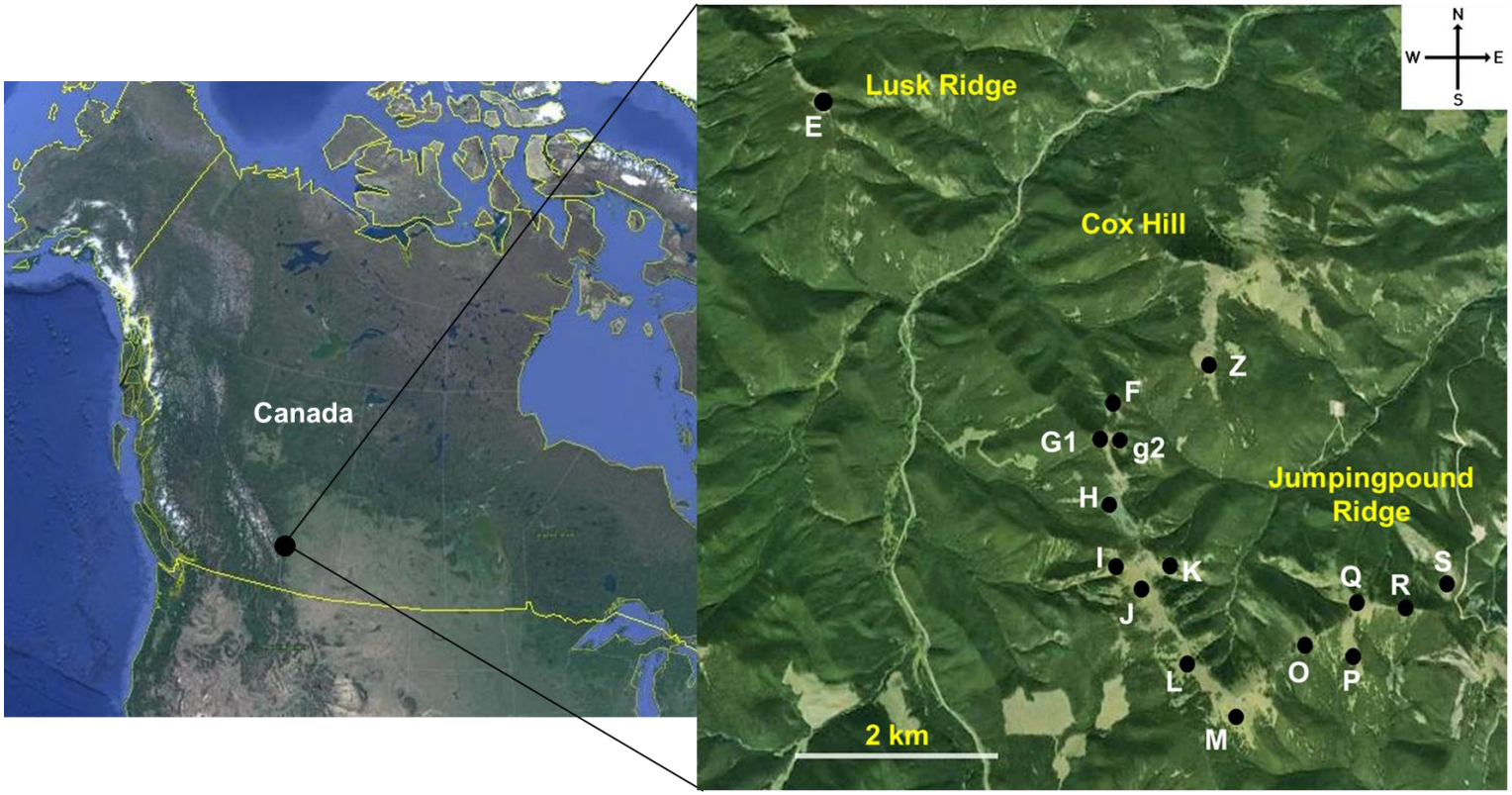


Figure 1.1 Map of the study patches occupied by a network of *Parnassius smintheus* populations, in Kananaskis, Alberta, Canada (50° 57'N, 114° 54'W). Black circles with letter labels show the location of 16 habitat patches sampled along ridge-tops in this study. Map data: Google, DigitalGlobe 2018**.**

The size of the adult population in *P. smintheus* varies from year to year (Roland et al. 2000; Matter et al. 2014). In the Jumpingpound network, Roland and Matter (2016) have shown that population dynamics of *P. smintheus* is largely driven by early winter weather, which they found was a strong descriptor of annual population growth. Specifically, increased mortality of overwintering, pharate larvae is associated with extreme weather, including both cold and warm temperatures, and reduced snowfall in November. In my study site, two severe network-wide demographic bottlenecks have been documented (starting in 2003 and 2010, respectively) and these were likely driven by low overwinter survival of larvae as a result of reduced early winter snow cover. In general, such bottlenecks are predicted to occur regularly in this system, on the order of every decade, as a result of year-to-year variation in winter weather conditions (Roland and Matter 2016).

Basic aspects of the dispersal of adult *P. smintheus* have been described in my study system (Roland et al. 2000; Matter et al. 2004). Mark-recapture data suggest most movements occur within habitat patches and indicate that within a flight season the mean net displacement is ~ 150 m and the maximum net displacement is ~ 2 km (Roland et al. 2000). Movement and dispersal decline exponentially with distance, but at a more rapid rate over forested areas than over open meadows (Roland et al. 2000; Matter et al. 2004; Keyghobadi et al. 2005). Thus, in this system, forest represents a barrier to movement and the amount of forest in the landscape determines functional connectivity (Roland et al. 2000; Roland and Matter 2007) with important implications for the level of synchrony in population dynamics (Roland and Matter 2007; Matter et al. 2014) and for gene flow and genetic structure (Keyghobadi et al. 1999). Adult females are more cryptic and harder to

capture than males, such that mark-recapture datasets are typically dominated by data from males. However, using data collected over multiple years, Goff et al. (2018) showed that although the sexes display similar mean dispersal distances, females are less sensitive to the effects of intervening forest than males. Population size and density also influence dispersal, as butterflies are more likely to leave lower density populations and immigrate to higher density populations (Roland et al. 2000; Matter et al. 2004).

1.8 Overview of thesis

In this thesis, I explore the dynamics of genetic variation in a network of interconnected populations that experience regular fluctuations in size. I take advantage of, and contribute to, a unique, long-term dataset comprising demographic, genetic, and movement data from the spatial population network of *P. smintheus*. My thesis consists of four data chapters, which were designed as separate studies for independent publication. Chapter 2 has been published (Jangjoo et al. 2016), Chapter 3 is under review (Jangjoo et al. submitted), and Chapters 4 and 5 will soon be submitted for publication.

In Chapters 2 and 3 I focus on the consequences of demographic bottlenecks for neutral genetic variation (assessed using microsatellite markers), taking into account the effects of landscape structure. In Chapter 2, I test the hypothesis that habitat patch connectivity contributes to the rescue of genetic diversity after a demographic bottleneck by facilitating immigration and gene flow. In Chapter 3, I investigate how patterns of genetic differentiation among populations respond to repeated demographic bottlenecks.

Specifically, I document changes in neutral genetic structure associated with two cycles of population size collapse and recovery across the entire network.

In Chapters 4 and 5 I focus on developing a better understanding of the genetics underlying dispersal in this species. In Chapter 4 I determine if there are differences in gene expression between individuals that moved between habitat patches (dispersers) and those that remained in the same habitat patch (non-dispersers). More specifically, I use RNA-Seq technology to assemble a transcriptome for *P. smintheus* thoracic muscle tissue and to compare gene expression patterns among individuals with differing dispersal histories and caught flying under different temperature conditions. In Chapter 5 I describe, for the first time, the coding sequence of the candidate gene *Pgi* in *P. smintheus*. I also develop assays for variation at SNPs in *Pgi* and assess whether *Pgi* genotype predicts variation in movement and dispersal among individuals.

Overall, I aim to contribute to the rich literature on the genetics of spatially structured populations, as well as to the literature exploring effects of habitat fragmentation and climate change on genetic diversity.

1.9 Literature cited

Allendorf FW (2017). Genetics and the conservation of natural populations: allozymes to genomes. *Molecular Ecology* 26: 420–430.

Avise JC, Giblin-Davidson C, Laerm J, Patton JC, Lansman RA (1979a). Mitochondrial DNA clones and matriarchal phylogeny within and among geographic populations of the pocket gopher, *Geomys pinetis*. *Proceedings of the National Academy of Sciences of the United States of America* 76: 6694–6698.

Avise JC, Lansman RA, Shade RA (1979b). 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.

Bax N, Williamson A, Aguero M, Gonzalez E, Geeves W (2003). Marine invasive alien species: a threat to global biodiversity. *Marine Policy* 27: 313–323.

Bensch S, Åkesson S, Irwin DE (2008). The use of AFLP to find an informative SNP: genetic differences across a migratory divide in willow warblers. *Molecular Ecology* 11: 2359–2366.

Benton MJ (2016). Origins of biodiversity. *PLOS Biology* 14: e2000724.

Berg J., Tymoczko J., Stryer L (2002). Biochemistry, 5th edition. W H Freeman, New York.

Bishop MJ, Mayer-Pinto M, Airoldi L, Firth LB, Morris RL, Loke LHL, et al. (2017). Effects of ocean sprawl on ecological connectivity: impacts and solutions. *Journal of Experimental Marine Biology and Ecology* 492: 7–30.

Blouin MS, Parsons M, Lacaille V, Lotz S. (1996). Use of microsatellite loci to classify individuals by relatedness. *Molecular Ecology* 5: 393–401.

Bohonak AJ (1999). Dispersal, gene flow, and population structure. *Quarterly Review of Biology*74: 21–45.

Bonte D, Van Dyck H, Bullock JM, Coulon A, Delgado M, Gibbs M, et al. (2012). Costs of dispersal. *Biological Reviews* 87: 290–312.

Bouzat JL, Cheng HH, Lewin HA, Westemeier RL, Brawn JD, Paige KN (1998). Genetic evaluation of a demographic bottleneck in the greater prairie chicken. *Conservation Biology* 12: 836–843.

Brook B, Sodhi N, Bradshaw C (2008). Synergies among extinction drivers under global change. *Trends in Ecology & Evolution* 23: 453–460.

Caillaud M., Boutin M, Braendle C, Simon J. (2002). A sex-linked locus controls wing polymorphism in males of the pea aphid, *Acyrthosiphon pisum* (Harris). *Heredity* 89: 346–52.

Cain A, King M, Sheppard P (1960). New data on the genetics of polymorphism in the snail *Cepaea nemoralis* L. *Genetics* 45: 393-411.

Candy JR, Campbell NR, Grinnell MH, Beacham TD, Larson WA, Narum SR (2015). Population differentiation determined from putative neutral and divergent adaptive genetic markers in Eulachon (*Thaleichthys pacificus*, Osmeridae), an anadromous Pacific smelt. *Molecular Ecology Resources* 15: 1421–1434.

Caplins SA, Gilbert KJ, Ciotir C, Roland J, Matter SF, Keyghobadi N (2014). Landscape structure and the genetic effects of a population collapse. *Proceedings of the Royal Society B* 281: 20141798.

Charlesworth B, Charlesworth D (2017). Population genetics from 1966 to 2016. *Heredity* 118: 2–9.

Chevin L-M, Billiard S, Hospital F (2008). Hitchhiking both ways: effect of two interfering selective sweeps on linked neutral variation. *Genetics* 180: 301–316.

Clarke B (1970). Darwinian evolution of proteins. *Science* 168: 1009–1011.

Clobert J, Baguette M, Benton TG, Bullock JM (2012). Dispersal ecology and evolution. Oxford University Press, Oxford, UK.

Conner JK, Hartl DL (2004). A primer of ecological genetics. Sinauer Associates, Sunderland, USA.

Crow JF (2010). Wright and Fisher on inbreeding and random drift. *Genetics* 184: 609–611.

Dahlhoff EP, Rank NE (2000). Functional and physiological consequences of genetic variation at phosphoglucose isomerase: Heat shock protein expression is related to enzyme genotype in a montane beetle. *Proceedings of the National Academy of Sciences of the United States of America* 97: 10056–10061.

Davey JW, Blaxter ML (2010). RADSeq: next-generation population genetics. *Briefings in Functional Genomics* 9: 416–423.

Davidson TA, Audet J, Jeppesen E, Landkildehus F, Lauridsen TL, Søndergaard M, et al. (2018). Synergy between nutrients and warming enhances methane ebullition from experimental lakes. *Nature Climate Change* 8: 156–160.

de Bono M, Bargmann CI (1998). Natural variation in a neuropeptide y receptor homolog modifies social behavior and food response in *C. elegans*. *Cell* 94: 679–689.

Debinski DM, Holt RD (2000). A survey and overview of habitat fragmentation experiments. *Conservation Biology* 14: 342–355.

Derycke S, Backeljau T, Moens T (2013). Dispersal and gene flow in free-living marine nematodes. *Frontiers in Zoology* 10: 1.

DiLeo MF, Wagner HH (2016). A landscape ecologist's agenda for landscape genetics. *Current Landscape Ecology Reports* 1: 115–126.

Dobzhansky T (1955). A review of some fundamental concepts and problems of population genetics. *Cold Spring Harbor Symposia on Quantitative Biology* 20: 1–15.

Du Q, Gong C, Wang Q, Zhou D, Yang H, Pan W, et al. (2016). Genetic architecture of growth traits in *Populus* revealed by integrated quantitative trait locus (QTL) analysis and association studies. *New Phytologist* 209: 1067–1082.

Edelsparre AH, Vesterberg A, Lim JH, Anwari M, Fitzpatrick MJ (2014). Alleles underlying larval foraging behaviour influence adult dispersal in nature. *Ecology Letters* 17: 333–339.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6: e19379.

Fahrig L, Baudry J, Brotons L, Burel FG, Crist TO, Fuller RJ, et al. (2011). Functional landscape heterogeneity and animal biodiversity in agricultural landscapes: Heterogeneity and biodiversity. *Ecology Letters* 14: 101–112.

Filatov DA, Charlesworth D (1999). DNA polymorphism, haplotype structure and balancing selection in the *Leavenworthia* PgiC Locus. *Genetics* 153: 1423–1434.

Ford EB (1975). Ecological genetics. Chapman and Hall, New York, US.

Forester BR, Jones MR, Joost S, Landguth EL, Lasky JR (2016). Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Molecular Ecology* 25: 104–120.

Fownes S, Roland J (2002). Effects of meadow suitability on female behaviour in the alpine butterfly *Parnassius smintheus*. *Ecological Entomology* 27: 457–466.

Fox C, Wolf J (2006). In: Evolutionary genetics: concepts and case studies, Oxford University Press. Oxford, NY, USA.

Frankham R, Briscoe DA, Ballou JD (2010). Introduction to conservation genetics. Cambridge University Press, Cambridge.

Franks SJ, Hoffmann AA (2012). Genetics of climate change adaptation. *Annual Review of Genetics* 46: 185–208.

Franks SJ, Sim S, Weis AE (2007). Rapid evolution of flowering time by an annual plant in response to a climate fluctuation. *Proceedings of the National Academy of Sciences of the United States of America* 104: 1278–1282.

Freeland JR, Heather K, Petersen S (2011). Molecular ecology. Wiley-Blackwell, UK.

Gardner JL, Peters A, Kearney MR, Joseph L, Heinsohn R (2011). Declining body size: a third universal response to warming? *Trends in Ecology & Evolution* 26: 285–291.

Gauffre B, Estoup A, Bretagnolle V, Cosson JF (2008). Spatial genetic structure of a small rodent in a heterogeneous landscape. *Molecular Ecology* 17: 4619–4629.

Goff J, Yerke C, Keyghobadi N, Matter SF (2018). Dispersing male *Parnassius smintheus* butterflies are more strongly affected by forest matrix than are females. *Insect Science*. doi: 10.1111/1744-7917.12592.

Haag CR, Saastamoinen M, Marden JH, Hanski I (2005). A candidate locus for variation in dispersal rate in a butterfly metapopulation. *Proceedings of the Royal Society B* 272: 2449–2456.

Hamrick JL, Allard RW (1972). Microgeographical variation in allozyme frequencies in *Avena barbata*. *Proceedings of the National Academy of Sciences* 69: 2100–2104.

Hanski I (1994). A practical model of metapopulation dynamics. *Journal of Animal Ecology* 63: 151–162.

Harris H (1966). Genetics of man enzyme polymorphisms in man. *Proceedings of the Royal Society B* 164: 298–310.

Harris H, Hopkinson DA (1978). Handbook of enzyme electrophoresis in human genetics[M]. North-Holland Publishing Company, Amsterdam.

Hartl DL, Clark AG (1989). Principles of population genetics, Fourth Edition. Sinauer Associate. Sunderland, Massachusetts, USA.

Hill JK, Griffiths HM, Thomas CD (2011). Climate change and evolutionary adaptations at species' range margins. *Annual Review of Entomology* 56: 143–159.

Hoffman R. (1981). Evolutionary genetics of Metridium senile. I. Kinetic differences in phosphoglucose isomerase allozymes. *Biochemical Genetics* 19: 129–144.

Holderegger R, Kamm U, Gugerli F (2006). Adaptive vs. neutral genetic diversity: implications for landscape genetics. *Landscape Ecology* 21: 797–807.

Hughes AR, Inouye BD, Johnson MTJ, Underwood N, Vellend M (2008). Ecological consequences of genetic diversity. *Ecology Letters* 11: 609–623.

IPCC (2014). Climate Change 2013: The Physical science basis: Working group I contribution to the fifth assessment report of the international panel on climate change. Cambridge University Press, London.

Jangjoo M, Matter SF, Roland J, Keyghobadi N (2016). Connectivity rescues genetic diversity after a demographic bottleneck in a butterfly population network. *Proceedings of the National Academy of Sciences of the United States of America* 113: 10914–10919.

Jangjoo M, Matter SF, Roland J, Keyghobadi N. Demographic fluctuations lead to rapid and cyclic shifts in genetic structure among populations of an alpine butterfly, (*Parnassius smintheus*), (submitted to *Heredity*).

Jones CM, Papanicolaou A, Mironidis GK, Vontas J, Yang Y, Lim KS, et al. (2015). Genomewide transcriptional signatures of migratory flight activity in a globally invasive insect pest. *Molecular Ecology* 24: 4901–4911.

Keyghobadi N (2007). The genetic implications of habitat fragmentation for animals. *Canadian Journal of Zoology* 85: 1049–1064.

Keyghobadi N, Roland J, Strobeck C (1999). Influence of landscape on the population genetic structure of the alpine butterfly *Parnassius smintheus* (Papilionidae). *Molecular Ecology* 8: 1481–1495.

Keyghobadi N, Roland J, Strobeck C (2005). Genetic differentiation and gene flow among populations of the alpine butterfly, *Parnassius smintheus*, vary with landscape connectivity. *Molecular Ecology* 14: 1897–1909.

Kimura M, Crow JF (1964). The number of alleles that can be maintained in a finite population. *Genetics* 49: 725–738.

Kliman, R, Sheehy B, Schultz J (2008). Genetic drift and effective population size. *Nature Education* 1: 3.

Kreitman M (1983). Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. *Nature* 304: 412–417.

Kvist J, Mattila ALK, Somervuo P, Ahola V, Koskinen P, Paulin L, et al. (2015). Flight-induced changes in gene expression in the Glanville fritillary butterfly. *Molecular Ecology* 24: 4886–4900.

Lamotte M (1959). Polymorphism of natural populations of *Cepaea nemoralis*. *Cold Spring Harbor Symposia on Quantitative Biology* 24: 65–86.

Lande R, Shannon S (1996). The role of genetic variation in adaptation and population persistence in a changing environment. *Evolution* 50: 434–437.

Levins R (1969). Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the Entomological Society of America* 15: 237–240.

Lewontin, RC, Hubby JL (1966). A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophzla pseudoobscura*. *Genetics* 54: 595–609.

Li YC, Korol AB, Fahima T, Beiles A, Nevo E (2002). Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review: Microsatellite evolution. *Molecular Ecology* 11: 2453–2465.

Loenen WAM, Dryden DTF, Raleigh EA, Wilson GG, Murray NE (2014). Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic Acids Research* 42: 3–19.

Lowe CD, Montagnes DJS, Martin LE, Watts PC (2010). High genetic diversity and fine-scale spatial structure in the marine flagellate *Oxyrrhis marina* (Dinophyceae) uncovered by Microsatellite loci. *PLoS ONE* 5: e15557.

Mahamdallie SS, Ready PD (2012). No recent adaptive selection on the apyrase of Mediterranean *Phlebotomus*: implications for using salivary peptides to vaccinate against canine leishmaniasis. *Evolutionary Applications* 5: 293–305.

Manel S, Holderegger R (2013). Ten years of landscape genetics. *Trends in Ecology & Evolution* 28: 614–621.

Manel S, Joost S, Epperson BK, Holderegger R, Storfer A, Rosenberg MS, et al. (2010). Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Molecular Ecology* 19: 3760–3772.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003). Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution* 18: 189–197.

Margotta J, Mancinelli G, Benito A, Ammons A, Roberts S, Elekonich M (2012). Effects of flight on gene expression and aging in the honey bee brain and flight muscle. *Insects* 4: 9–30.

Masel J (2011). Genetic drift. *Current Biology* 21: R837–R838.

Matter SF, Doyle A, Illerbrun K, Wheeler J, Roland J (2011). An assessment of direct and indirect effects of climate change for populations of the Rocky Mountain Apollo butterfly (*Parnassius smintheus* Doubleday): Effects of climate change for *P. smintheus*. *Insect Science* 18: 385–392.

Matter SF, Keyghobadi N, Roland J (2014). Ten years of abundance data within a spatial population network of the alpine butterfly, *Parnassius smintheus*. *Ecology* 95: 2985–2985.

Matter SF, Roland J, Moilanen A, Hanski I (2004). Migration and survival of *Parnassius smintheus*: Detecting effects of habitat for individual butterflies. *Ecological Applications* 14: 1526–1534.

Metcalf JL, Turney C, Barnett R, Martin F, Bray SC, Vilstrup JT, et al. (2016). Synergistic roles of climate warming and human occupation in *Patagonian megafaunal* extinctions during the last deglaciation. *Science Advances* 2: e1501682–e1501682.

Meudt HM, Clarke AC (2007). Almost forgotten or latest practice? AFLP applications, analyses and advances. *Trends in Plant Science* 12: 106–117.

Muller HJ (1950). Our load of mutations. *American Journal of Human Genetics* 2: 111–176.

Narum SR, Hess JE (2011). Comparison of F_{ST} outlier tests for SNP loci under selection. *Molecular Ecology Resources* 11: 184–194.

Niitepold K (2010). Genotype by temperature interactions in the metabolic rate of the Glanville fritillary butterfly. *Journal of Experimental Biology* 213: 1042–1048.

Niitepõld K, Smith AD, Osborne JL, Reynolds DR, Carreck N, Martin AP, et al. (2009). Flight metabolic rate and *Pgi* genotype influence butterfly dispersal rate in the field. *Ecology* 90: 2223–2232.

Orsini L, Wheat CW, Haag CR, Kvist J, Frilander MJ, Hanski I (2009). Fitness differences associated with *Pgi* SNP genotypes in the Glanville fritillary butterfly (*Melitaea cinxia*): Reduced fitness in a butterfly metapopulation. *Journal of Evolutionary Biology* 22: 367–375.

Ortego J, Aguirre MP, Noguerales V, Cordero PJ (2015). Consequences of extensive habitat fragmentation in landscape-level patterns of genetic diversity and structure in the Mediterranean esparto grasshopper. *Evolutionary Applications* 8: 621–632.

Parmesan C, Root TL, Willig MR (2000). Impacts of extreme weather and climate on Terrestrial Biota. *Bulletin of the American Meteorological Society* 81: 443–450.

Pauls SU, Nowak C, Bálint M, Pfenninger M (2013). The impact of global climate change on genetic diversity within populations and species. *Molecular Ecology* 22: 925–946.

Pélabon C, Hansen TF, Carter AJR, Houle D (2010). Evolution of variation and variability under fluctuating, stabilizing, and disruptive selection. *Evolution* 64: 1912–1925.

Pianka E. (2000). Evolutionary Ecology. Harper and Row, New York.

Rao VR, Hodgkin T (2002). Genetic diversity and conservation and utilization of plant genetic resources. *Plant Cell, Tissue and Organ Culture* 68: 1–19.

Rich SS, Bell AE, Wilson SP (1979). Genetic drift in small populations of *Tribolium*. *Evolution* 33: 579-584.

Roff DA (1986). The evolution of wing dimorphism in insects. *Evolution* 40: 1009–1020.

Roff DA, Fairbairn DJ (2001). The genetic basis of dispersal and migration, and its consequences for correlated traits. Oxford University Press, Oxford.

Roff DA, Fairbairn DJ (2007). The evolution and genetics of migration in insects. *BioScience* 57: 155–164.

Roland J, Keyghobadi N, Fownes S (2000). Alpine *Parnassius* butterfly dispersal: Effects of landscape and population size. *Ecology* 81: 1642–1653.

Roland J, Matter SF (2007). Encroaching forests decouple alpine butterfly population dynamics. *Proceedings of the National Academy of Sciences of the United States of America* 104: 13702–13704.

Roland J, Matter SF (2013). Variability in winter climate and winter extremes reduces population growth of an alpine butterfly. *Ecology* 94: 190–199.

Roland J, Matter SF (2016). Pivotal effect of early-winter temperatures and snowfall on population growth of alpine *Parnassius smintheus* butterflies. *Ecological Monographs* 86: 412–428.

Ronce O (2007). How does it feel to be like a rolling stone? Ten questions about dispersal evolution. *Annual Review of Ecology, Evolution, and Systematics* 38: 231–253.

Ross JA, Matter SF, Roland J (2005). Edge avoidance and movement of the butterfly *Parnassius smintheus* in matrix and non-matrix habitat. *Landscape Ecology* 20: 127–135.

Rowan BA, Weigel D, Koenig D (2011). Developmental genetics and new sequencing technologies: The rise of non-model organisms. *Developmental Cell* 21: 65–76.

Saastamoinen M, Bocedi G, Cote J, Legrand D, Guillaume F, Wheat CW, et al. (2018). Genetics of dispersal. *Genetic of dispersal. Biological Reviews* 93: 574–599.

Saccheri I, Kuussaari M, Kankare M, Vikman P, Fortelius W, Hanski I (1998). Inbreeding and extinction in a butterfly metapopulation. *Nature* 392: 491–494.

Sala O., Chapin FS, Armesto J, Berlow E, Bloomfield J, Dirzo R, et al. (2000). Global Biodiversity Scenarios for the Year 2100. *Science,* 287: 1770–1774.

Schierenbeck KA (2017). Population-level genetic variation and climate change in a biodiversity hotspot. *Annals of Botany* 119: 215–228.

Schmitt MW, Kennedy SR, Salk JJ, Fox EJ, Hiatt JB, Loeb LA (2012). Detection of ultra-rare mutations by next-generation sequencing. *Proceedings of the National Academy of Sciences* 109: 14508–14513.

Schwander T, Leimar O (2011). Genes as leaders and followers in evolution. *Trends in Ecology & Evolution* 26: 143–151.

Slatkin M (1987). Gene flow and the geographic structure of natural populations. *Science* 236: 787–792.

Sokolowski MB (1980). Foraging strategies of *Drosophila melanogaster*: A chromosomal analysis. *Behavior Genetics* 10: 291–302.

Somervuo P, Kvist J, Ikonen S, Auvinen P, Paulin L, Koskinen P, et al. (2014). Transcriptome Analysis Reveals Signature of Adaptation to Landscape Fragmentation. *PLoS ONE* 9: e101467.

Spielman D, Brook D., Frankham R (2004). Most species are not driven to extinction before genetic factors impact them. *Proceedings of the National Academy of Sciences of the United States of America* 101: 15261–15264.

Takami Y, Koshio C, Ishii M, Fujii H, Hidaka T, Shimizu I (2004). Genetic diversity and structure of urban populations of Pieris butterflies assessed using amplified fragment length polymorphism: Population genetics of urban butterflies. *Molecular Ecology* 13: 245–258.

Thomas CD, Bodsworth EJ, Wilson RJ, Simmons AD, Davies ZG, Musche M, et al. (2001). Ecological and evolutionary processes at expanding range margins. *Nature* 411: 577–581.

Tischendorf L, Fahrig L (2000). On the usage and measurement of landscape connectivity. *Oikos* 90: 7–19.

Travis JMJ, Delgado M, Bocedi G, Baguette M, Bartoń K, Bonte D, et al. (2013). Dispersal and species' responses to climate change. *Oikos* 122: 1532–1540.

Turner MG (1989). Landscape ecology: The effect of pattern on process. *Annual Review of Ecology, Evolution, and Systematics* 20: 171–197.

Turner MG (1990). Spatial and temporal analysis of landscape patterns. *Landscape Ecology* 4: 21–30.

Young A, Boyle T, Brown T (1996). The population genetic consequences of habitat fragmentation for plants. *Trends in Ecology and Evolution* 11: 413–418.

Wagner AP, Creel S, Kalinowski ST (2006). Estimating relatedness and relationships using microsatellite loci with null alleles. *Heredity* 97: 336–345.

Watt WB (1977). Adaptation at specific loci. I. Natural selection on phosphoglucose isomerase of *Colias* butterflies: biochemical and population aspects. *Genetics* 87: 177-194.

Watt WB (1977). Adaptation at specific loci. I. Natural selection on phosphoglucose isomerase of *Colias* butterflies: biochemical and population aspects. *Genetics* 87: 177-194.

Watt WB (1983). Adaptation at specific loci. II. Demographic and biochemical elements in the maintenance of the *Colias* PGI polymorphism. *Genetics* 103: 691–724.

Watt WB (2003). Mechanistic studies of butterfly adaptation. In Butterflies: Ecology and evolution taking flight. University of Chicago, Chicago.

Watt WB, Cassin RC, Swan (1983). Adaptation at specific loci. III. Field behavior and survivorship differences among *Colias Pgi* genotypes are predictable from in vitro biochemistry. *Genetics* 103: 725–739.

Wheat CW (2010). Phosphoglucose isomerase (*Pgi*) performance and fitness effects among Arthropods and its potential role as an adaptive marker in conservation genetics. *Conservation Genetics* 11: 387–397.

Wheat CW, Haag CR, Marden JH, Hanski I, Frilander MJ (2010). Nucleotide polymorphism at a gene (*Pgi*) under balancing selection in a butterfly metapopulation. *Molecular Biology and Evolution* 27: 267–281.

Wheat CW, Hill J (2014). *Pgi*: the ongoing saga of a candidate gene. *Current Opinion in Insect Science* 4: 42–47.

Wheat CW, Watt WB, Pollock DD, Schulte PM (2006). From DNA to fitness differences: sequences and structures of adaptive variants of *Colias* phosphoglucose isomerase (PGI). *Molecular Biology and Evolution* 23: 499–512.

Whitehead A, Crawford DL (2006). Neutral and adaptive variation in gene expression. *Proceedings of the National Academy of Sciences of the United States of America* 103: 5425–5430.

Witkowski Z, Adamski P, Kosior A, Plonka PP (1997). Extinction and reintroduction of *Parnassius apollo* in the Pieniny National Park (Polish Carpathians). *Biologia* 52: 199–208.

Zera AJ, Brisson JA (2012). Quantitative, physiological, and molecular genetics of dispersal/migration. In Dispersal Ecology and Evolution. Oxford University Press, Oxford pp 63–82.

Chapter 2

2 Connectivity rescues genetic diversity after a demographic bottleneck: empirical evidence from a butterfly population network[1]

2.1 Introduction

Genetic diversity is the most fundamental level of biological diversity. Loss of genetic diversity is a central concern in conservation biology because populations with low genetic diversity may suffer from inbreeding and reduced fitness, lack the potential to adapt to future environmental change, and be more vulnerable to extinction (Saccheri et al. 1998; Spielman et al. 2004). Genetic diversity can be lost from populations through various mechanisms, with random drift in finite populations and demographic bottlenecks (temporary but severe reductions in population size), being of greatest relevance in conservation (Lacy 1987; Bouzat et al. 1998).

Immigration into a genetically impoverished population can rescue genetic diversity, and can be achieved artificially through translocations or through natural movement of individuals (Ehrich and Jorde 2005; Frankham 2015). Natural immigration requires connectivity within the landscape, where connectivity measures the extent to which movement and gene flow can occur among populations (Tischendorf and Fahrig

---

[1] A version of this chapter has been published in Proceedings of the National Academy of Sciences of the United States of America

Citation: Jangjoo M, Matter SF, Roland J, Keyghobadi N (2016). Connectivity rescues genetic diversity after a demographic bottleneck in a butterfly population network. *PNAS* 113: 10914–10919.

2000). Connectivity can be defined at the level of the landscape or individual habitat patches, and is a function of structural elements of the landscape in combination with the movement behavior of individual species (Hanski 1994; Tischendorf and Fahrig 2000).

There is considerable interest in managing landscapes to improve connectivity among natural populations as this provides a variety of ecological and genetic benefits (Luque et al. 2012), including the potential for natural genetic rescue. Correlations between connectivity and genetic diversity shown in numerous systems suggest that connectivity contributes to maintenance of genetic diversity on some time scale (Keyghobadi et al. 2005; Vandewoestijne et al. 2008). However, the temporal scales involved in the establishment of such correlations are poorly understood. Although predicted by theory, the ability of connectivity to rescue genetic diversity rapidly after a demographic bottleneck has not, to our knowledge, been demonstrated in a natural system. Consequently, the extent to which connectivity may contribute directly to genetic diversity via immigration of novel alleles, versus indirectly via effects on population size and stability is also not well understood.

A network of populations of the Rocky Mountain Apollo butterfly, *Parnassius smintheus*, occupying patches of alpine meadow habitat in Alberta, Canada has been monitored and studied continuously since 1995 and effects of both landscape structure and climate variation on population dynamics and dispersal have been described (Roland et al. 2000; Roland and Matter 2013) (Figure 1.1). Population dynamics of *P. smintheus* are influenced by climate variation, with the Pacific Decadal Oscillation (PDO) index being a strong descriptor of annual population growth. More frequent extremely cold or warm winters, which can be expected as a result of climate change, are predicted to

increase the occurrence of years with negative population growth for *P. smintheus* (Roland and Matter 2013).

In this study site, two severe demographic bottlenecks have been documented and linked to poor overwintering weather conditions: in 2003 and 2010 (Figure 2.1). The bottleneck that began in 2010 was more protracted with population sizes remaining low in 2011 and recovering more slowly than after the 2003 bottleneck (Figure 2.1). The reason for the differing nature of these two events may be more severe overwintering conditions in 2010, leading to higher egg mortality (Roland and Matter 2013). A study comparing genetic diversity prior to and after the 2003 demographic bottleneck (Caplins et al. 2014) yielded a key result: no overall loss of diversity across the network was detected, but an interaction between patch connectivity and severity of the demographic collapse affected loss of allelic richness within individual habitat patches. This result suggested some role of connectivity in maintaining genetic diversity in populations experiencing demographic bottlenecks**.**

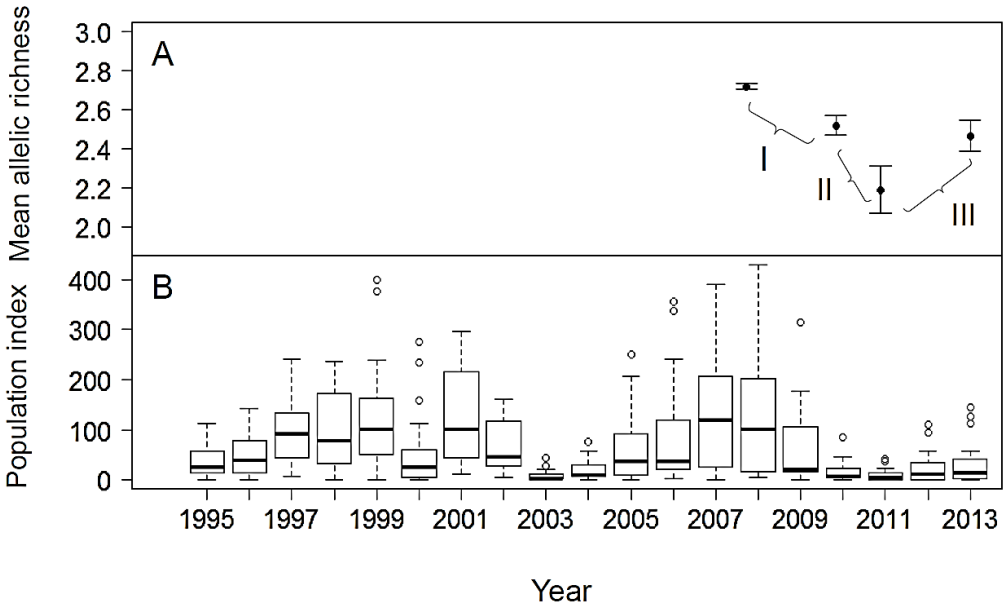


Figure 2.1 Changes in *Parnassius smintheus* population size and allelic richness (A_R) in the network over different years. (A) Mean A_R over seven microsatellite loci in 2008, 2010, 2011 and 2013 (with rarefaction to four genes). Significant predictors of A_R change were: 2010 population size during phase (I), none during phase (II), and 2012 connectivity during phase (III). (B) Boxplots of yearly *P. smintheus* abundance estimates for all populations, showing interquartile range (IQR; boxes), maximum and minimum estimates up to 1.5 × IQR (whiskers), and outliers beyond 1.5 × IQR (open circles).

Here, I assess effects of the more protracted 2010 demographic bottleneck on genetic diversity by comparing samples collected prior to (in 2008), during (in 2010 and 2011) and after (in 2013) the event. I show that across this bottleneck, connectivity plays a clear and significant role in recovery of genetic diversity, highlighting the importance of conserving connectivity in fluctuating populations.

2.2 Materials and Methods

2.2.1 Sample collection and study region

Wing-clips from adult *P. smintheus* have been collected since 1995, and yearly since 2005, from populations along Jumpingpound Ridge, in the Kananaskis region of Alberta, Canada (50° 57'N, 114° 54'W; Figure 1.1). Here, *P. smintheus* occupies meadows above treeline (~2100 m), and Roland et al. (2000) delineated habitat patches that range in area from 0.2 ha to 22.7 ha and are separated from each other by either intervening forest or open meadow habitat. Butterflies were captured with hand-nets, individually marked, and approximately 0.2 cm^2 of wing tissue was removed using forceps or iris scissors and stored immediately in 95-100% ethanol. To assess genetic diversity before and after the 2010 demographic bottleneck, I focused on individuals sampled from the same patches in 2008 (pre-bottleneck) and 2013 (post-bottleneck). These are the years closest to the bottleneck in which larger samples from several patches were available (Figure 2.1A). Supporting analyses also included individuals sampled during the low population size years of 2010 and 2011. The number of individuals sampled per population was by

necessity small in the bottleneck years because of the low population sizes in those years. However, an equal or even larger proportion of the total population was sampled in these years, as well as in 2013, compared to 2008, as reflected in the ratios of sample size to index of population size for the different years (Table A2).

2.2.2 Mark-recapture study and estimates of patch connectivity

Mark-recapture studies of *P. smintheus* have been conducted in the population network since 1995, and are described extensively elsewhere (Roland et al. 2000; Roland and Matter 2013). Adults were individually marked, and spatial locations of captures and recaptures were recorded. An index of population size in each habitat patch was determined using Craig's method, which provides an estimate of the number of adults in the population on a single day of sampling (Craig 1953; Matter and Roland 2004). I used the maximum Craig's estimate from three to five different sampling days per year as an index of population size in each patch, each year. Rates of movement among patches were estimated with the virtual migration model (VMM) (Hanski et al. 2000).

I defined the connectivity of patch *k* as its relative attractiveness and accessibility to emigrants from all other patches in the network. This was estimated as the sum, over all other patches, of the probabilities of individuals leaving each other patch and reaching *k* (Hanski et al. 2000):

$$\psi_k = \sum_{\substack{j=1 \\ j \neq k}}^{n} \frac{A_k^{\zeta} e^{(-\alpha_f d_{jkf} - \alpha_m d_{jkm})}}{\frac{\lambda}{S_j} + S_j}$$

where

$$S_j = \sum_{\substack{k=1 \\ k \neq j}}^{n} A_k^{\zeta} e^{(-\alpha_f d_{jkf} - \alpha_m d_{jkm})}$$

A_k is the area of patch *k*; d_{jkf} and d_{jkm} are the distances through forest and meadow habitat between patches *j* and *k*, respectively; and λ is mortality during dispersal. Additional parameters describe the effect of forest and meadow on movement (α_f and α_m, respectively), and the scaling of immigration with patch area (ζ). Connectivity for each patch, each year was calculated using parameters estimated from the mark-recapture data for that year. Since movements of *P. smintheus* are restricted along ridge-tops (Roland et al. 2000), I calculated pairwise distances between patches along ridge-tops between centroids of butterfly capture in each patch (Figure 1.1). The total distance between any two patches was divided into two components, the distance comprised of forest and distance comprised of open meadow, which were estimated from digitized aerial photos.

2.2.3 DNA extraction and Microsatellite analysis

DNA was extracted from wing-clips using a DNeasy Blood and Tissue Kit (QIAgen, Germantown, MD), with a final elution volume of 200 µl. Each sample was genotyped at seven highly variable microsatellite loci ((Ps50, Ps76, Ps81, Ps85, Ps163, Ps165 and Ps262; Keyghobadi et al. 1999, 2002). PCR amplification of microsatellites occurred in two multiplex amplifications (multiplex 1: Ps 50, Ps 81and Ps 85, and multiplex 2: Ps 76 and Ps 163), and two individual locus amplifications (Ps 262 and Ps 165). PCR reactions occurred in a final solution of volume 10µL; each amplification contained 1× AmpliTaq buffer (10 mM Tris, pH 8.8, 0.1% Triton X100, 50 mM KCl), 3.125 mM $MgCl_2$, 0.075 to

0.275 µM of each primer, 0.25 mM dNTP, 0.0625 U of AmpliTaq DNA polymerase (Applied Biosystems, Foster City, CA), 0.15 µg bovine serum albumin (BSA) and 3µl of genomic DNA. PCR amplifications were performed in a PTC 0200 DNA Engine Cycler (BioRad, Hercules, CA). One of the two primers (forward primer) was labeled with a fluorescent dye to allow visualization of PCR products.

Thermal cycling profiles followed one of two protocols: 1) multiplex amplification: Denaturation for 60 s at 94 °C; followed by 3 cycles at of 30 s at 94°C, 30 s at 56°C annealing, and 30 s at 72°C; 10 touchdown (TD) cycles when annealing temperature was reduced 0.5 °C per cycle and all hold times were reduced to 15 s; 27 additional cycles of 15 s at 94°C, 15 s at 51°C, and 15 s at 72°C; and final elongation at 72°C for 180 s.; and 2) individual locus amplification: Denaturation for 60 s at 94 °C; followed by 3 cycles of 30 s at 94°C; 20 s at 54°C, and 10 s at 72°C; followed by 32 cycles of 15 s at 94°C; 20 s at 54°C; 5 s at 72°C; and final elongation at 72°C for 30s. Ramp speed was set to 1 °C per second for all thermal cycling.

PCR products were visualized and sized on an Applied Biosystems® 3730S capillary DNA analyzer, using LIZ-500 size standard. All loci (the PCR products) for each individual were multi-loaded in a single lane of the DNA analyzer. Electropherograms generated by the DNA analyzer were viewed and processed using GeneMapper software ver. 4.0 (Applied Biosystems) to score microsatellite genotypes. All genotypes were checked manually and loci that failed to amplify were re-run up to two more times. If a locus failed to amplify in an individual after three attempts, the individual was considered null homozygous for that locus. However, any individuals with two or more failed loci were removed altogether from the dataset.

2.2.4 Linkage disequilibrium and Hardy–Weinberg tests

For each of the four years separately (2008, 2010, 2011, 2013), linkage disequilibrium and conformity to Hardy-Weinberg proportions were tested for each locus in each population using Genepop v.4.2 (Raymond and Rousset 1995). For linkage tests, significance was assessed using a Markov chain method of 100 batches of 1,000 iterations per batch. Hardy-Weinberg tests used the Markov chain method and approximation of Fisher's exact test implemented in Genepop (Guo and Thompson 1992). Consistent with previous analyses of these loci (Keyghobadi et al. 1999, 2002, 2005; Caplins et al. 2014) there was no evidence for linkage disequilibrium but there were significant deviations from expected Hardy-Weinberg genotypic proportions. Of a total of 203 tests of conformity to Hardy-Weinberg proportions, 115 indicated significant homozygote excess, which was observed at all loci and in each time period (in 46 of 63 tests for 2008, 21 of 49 tests for 2010, 16 of 28 tests for 2011, and 32 of 63 tests for 2013). Homozygote excess at these loci is known to be a result of null alleles (Keyghobadi et al. 1999, 2002), which are non-amplifying alleles that result from variation in microsatellite flanking regions. Null allele frequencies were estimated, and frequencies of other alleles simultaneously re-estimated, using the 'ENA' method in the software FreeNA (Chapuis and Estoup 2007).

2.2.5 Changes in genetic diversity

For each year, within-population genetic diversity was quantified using two metrics: 1) unbiased expected heterozygosity (H_E) calculated using null-corrected allele frequencies

for each locus (Nei and Roychoudhury 1974) and then averaged over all loci in each population; and 2) allelic richness (A_R) estimated in HP-Rare software (Petit et al. 1998; Kalinowski 2005), also averaged over all loci in each population. Allelic richness is a count of visible alleles at each locus, corrected for the number of sampled gene copies by rarefaction to the smallest sample size in the dataset (Leberg 2002), and is expected to show a stronger response to demographic bottlenecks than heterozygosity (Nei et al. 1975). Allelic richness also allows for robust comparisons of genetic diversity despite very unequal sample sizes (Pruett and Winker 2008). Because null allele frequency was consistent between sampling periods (mean of 10.8% in 2008 and 7.9% in 2013, with overlapping confidence intervals; Table A1), the presence of null alleles should not affect temporal changes in allelic richness, which are estimated using the visible alleles (Chapuis et al. 2008). For comparisons between 2008 and 2013, I focused my analyses on the nine populations in which a minimum of five individuals were sampled in each year (Table A1), thus allowing us to estimate A_R with rarefaction to ten gene copies. For supplementary analyses involving samples from the bottleneck years of 2010 and 2011, the number of available samples was necessarily very small (Table A3). Inclusion of only those populations in which I could rarefy to ten, or even as few as six, gene copies left me with very few populations (two or three for some pairs of years) to examine effects of connectivity and crash severity on A_R change. Rarefaction to two gene copies allowed me to include more populations, but rarefaction to this small number of samples produced high variability in the A_R estimates and patterns of change that were not consistent with those detected when rarefaction was to four or more gene copies. Therefore, for the supplementary analyses involving 2010 and 2011 samples, I included populations with two or more individuals sampled in those years, and estimated A_R for all years based on

rarefaction to four gene copies. In comparisons of A_R between 2008 and 2013, rarefaction to four or ten gene copies gave highly consistent results although linear models for A_R change had slightly less explanatory power with rarefaction to four gene copies. I used the nonparametric Wilcoxon-Mann-Whitney test to determine whether levels of genetic diversity differed between years.

I tested for evidence of genetic bottleneck signatures, separately for each population, using the software BOTTLENECK v.1.2.02 (Cornuet and Luikart 1997; Piry et al. 1999). Wilcoxon signed-rank tests (WSR) were used to compare the observed heterozygosity to that expected from the observed number of alleles, given the sample size, under the assumption of mutation-drift equilibrium, for each locus in each population. The infinite allele model (IAM) and the two-phase model (TPM) of mutation were used to simulate mutation-drift equilibrium. For TPM, two values (10 and 30) were tested for the variance of the geometric distribution with a low probability of single-step mutations (70%). I did not include the strictly stepwise mutational model (SMM), because it is inappropriate here due to the occurrence of flanking-sequence insertions or deletions (Keyghobadi et al. 1999, 2002).

2.2.6 Relationship to patch connectivity and severity of the demographic bottleneck

I examined whether the level of genetic diversity within populations before the bottleneck (in 2008) affected bottleneck severity, by separately testing each of A_R and H_E as predictors of the 2010 and 2011 population size indices, using linear regression.

To examine factors affecting changes in allelic richness between years, I first quantified the proportional loss or gain of A_R between time periods. To improve interpretability of the results and minimize reference to negative changes in A_R, I quantified changes in A_R between years as either proportional loss or gain of A_R depending on whether allelic diversity, on average across all populations, decreased or increased between the time periods considered. For pairs of years where A_R had, on average, declined between the two time periods I estimated proportional loss of A_R as:

$$\frac{(A_R \text{ in first period} - A_R \text{ in second period})}{A_R \text{ in first period}}$$

Where A_R had, on average, increased between the two time periods I estimated proportional gain of A_R as:

$$\frac{(A_R \text{ in second period} - A_R \text{ in first period})}{A_R \text{ in first period}}$$

I examined the effect of connectivity on A_R change using linear regression. I also included severity of the demographic bottleneck and the two-way interaction term as predictors in my models. Each population was considered a random effect. Proportional changes in A_R were arcsine transformed before being included as response variables in the regression analyses. For changes in A_R between 2008 and 2013, I tested separately the effects of connectivity in each of the years 2010, 2011 and 2012. Since I was interested in whether connectivity facilitated genetic rescue after the demographic bottleneck, it was most relevant to use connectivity estimated for these years, starting with the initial year in which population size crashed (2010) and up to the year before the 2013 samples were collected. The population size indices in the two years of lowest abundance (2010 and 2011) were used separately as measures of the severity of the demographic bottleneck in

each population. I therefore tested models that included all possible combinations of connectivity for one of the three years, and bottleneck severity for one of the two years, and compared the performance of models based on the corrected Akaike Information Criterion (AICc) (Table A4). For supplementary analyses examining changes in A_R in the distinct phases of demographic decline (2008 to 2010) and recovery (2011 to 2013), the same sets of predictors were used. While I would not expect causal relationships between some of the predictors and changes in A_R during these phases (e.g., I would not expect an effect of 2012 connectivity on A_R change between 2008 and 2010), I nonetheless examined all models so that I could have more confidence in the interpretation of factors affecting A_R change across the entire period of 2008 to 2013; for example, I wanted to confirm that the significant effect of 2012 connectivity on A_R change from 2008 to 2013 was indeed via immigration during the recovery stage and did not represent an artefact of processes operating during the decline phase (Table A5). The predictors used in each model were not collinear ($r^2 < 0.4$). All linear regressions were performed using the 'lm' function of the 'Stats' package in R v.3.1.2 (R Development Core Team 2013). I confirmed that model residuals were not spatially autocorrelated using Moran's I (all $P > 0.05$) executed in the 'ape' package in R, based on the coordinates of the centroid of each sampled patch.

2.3 Results and Discussion

I assessed genetic diversity using a panel of seven highly variable microsatellite loci. Prior to the demographic bottleneck, I observed a total of 132 different alleles across all these loci, and 123 cases of an allele within a patch having an observed frequency below 2%. Thus, with a large number of alleles per locus, these markers would have been particularly sensitive to changes in allelic diversity (Spencer et al. 2000). I first established there were no effects of pre-bottleneck allelic richness (A_R) or expected heterozygosity (H_E) on 2010 or 2011 population size, indicating that levels of neutral genetic diversity did not predict or determine the severity of the demographic decline in each patch (2010: A_R, P = 0. 0.33, H_E, P=0.86; 2011: A_R, P =0.39, H_E, P = 0.95).

Allelic diversity across the population network declined after the 2010 demographic bottleneck. Averaging over all patches and loci, A_R (rarefied to ten genes) was reduced significantly from 4.76 (S.E, ± 0.026) in 2008 to 4.08 (S.E, ± 0.17) in 2013 (Wilcoxon-Mann-Whitney test, W= 77.5, P = 0.001). This represents a mean loss of 14% (0.14 ± 0.039; Table A1) of the allelic richness present before the bottleneck. These results contrast with the 2003 demographic bottleneck where no overall loss of allelic diversity across the network occurred (Caplins et al. 2014), and reflect differences in duration and recovery from the two events; a longer duration at low abundance has a stronger negative effect on genetic diversity (Williamson-Natesan 2005). Expected heterozygosity also declined after the demographic bottleneck, from a mean of 0.71 (S.E, ± 0.005) across patches in 2008 to 0.67 (S.E, ± 0.016) in 2013, although the difference was non-significant (Wilcoxon-Mann-Whitney test, W= 62, P = 0.061). The weaker response of heterozygosity is consistent with theoretical expectations that allelic diversity

should respond more strongly and rapidly to a demographic bottleneck than heterozygosity (Nei et al. 1975).

Despite the clearly documented demographic bottlenecks in this system, and the significant reduction in allelic diversity across the 2010 event, significant signatures of genetic bottlenecks using the program BOTTLENECK were not detected in samples collected before or after the 2010 event, using either the infinite allele model or two-phase model (all $P > 0.05$ for Wilcoxon tests). The low power of single sampling-period bottleneck detection methods has been previously noted (Berthier et al. 2005; Peery et al. 2012), and in my study system, the effectiveness of such methods may be particularly limited by immigration, which can erase a genetic bottleneck signature in two to three generations (Keller et al. 2001; Busch et al. 2007).

I observed substantial variation among populations occupying different habitat patches in the amount of allelic diversity lost from 2008 to 2013, with the proportional loss of A_R (averaged across loci) ranging from -0.03 (A_R increased slightly in one site) to 0.31 (31% of A_R lost). I examined the ability of patch connectivity to explain these changes in A_R using a connectivity measure that accounts for both landscape structure (areas and distances between patches, and the nature of the intervening matrix) and movement parameters estimated from mark-recapture data. Severity of the demographic bottleneck, measured as population size during the lowest abundance years of 2010 or 2011, and the interaction between connectivity and bottleneck severity were included in the models analyzed. A measure of patch connectivity that was based on movement parameters from 2012 was the single best predictor of change in A_R from 2008 to 2013 ($r^2 = 0.81$, $F_{1, 7} = 30.38$, $P = 0.0009$). Populations in patches with greater connectivity

retained more allelic diversity through the demographic bottleneck (Figure 2.2). There were no significant effects of bottleneck severity, measured as either 2010 or 2011 population size, on the loss of A_R. Compared to the 2003 demographic bottleneck therefore, where the effect of connectivity on loss of A_R was complicated by an interaction with severity of local population size decline (Caplins et al. 2014), across this more protracted event I observed a very distinct and clear effect of connectivity on the loss of A_R.

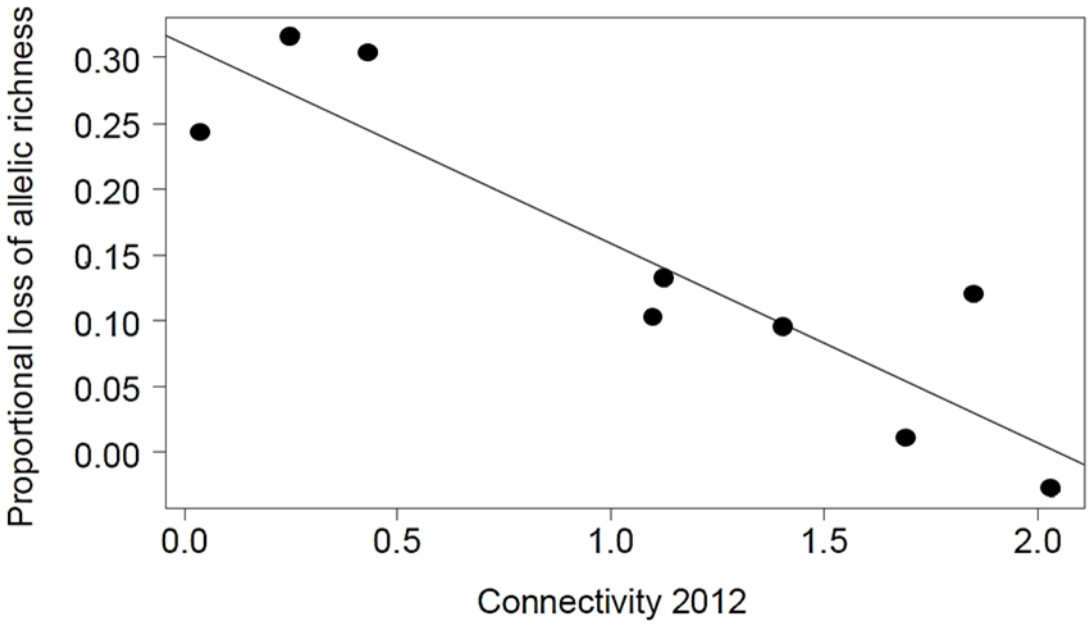


Figure 2.2 Relationship between the proportional loss of allelic richness (A_R) across a demographic bottleneck and habitat patch connectivity in *Parnassius smintheus*. Proportional loss of allelic richness was measured from 2008 and 2013, and was best explained by connectivity in 2012. Solid dots indicate individual patches. Least-square line of best fit is shown.

Connectivity in 2012 was also a stronger predictor of A_R change from 2008 to 2013 than connectivity estimated using movement parameters from either 2010 (r^2 = 0.62, $F_{1, 7}$ = 11.49, P = 0.011) or 2011 (r^2 = 0.71, $F_{1, 7}$ = 17.17, P = 0.004). This result supports the hypothesis that the effect of connectivity on A_R change was via facilitation of immigration into patches, since this process should be acting most strongly during the demographic recovery phase (2011 to 2013) than during the initial decline (2008 to 2010). To examine this further, and to characterize the behaviour of A_R through the demographic bottleneck, I genotyped samples collected in 2010 and 2011; sample sizes in these years were necessarily small, therefore a smaller number of patches and fewer individuals per patch could be analyzed. I then examined effects of patch connectivity and population size in various years, and their interactions, on changes in A_R (rarefied to four gene copies because of smaller sample sizes in 2010 and 2011) during the demographic decline phase and the recovery phase, separately.

Mean A_R declined from 2008 to 2010, accompanying the demographic bottleneck. Mean A_R dropped even further in 2011 with the continued time at low population size, and then increased from 2011 to 2013 as population sizes recovered, although not fully returning to pre-bottleneck (2008) levels (Figure 2.1). Using the same number of populations as in my comparisons between 2008 and 2013, I found that loss of A_R from 2008 to 2010 was best explained by population size in 2010 (r^2 = 0.57, $F_{1, 7}$ = 9.38, P = 0.018; Figure 2.1). This indicates that severity of the demographic bottleneck, but not connectivity, determined the loss of genetic diversity during the demographic decline phase. In contrast, the increase in A_R from 2011 to 2013 was best explained by connectivity in 2012 (r^2 = 0.91, $F_{1, 2}$ = 19.46, P = 0.047; Figure 2.1). While this latter analysis of A_R gain was based on only four populations, overall my results do suggest that

the effect of connectivity on A_R change observed across the entire demographic bottleneck (2008 to 2013) reflects the importance of connectivity during the recovery phase, rather than through any effect of connectivity on the loss of A_R during the initial population decline. The results are therefore consistent with the hypothesis that connectivity rescued genetic diversity via immigration and gene flow during the recovery phase. No predictors explained the additional losses of A_R between 2010 and 2011 (Figure 2.1) that likely resulted from genetic drift. Compared to the contracted episode of population decline and recovery that occurred in 2003, it appears that across this more protracted, recent episode I was able tease apart the effects of the severity of demographic decline, which affects loss of A_R from populations, and connectivity, which affects recovery of A_R, that were previously found to interact.

Immigration accompanied by gene flow is a key process leading to recovery of genetic diversity after a demographic bottleneck (Keller et al. 2001; McEachern et al. 2011), allowing populations to maintain genetic diversity despite fluctuating dynamics (Ehrich and Jorde 2005). While immigration is mediated by patch or landscape connectivity, empirical evidence for a direct effect of connectivity in rescuing genetic diversity had been lacking. My study provides evidence in a natural system for a direct effect of connectivity in recovery of genetic diversity following a demographic bottleneck. My study also underlines the importance of maintaining connectivity in the face of climate change, as natural populations are expected to experience more frequent and severe fluctuations in size as result of increasing climatic instability (Vandenbosch 2003; Roland and Matter 2013). The two global change factors of loss of habitat connectivity and climate change may act together in this and other systems to reduce genetic diversity of populations.

2.4 Literature Cited

Berthier K, Galan M, Foltete JC, Charbonnel N, Cosson JF (2005). Genetic structure of the cyclic fossorial water vole (*Arvicola terrestris*): Landscape and demographic influences. *Molecular Ecology* 14: 2861–2871

Bouzat JL, Cheng HH, Lewin HA, Westemeier RL, Brawn JD, Paige KN (1998). Genetic evaluation of a demographic bottleneck in the greater prairie chicken. *Conservation Biology* 12: 836–843.

Busch JD, Waser PM, DeWoody JA (2007). Recent demographic bottlenecks are not accompanied by a genetic signature in banner-tailed. *Molecular Ecology* 16: 2450–2462.

Caplins SA, Gilbert KJ, Ciotir C, Roland J, Matter SF, Keyghobadi N (2014). Landscape structure and the genetic effects of a population collapse. *Proceedings of the Royal Society B* 281: 20141798.

Chapuis MP, Lecoq M, Michalakis Y, Loiseau A, Sword GA, Piry S, et al. (2008). Do outbreaks affect genetic population structure? A worldwide survey in *Locusta migratoria*, a pest plagued by microsatellite null alleles. *Molecular Ecology* 17: 3640–3653.

Chapuis MP, Estoup A (2007). Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* 24: 621–631.

Cornuet JM, Luikart G (1997). Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* 144: 2001–2014.

Craig CC (1953). On the utilization of marked specimens in estimating populations of flying insects. *Biometrika* 40: 170–176.

Ehrich D, Jorde E (2005). High genetic variability despite high amplitude population cycles in lemmings. *Mammalogy* 86: 380–385.

Frankham R (2015). Genetic rescue of small inbred populations: meta-analysis reveals large and consistent benefits of gene flow. *Molecular Ecology* 24: 2610–2618.

Guo S, Thompson EA (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48: 361–372.

Hanski I, Alho J, Moilanen A (2000). Estimating the parameters of survival and migration of individuals in metapopulations. *Ecology* 81: 239 – 251.

Hanski I (1994). A practical model of metapopulation dynamics. *Journal of Animal Ecology* 63: 151–162.

Kalinowski ST (2005). HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology* 5: 187-189.

Keller LF, Jeffery KJ, Arcese P, Beaumont MA, Hochachka WM, Smith JNM, et al. (2001). Immigration and the ephemerality of a natural population bottleneck:

evidence from molecular markers. *Proceedings of The Royal Society B* 268: 1387–1394.

Keyghobadi N, Roland J, Strobeck C (1999). Influence of landscape on the population genetic structure of the alpine butterfly *Parnassius smintheus* (Papilionidae). *Molecular Ecology* 8: 1481–1495.

Keyghobadi N, Roland J, Strobeck C (2002). Isolation of novel microsatellite loci in the Rocky Mountain Apollo butterfly, *Parnassius smintheus*. *Hereditas* 136: 247–250.

Keyghobadi N, Roland J, Strobeck C (2005). Genetic differentiation and gene flow among populations of the alpine butterfly, *Parnassius smintheus*, vary with landscape connectivity. *Molecular Ecology* 14: 1897–1909.

Lacy RC (1987). Loss of genetic diversity from managed populations: interacting effects of drift, mutation, immigration, selection, and population subdivision. *Conservation Biology* 1: 143–158.

Leberg PL (2002). Estimating allelic diversity: Effects of sample size and bottlenecks. *Molecular Ecology* 11: 2445–2449.

Luque S, Santiago S, Fortin MJ (2012). Landscape connectivity analysis for conservation: insights from combining new methods with ecological and genetic data. *Landscape Ecology* 27: 153–157.

Matter SF, Roland J (2004). Relationships among population estimation techniques: an examination for *Parnassius smintheus*. *Journal of the Lepidopterists' Society* 58: 189–195.

McEachern MB, Van Vuren DH, Floyd CH, May B, Eadie JM (2011). Bottlenecks and rescue effects in a fluctuating population of golden-mantled ground squirrels (*Spermophilus lateralis*). *Conservation Genetics* 12: 285–296.

Nei M, Roychoudhury AK (1974). Sampling variances of heterozygosity and genetic distance. *Genetics* 76: 379–390.

Nei M, Maruyama T, Chakraborty R (1975). Bottleneck effect and genetic-variability in populations. *Evolution* 29: 1–10.

Peery MZ, Kirby R, Reid BN, Stoelting R, Doucet-Bëer E, Robinson S, et al. (2012). Reliability of genetic bottleneck tests for detecting recent population declines. *Molecular Ecology* 21: 3403–3418.

Petit RJ, Mousadik AE, Pons O (1998). Identifying populations for conservation on the basis of genetic markers. *Conservation Biology* 12: 844-855.

Piry S, Luikart G, Cornuet JM (1999). BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. *Heredity* 90: 502–503.

Pruett CL, Winker K (2008). The effects of sample size on population genetic diversity estimates in song sparrows *Melospiza melodia*. *Avian Biology* 39: 252–256.

Raymond M, Rousset F (1995). GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. *Heredity* 86: 248–249.

Roland J, Matter SF (2007). Encroaching forests decouple alpine butterfly population dynamics. *Proceedings of the National Academy of Sciences of the United States of America* 104: 13702–13704.

Roland J, Matter SF (2013). Variability in winter climate and winter extremes reduces population growth of an alpine butterfly. *Ecology* 94: 190–199.

Roland J, Keyghobadi N, Fownes S (2000). Alpine *Parnassius* butterfly dispersal: Effects of landscape and population size. *Ecology* 81: 1642–1653.

Saccheri I, Kuussaari M, Kankare M, Vikman P, Fortelius W, Hanski I (1998). Inbreeding and extinction in a butterfly metapopulation. *Nature* 392: 491–494.

Spencer CC, Neigel JE, Leberg PL (2000). Experimental evaluation of the usefulness of microsatellite DNA for detecting demographic bottlenecks. *Molecular Ecology* 9: 1517–1528.

Spielman D, Brook DW, Frankham R (2004). Most species are not driven to extinction before genetic factors impact them. *Proceedings of the National Academy of Sciences of the United States of America* 101: 15261–15264.

Tischendorf L, Fahrig L (2000). On the usage and measurement of landscape connectivity. *Oikos* 90: 7–19.

Vandenbosch R (2003). Fluctuations of Vanessa cardui butterfly abundance with El Niño and Pacific Decadal Oscillation climatic variables. *Global Change Biology* 9: 785–790.

Vandewoestijne S, Schtickzelle N, Baguette M (2008). Positive correlation between genetic diversity and fitness in a large, well-connected metapopulation. *BMC Biology* 6: 1741–7007–6–46.

Williamson-Natesan EG (2005). Comparison of methods for detecting bottlenecks from microsatellite loci. *Conservation Genetics* 6: 551–562.

Chapter 3

3 Demographic fluctuations lead to rapid and cyclic shifts in genetic structure among populations of an alpine butterfly, (*Parnassius smintheus*)

3.1 Introduction

Genetic structure, the distribution of genetic variation within and among populations (Epperson and Allard 1989), is an important property of population networks that reflects their potential to respond to environmental change through local adaptation or migration (Balloux and Lugon-Moulin 2002; Toro and Caballero 2005). Genetic variation among populations typically arises when gene flow is at least somewhat limited, providing the opportunity for divergence in allele frequencies, primarily through drift or difference in selection pressure among populations (Wright 1978). Under stable conditions, genetic structure reaches a state of equilibrium where these evolutionary processes are balanced, and patterns of variation within and among populations are constant across generations (Varvio et al. 1986). However, conditions are often not stable and genetic structure among many natural populations may not be at equilibrium (Whitlock 1992).

Population size is an important determinant of genetic structure, primarily through its influence on the effective number of breeding individuals (Slatkin 1987; Gauffre et al. 2008), and thereby the levels of genetic drift (Kalinowski and Waples 2002). Population size is rarely constant, and may fluctuate considerably, even on very short time scales. Fluctuations in population size can have a variety of causes, including both density-dependent factors such as disease and predator-prey interactions, and

density-independent factors such as extreme weather events (Hansen et al. 1999; Bjørnstad and Grenfell 2001). When population size fluctuates, the lowest population sizes experienced exert the strongest influence on levels of genetic drift (Rich et al. 1979; Bouzat et al. 1998). As a result, demographic bottlenecks (severe but temporary reductions in population size) can have a strong effect on genetic variation within and among populations (Bouzat et al. 1998; Spielman et al. 2004).

The effects of demographic bottlenecks on genetic variation have received considerable attention. Specifically, many studies document the effects of bottlenecks on genetic variation within single populations. These studies demonstrate that demographic bottlenecks can erode genetic diversity, and result in inbreeding and reduced fitness (Hoelzel et al. 2002; Spielman et al. 2004). Some such studies invoke an effect of immigration from other populations in rescuing genetic variation that might otherwise have been lost from the focal population (Keller et al. 2001; Pilot et al. 2010; McEachern et al. 2011). However, relatively few studies address the effects of demographic bottlenecks on genetic structure among populations, or examine directly how immigration and gene flow interact with bottlenecks in determining patterns of genetic variation (e.g., Le Gouar et al. 2008; Ehrich et al. 2009; Chapter 2). Furthermore, the effects of repeated cycles of demographic decline and recovery on the genetic structure of natural population networks have not been widely described. There is therefore a significant gap in evolutionary research as theoretical analyses indicate that fluctuations in demographic parameters such as population size and gene flow can have important evolutionary consequences (Whitlock 1992). Fluctuations in these parameters can lead to fluctuations in the genetic variation among populations, potentially creating temporary opportunities

for certain evolutionary scenarios such as group selection or shifting-balance dynamics (Wright 1978; Whitlock 1992).

Here, I investigate changes in genetic structure in response to fluctuations in population size and repeated demographic bottlenecks in an interconnected network of populations of the Rocky Mountain Apollo butterfly, *Parnassius smintheus*. These populations have been studied continuously since 1995, providing insight into the effects of both landscape structure and climate variation on dispersal and population dynamics (Figure 1.1, Roland et al. 2000; Roland and Matter 2007). Population dynamics of *P. smintheus* in this network are largely driven by early winter weather, which is a strong descriptor of annual population growth. Specifically, increased mortality of overwintering, pharate larvae is associated with extreme weather, including both cold and warm temperatures, and reduced snowfall in November (Roland and Matter 2016).

In this study site, two severe network-wide demographic bottlenecks have been documented and attributed to low overwinter survival of larvae as a result of reduced early winter snow cover: in 2003 and also 2010 (Matter and Roland 2010; Roland and Matter 2016, Figure 3.1). In general, such bottlenecks are expected to occur regularly in this system in response to highly variable early winter conditions (Roland and Matter 2016). The effects of the 2003 demographic bottleneck on genetic structure were described by Caplins et al. (2014) who compared samples collected before (in 1995) and after (in 2005) the event. The bottleneck resulted in increased genetic differentiation among populations and loss of spatial patterns of genetic structure measured two years (generations) after the collapse in population size. Specifically, the bottleneck led to a breakdown of isolation-by-distance and disrupted associations between genetic

differentiation and both landscape variables and contemporaneous movement. These effects were largely attributable to genetic drift during the rapid collapse in population sizes. During the second bottleneck that began in 2010, the time spent at low population size was longer than in the 2003 event. Populations stayed at very low size for two years, and recovered relatively slowly thereafter, in contrast to the 2003 bottleneck in which population sizes immediately rebounded the following year (Figure 3.1). An important effect of patch connectivity in facilitating recovery of within-population genetic diversity following the second bottleneck has been demonstrated (Chapter 2). The latter result highlights the key role of dispersal and immigration in maintaining genetic diversity in this natural system, despite the occurrence of regular demographic bottlenecks

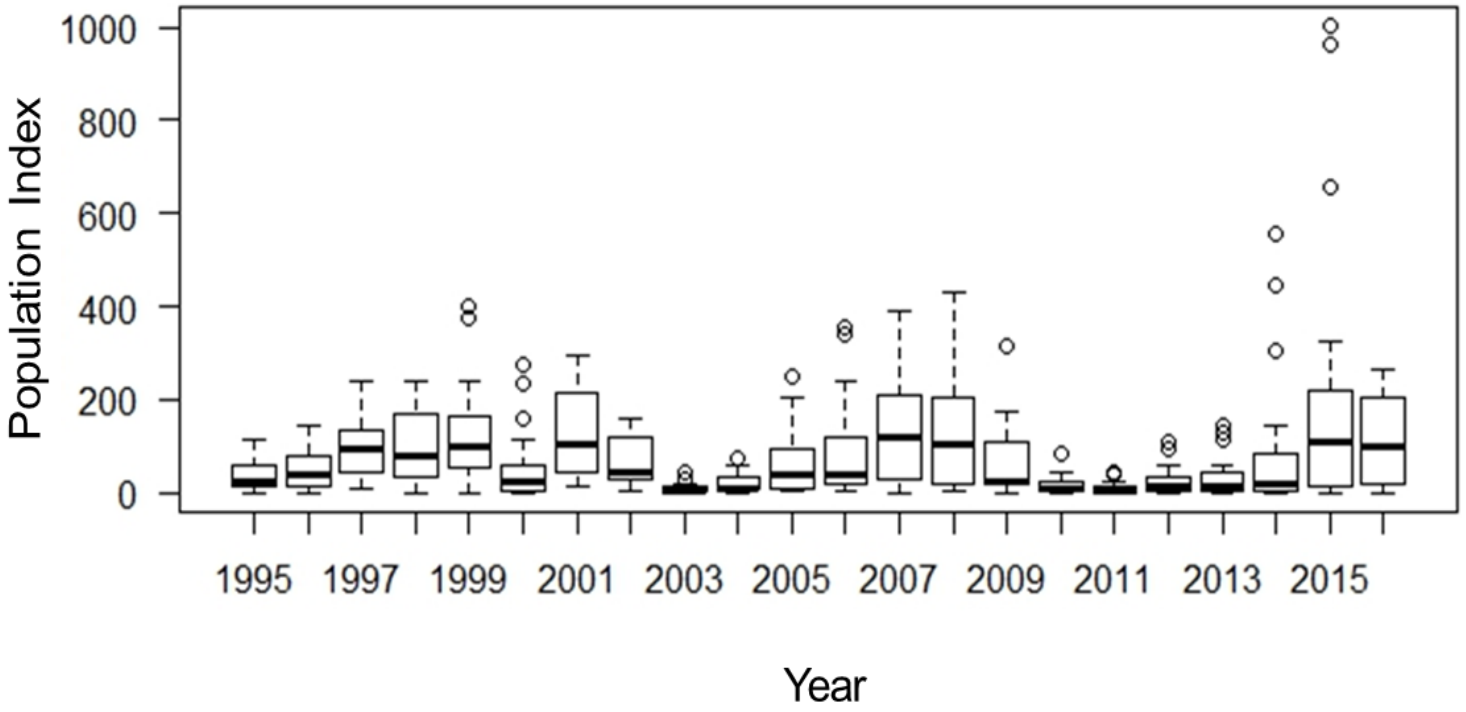


Figure 3.1 *Parnassius smintheus* population size over time. The population index is estimated based on mark-recapture of uniquely marked adults. For each day on which a particular habitat patch is visited during the flight period, the adult population size on that day was estimated using Craig's method (Craig 1953). The maximum daily estimate for each patch in each year was then used as the population size index for that patch in that year (Matter et al. 2014) Boxplots display yearly *P. smintheus* abundance indices for all populations in the network, showing interquartile range (IQR;boxes), maximum and minimum estimates up to 1.5 × IQR (whiskers), and outliers beyond 1.5 × IQR (open circles).

In systems where demographic parameters fluctuate over time, corresponding responses in genetic structure may be marked by temporal lags. This is because genetic variation does not necessarily respond immediately to changes in population size and demography, but can take several generations to approach new equilibrium values (Varvio et al. 1986). As a consequence, current genetic structure may reflect the influence of past population sizes and demographic events (Keyghobadi et al. 2005; Orsini et al. 2008). The rate at which genetic structure responds to changes in population size depends on the population size itself, as well as several additional factors such as generation time, size of the population network, direction of population size change (increase or decrease), and rates of gene flow (Epps and Keyghobadi 2015). In some situations, patterns of genetic variation require hundreds or thousands of generations to approach a new equilibrium (Varvio et al. 1986). In other cases, however, time lags are very short and genetic structure responds quickly to changes in demography. For example, Orsini et al. (2008) demonstrated a lag of only 6-7 years (generations) in the response of genetic structure to changes in demographic structure in the Glanville fritillary butterfly.

In my study system, Caplins et al. (2014) have already documented a rapid increase in genetic differentiation, and loss of spatial patterns of genetic structure, immediately after a bottleneck, due to the effect of drift. If bottlenecks erase spatial patterns of genetic structure in this way, and are expected to occur regularly and frequently in this system (on the order of approximately every decade; Roland and Matter 2016), then a question that arises is why spatial genetic structure is ever observed (as in Keyghobadi et al. 1999), given the potential for time lags to affect the recovery of spatial patterns? I hypothesize that in this population network the recovery of spatial patterns of

genetic structure after bottlenecks, through the effects of immigration and gene flow, occurs very quickly (Caplins et al. 2014). Here, I test this hypothesis and characterize the recovery of spatial patterns of genetic structure following the first documented demographic bottleneck. I also examine whether the effects of the second demographic bottleneck on genetic structure are consistent with those of the first. The central questions I address are what changes in genetic structure accompany recovery from a demographic bottleneck, and whether patterns of genetic variation among populations can return rapidly to pre-bottleneck levels under some circumstances? I take advantage of a unique, long-term dataset comprising demographic, genetic, and movement data from a spatial population network to analyze how immigration and gene flow interact with fluctuating population size to shape patterns of genetic structure over time.

3.2 Material and Methods

3.2.1 Study site and sample collection

My study was conducted in a network of populations that occupy patches of alpine meadow along three ridge-tops in the Kananaskis region of Alberta, Canada (50° 57'N, 114° 54'W; Figure 1.1). The meadows are located above treeline (2100 m), range in area from 0.2 ha to 22.7 ha, and are separated by either intervening forest or non-forested habitat (Roland and Matter 2007).

Since 1995, adult butterflies from these populations have been captured and their wing tissue sampled for genetic analysis. Keyghobadi et al. (1999) and Caplins et al.

(2014) previously described genetic structure among these populations at specific time points, 1995 and 2005, the latter being two years after a demographic bottleneck in 2003. Here, I describe genetic structure at two additional time points: in 2008 and 2013. The 2008 samples allow us to assess change in genetic structure as the populations continued to recover from the demographic bottleneck that began in 2003. The 2013 samples allow us to test the effects on genetic structure of a second, more protracted bottleneck that began in 2010. Combining my new data with those from Caplins et al. (2014), I therefore consider changes in genetic structure across two demographic transition periods: the period of continued demographic recovery and stability after the first bottleneck (comparing data from 2005 and 2008), and the period spanning the second bottleneck (comparing 2008 and 2013). The years 2008 and 2013 are the ones closest to the 2010 bottleneck in which sufficient numbers of samples were available for a robust analysis of genetic structure (Figure 3.1). Population sizes in 2013 were still relatively low however, and therefore a smaller number of patches and fewer individuals per patch could be analyzed compared to the earlier sampled time points (Table 3.1).

Table 3.1 Sample size and genetic diversity for populations of *Parnassius smintheus* sampled at four different time points**.** Data for 1995 from Keyghobadi et al. (1999, 2005) and for 2005 from Caplins et al. (2014) and are included here to provide further context and show changes over time.

Patch/ Population	**NO. genotyped individuals**				**H_E**			
	1995	2005	2008	2013	1995	2005	2008	2013
E	40	31	28	-	0.70	0.69	0.75	-
F	41	11	27	-	0.73	0.75	0.72	-
G1	40	20	51	16	0.69	0.73	0.70	0.65
g2	40	5	17	9	0.74	0.71	0.74	0.71
I	21	9	-	11	0.76	0.72	-	0.62
J	31	15	46	17	0.72	0.67	0.69	0.71
K	40	11	20	16	0.75	0.74	0.73	0.68
L	40	26	72	12	0.71	0.75	0.70	0.67
M	38	56	41	37	0.71	0.73	0.69	0.69
O	12	6	30	6	0.67	0.71	0.70	0.64
R	24	6	13	-	0.72	0.70	0.75	-
S	14	15	17	-	0.63	0.76	0.68	-
Z	41	54	33	7	0.70	0.73	0.70	0.67

No. genotyped individuals is the number of genotyped individuals that amplified at seven microsatellite loci. Expected heterozygosity, H_E, is averaged across loci.

3.2.2 Tissue sampling and genotyping

Tissue samples were collected from adult butterflies using non-lethal sampling (Koscinski et al. 2011). All samples were small wing clippings (approximately 0.2cm^2) removed from either the hind- or fore-wings, and individually stored in 95-100% (vol/vol) ethanol. DNA was extracted from wing samples using a DNeasy Blood and Tissue Kit (QIAgen, Germantown, MD) following the manufacturer's protocol. Samples were genotyped at seven highly variable microsatellite loci (Ps50, Ps76, Ps81, Ps85, Ps163, Ps165 and Ps262) (Keyghobadi et al. 1999, 2002), as described previously (Chapter 2). Microsatellite genotypes were scored using GeneMapper software Ver. 4.0 (Applied Biosystems). I checked all genotypes manually, and re-ran at least two additional times any loci that initially failed to amplify. If a locus failed a third time, but all other loci in the same individual amplified clearly, then I considered the individual null homozygous at the failed locus. I omitted from the dataset any individuals that failed at two or more loci.

3.2.3 Linkage disequilibrium and Hardy–Weinberg tests

For each locus in each population, linkage disequilibrium and conformity to Hardy–Weinberg proportions were tested for each year separately (2008 and 2013) using Genepop v.4.2 (Raymond and Rousset 1995). No linkage disequilibrium was detected based on 189 pairwise comparisons in the years 2008 and 2013, but significant deviations from expected Hardy–Weinberg genotypic proportions occurred at all loci for both time periods. Departures from equilibrium were all due to heterozygote deficiencies (in 46 of

63 tests for 2008, and 32 of 63 tests for 2013). Homozygote excess at these loci has previously been shown to be a result of null alleles, which are non-amplifying alleles that result from variation in microsatellite flanking regions (Keyghobadi et al. 1999, 2002). Null allele frequencies were estimated, and frequencies of other alleles simultaneously re-estimated, using the software FreeNA (Chapuis and Estoup 2007).

3.2.4 Spatial genetic structure over time

I estimated global and pairwise F_{ST}, corrected for presence of null alleles, within sampling periods using the software FreeNA (Chapuis, and Estoup 2007). In this system, pairwise F_{ST} displays the strongest patterns of isolation by distance compared to alternative genetic distance measures (Caplins et al. 2014). To evaluate isolation by distance, I considered geographic distances between pairs of habitat patches based on the centroids of butterfly capture within each patch. These distances were measured along the ridge-tops (Figure 1.1), rather than 'as the crow flies', since movements of *P. smintheus* are largely restricted to ridge-tops (Roland et al. 2000). Furthermore, to evaluate relationships between intervening land cover and genetic differentiation of populations, two different types of distances between patches were determined from digitized aerial photos: distance over non-forested habitat, mainly alpine meadow, and distance over forest. That is, the total distance between any two patches was measured along the ridge-tops (called 'total distance' from here on), and partitioned into that distance occurring over forest cover and that over open meadow (Roland et al. 2000).

For each year separately, I tested for isolation by distance by correlating pairwise F_{ST} to total distance using Mantel tests (Mantel 1967). I examined the effects of intervening land cover on genetic differentiation using partial Mantel tests (Smouse et al. 1986): I correlated F_{ST} to the distance through forest, controlling for the distance through meadow, and vice-versa, as in previous studies in this system (Keyghobadi et al. 1999; Caplins et al. 2014). I executed all Mantel and partial Mantel tests in the vegan package in R (Oksanen et al. 2013), with 10,000 permutations and using Pearson correlations. After demographic bottlenecks, random divergence of allele frequencies among populations is expected to lead to higher variance in pairwise F_{ST} estimates and a reduced association of pairwise F_{ST} with geographic distance and intervening landscape. Such effects are described well by changes in Mantel correlation coefficients across time periods. However, Mantel and partial Mantel tests have received criticism as tests of the statistical significance of those correlation coefficients (*i.e.*, H_0: $r = 0$; (Raufaste and Rousset 2001; Legendre and Fortin 2010)). I therefore also used maximum likelihood population effects (MLPE) models as a supplementary approach to assess the significance of relationships between my pairwise variables. MLPE models are linear mixed models with a covariance structure that accounts for the pairwise nature of the data in distance or similarity matrices (Clarke et al. 2002). For each year separately, I used MLPE to model F_{ST} as a function of total distance, and F_{ST} as a function of forest distance only, meadow distance only, and both forest and meadow distance. I first scaled all predictors and then fit MLPE models by either maximum likelihood (ML) or restricted log-likelihood (REML) estimation, using the 'gls' function in the R package 'lme4' (Bates et al. 2015) and using the correlation structure implemented by the 'corMLPE' package (Pope 2014). I used REML to obtain estimates of unbiased regression coefficients. To examine the effects of

intervening landscape on genetic differentiation, I used ML to compare the three landscape models (forest distance only, meadow distance only, or both forest and meadow distance) based on the corrected Akaike Information Criterion (AICc) to determine which model best explained genetic differentiation between sites.

As a result of low population sizes in 2005 and 2013, sample sizes for most patches were smaller in those years as compared to 2008 (Table 3.1). To ensure that any observed differences in spatial genetic structure between years were not driven by the larger sample sizes in 2008, I evaluated genetic structure and spatial patterns using subsampled individuals from the 2008 dataset. Because my central focus was on the change in genetic structure accompanying demographic recovery from 2005 to 2008, 25 datasets were randomly subsampled, without replacement, from the 2008 dataset with within-patch sample sizes matched to those in 2005. Global and pairwise F_{ST} values were estimated for each subsampled dataset. I also conducted the Mantel and partial Mantel tests for each subsampled dataset, and determined the significance of the median correlation coefficient from all subsampled datasets using Wilcoxon signed rank (WSR) tests.

3.2.5 Genetic structure and movement patterns

Mark-recapture studies of *P. smintheus* have been conducted in the population network since 1995 to determine indices of population size and estimate movement parameters (Matter et al. 2014). Adults were captured with hand nets, individually marked, and spatial locations of captures and recaptures were recorded. Following Caplins et al.

(2014), I used the Virtual Migration Model (VMM; Hanski et al. 2000) to obtain maximum likelihood estimates of movement between patches in each year using the mark-recapture data from that year. The estimated number of individuals moving in both directions between each pair of patches (e.g., the number moving from L to M, plus the number moving from M to L) was used as an index of total flow of individuals between patches. To assess the link between genetic structure and contemporaneous movement patterns, within each sampling year I determined the relationship between pairwise F_{ST} and the log-transformed estimate of total flow of individuals between pairs of patches, using Mantel tests with 10,000 permutations. As for the isolation by distance analyses, 25 random subsamples of the 2008 dataset were used to assess the effect of sample size on my conclusions. As a supplementary approach to the Mantel test, I also used MLPE models, as described above. I removed population E from these analyses due to missing mark–recapture data from that site.

3.2.6 Direct test of change in spatial genetic structure over time

In addition to characterizing differences in spatial genetic structure among separate years, I performed direct tests of the effect of year on genetic differentiation and spatial structure, using data for all available years (1995, 2005, 2008 and 2013) simultaneously. I fit three separate MLPE models to pairwise F_{ST}. Each model included one of total distance, intervening forest distance, or estimated movement as a focal predictor, and also included year, and the interaction between the focal predictor and year, as additional factors. The model for forest distance also included intervening meadow distance as a

control variable. Year was treated as a categorical factor and REML estimates of year effects used 1995 as a reference for contrasts.

3.3 Results

3.3.1 Recovery of spatial genetic patterns following a demographic bottleneck

Across the demographic recovery period, from 2005 to 2008, global F_{ST} for the population network decreased from 0.018 (95% confidence interval CI: 0.013–0.024) to 0.013 (95% CI: 0.009-0.018). Although overall genetic differentiation declined only slightly across the recovery period, with overlap of 95% CIs, there was a marked change in the spatial patterning of genetic structure. As previously reported by Caplins et al. (2014), the correlation between pairwise F_{ST} and total distance in 2005 was weak and non-significant (Mantel $r = 0.15$, $P = 0.20$; Figure 3.2). Using MLPE fit by REML to evaluate the same relationship in 2005, I found that the estimated effect of total distance on F_{ST} was also not significant ($P= 0.26$; Table 3.2). By contrast, a significant positive correlation between pairwise F_{ST} and total distance was re-established in 2008 (Mantel $r = 0.53$, $P = 0.001$; Figure 3.2), indicating a recovery of isolation by distance after the bottleneck (Table 3.2). For the 25 subsampled datasets from 2008, the median correlation coefficient (r) between pairwise F_{ST} and total distance was 0.36, and was significantly greater than zero (WSR test: $V = 325$, $P < 0.001$; Table B1). Using MLPE fit by REML

to evaluate isolation by distance in 2008, the effect of total distance on F_{ST} was also significant (P= 0.002; Table 3.2).

Across the demographic recovery period, from 2005 to 2008, associations between intervening forest cover and genetic differentiation re-established, as did a significant correlation between genetic differentiation and estimated contemporaneous movement between populations. In 2005 there had been no correlation between forest distance and pairwise F_{ST}, controlling for meadow distance (Mantel r = -0.04, P = 0.36; Caplins et al. 2014), or vice-versa (Mantel r = 0.08, P = 0.26; Caplins et al. 2014). My evaluation of intervening land cover effects using MLPE corroborated this result. In 2005, models with forest distance and meadow distance only were equally well supported (Δ AICc < 0.3; models fit using ML), and neither forest distance nor meadow distance were significant predictors of pairwise F_{ST} (P for forest = 0.33 and P for meadow = 0.27; models fit using REML). In contrast, in 2008, there was a significant positive correlation between forest distance and pairwise F_{ST} after controlling for meadow distance (Mantel r = 0.40, P = 0.03; Table 3.3), but not vice versa (Mantel r = 0.10, P = 0.22). Across 25 subsampled datasets from 2008, the median partial correlation coefficient (r) between forest distance and F_{ST}, controlling for meadow distance, was 0.22 and was significantly greater than zero (WSR test: V= 326, P < 0.001; Table B1). In contrast, the median partial correlation (r) between meadow distance and F_{ST}, controlling for forest distance, was 0.025 and was not significantly different from zero (WSR test: V = 210, P =0.10; Table B1). MLPE analyses corroborated these results for 2008. The model with forest distance as the only predictor was the best model explaining pairwise F_{ST} (Δ AICc > 2; models fit using ML; Table 3.4) and the estimated effect of forest distance was significant (P< 0.001; model fit using REML).

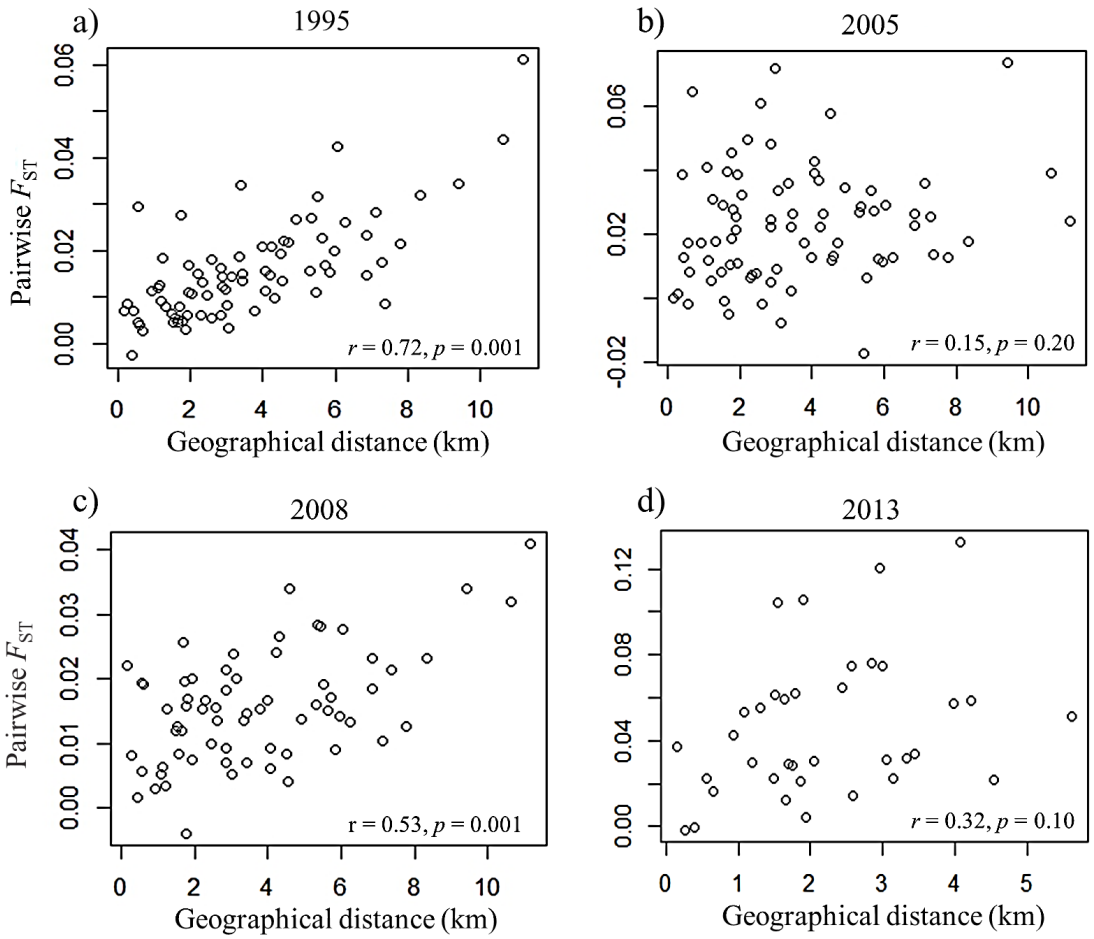


Figure 3.2 Change in the relationship between pairwise genetic differentiation (F_{ST}) and total geographical distance between populations over time, in a network of populations of *Parnassius smintheu*s. (a) a pattern of isolation by distance was observed before a demographic bottleneck in 1995 (b) no pattern of isolation by distance is detected in 2005, two years after the beginning of the demographic bottleneck, (c) a pattern of isolation by distance is re-established only three years later, in 2008, and (d) the next bottleneck led to a breakdown of isolation by distance in 2013. I have shown the results of Mantel tests of correlation for each year.

Table 3.2 Summary of Mantel tests and maximum likelihood population effects (MLPE) models testing the relationship between pairwise genetic distance (F_{ST}) and total geographical distance between populations of *Parnassius smintheus* at four different time points. Data for 1995 from Keyghobadi et al. (1999, 2005) and for 2005 from Caplins et al. (2014) and are included here to provide further context and show changes over time.

	1995		2005		2008		2013	
Model type	Mantel	MLPE	Mantel	MLPE	Mantel	MLPE	Mantel	MLPE
	r (*P*)	β ±SE (*P*)	*r* (*P*)	β ±SE (*P*)	*r* (*P*)	β ±SE (*P*)	*r* (*P*)	β ±SE (*P*)
Total distance	0.72 (**0.001**)	0.0013 ± 0.0004 (**0.002**)	0.15 (0.20)	0.001 ± 0.001 (0.26)	0.53 (**0.001**)	0.001±0.0004 (**0.002**)	0.32 (0.10)	0.004±0.003 (0.22)

r: Mantel correlation coefficients; β: MLPE regression coefficient, ± SE. Significant values are in bold typeface

Similarly, in 2005 there had been no correlation between pairwise F_{ST} and estimated movement between populations in that year (Mantel r = -0.08, P = 0.31; Caplins et al. 2014). The MLPE model confirmed that the estimated effect of contemporaneous movement in 2005 was not significant (P= 0.504; model fit using REML; Table 3.5). In 2008, pairwise F_{ST} was significantly negatively correlated with the number of estimated individuals moving between each pair of populations that year (Mantel r = -0.34, P = 0.04; Table 3.5). Across all 25 subsamples, the median correlation coefficient (r) was -0.25 and was significantly less than zero (WSR test: V = 0, P< 0.001; Table B1). In this one case however, results of MLPE analysis were not consistent with the Mantel test results. The MLPE estimate of the effect of contemporaneous movement on pairwise F_{ST} for 2008 was stronger than for 2005, but still not significant (P= 0.11; model with using REML; Table 3.5).

3.3.2 Genetic consequences of the next demographic bottleneck

Across the demographic bottleneck that started in 2010, I observed similar and even stronger changes in genetic structure as documented for the 2003 bottleneck (Caplins et al. 2014). Global F_{ST} was significantly higher after the bottleneck (2013 global F_{ST} = 0.041, 95% CI: 0.029–0.055) than before (2008 global F_{ST} = 0.013, 95% CI: 0.009–0.018). The second bottleneck also affected spatial patterning of genetic structure in the same manner as the first bottleneck. Specifically, I observed a breakdown of isolation by distance, reflected in lack of correlation between pairwise F_{ST} and total distance in 2013 (Mantel r = 0.32, P = 0.10; MLPE effect of total distance: P= 0.22; model fit using REML; Table 3.2). In 2013, I also saw a loss of any observable effect of intervening

forest cover on genetic differentiation. The partial Mantel test between pairwise F_{ST} and forest distance, controlling for meadow distance, was no longer significant (Mantel r = -0.01, P = 0.40; Table 3.3). MLPE analyses supported the partial Mantel test results. The model with meadow distance as the only predictor was marginally better in explaining pairwise F_{ST} than the model with forest distance only (ΔAICc = 1.9), although all land cover models were equally well supported (ΔAICc < 2, models fit by ML; Table 3.4). Furthermore, coefficient estimates for all land cover models in 2013 were not significant (P > 0.05, models fit by REML; Table 3.4). Finally, similar to the first documented bottleneck in 2003, the next bottleneck in 2010 disrupted any association between pairwise F_{ST} and estimated contemporaneous movement between populations (Mantel test for 2013: r = -0.33, P = 0.13; MLPE results: P= 0.30; Table 3.5).

Table 3.3 Summary of partial Mantel tests results showing the effects of intervening land cover on genetic differentiation among populations of *Parnassius smintheus* at four different time points. Data for 1995 from Keyghobadi et al. (1999, 2005) and for 2005 from Caplins et al. (2014) are also included here to provide further context and show changes over time. Partial Mantel tests were conducted for pairwise genetic distance (F_{ST}) against distance through forest controlling for the distance through meadow (Forest effect), and vice versa (Meadow effect).

Model type	**1995**	**2005**	**2008**	**2013**
	r (*P*)	*r* (*P*)	*r* (*P*)	*r* (*P*)
Forest effect	0.59 (**0.01**)	-0.04 (0.36)	0.40 (**0.03**)	-0.01 (0.40)
Meadow effect	0.15(0.14)	0.08 (0.26)	0.10 (0.22)	0.32 (0.07)

r: Partial mantel correlation coefficients. Significant values are in bold typeface.

Table 3.4 Summary of maximum likelihood population effects (MLPE) models explaining genetic differentiation (F_{ST}) between populations of *Parnassius smintheus* as a function of intervening land cover at four different time points. Data for 1995 from Keyghobadi et al (1999, 2005) and for 2005 from Caplins et al. (2014) are also included here to provide further context and show changes over time.

Models	**1995**		**2005**		**2008**		**2013**	
	ΔAIC_c	β ±SE (*P*)	ΔAICc	β ±SE (*P*)	ΔAICc	β ±SE (*P*)	ΔAICc	β ±SE (*P*)
Forest distance only	0	0.006 ±0.001 (<**0.001**)	0.26	0.0027±0.002 (0.33)	0	0.004±0.001 (**<0.001**)	1.97	0.003±0.014 (0.80)
Meadow distance only	13.55	0.001 ±0.0005 (**0.03**)	0	0.0011±0.001 (0.27)	7.53	0.0012±0.0006 (**0.034**)	0	0.005±0.003 (0.16)
Forest+ Meadow	2.28	Forest effect: 0.006± 0.001 (<**0.001**) Meadow effect: 0.0001±0.001 (0.90)	2.19	Forest effect: 0.001± 0.003 (0.77) Meadow effect: 0.0010±0.001 (0.56)	2.28	Forest effect: 0.004± 0.001 (**0.007**) Meadow effect: 0.0001±0.001 (0.86)	2.21	Forest effect: -0.015± 0.018 (0.43) Meadow effect: 0.007±0.004 (0.12)

Δ AICc is the difference in corrected Akaike Information Criterion from the top ranked model. β: MLPE regression coefficient ± SE. Significant values are in bold typeface. Models were fit using ML for model comparisons and using REML for coefficient estimation.

Table 3.5 Summary of Mantel tests and maximum likelihood population effects (MLPE) models testing the relationship between pairwise genetic distance (F_{ST}) and estimated contemporaneous movement between populations of *Parnassius smintheus* at four different time points. Data for 1995 from Keyghobadi et al. (1999, 2005) and for 2005 from Caplins et al. (2014) are included here to provide further context and show changes over time.

	1995		**2005**		**2008**		**2013**	
Model type	Mantel	MLPE	Mantel	MLPE	Mantel	MLPE	Mantel	MLPE
	r (*P*)	β ±SE (*P*)	*r* (*P*)	β ±SE (*P*)	*r* (*P*)	β ±SE (*P*)	*r* (*P*)	β ±SE (*P*)
Movement	–0.47 (**0.005**)	-0.0003±0.0001 (**0.05**)	–0.08 (0.31)	-0.0002 ± 0.0003 (0.50)	-0.34 (**0.04**)	-0.0003±0.0002 (0.11)	-0.33 (0.13)	-0.0003±0.0003 (0.30)

r: Mantel correlation coefficient; β: MLPE regression coefficient ± SE. Significant values are in bold typeface

3.3.3 Direct test of change in spatial genetic structure over time

Combining all years of available data, MLPE models fit by REML indicated that pairwise F_{ST} values in 2005 and 2013 were significantly higher than in 1995 and 2008 (Table B2). There was a significant interaction between total distance and year on pairwise F_{ST}, which indicated that isolation by distance patterns were also significantly different in 2005 and 2013 as compared to 1995 and 2008 ($P \leq 0.03$; Table B2). Similarly, there was a significant interaction between intervening forest distance and year on pairwise F_{ST} (controlling for meadow distance in the model); in this case, the effect of intervening forest on F_{ST} in 2005 was different than in 1995 ($P = 0.01$), but the effect in 2008 and 2013 was not different than in 1995 ($P > 0.05$). No interaction between estimated contemporaneous movement and year was detected with this analysis ($P > 0.05$; Table B2).

3.4 Discussion

Demographic bottlenecks can have significant effects on genetic variation, leading to a loss of genetic diversity (Nei et al. 1975), particularly allelic diversity (Maruyama and Fuerst 1985; Osborne et al. 2016), and increased differentiation among populations (Kekkonen et al. 2011). While many empirical studies have characterized the genetic effects of a single bottleneck, typically in a single population, I tracked changes in the genetic structure of a population network across two demographic bottlenecks, as well as through the intervening period of demographic recovery. Previous work in this system

has shown that the first demographic bottleneck, in 2003, led to the loss of spatial patterns of genetic variation in the population network of the butterfly *P. smintheus* (Caplins et al. 2014). Here, I show that spatial genetic structure recovered rapidly, within five years, as population sizes rebounded. Specifically, isolation by distance and a significant correlation between genetic differentiation and intervening land cover could be detected in the network by 2008. I also show that the second demographic bottleneck, which began in 2010, had similar effects to the first bottleneck, and that these effects were even stronger, concordant with the populations staying at a very low size for a longer time during the second bottleneck.

3.4.1 Temporally dynamic interplay of drift and gene flow

Network-wide demographic bottlenecks in this system appear to consistently drive random divergence of allele frequencies among populations, resulting in increased differentiation of populations and loss of spatial pattern such as isolation by distance (Caplins et al. 2014). My results indicate that subsequent rapid recovery of spatial genetic patterns occurs. This recovery is most likely driven by gene flow among populations, countering the effects of genetic drift that occurs when population sizes suddenly and dramatically collapse. As population sizes rebound, dispersal and accompanying gene flow redistribute genetic variation across the network so as to reduce differentiation among nearby populations. Furthermore, because dispersal and gene flow in this system are known to be spatially limited and strongly affected by intervening landscape (Keyghobadi et al. 1999; Roland et al. 2000), their action after a bottleneck event also

results in re-establishment of spatial patterns of genetic structure, specifically isolation by distance, as well as correlation between intervening land cover and genetic differentiation. At the same time, a correlation between estimated contemporaneous movement rates and genetic differentiation may also be re-established. Previously I showed that bottlenecks in this system reduce allelic diversity of populations (Chapter 2) and that recovery of allelic diversity immediately after bottlenecks can be explained by population connectivity within the network (Chapter 2). This effect of connectivity supports the hypothesis that dispersal and gene flow are the key processes driving the restoration of genetic variation, and patterns of genetic structure, after bottlenecks.

In my study system, rapid and cyclic changes in genetic structure occur because genetic drift and gene flow are continually shifting in dominance as populations experience repeated, dramatic fluctuations in size. As a result, this system is likely never in a state of gene flow-drift equilibrium. My results confirm theoretical expectations that genetic structure among populations (e.g., as measured by F_{ST}) fluctuate over time as a result of fluctuations in demographic parameters (Whitlock 1992). Such cycling of genetic structure could, in turn, have interesting evolutionary consequences. For example, such dynamics may create the opportunity for a type of shifting-balance evolution where a combination of drift and selection allow beneficial genetic variants to increase in frequency in some populations during and immediately after bottlenecks, while gene flow in the periods between bottlenecks allow those variants to spread through the network.

3.4.2 Rapid recovery of spatial genetic patterns

A key aspect of my results is the very rapid return of spatial genetic structure after a bottleneck. While there was no spatial pattern detectable in 2005, two years after the initial collapse in population sizes, within only three additional years (i.e., generations) spatial patterns were largely re-established. I had hypothesized that re-establishment of spatial genetic structure would occur very rapidly. Given that bottlenecks are expected to occur regularly in this system (Roland and Matter 2016), and that they consistently act to erase spatial patterns of genetic structure (Caplins et al. 2014 and this study), only a very rapid recovery of these patterns could explain my ability to nonetheless observe isolation by distance and effects of intervening landscape on genetic differentiation at some timepoints (Keyghobadi et al. 1999).

Several factors may facilitate a particularly rapid recovery of spatial genetic patterns in this system. Short generation time (in this case, one generation per year) clearly makes an important contribution. However, even when measured in numbers of generations, the rate at which spatial genetic structure returned after the bottleneck is remarkable. Additional characteristics of the population network also contribute to this rapid response. First, the populations are essentially distributed linearly, or in one dimension, along the main ridge top (Figure 1.1), and isolation by distance patterns are known to develop more quickly in one- versus two-dimensional networks (Slatkin 1993). Furthermore, this particular network is characterized by moderate to high levels of gene flow that drop off rapidly with distance, fluctuating and relatively small local population sizes, and a moderately large number of populations (Keyghobadi et al. 1999; Roland et al. 2000; Roland and Matter 2013). All of these characteristics can contribute to short

genetic time lags and more rapid shifts in spatial genetic structure (Epps and Keyghobadi 2015). For example, Varvio et al. (1986) simulated that a larger number of populations of small effective size reach gene flow-drift equilibrium more quickly than do fewer, large populations. In other systems where time lags may be expected to be longer (e.g., where effective population size is larger or gene flow is more limited), demographic bottlenecks likely have longer-lasting effects on genetic structure and its spatial patterning.

Researchers often assume that genetic structure changes only very slowly over time, particularly when it is measured at the population level and using allele frequency based measures of genetic structure and differentiation such as F_{ST} (Landguth et al. 2010). However, my results clearly indicate that genetic structure, even when measured using F-statistics, can be dynamic on very short time scales. These short-term changes could lead to inaccurate, or at least incomplete, inferences about genetic structure and its relationships to landscape variables or estimated movement rates if genetic variation is only characterized at a single time point. Among populations that show considerable fluctuations in size in particular, spatial genetic structure may not remain constant over time. In the context of landscape genetic studies specifically, my results show that associations between landscape variables and genetic variation can change quickly in response to population size fluctuations. Therefore, the observed genetic structure among populations may reflect how landscape variables actually affect movement more or less accurately, depending on the temporary predominance of drift versus gene flow at a particular point in time. For example, in my study system sampling in the years immediately after bottleneck events would lead to a failure to detect effects of intervening forest on genetic differentiation which are observable at other time points (Keyghobadi et

al. 1999) and which reflect known effects of land cover on movement (Roland et al. 2000). My results therefore suggest a certain degree of caution in the interpretation of spatial patterns of genetic structure that have been measured at a single point in time, which is currently the case in most studies.

With increased variation in yearly weather conditions as a result of climate change, specifically winter weather extremes affecting overwintering eggs, *P. smintheus* populations may experience more frequent demographic bottlenecks (Roland and Matter 2013, 2016). At the same time, rising tree line in alpine areas, also driven by climate change (Gehrig-Fasel et al. 2007), and the resulting increased isolation of habitat patches would reduce gene flow across the network. I therefore expect that over time, although drift and gene flow will continue to cycle in dominance as population sizes fluctuate, the relative influence of drift in these populations will increase overall. Increased isolation of populations and lower levels of gene flow could also mean a longer time for spatial patterns to re-establish after bottlenecks, although these effects could be counteracted by lower effective population sizes resulting from more severe bottlenecks and smaller habitat patch sizes (Epps and Keyghobadi 2015).

3.5 Conclusion

In conclusion, I have demonstrated that spatial genetic structure, and the degree of correlation between genetic differentiation and both landscape and movement patterns, can be highly dynamic over short time periods due to the cyclical predominance of gene

flow versus drift in a network of populations of fluctuating size. My results show the potential for genetic structure and its spatial patterning, as well as the underlying neutral processes of drift and gene flow, to fluctuate regularly and rapidly in natural systems.

3.6 Literature Cited

Balloux F, Lugon-Moulin N (2002). The estimation of population differentiation with microsatellite markers. *Molecular Ecology* 11: 155-65.

Bates D, Maechler M, Bolker B, Walker S (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67: 1-48.

Bjørnstad, ON, Grenfell BT (2001). Time series analysis of population fluctuations in animals. *Science* 293: 638–643.

Bouzat JL, Cheng HH, Lewin, HA, Westemeier RL, Brawn JD, Paige, KN (1998). Genetic evaluation of a demographic bottleneck in the greater prairie chicken. *Conservation Biology* 12: 836–843.

Caplins SA, Gilbert KJ, Ciotir C, Roland J, Matter SF, Keyghobadi N (2014). Landscape structure and the genetic effects of a population collapse. *Proceedings of the Royal Society B* 281(1796): 20141798.

Chapuis M, Estoup A (2007). Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* 24*:* 621–631.

Clarke RT, Rothery P, Raybould AF (2002). Confidence limits for regression relationships between distance matrices: Estimating gene flow with distance. *Journal of Agricultural, Biological, and Environmental Statistics* 7: 361–372.

Craig C (1953). On the utilization of marked specimens in estimating populations of flying insects. *Biometrika* 40:170–176

Ehrich D, Yoccoz NG, Ims RA (2009). Multi-annual density fluctuations and habitat size enhance genetic variability in two northern voles. *Oikos* 118: 1441–1452.

Epperson BK, Allard RW (1989). Spatial autocorrelation analysis of the distribution of genotypes within populations of lodgepole pine. *Genetics* 121: 369–377.

Epps CW, Keyghobadi N (2015). Landscape genetics in a changing world: Disentangling historical and contemporary influences and inferring change. *Molecular Ecology* 24: 6021–6040.

Gauffre B, Estoup A, Bretagnolle V, Cosson JF (2008). Spatial genetic structure of a small rodent in a heterogeneous landscape. *Molecular Ecology* 17: 4619–4629.

Gehrig-Fasel J, Guisan A, Zimmermann NE (2007). Tree line shifts in the Swiss Alps: climate change or land abandonment? *Journal of Vegetation Science* 18: 571–582.

Hansen TF, Stenseth NC, Henttonen H, Tast J (1999). Interspecific and intraspecific competition as causes of direct and delayed density dependence in a fluctuating vole population. *Proceedings of the National Academy of Sciences of the United States of America* 96: 986–991.

Hanski I, Alho J, Moilanen A (2000). Estimating the parameters of survival and migration of individuals in metapopulations. *Ecology* 81: 239 – 251.

Hoelzel AR, Fleischer RC, Campagna C, Le Boeuf BJ, Alvord G (2002). Impact of a population bottleneck on symmetry and genetic diversity in the northern elephant seal. *Journal of Evolutionary Biology* 15: 567–575.

Kalinowski ST, Waples RS (2002). Relationship of effective to census size in fluctuating populations. *Conservation Biology* 16: 129–136.

Kekkonen J, Hanski IK, Jensen H, Väisänen RA, Brommer JE (2011). Increased genetic differentiation in house sparrows after a strong population decline: From panmixia towards structure in a common bird. *Biological Conservation* 144: 2931–2940.

Keller L, Jeffery K, Arcese P, Beaumont M, Hochachka W, Smith J, et al. (2001). Immigration and the ephemerality of a natural population bottleneck: Evidence from molecular markers. *Proceedings of the Royal Society B* 268: 1387–1394.

Keyghobadi N, Roland J, Strobeck C (1999). Influence of landscape on the population genetic structure of the alpine butterfly *Parnassius smintheus* (Papilionidae). *Molecular Ecology* 8: 1481–1495.

Keyghobadi N, Roland J, Strobeck C (2002). Isolation of novel microsatellite loci in the Rocky Mountain Apollo butterfly, *Parnassius smintheus*. *Hereditas* 136: 247–250.

Keyghobadi N, Roland J, Matter SF, Strobeck C (2005). Among- and within-patch components of genetic diversity respond at different rates to habitat fragmentation: an empirical demonstration. *Proceedings of the Royal Society B* 272: 553–560.

Koscinski D, Crawford L. A, Keller HA, Keyghobadi N (2011). Effects of different methods of non-lethal tissue sampling on butterflies. *Ecological Entomology* 36: 301–308.

Landguth EL, Cushman SA, Schwartz MK, Mckelvey KS, Murphy M, Luikart G (2010). Quantifying the lag time to detect barriers in landscape genetics. *Molecular Ecology* 19: 4179–4191.

Le Gouar P, Rigal F, Boisselier-Dubayle MC, Sarrazin F, Arthur C, Choisy J, et al. (2008). Genetic variation in a network of natural and reintroduced populations of Griffon vulture (*Gyps fulvus*) in Europe. *Conservation Genetics* 9: 349–359.

Legendre P, Fortin MJ (2010). Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources* 10: 831–844.

Mantel N (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research* 27: 209–220.

Maruyama T, Fuerst PA (1985). Population bottlenecks and nonequilibrium models in population genetics. II. Number of alleles in a small population that was formed by a recent bottleneck. *Genetics* 111: 675–689.

Matter SF, Keyghobadi N, Roland J (2014). Ten years of abundance data within a spatial population network of the alpine butterfly *Parnassius smintheus*. *Ecology* 95: 2985–2985.

Matter SF, Roland J. (2010). Effects of experimental population extinction for the spatial population dynamics of the butterfly *Parnassius smintheus*. *Oikos* 119: 1961–1969.

McEachern B, Van Vuren DH, Floyd CH, May B, Eadie JM (2011). Bottlenecks and rescue effects in a fluctuating population of golden-mantled ground squirrels (*Spermophilus lateralis*). *Conservation Genetics* 12: 285–296.

Nei M, Maruyama T, Chakraborty R (1975). Bottleneck effect and genetic-variability in populations. *Evolution* 29: 1–10.

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. (2013). Vegan: Community Ecology Package. R package version 2.0-10. https://CRAN.R-project.org/package=vegan.

Orsini L, Corander J, Alasentie A, Hanski I (2008). Genetic spatial structure in a butterfly metapopulation correlates better with past than present demographic structure. *Molecular Ecology* 17: 2629–2642.

Osborne AJ, Negro SS, Chilvers BL, Robertson BC, Kennedy MA, et al. (2016). Genetic evidence of a population bottleneck and inbreeding in the endangered New Zealand sea lion, *Phocarctos hookeri*. *Journal of Heredity* 107: 392–402.

Pilot M, DąBrowski MJ, Jancewicz E, Schtickzelle N, Gliwicz J (2010). Temporally stable genetic variability and dynamic kinship structure in a fluctuating population of the root vole *Microtus oeconomus*. *Molecular Ecology* 19: 2800–2812.

Pope N (2014). corMLPE: A correlation structure for symmetric relational data.

Raufaste N, Rousset F (2001). Are partial mantel tests adequate? *Evolution* 55: 1703–1705.

Raymond M, Rousset F (1995). GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* 86: 248–249.

Rich SS, Bell AE, Wilson SP (1979). Genetic drift in small populations of *Tribolium*. *Evolution* 33: 579-584.

Roland J, Keyghobadi N, Fownes S (2000). Alpine *Parnassius* butterfly dispersal: Effects of landscape and population size. *Ecology* 81: 1642-1653.

Roland J, Matter SF (2007). Encroaching forests decouple alpine butterfly population dynamics. *Proceedings of the National Academy of Sciences of the United States of America* 104: 13702–13704.

Roland J, Matter SF (2013). Variability in winter climate and winter extremes reduces population growth of an alpine butterfly. *Ecology* 94: 190–199.

Roland J, Matter SF (2016). Pivotal effect of early-winter temperatures and snowfall on population growth of alpine *Parnassius smintheus* butterflies. *Ecological Monographs* 86: 412–428.

Slatkin M (1987). Gene flow and the geographic structure of natural populations. *Science* 236: 787–792.

Slatkin M (1993). Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* 47: 264-279.

Smouse PE, Long JC, Sokal RR (1986). Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology* 35: 627–632.

Spielman D, Brook D, Frankham R (2004). Most species are not driven to extinction before genetic factors impact them. *Proceedings of the National Academy of Sciences of the United States of America* 101: 15261–15264.

Toro MA, Caballero A (2005). Characterization and conservation of genetic diversity in subdivided populations. *Philosophical Transactions of the Royal Society B: Biological Sciences* 360: 1367–1378.

Varvio SL, Chakraborty R, Nei M (1986). Genetic variation in subdivided populations and conservation genetics. *Heredity* 57: 189-198.

Whitlock MC (1992). Temporal fluctuations in demographic parameters and the genetic variance among populations. *Evolution* 46: 608–615.

Wright S (1978). Evolution and the genetics of population variability within and among natural populations. University of Chicago Press, Chicago.

Chapter 4

4 Gene expression associated with dispersal history in the alpine butterfly, *Parnassius smintheus*

4.1 Introduction

Dispersal, the movement and settlement of individuals away from their natal habitat patch, drives the dynamics, persistence and evolutionary trajectories of spatially-structured populations (Bowler and Benton 2005; Clobert et al. 2012; Travis et al. 2013). Dispersal among patches is advantageous because it can reduce competition among relatives for resources, facilitate escape from natural enemies or poor environmental conditions, and help avoid inbreeding (Bowler and Benton 2005; Ronce 2007). However, dispersal requires longer, more sustained movements than the foraging, mating or other routine movements within a habitat patch. Thus, dispersal is also physically and energetically costly, and increases the risk of injury and mortality (Bonte et al. 2012); these costs of dispersal increase with travel distance, particularly if inter-patch habitats are inhospitable (Sekar 2012). Consequently, dispersal exerts strong selection pressure: dispersal tendency or ability responds to changes in population size, habitat quality (i.e. resource availability) and landscape structure (Fraser et al. 2001; Meylan et al. 2009; Edelsparre et al. 2014). For example, habitat fragmentation can induce strong selection for adaptions that alter rates of emigration and dispersal (Gibbs and Dyck 2010; Gomez and van Dyck 2012).

In flying animals, flight capacity is an important determinant of whether individuals disperse or not (Niitepõld et al. 2009). In poikilothermic insects, in turn, flight

and therefore dispersal is sensitive to environmental and body temperature (Niitepõld et al. 2009; Jones et al. 2015). Because the thorax contains the flight muscles, thoracic temperature plays a central role in flight and needs to be regulated to meet flight requirements (Wickman 2009). A minimum thoracic temperature is necessary for the initiation and continuation of flight (Vogt and Heinrich 1983; Heinrich 1993). At the same time, the thoracic muscles have extremely high metabolic rates during flight (Heinrich 1995; Mattila 2015) such that there is a risk of overheating of the muscles, which can result in reduced flight ability, injury or mortality (Mattila 2015). If thoracic temperature exceeds either the upper or lower limit, enzymatic activities involved in flight metabolism and flight muscle function may be impaired. Thus, insects have a critical thoracic temperature range within which flight can be sustained. This range is highly variable among species. For example, during flight the butterfly genera *Papilio* and *Colias* experience thoracic temperatures between 28 °C and 42 °C (Kingsolver 1985; Srygley and Chai 1990), while in the alpine genus *Parnassius* much lower flight thoracic temperatures of 17 °C to 20 °C have been documented (Guppy 1986).

Insects use a number of different mechanisms to regulate thoracic temperature in preparation for, and during, flight. To gain heat at low ambient temperatures and prior to flight, insects may use muscular shivering or basking in sunlight (Masters et al. 1988; Heinrich 1995). For example, in the monarch butterfly (*Danaus plexippus*) both behaviors cause rapid warming of the thoracic muscles to the flight threshold (12.7 °C - 16.0 °C) at ambient temperatures as low as 9 °C (Masters et al. 1988). In *Colias* butterflies in contrast, most heat gained both in preparation for flight occurs via solar flux (Tsuji et al. 1985). Solar heat flux and muscular activity can also contribute to thoracic heat gain

while in flight. During flight, excessive heat gain is a primary concern for insects and to avoid overheating they may minimize solar exposure or use evaporative cooling. Both of these activities can be augmented by appropriate wing opening and closing behaviours, and by adjustment of wing and body position relative to the sun or wind (Masters et al. 1988; Prange 1995; Berwaerts et al. 2001). Since the thoracic temperature of insects may be influenced by ambient temperature, solar radiation and muscular activity (Wickman 2009; Mattila 2015), its regulation is an important potential cost of flight, particularly during long flights associated with dispersal events.

In many organisms, individuals display variation in dispersal ability or propensity. In some cases, dispersal variation is manifested via obvious morphological differences, such as winged and wingless morphs in insects (Roff 1986; Schwander and Leimar 2011). In other cases, dispersal variation is associated with more subtle differences among individuals in morphology, behavior, or physiology (Bonte et al. 2012). This variation has, in several cases, been shown to be heritable and in some animals specific genes associated with dispersal have been identified (e.g., in *Drosophila* and in the Glanville fritillary butterfly, *Melitaea cinxia*; Edelsparre et al. 2014; Saastamoinen et al. 2018). Furthermore, in insects, variation in dispersal and flight ability is often also linked to variation in thermoregulation. For example, in the Glanville fritillary, polymorphism in the thermal stress-related *heat-shock protein* (Hsp) gene is associated with variation in male flight metabolic rate and thoracic temperature at take-off (Mattila 2015). Variation in dispersal ability or tendency among individuals, and related thermoregulatory traits, may be reflected in differences in gene expression. In the Glanville fritillary, baseline differences in gene expression among individuals from

different populations are related to differences in flight metabolic rate and dispersal ability (Kvist et al. 2015). However, flight activity in insects also induces significant short and long term changes in gene expression (Margotta et al. 2012; Kvist et al. 2015), so that differences in gene expression patterns could potentially be either a cause of variation in dispersal propensity or a consequence of dispersal at some point in the past.

The Rocky Mountain Apollo, *Parnassius smintheus*, is an alpine butterfly that occupies naturally patchy, high-altitude habitats and in which some important population dynamic and genetic consequences of dispersal have been documented (Roland et al. 2000; Keyghobadi, et al. 2005; Chapter 2). Here, I examine potential differences in gene expression between dispersing and non-dispersing individuals in this system. Specifically, I used mark-recapture data to identify individuals that moved between habitat patches (dispersers) and those that remained in the same habitat patch (non-dispersers). I then used RNA−sequencing to profile a *de novo* transcriptome for this species and to perform a comparative transcriptomics analysis of thoracic gene expression changes associated with dispersal history (i.e., dispersers versus non-dispersers). I focused specifically on gene expression in the thorax because of the thorax's central role in insect flight. Because of the importance of regulating thoracic temperature during insect flight, I also measured thoracic temperature (relative to ambient) for all sampled individuals.

I hypothesized that expression of genes linked to physiological and morphological traits, specifically energy mobilization, thermoregulation, and muscle regulation, could either be a consequence or cause of long-distance flight associated with dispersal. In terms of potential gene expression differences that are a consequence of dispersal, I expected to detect only those flight-associated gene expression changes that

are relatively long lasting (i.e., on the order of one or more days), since I was not capturing dispersing individuals during or immediately after the dispersal event.

I had two competing hypotheses for the role of thoracic temperature in dispersal, related to whether long-distance flight in this species is limited primarily by the ability to maintain a high enough thoracic temperature or whether it is limited by the ability to dissipate the heat that accumulates as a result of muscular activity and elevated metabolism. First, low temperatures in this temperate, high-altitude environment may be a key factor limiting activity, including flight, in *P. smintheus*. Those individuals that are generally able to maintain a higher thoracic temperature could therefore be more likely to initiate and complete dispersal movements (Mattila 2015; Wong et al. 2016). Under this hypothesis, I predicted higher thoracic temperature (relative to ambient temperature) in those individuals that were dispersers. On the other hand, once flight is initiated, the ability to prevent overheating in the thoracic muscles may be the key determinant of whether an individual can successfully complete a long-distance displacement. In that case, I predicted lower thoracic temperature (relative to ambient temperature) in those individuals that were dispersers. Both hypotheses also led to a prediction of potential differences in gene expression between thoraxes of individuals with high versus low thoracic temperature (relative to ambient temperature). Such differences in gene expression could represent baseline differences among individuals that underlie differences in dispersal ability or tendency, as have been documented in other butterflies (Somervuo et al. 2014; Kvist et al. 2015), and represent a cause rather than consequence of dispersal.

4.2 Material and methods

4.2.1 Collection of samples and field data

I collected 12 adult individuals, three females and nine males, of *P. smintheus* from seven different patches of alpine meadow along Jumpingpound Ridge, in the Kananaskis region of Alberta, Canada (50°57' N, 114°54' W) in the summer 2015 (Figure 1.1; Table 4.1). The meadows are located above treeline (2100 m), range in area from 0.2 ha to 22.7 ha, and are separated by either intervening forest or open meadow habitat (Roland and Matter 2007).

I captured individuals in flight using a hand net and measured thoracic temperature for each individual within 5 seconds of capture using a digital thermometer (OMEGA HH91, Norwalk CT) attached to a copper thermocouple (OMEGA type T, Norwalk CT) housed within a disposable hypodermic needle (precisionGlide 19-Gauge needle, BD Medical, Franklin Lakes NJ). To avoid any direct contact between the butterfly's body and my hands, I inserted the needle into the thorax through the net to measure body temperature. I used the same thermometer to measure ambient air temperature at the time and location of capture. In my subsequent analyses, I corrected for variation in ambient air temperature by using the difference between thoracic temperature and ambient temperature for each individual as a variable of interest. Immediately after recording thoracic temperature, I separated the thorax from the rest of the body, and from the wings and legs, using a clean scalpel and placed the entire thorax, which contains the flight muscles, into a 1.7 ml microcentrifuge tube containing 1.5 ml RNA*later* solution (Qiagen, Germantown MD). I stored all samples at -20°C until I conducted RNA extraction

Sampling for the current study was concurrent with a larger mark-recapture study across Jumpingpound Ridge (Matter et al. 2014). Every individual sampled for my study had already been uniquely marked, and the spatial coordinates of all capture locations had been recorded (Matter et al. 2014), such that each individual could be classified as either as a disperser (having been re-captured in a patch different from the one in which it was originally marked) or a non- disperser (re-captured in the same patch in which it was originally marked).

4.2.2 Experimental designs

With both dispersal history and thoracic temperature available for each individual (Table 4.1), I was able to set up two different experimental designs: **a**) All 12 individuals sorted into two groups according to their dispersal histories: dispersers and non-dispersers. I had six biological replicates (different individuals) for each group. **b**) All 12 individuals sorted into two groups according to the difference between their thoracic temperature and ambient air temperature: thoracic temperature either higher or lower than ambient temperature. In this design, I had eight biological replicates in the first category (thoracic higher than ambient temperature) and four biological replicates in the second category (thoracic lower than ambient temperature).

Table 4.1 Information about the samples used for gene expression analysis in this study**.** Patch indicates the individual's final capture location (Figure 1.1), when tissue for RNA analysis was sampled. Every individual was either classified as a disperser or non-disperser based on their mark-recapture history. For dispersers, the patch where the individual was initially marked is given in brackets. The difference between thoracic and ambient temperature at the time of sampling was calculated for each individual. The number of days between initial marking of the individual and sampling for this study (Time since marking) are also provided.

Patch	Disperser/ non-disperser	Sex	Thoracic temp. (°C)	Ambient temp. (°C)	Thoracic- ambient temp. (°C)	Time since marking
F	Non-disperser	male	25.8	13.3	12.5	19
Z	Non-disperser	female	22.3	15.8	6.5	13
Q	Non-disperser	female	30.8	24.7	6.1	14
Q	Non-disperser	male	32.6	22	10.6	27
Q	Non-disperser	female	26	23	3	0
Q	Non-disperser	male	29	21	8	12
M	Disperser (P)	male	21	20	1	15
L	Disperser (Q)	male	19.5	22.4	-2.9	12
L	Disperser (J)	male	21.2	22.6	-1.4	7
J	Disperser (K)	male	20	24.8	-4.8	5
L	Disperser (P)	male	19.1	23.4	-4.3	10
Q	Disperser (M)	male	24.2	22.3	1.9	2

4.2.3 RNA extractions, and mRNA sequencing

Total RNA was extracted from each thorax sample using the RNeasy Mini Kit (Qiagen) according to the manufacturer's protocol. An aliquot of total RNA for each individual, containing from 2–6 µg of RNA in a maximum volume of 20 µL, was sent to the Next-Generation Sequencing Services at Genome Québec (McGill University, Montréal, Québec; MGU-GQ) for 100 bp paired-end mRNA sequencing on an Illumina HiSeq platform (Illumina, San Diego CA). All RNA samples successfully passed the Quality Control with an RNA integrity number greater than '8' (Bioanalyzer 2100, Agilent Technologies). For each sample, mRNA was purified from the total RNA and indexed with a unique barcode used in library preparation at the MGU-GQ facilities using the Illumina TruSeqmRNA Library Prep Kit v2. All 12 mRNA libraries were sequenced on four lanes of an Illumina HiSeq 2000 flow cell.

4.2.4 Sequence data processing & *de novo* transcriptome assembly

I trimmed raw reads of adapter sequence and removed any reads that had low base quality scores (<30) or that were shorter than 36 bp using TRIMMOMATIC v.0.36 (Bolger et al. 2014). The remaining pair-matched reads were assessed for overall quality in FASTAQC v.0.11.5 (Andrews 2010). All clean, high-quality reads were normalized with a maximum of 30 reads coverage per contig using *in silico* normalization in TRINITY v.2.5.0 (Grabherr et al. 2011) to improve efficiency of the transcriptome assembly by reducing the quantity of input reads.

I pooled the resulting 998,970,648 million normalized reads from all 12 individuals to assemble a *de novo* reference transcriptome using three different assemblers, TRINITY v2.5.0, CLC Genomics Workbench v8.5 (CLC Bio-Qiagen) and Oases v0.2.08 (Schulz et al. 2012), under the default settings. I then carried out a comparison of the three assemblers for construction of an optimal reference transcriptome for downstream analyses. I compared N50 and average contig length between assemblies, which are commonly used metrics of assembly quality, to assess the effectiveness of each *de novo* assembly. The completeness of the three *de novo* assemblies was also evaluated via Benchmarking Universal Single-Copy Orthologs (BUSCO v2; Simão et al. 2015). BUSCO quantitates assembly completion by determining whether assembled transcripts align to highly conserved, single-copy amino acid sequences within its database, and by classifying matches as complete or fragmented.

4.2.5 Functional annotation of genes

I performed sequence homology searches via BLASTX against well-annotated sequences from a custom-made insect gene database (Table 4.2) to verify insect origin of genes in my reference assembly. BLASTX translates a given nucleotide query sequence and compares it to the database sequences using all six possible reading frames (three in each direction). Only contigs that showed a minimum 70% amino acid identity (E-value threshold 10^{-3}) with at least one other insect gene were kept in the transcriptome assembly. I identified genes in my final reference transcriptome assembly and assigned putative functions to them using the following steps. First, contigs in the final reference transcriptome were translated to predicted amino acid sequences using

TRANSDECODER (v2.0.2; Haas et al. 2013). Then, BLASTX and BLASTP were used to identify annotated homologs (significant thresholds of e-value < $1e^{-5}$) within Swiss-Prot database (UniProt 2014). HMMER (v3.1; Eddy 2011) was also used to determine homology of protein domains against those in the Protein Family Database (Pfam v29.0; Finn et al. 2015). I integrated the results from the gene- and domain-level analyses of amino acid sequences in the TRINOTATE program (v.3.0.2; Haas et al. 2013) to create a comprehensive annotation report for the reference transcriptome.

4.2.6 Differential gene expression analysis

Trimmed paired-end reads (not normalized) from each library were mapped back to the reference assembly using the splice-aware aligner Bowtie2 (v2.3.0; Langmead and Salzberg 2012) with default settings. I then used the program RSEM (v1.2.25; Li and Dewey 2011) to count the number of raw reads that aligned to each contig for each individual library. Before estimating differentially expressed genes (DEGs), the mapped read counts were normalized for differences in gene length and library size (total reads) using the Trimmed Mean of M-values (TMM) method which is implemented in the R Bioconductor package edgeR (v 3.5; Robinson et al. 2010).

According to my two different potential experimental designs, I could compare differential expression patterns between disperser and non-disperser individuals (grouped regardless of thoracic temperature), as well as between individuals with thoracic temperature higher versus lower than ambient temperature (grouped regardless of dispersal history). All differential gene expression analyses were performed using the package edgeR. Expression differences between groups were considered significant after

correction for multiple testing using a false discovery rate (FDR) of less than 0.05, and also a minimum four-fold expression change. In each experimental condition, I estimated Pearson's correlation for each possible pair of individuals to examine how individuals sorted based on similarity in their expression patterns. Hierarchical clustered heatmaps were also constructed to visualize how DEGs and individuals were related according to their gene expression patterns. I visualized the number of DEGs between dispersers and non-dispersers, and different body temperatures, by Venn diagrams generated using Venny (accessible at http://bioinfogp.cnb.csic.es/tools/venny/index.html).

Since all three female individuals sampled were non-dispersers, to ensure that any observed differential expression patterns were not influenced by sex-specific gene expression, I also conducted the analyses for both experimental designs (i.e., based on dispersal history and relative thoracic temperature) without females.

4.2.7 Enrichment analysis

To determine the potential function of DEGs, I performed gene ontology (GO) enrichment analysis using the GOseq R package (1.22.0; Young et al. 2010). I considered a FDR-corrected *P*- value less than 0.05 as the threshold to determine significantly enriched GO terms. To gain insight into the functional categories associated with DEGs in each gene set, I specifically focused on only GO terms assigned to 'biological processes'. I also used Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis (Kanehisa and Goto 2000) to identify potential pathways involved in the

differentially expressed gene sets. I then visualized differentially expressed pathways using ClustVis (Metsalu and Vilo 2015).

4.3 Results

4.3.1 Thoracic temperature and dispersal history

I observed flight thoracic temperatures in *P. smintheus* ranging from 19.1 °C to 32.6 °C. Recorded ambient temperatures ranged from 13.3 °C to 24.8 °C, and the difference in thoracic and ambient temperature for individual butterflies ranged from – 4.8 °C to + 12.5°C. Individuals classified as dispersers tended to have thoracic temperatures lower than ambient (mean difference from ambient = -2.01 ± 2.3 °C), while individuals classified as non-dispersers consistently had thoracic temperatures higher than ambient (mean difference from ambient = 7.78 ± 3.4 °C). The difference in mean thoracic temperature of dispersers and non-dispersers was significant ($t_{9.10}$= 5.83, $P < 0.001$).

4.3.2 Transcriptome sequencing & *de novo* reference assembly

Illumina sequencing yielded a total of 1,267,460,876 raw reads from 12 libraries. After trimming and quality control, 1,172,691,325 high-quality reads were retained with the average quality score greater than 30, for use in *de novo* transcriptome assembly. Among the three assemblers (Trinity, Oases and CLC), Trinity produced the assembly with longest N50 length and highest average contig length (Table 4.3). An analysis of

transcriptome completeness by BUSCO also indicated that Trinity performed best in terms of having the highest complete arthropod BUSCOs and fewest missing and fragmented arthropod BUSCOs, followed by Oases and CLC (Table 4.3). I therefore selected the Trinity assembly as the reference for all downstream gene expression analyses.

The raw Trinity assembly generated 344,367 ‘contigs’ that represented 508,833 transcripts, and after cross-referencing to my insect gene database, the final assembled transcriptome had a total of 33,165 Trinity ‘genes’ that represented 72,469 transcripts (Table 4.4). Cross-referencing for homology to insect genomes improved the quality of the assembly, resulting in a higher N50 value (from 847 to 1976 bp), a longer average contig length (from 585 to 1217 bp), and fewer missing genes identified by BUSCO (from 288 to 179). Queries against the Swiss-Prot database identified matches with annotated proteins for 15,118, or 47%, of total genes.

Table 4.2 The insect gene database used to reduce the number of spurious genes and verify the insect origin of transcripts in the reference transcriptome.

Order	**Species**	**Reference**
Lepidoptera	*Melitaea cinxia*	Ahola et al. 2014
	Heliconius numata	Wallbank et al. 2016
	Papilio machaon	Li et al.2015
	Danaus plexippus	Zhan et al. 2011
Diptera	*Drosophila melanogaster*	Attrill et al. 2016
Hymenoptera	*Bombus terrestris*	Sadd et al. 2015
	Apis mellifera	Weinstock et al. 2006

Table 4.3 Summary of *de novo* assemblies' quality using common statistical metrics: N50 (which represents the distribution of contig lengths in an assembly), average contig length, and BUSCO's evaluation metrics that quantify assembly completion by aligning all transcripts to highly conserved proteins within its dataset.

Assembler	N50 length	Average contig length (bp)	Missing BUSCOs%	Complete BUSCOs%	Fragmented BUSCOs%
Trinity	1976	1217.61	6.7	80.7	6
Oases	386	300.28	11.36	73	15
CLC	272	261.72	26.7	54.2	20

Table 4.4 Summary statistics of sequencing and *Parnassius smintheus de novo* reference transcriptome assembly.

Raw reads	1,267,460,876
Trimmed reads	1,172,691,325
Normalized reads	998,970,648
Trinity assembly statistics	
Number of contigs	344,367
Number of transcripts	508,833
N50 length	847
Average contig length (bp)	585.25
GC content (%)	37.62
After crossing reference to insect gene database (≥70% amino acid identity)	
Number of contigs	33,165
Number of transcripts	72,469
N50 length	1976
Average contig length (bp)	1217.61
GC content (%)	40.8

4.3.3 Differential gene expression as a function of dispersal history

I identified 94 genes differentially expressed (FDR-corrected *P*-value < 0.05) between dispersers and non-dispersers. These unique DEGs were strongly associated with dispersal as all the individuals clustered based on their dispersal history (disperser or non-disperser) (Figure 4.1). Furthermore, hierarchical clustered heatmaps detected two main clusters encompassing uniquely co-regulated genes associated with dispersal history (Figure 4.2). A larger cluster included genes upregulated in dispersers and conspicuously down-regulated in non-dispersers (gene set I; n= 56 or 60 % of DEGs), while a second, smaller cluster of genes were upregulated in non-dispersers and conspicuously down-regulated in dispersers (gene set II; n= 38 or 40% of DEGs). In gene set I, I found 'ubiquinol-cytochrome c reductase core protein 2' (UQCRC2), involved in the mitochondrial respiratory chain, 'glycogen synthase' (GYS), involved in glycogenesis, and 'isocitrate dehydrogenase1, 2' (IDH1/IDH2), involved in the citric acid cycle to be the most highly significantly upregulated genes in dispersers (FDR < 0.00001). In gene set II, the three top genes showing highly significant upregulation in non-dispersers were F-type H+-transporting ATPase' (ATPeF0F6), involved in ion transporting, 'diacylglycerol O-acyltransferase' (DGAT), involved in triacylglycerol synthesis, and 'phospholipid: diacylglycerol acyltransferase' (PDAT), involved glycerolipid metabolism (FDR < 0.00001). None of the DEGs were shared between disperser and non-disperser individuals (Venn: Figure C1a).

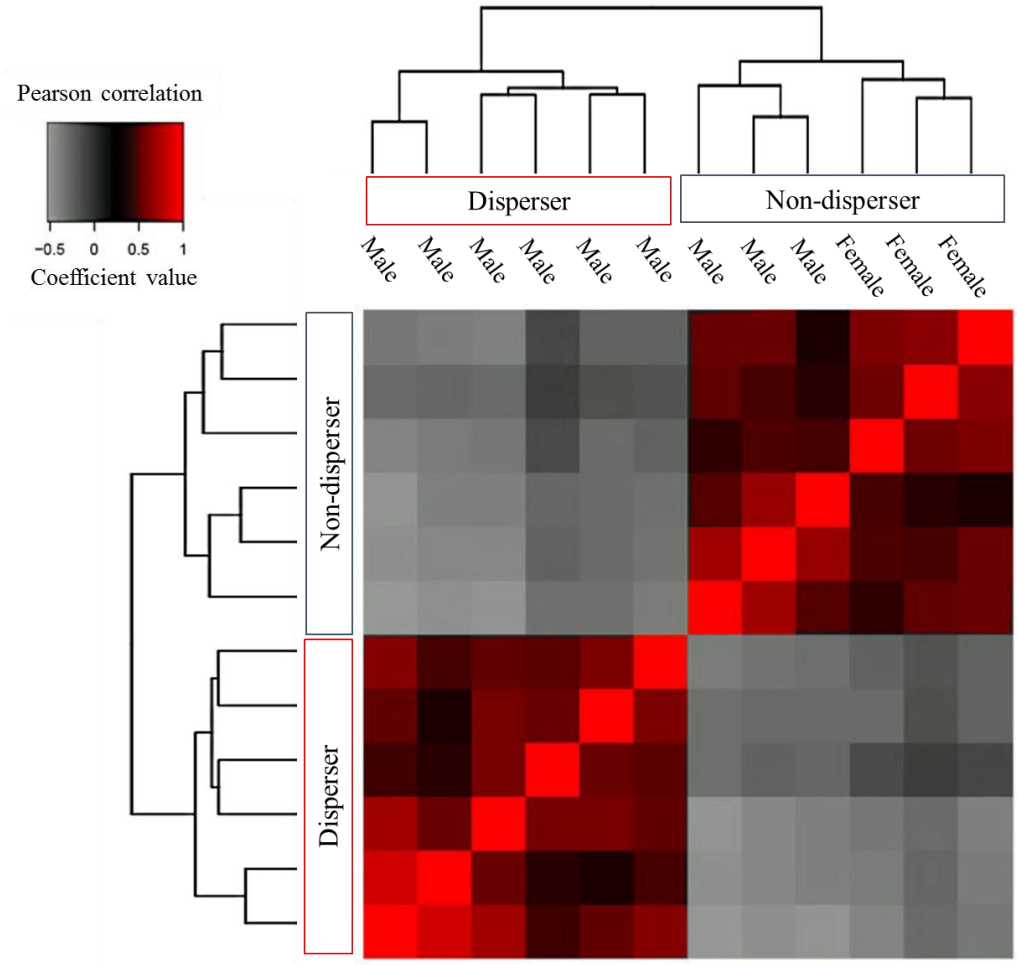


Figure 4.1 Cluster analysis based on Pearson correlation coefficients showing similarity of gene expression profiles in *Parnassius smintheus* samples. Samples are classified based on their dispersal history as inferred by mark-recapture data. Red is indicative of similarity, while grey is indicative of dissimilarity in the level of gene expression. Each cell represents the average correlation coefficient of a set of n = 94 dispersal-related genes. The diagram is symmetric across the red-cell diagonal.

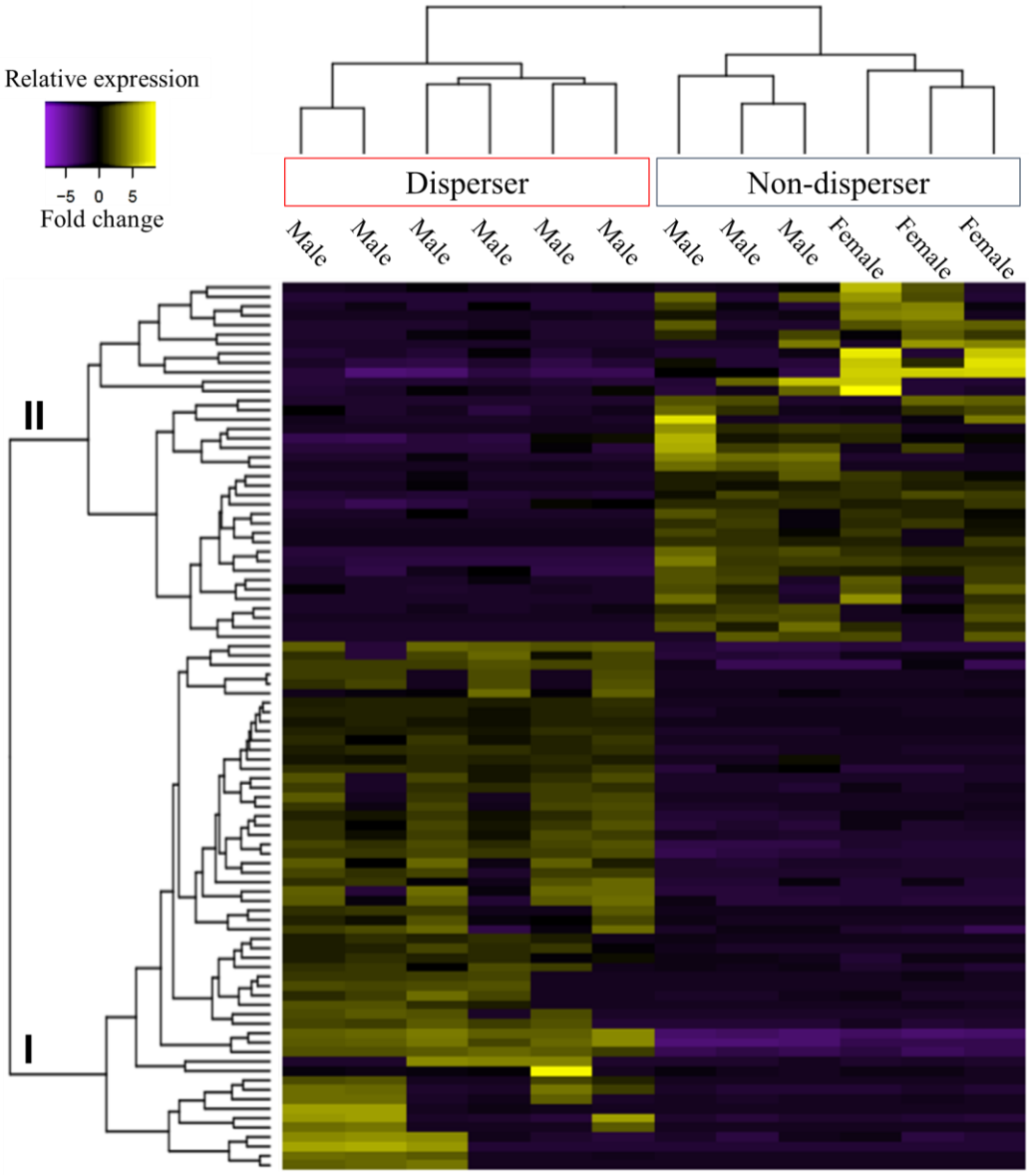


Figure 4.2 Heat map matrix of 94 genes differentially expressed between disperser and non-disperser *Parnassius smintheus* (FDR <0.05 and minimum four-fold change). The colour code represents the relative expression, where yellow represents upregulation, purple represents down-regulation, and black represents no change in expression. Genes were clustered by means of a hierarchical clustering algorithm presenting two gene sets, I and II (vertical axis).

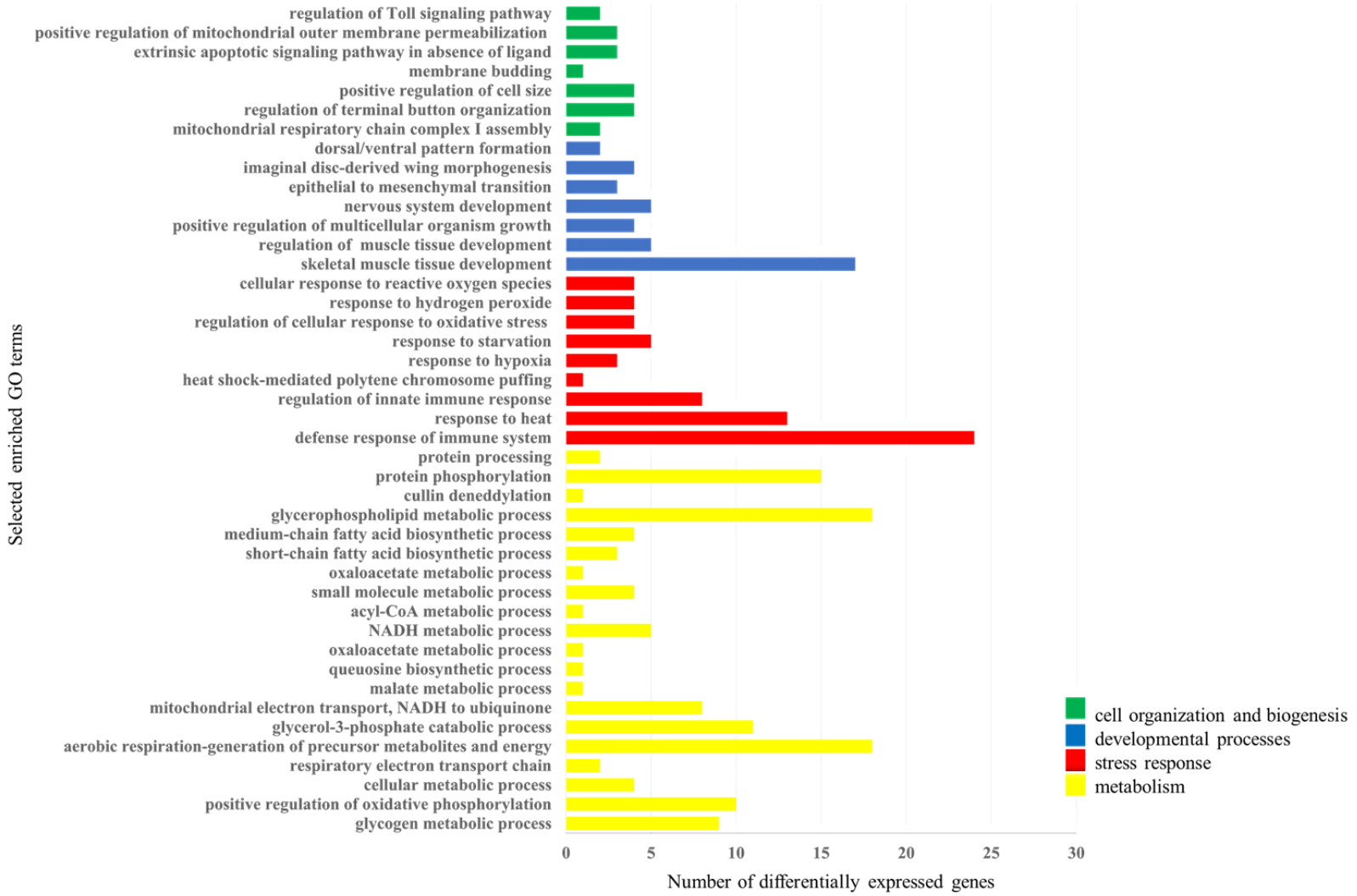
After removing the females from the analysis, the results still showed significant differential gene expression among males with different dispersal histories (Figures C2 and C3). I identified a total of 90 DEGs between male dispersers and non-dispersers, the majority of which overlapped with the 94 DEGs previously identified using all dispersers (overlap of 56 genes). Of these 90 DEGs, 30 (33% of DEGs) were upregulated in dispersers and 60 (66% of DEGs) were upregulated in non-dispersers. Thus, the gene expression profiles of dispersers and non-dispersers were distinct, regardless of sex.

Gene set I (56 DEGs) was significantly enriched for 122 GO terms related to "biological processes', which was over half (70%) of all GO terms associated with this gene set. Metabolic genes accounted for many of the upregulated transcripts in dispersers. Over 36% of transcripts upregulated in these individuals were involved in metabolism of carbohydrate, protein and lipid (e.g., glycerophospholipid metabolism, protein phosphorylation and oxidative phosphorylation), while 15% were involved in cellular regulation of stress response (e.g., immune response, oxidative stress, heat and hyperoxia responses), and 13% were involved in developmental processes (e.g., development of muscle tissue and nervous system). In the latter category of developmental processes, the largest number of genes upregulated in dispersers were associated with the specific GO term 'skeletal muscle development', which in the context of insects can be interpreted as striated muscle development. Under stress responses, both 'defense response of immune system' and 'response to heat' were represented by a large number of upregulated genes.

Gene set II (38 DEGs) was enriched for 65 biological process (BP) GO terms, which was over 60% of all GO terms associated with non-dispersers. Of the upregulated biological processes in non-dispersers, approximately 30% of transcripts were transport-

related (e.g., lipid transport and ion transport), 22% were involved in metabolism (mostly enriched kinase activity and mitochondrial electron transport), and 18% were involved in cell organization and biogenesis (Figure 4.3 a, b).

a)


regulation of Toll signaling pathway
positive regulation of mitochondrial outer membrane permeabilization
extrinsic apoptotic signaling pathway in absence of ligand
membrane budding
positive regulation of cell size
regulation of terminal button organization
mitochondrial respiratory chain complex I assembly
dorsal/ventral pattern formation
imaginal disc-derived wing morphogenesis
epithelial to mesenchymal transition
nervous system development
positive regulation of multicellular organism growth
regulation of muscle tissue development
skeletal muscle tissue development
cellular response to reactive oxygen species
response to hydrogen peroxide
regulation of cellular response to oxidative stress
response to starvation
response to hypoxia
heat shock-mediated polytene chromosome puffing
regulation of innate immune response
response to heat
defense response of immune system
protein processing
protein phosphorylation
cullin deneddylation
glycerophospholipid metabolic process
medium-chain fatty acid biosynthetic process
short-chain fatty acid biosynthetic process
oxaloacetate metabolic process
small molecule metabolic process
acyl-CoA metabolic process
NADH metabolic process
oxaloacetate metabolic process
queuosine biosynthetic process
malate metabolic process
mitochondrial electron transport, NADH to ubiquinone
glycerol-3-phosphate catabolic process
aerobic respiration-generation of precursor metabolites and energy
respiratory electron transport chain
cellular metabolic process
positive regulation of oxidative phosphorylation
glycogen metabolic process
Selected enriched GO terms
0
5
10
15
20
25
30
Number of differentially expressed genes
cell organization and biogenesis
developmental processes
stress response
metabolism

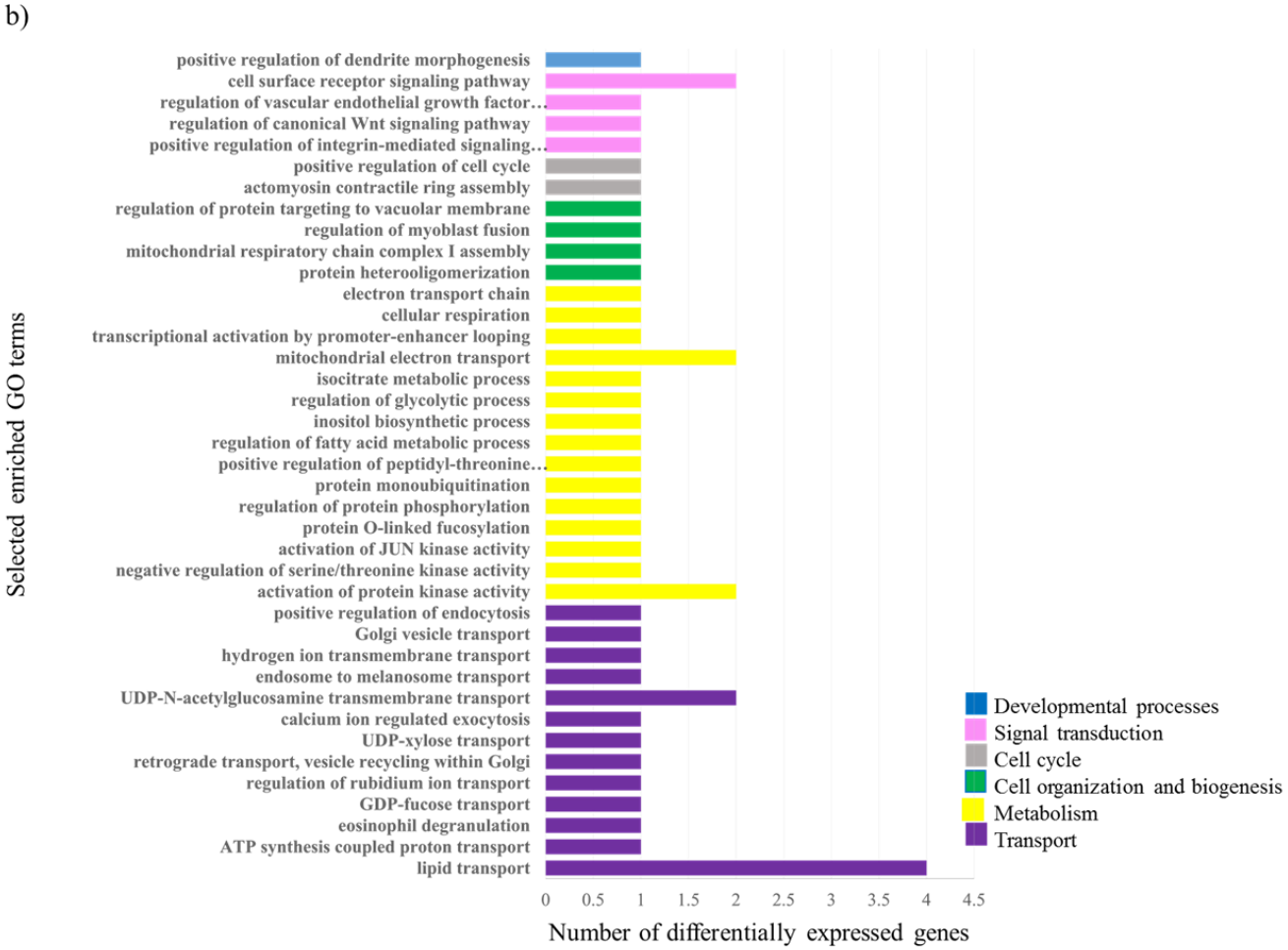


Figure 4.3 Gene ontology (GO) terms associated with upregulated differentially expressed genes (DEGs; FDR < 0.05) assigned to the biological process (BP) category in a) disperser and b) non-disperser individuals. The vertical axis shows the BP-GO terms classified by color; the horizontal axis represents the number of the DEGs annotated in the GO terms.

I identified 27 unique KEGG pathways associated with the set of DEGs upregulated in dispersers and conspicuously downregulated in non-dispersers (i.e., gene set I; Figure 4.4). The most frequently occurring pathways were related to ‘lipid metabolism’, ‘carbohydrate metabolism’,’ nucleotides metabolism’, and ‘energy metabolism’. Of the genes upregulated in dispersers, 45% were involved in pathways related to ‘citrate cycle’, ‘glycolysis / gluconeogenesis’, ‘oxidative phosphorylation’, ‘glycerophospholipid metabolism‘ and ‘fatty acid biosynthesis’. Most of the DEGs involved in these pathways are associated with processes in which stored energy is released and ATP is formed. The second largest category of genes upregulated in dispersers, encompassing 32% of the DEGs, was related to environmental information processing (signal transduction and membrane transport). In this category, I identified five signaling pathways (VEGF, cAMP, AMPK, MAPK and PI3K-Akt). Furthermore, two important pathways were identified, namely ‘thermogenesis’ and ‘insulin signaling’ in addition to pathways related to ‘immune system’, ‘Toll-like receptor signaling’ and ‘T cell receptor signaling’.

Figure 4.4 Heat map of KEGG pathways associated with significant DEGs (FDR < 0.05) enriched between dispersers and non-dispersers, The intensity of color indicates the level of regulation of pathways: Darker color represents higher upregulation of pathways and lighter color represents higher downregulation of pathways. For detailed information for each pathway, see the KEGG online resource (http://www.genome.jp/kegg/)

In gene set II (upregulated in non-dispersers and conspicuously down-regulated in dispersers), many of the 14 identified KEGG pathways (Figure 4.4) were related to metabolism (70%), namely, ‘oxidative phosphorylation’, ‘Inositol phosphate metabolism’, ‘galactose metabolism’ and ‘starch and sucrose metabolism’ and ’glycolipid metabolism’. I identified three signaling pathways including FoxO, TNF and MAPK as well as the ‘thermogenesis’ pathway. While some of the upregulated pathways in non-disperser individuals were similar to those in dispersers, the DEGs involved in each pathway differed between the two groups.

4.3.4 Differential gene expression as a function of thoracic temperature

I identified a total of 26 genes that were differentially expressed in individuals with thoracic temperature higher than ambient, compared to individuals with thoracic temperature lower than ambient (grouped regardless of dispersal history). All the individuals clustered by their thoracic temperature relative to ambient (Figure 4.5). Moreover, hierarchical clustering of these temperature-biased genes revealed two main clusters of co-regulated genes (Figure 4.6). The first gene set (hereafter ‘gene set TI’), with five genes (19% of DEGs), is uniquely upregulated in individuals with thoracic temperature higher than ambient, and conspicuously downregulated in individuals with thoracic temperature lower than ambient. A second, larger gene set with 21 genes (hereafter ‘gene set TII’; 81% of DEGs) was upregulated in individuals with thoracic temperature lower than ambient.

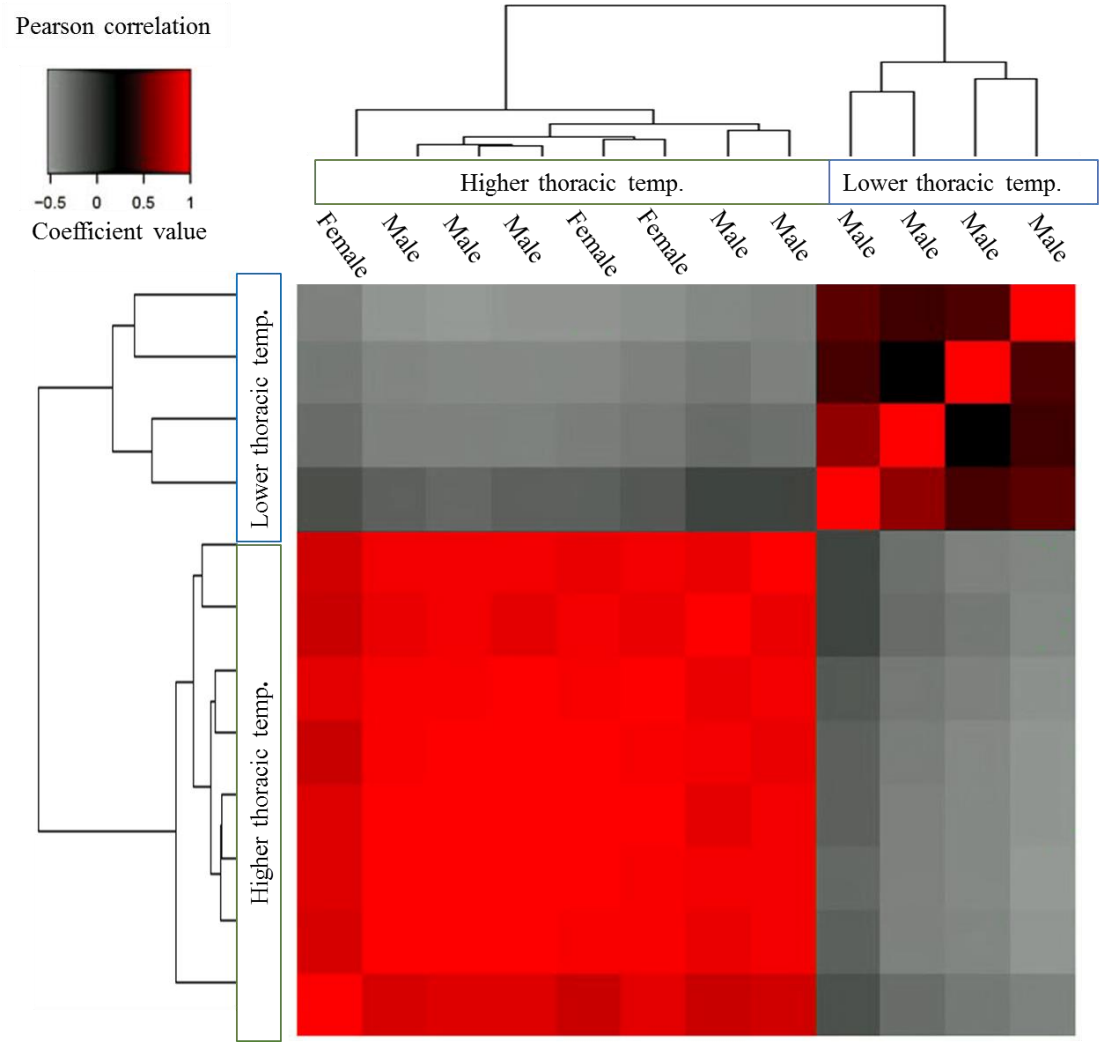


Figure 4.5 Cluster analysis based on Pearson correlation coefficients showing similarity of gene expression profiles in *Parnassius smintheus* samples. Samples are classified based on whether thoracic temperature was higher or lower than ambient temperature. Red is indicative of similarity, while grey is indicative of dissimilarity in the level of gene expression. Each cell represents the average correlation coefficient of a set of n = 26 genes. The diagram is symmetric across the red-cell diagonal.

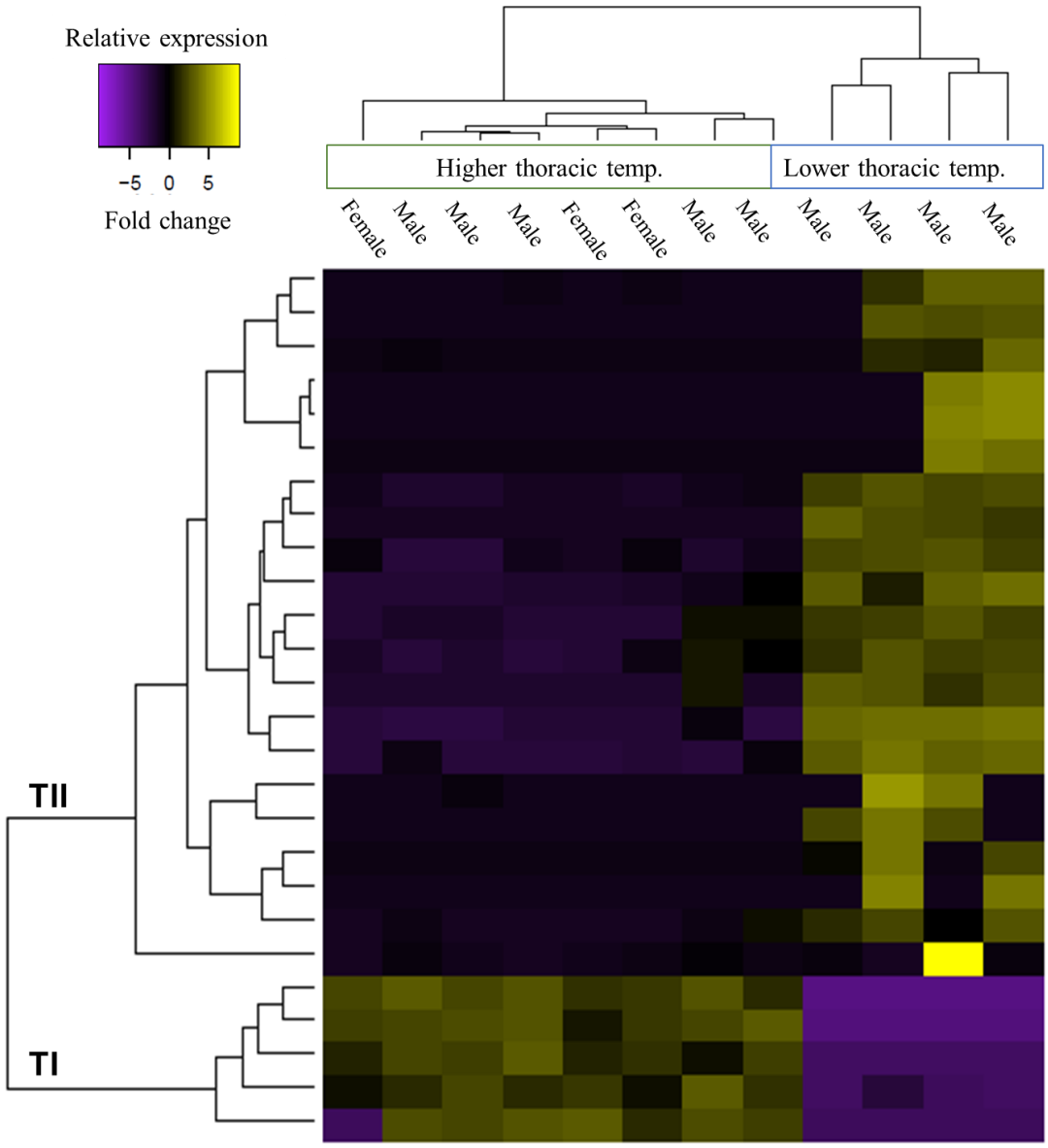


Figure 4.6 Heat map matrix of 26 genes differentially expressed between individuals with thoracic temperature higher and lower than ambient temperature (FDR < 0.05 and minimum four-fold change). The colour code represents the relative expression, where yellow represents upregulation, purple represents down-regulation, and black represents no change in expression. Genes were clustered by means of a hierarchical clustering algorithm presenting two gene sets, TI and TII (vertical axis).

The genes showing highly statistically significant upregulation in gene set TI were ’ornithine decarboxylase 1’, involved in polyamine biosynthesis processes and ‘growth hormone-inducible transmembrane’, involved in signaling and cellular processes, while in gene set TII ‘heat shock protein family A Member 9 (HSPA9), involved in Protein folding/chaperone, ‘xanthine dehydrogenase’ (XDH), involved in the oxidative metabolism of purines and ‘p38 mitogen-activated protein kinases‘ (p38 MAPK), involved in a signaling cascade controlling cellular responses to stress were the three top significant upregulated genes (FDR<0.00001). None of the DEGs were shared between these two groups of individuals (Venn: Figure C1b), although, I found five upregulated genes shared between dispersers and individuals that had lower thoracic than ambient temperature - lldD; MAPK; HSPA9; GSK3B; and XDH (Venn: Figure C1c).

After removing females from the analysis, I still observed a similar pattern of differential gene expression between males that had higher thoracic temperature than ambient, compared to males that had lower thoracic temperature than ambient (Figure C4 and C5). I found nine upregulated genes (out of 24; 37% of DEGs) in males with higher thoracic temperature than ambient and 15 upregulated (out of 24; 63% of DEGs) in those with lower thoracic temperature than ambient. All these DEGs identified in males were a subset of the 26 DEGs previously reported for all individuals in the full dataset.

I observed only ten biological process (BP) GO annotations assigned to upregulated DEGs in gene set TI (Figure 4.7a). The most abundant annotations were related to ‘cell organization and biogenesis’ and ‘metabolism’, which accounted for approximately 40% and 30 % of all BP-GO terms associated with this gene set, respectively. Gene set TII was significantly enriched for 66 BP-GO terms (Figure 4.7b).

Similar to the pattern in dispersers, genes pertaining to metabolism (of carbohydrate, lipid and protein) accounted for most of the upregulated transcripts (30%) in gene set TII, followed by genes related to cell organization and biogenesis (e.g., regulation of cell size and myofibril assembly), developmental processes (particularly striated muscle development) and stress responses (e.g., response to heat and hypoxia, and immune response), with approximately 17% of upregulated transcripts associated with each of these latter categories.

Using enrichment analysis of differentially expressed KEGG pathways, only five pathways ('biosynthesis of antibiotics', 'biosynthesis of secondary metabolites', 'arginine and proline metabolism', 'glutathione metabolism', and 'ribosome') showed a high degree of enrichment in gene set TI (Figure 4.8). Upregulated DEGs in gene set TII were distributed across 16 different pathways (Figure 4.8). Among these 16 pathways, the signaling transduction pathways were prominent, with 40% of upregulated genes involved in VEGF, MAPK, and calcium signaling pathways. I also found enrichment of 'insulin signaling', 'neurotrophin signaling' and 'Toll-like receptor signaling' pathways in gene set TII, along with multiple pathways related to metabolism of carbohydrates, energy, and nucleotides ('glycolysis/gluconeogenesis', 'glycogen and sucrose metabolism', and 'nucleotides metabolism').

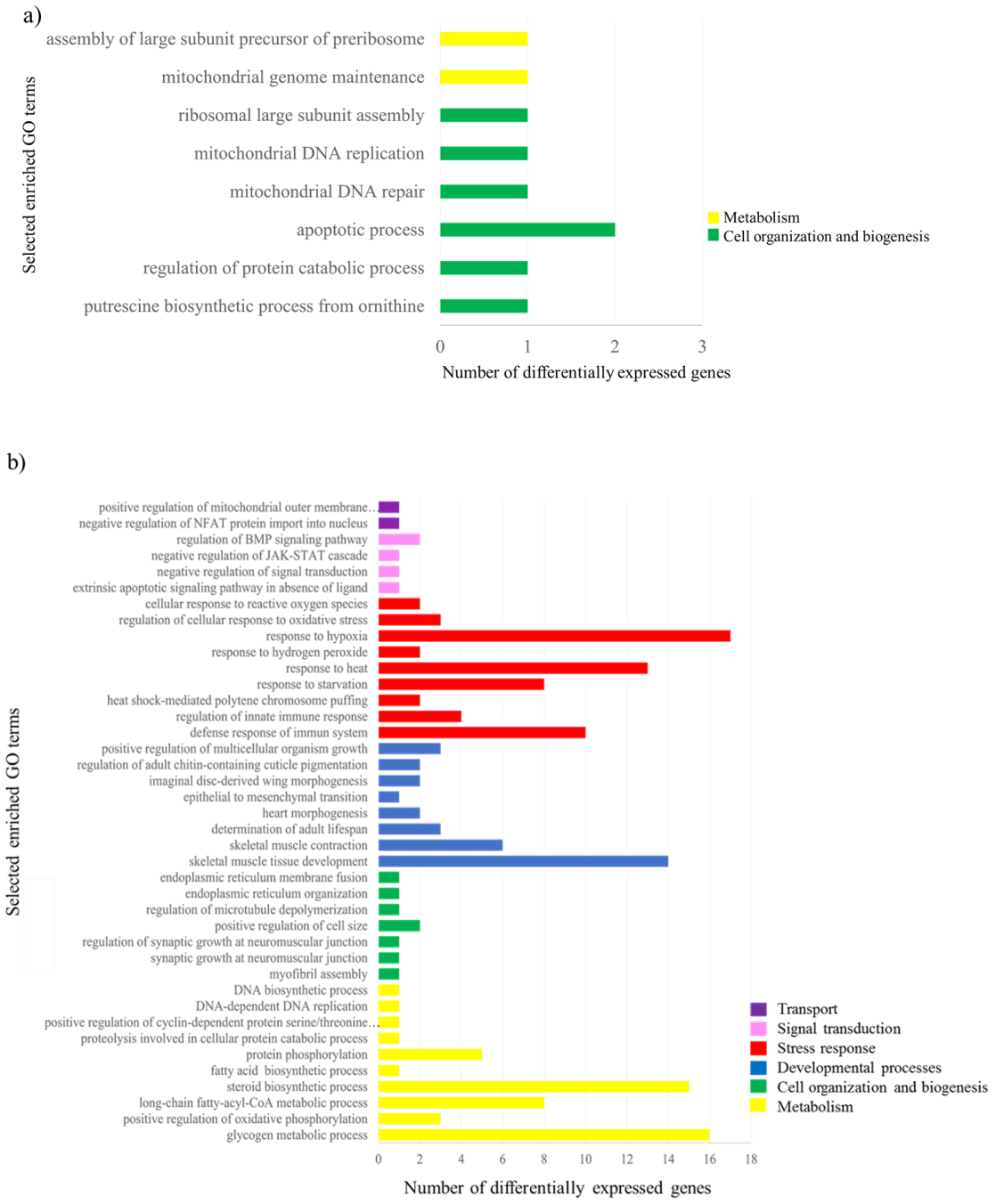


Figure 4.7 Gene ontology (GO) terms associated with upregulated differentially expressed genes (DEGs; FDR < 0.05) assigned to the biological process (BP) category in a) individuals with higher thoracic temperature than ambient and b) with lower thoracic temperature than ambient. The vertical axis shows the BP-GO terms classified by color; the horizontal axis represents the number of the DEGs annotated in the GO terms**.**

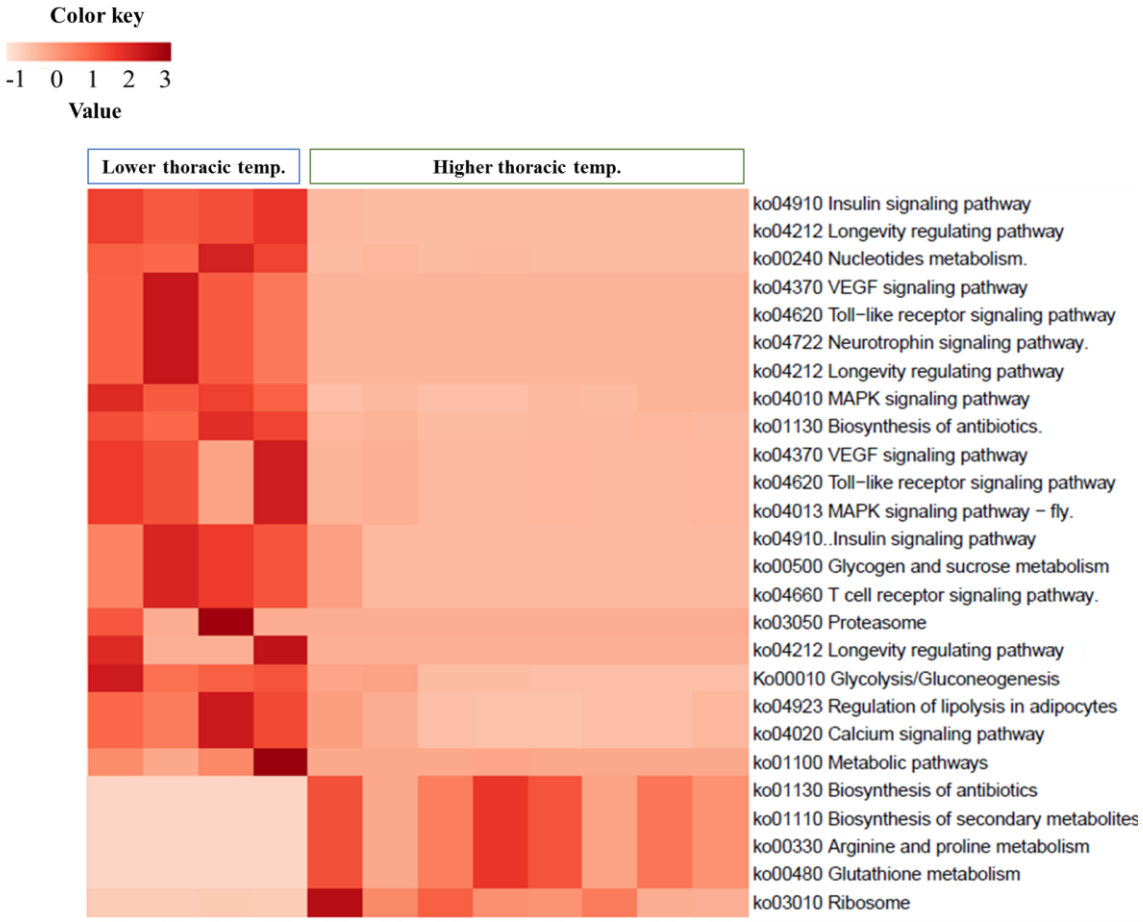


Figure 4.8 Heat map of KEGG pathways associated with significant DEGs (FDR < 0.05) enriched between individuals with higher and lower body temperature than ambient. The intensity of color indicates the level of regulation of pathways: Darker color represents higher upregulation of pathways and lighter color represents higher downregulation of pathways. For detailed information for each pathway, see the KEGG online resource (http://www.genome.jp/kegg/).

4.4 Discussion

Dispersal is an important behaviour that can be characterized at ecological, physiological and genetic levels (Clobert et al. 2012). Here, I attempt to integrate across all three levels. In a system in which the ecological consequences of dispersal are relatively well understood (Roland et al. 2000; Matter et al. 2004), I assessed differences in gene expression between dispersers and non-dispersers and in the context of ambient and body temperature.

The thoracic temperatures I observed in *P. smintheus* are generally consistent with a previous study by Guppy (1986), which documented thoracic temperatures of 17 $^{\circ}$C - 20 $^{\circ}$C during flight in the early morning in this species. While I also observed some higher thoracic temperatures, up to 32.6 $^{\circ}$C, this likely reflects the fact that I did not restrict my measurements to the early morning flight period as Guppy (1986) had. The range of flight thoracic temperatures I observed is also similar to that seen in monarch butterflies in their wintering grounds (~ 17 $^{\circ}$C – 28 $^{\circ}$C; Masters et al. 1988).

I found that thoracic temperature, relative to ambient, was significantly lower in dispersers compared to non-dispersers. Indeed, all non-dispersers had thoracic temperatures higher than ambient, while the majority of dispersers had thoracic temperatures lower than ambient. This result is consistent with the hypothesis that dispersers may have a better ability to dissipate the heat that accumulates during flight as a result of increased flight muscle activity and metabolic rate. In the long-distance flights associated with dispersal, maintenance of a lower thoracic temperature may be important to avoid overheating and dehydration (Kingsolver and Watt 1983; Masters et al. 1988;

Neve and Hall 2016). In the speckled wood butterfly, *Pararge aegeria*, Van Dyck and Matthysen (1988) found that patrolling males, which tend to fly longer distances, had lower thoracic temperatures than perching males, who tended to have short flights, even when engaged in the same activity such as basking. Thoracic temperature between the two types of males did not differ at the initiation of flight however, suggesting that the patrollers were more effectively dissipating heat from their thoraxes. Masters et al. (1988) also found that the monarch butterfly, *Danaus plexippus* (L.), adopts thermoregulatory behaviors to reduce thoracic temperature during flight, specifically alternating periods of powered flight with gliding.

A second, related hypothesis that can explain the lower relative thoracic temperatures of dispersers is that dispersers maintain low body temperatures as a general strategy to conserve energetic resources. In insects, higher body temperature results in greater energy utilization (Chaplin and Wells 1982). At the same time, long-distance flight is a very energetically demanding activity requiring stored energy reserves. Lipids in particular are an important fuel source for insect flight, and individuals with greater lipid reserves may be more likely to be successful in dispersal (Chaplin and Wells 1982). Migrating monarch butterflies maintain low body temperatures during flight to avoid rapid depletion of lipid reserves (Calvert and Brower 1986; Masters et al. 1988). Thus, in *P. smintheus* dispersers may be those individuals that are better able to build up or maintain energy reserves by reducing their body temperature while in flight and possibly during other activities. If maintenance of lower body temperature is important for dispersers to either avoid overheating or reduce energy consumption, then I would predict

that dispersers may have additional adaptations to reduce body temperature including lighter wing color, or less hairy bodies and wing bases, as compared to non-dispersers.

The initiation and completion of dispersal and long-distance flight, as well as the subsequent recovery, require integration of a number of physiological mechanisms, including control of movement, body temperature, and energy utilization, which can be mediated by the neuroendocrine and hormonal systems (Ramenofsky and Wingfield 2007). These physiological mechanisms, in turn, may be modulated through the regulation of gene expression (Margotta et al. 2013; Somervuo et al. 2014, Kvist et al. 2015). Using an RNA-seq approach, I observed marked differences in gene expression between individuals of *P. smintheus* classified as dispersers and non-dispersers. I obtained 94 DEGs between dispersers and non-dispersers, even though I was not capturing individuals during or immediately after dispersal. I also observed differential gene expression between individuals with thoracic temperature higher versus lower than ambient temperature, but these effects were not as strong as the patterns observed between dispersers and non-dispersers and involved fewer DEGs. I identified 26 DEGs between individuals with thoracic temperature higher versus lower than ambient. Since dispersers had lower thoracic than ambient temperature during flight, DEGs associated with lower body temperature were a subset of DEGs identified in dispersers and notably included a number of genes associated with the GO term ‘response to heat’.

Those individuals I classified as non-dispersers (i.e., recaptured in the same meadow where they were first marked) could potentially have been dispersers that were simply not identified as such based on their recapture history (i.e., they had already immigrated to a new patch before their first capture, or they went to a different patch and

returned between capture events). However, the large majority of individuals captured and marked in this system are not dispersers (Roland et al. 2000), so the likelihood of these individuals being unidentified dispersers is very low. Indeed, the striking differences in gene expression between the two groups, and consistency in gene expression among individuals within each group, suggest that the individuals that I classified as 'non-dispersers' were a homogeneous group.

The DEGs I identified in dispersers, who also had lower thoracic than ambient temperature, relate to a variety of physiological functions. My results suggest upregulation of a suite of genes involved in metabolism, stress responses (including response to heat) and striated muscle development in individuals that have moved between habitat patches. These findings are in accordance with my hypothesis that differences in expression of genes linked to processes including energy mobilization and thermoregulation may be either a cause or consequence of dispersal. More specifically, the nature of the genes and associated pathways upregulated in dispersers suggest that these individuals might be attempting to recover from the rigours of dispersal by replenishing energy stores, growing or repairing muscle, and regulating heat and hypoxia stress.

Most of the energy required during long term and sustained insect flight is obtained from stores of glycogen and triglycerides (Mordue et al. 1980; Arrese and Soulages 2010). I found that dispersers had elevated expression of genes involved in glycerophospholipid and glycerolipid metabolism, glycogen metabolism, and oxidative phosphorylation. Some of these genes related to fat and carbohydrate accumulation. For example, the gene *Gdp1* encoding glycerol-3-phosphate dehydrogenase, involved in

glycerophospholipid metabolism, showed significantly higher upregulation in dispersers. The encoded enzyme has a critical role in carbohydrate and lipid metabolism by catalyzing the reversible conversion of dihydroxyacetone phosphate (DHAP) to glycerol-3-phosphate (G3P), or triglyceride, which is stored in fat body cells (Nye et al. 2008; Mráček et al. 2013). I also observed increased expression of the gene encoding glycogen synthase, which is a key enzyme in glycogenesis and mediates the conversion of glucose into glycogen, in dispersers. These results suggest a shift to increased fat and carbohydrate accumulation in dispersers. In the Glanville fritillary butterfly, genes related to utilization of energy stores, specifically oxidative phosphorylation, glycolysis, TCA cycle and ATP metabolism were down-regulated at 20 hours after an intense and long bout of flight, reflecting potential depletion of energy stores and a resulting decrease in the rate of energy metabolism (Kvist et al. 2015). Such a depletion of energy stores likely also occurs in *P. smintheus* during dispersal and long flights, and here I may have observed gene expression profiles reflecting the longer-term post-flight replenishment of energy reserves. However, I cannot say at this point whether dispersers were indeed replenishing lost carbohydrate and lipid reserves from a previous long-term flight or are simply always better able to accumulate such reserves relative to non-dispersers.

I also observed that dispersers had higher expression of some genes involved in catabolism and releasing stored energy to cells. For example, dispersers displayed elevated expression of the gene encoding the enzyme triacylglycerol lipase, which catalyzes the hydrolysis of triacylglycerol to glycerol and fatty acids. Dispersers also upregulated genes involved in TCA cycle and oxidative phosphorylation. The release of stored energy may be coupled to the other key processes that were upregulated in

dispersers, notably muscle development and stress responses. Upregulation of genes related to development, cell organization, and biogenesis, particularly development of striated muscle, is consistent with a recovery response to intense and sustained exercise. These processes, and the regeneration of muscular tissue in particular, may demand considerable energetic resources.

The ability to cope with thermal stress is also critical for sustaining a long flight. Dispersers exhibited upregulation of a number of stress-related genes, including many involved in response to heat. For example, dispersers displayed upregulation of the gene that encodes heat shock protein family A (HSP70A), representing a potential response to thermal stress. Heat shock proteins function as molecular chaperones to maintain correct protein folding, and assist in refolding of damaged proteins following heat shocks or other stressful conditions (Wang and Lindquist 1998; Štětina et al. 2015). The heat shock protein HSP70s is known to be important in protecting insects from many aspects of thermal stress (Parsell and Lindquist 1993; Wang and Kang 2005; Luo et al. 2015). Indeed, *Hsp70* expression is considered an indicator of the intensity of stress (Iwama et al. 1998; Loeschcke and Hoffmann 2007). A second large set of stress-related genes that were upregulated in dispersers were genes involved in immune function. These included, for example, the gene encoding p38 MAPK, which is involved in VEGF and Toll-like receptor signaling pathways implicated in stress signals and regulation of immune response. MAPK is an important protein kinase that plays an essential role in regulating cellular processes, including apoptosis, cell fate determination, and immune function in response to various environmental stresses (Pearson et al. 2001; Zarubin and Han 2005). In *Drosophila*, it has been demonstrated that immune stimulation, heat shock, oxidative

stress and starvation all activate MAPK kinase signaling pathways (Han et al. 1998), and flies lacking D-p38 are susceptible to environmental stresses (Craig et al. 2004). Margotta et al. (2013) demonstrated different gene expression patterns in flight muscles of forager honeybees that engaged in intense flight and those of nurse honeybees that engaged in little or no flight. One key finding was that the expression of genes involved in immune signaling pathways (*toll-like receptor* and *hopscotch*) were upregulated in foragers. Previous studies on the Glanville fritillary also showed that intensive flight activity enhanced the immune response (Saastamoinen and Rantala 2013) and led to increased expression of immune genes (Kvist et al. 2015). My data suggest that the thermal and metabolic pressures, and potentially reactive oxygen species (ROS) production, induced by prolonged flight bouts also stimulate immune-related pathways in dispersing *P. smintheus*.

Overall, I was able to generate a high-quality and clean reference transcriptome for adult *P. smintheus* thoracic tissue, which can be used for future comparative analyses. My results show that *P. smintheus* individuals that have dispersed between habitat patches have lower relative thoracic temperatures and distinct gene expression profiles compared to non-dispersers. The fact that many of the DEGs relate to replenishment of energy reserves, muscle development and stress responses, suggests that the gene expression patterns I observed in dispersers represent a long-term ‘recovery’ response to long-distance flight and dispersal. My results suggest that dispersal may indeed be quite costly, since individuals are upregulating the expression of genes for replenishing lipid and carbohydrate reserves, tissue re-organization, as well as genes involved in stress responses, potentially days after having dispersed.

I cannot entirely rule out, that some of the gene-expression differences I observed may be inherent to dispersing and non-dispersing individuals, independent from the effect of flight activity (Kvist et al. 2015). That is, given that the other studies have demonstrated a heritable genetic basis of dispersal traits (Kent et al. 2009; Saastamoinen et al. 2012; Edelsparre et al. 2014), it is possible that some differences I observed between dispersers and non-dispersers reflect a certain degree of functional specialization, and may in fact be a cause rather than consequence of dispersal. Baseline differences in gene expression between individuals from populations showing different mean dispersal behaviours (Somervuo et al. 2014; Kvist et al. 2015) support this possibility. Manipulation in expression of the foraging gene (*for*) in larval fruit flies influences their foraging activity and dispersal tendencies as adults (Kent et al. 2009), providing further evidence that differential gene expression can lead to different dispersal behaviours. Controlled flight trial studies in *P. smintheus* are needed to determine short- and long-term effects of sustained flight on gene expression, and to shed light on those gene expression differences that may be a cause versus a consequence of dispersal. These and additional studies investigating differences between dispersers and non-dispersers, including morphological and colour differences, the thoracic temperature required to initiate flight, and flight metabolic rates will yield further insights into the costs and constraints associated with dispersal in this system.

4.5 Literature Cited

Andrews S (2010). FastQC: A quality control tool for high throughput sequence data.

Arrese EL, Soulages JL (2010). Insect fat body: Energy, metabolism, and regulation. *Annual Review of Entomology* 55: 207–225.

Berwaerts K, Dyck HV, Vints E, Matthysen E (2001). Effect of manipulated wing characteristics and basking posture on thermal properties of the butterfly *Pararge aegeria* (L.). *Journal of Zoology* 255: 261–267.

Bolger AM, Lohse M, Usadel B (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.

Bonte D, Van Dyck H, Bullock JM, Coulon A, Delgado M, Gibbs M, et al. (2012). Costs of dispersal. *Biological Reviews* 87: 290–312.

Bowler DE, Benton TG (2005). Causes and consequences of animal dispersal strategies: Relating individual behaviour to spatial dynamics. *Biological Reviews* 80: 205–225.

Calvert WH, Brower LP (1986). The location of monarch butterfly (*Danaus plexippus* L.) overwintering colonies in Mexico in relation to topography and climate. *Journal of the Lepidopterists' Society* 40: 164–187.

Chaplin SB, Wells PH (1982). Energy reserves and metabolic expenditures of monarch butterflies overwintering in southern California. *Ecological Entomology* 7: 249–256.

Clobert J, Baguette M, Benton TG, Bullock JM (2012). Dispersal ecology and evolution Oxford University Press. Oxford, UK.

Craig CR, Fink JL, Yagi Y, Ip YT, Cagan RL (2004). A *Drosophila* p38 orthologue is required for environmental stress responses. *EMBO Reports* 5: 1058–1063.

Eddy SR (2011). Accelerated profile HMM searches. *PLoS Computational Biology* 7: e1002195.

Edelsparre AH, Vesterberg A, Lim JH, Anwari M, Fitzpatrick MJ (2014). Alleles underlying larval foraging behaviour influence adult dispersal in nature. *Ecology Letters* 17: 333–339.

Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, et al. (2015). HMMER web server: 2015 update. *Nucleic Acids Research* 43: W30-38.

Fraser DF, Gilliam JF, Daley MJ, Le AN, Skalski GT (2001). Explaining leptokurtic movement distributions: Intrapopulation variation in boldness and exploration. *American Naturalist* 158: 124–135.

Gibbs M, Dyck HV (2010). Butterfly flight activity affects reproductive performance and longevity relative to landscape structure. *Oecologia* 163: 341–350.

Gomez GSY, Van Dyck H (2012). Ecotypic differentiation between urban and rural populations of the grasshopper *Chorthippus brunneus* relative to climate and habitat fragmentation. *Oecologia* 169: 125–133.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29: 644–652.

Guppy CS (1986). The Adaptive significance of alpine melanism in the butterfly *Parnassius phoebus* F. (Lepidoptera: Papilionidae). *Oecologia* 70: 205–213.

Haas BJ, Papanicolaou A, Yassour M, Levin JZ, Thompson DA, Amit l, et al (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* 8: 1494–1512.

Han SJ, Choi KY, Brey PT, Lee WJ (1998). Molecular cloning and characterization of a *Drosophila* p38 mitogen-activated protein kinase. *Journal of Biological Chemistry* 273: 369–374.

Heinrich B (1993). The hot-blooded insects. Strategies and mechanisms of thermoregulation. Springer-Verlag, Berlin.

Heinrich B (1995). Insect thermoregulation. *Endeavour* 19: 28–33.

Iwama GK, Thomas PT, Forsyth RB, Vijayan MM (1998). Heat shock protein expression in fish. *Reviews in Fish Biology and Fisheries* 8: 35–56.

Jones CM, Papanicolaou A, Mironidis GK, Vontas J, Yang Y, Lim KS, Oakeshott JG, Bass C, Chapman JW (2015). Genomewide transcriptional signatures of migratory flight activity in a globally invasive insect pest. *Molecular Ecology* 24: 4901–4911.

Kanehisa M, Goto S (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28: 27–30.

Kent CF, Daskalchuk T, Cook L, Sokolowski MB, Greenspan RJ (2009). The *Drosophila* foraging gene mediates adult plasticity and gene–environment interactions in behaviour, metabolites, and gene expression in response to food deprivation. *PLoS Genetics* 5: e1000609.

Keyghobadi N, Roland J, Strobeck C (2005). Genetic differentiation and gene flow among populations of the alpine butterfly, *Parnassius smintheus*, vary with landscape connectivity: landscape connectivity and gene flow. *Molecular Ecology* 14: 1897–1909.

Kingsolver JG, Watt WB (1983). Thermoregulatory strategies in *Colias* butterflies: Thermal stress and the limits to adaptation in temporally varying environments. *American Naturalist* 121: 32–55.

Kingsolver JG (1985). Butterfly thermoregulation: Organismic mechanisms and population consequence. *Journal of Research on the Lepidoptera* 24: 1–20.

Kvist J, Mattila ALK, Somervuo P, Ahola V, Koskinen P, Paulin L, et al. (2015). Flight-induced changes in gene expression in the Glanville fritillary butterfly. *Molecular Ecology* 24: 4886–4900.

Langmead B, Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9: 357–359.

Li B, Dewey CN (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12: 323.

Loeschcke V, Hoffmann AA (2007). Consequences of heat hardening on a field fitness component in *Drosophila* depend on environmental temperature. *The American Naturalist* 169: 175–183.

Luo S, Ahola V, Shu C, Xu C, Wang R (2015). Heat shock protein 70 gene family in the Glanville fritillary butterfly and their response to thermal stress. *Gene* 556: 132–141.

Margotta J, Mancinelli G, Benito A, Ammons A, Roberts S, Elekonich M (2012). Effects of flight on gene expression and aging in the honey bee brain and flight muscle. *Insects* 4: 9–30.

Masters AR, Malcolm SB, Brower LP (1988). Monarch butterfly (*Danaus Plexippus*) thermoregulatory behavior and adaptations for overwintering in Mexico. *Ecology* 69: 458–467.

Matter SF, Keyghobadi N, Roland J (2014). Ten years of abundance data within a spatial population network of the alpine butterfly, *Parnassius smintheus*. *Ecology* 95: 2985–2985.

Matter SF, Roland J (2004). Relationships among population estimation techniques: an examination for *Parnassius smintheus*. *Journal of the Lepidopterists' Society* 58: 189–195.

Mattila ALK (2015). Thermal biology of flight in a butterfly: genotype, flight metabolism, and environmental conditions. *Ecology and Evolution* 5: 5539–5551.

Metsalu T, Vilo J (2015). ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Research 43*: W566–W570.

Meylan S, de Fraipont M, Aragon P, Vercken E, Clobert J (2009). Are dispersal-dependent behavioral traits produced by phenotypic plasticity? *Journal of Experimental Zoology-A* 311: 377–388.

Mordue W, Goldsworthy G, Brady J, Blaney WM (1980). Insect physiology. Blackwell Science Ltd. Oxford, UK.

Mráček T, Drahota Z, Houštěk J (2013). The function and the role of the mitochondrial glycerol-3-phosphate dehydrogenase in mammalian tissues. *Biochimica et Biophysica Acta* 1827: 401–410.

Neve G, Hall C (2016). Variation of thorax flight temperature among twenty Australian butterflies (Lepidoptera: Papilionidae, Nymphalidae, Pieridae, Hesperiidae, Lycaenidae). *European Journal of Entomology 113*: 571–578.

Niitepõld K, Smith AD, Osborne JL, Reynolds DR, Carreck N, Martin AP, et al. (2009). Flight metabolic rate and *Pgi* genotype influence butterfly dispersal rate in the. *Ecology* 90: 2223–2232.

Nye CK, Hanson RW, Kalhan SC (2008). Glyceroneogenesis is the dominant pathway for triglyceride glycerol synthesis *in vivo* in the Rat. *Journal of Biological Chemistry* 283: 27565–27574.

Parsell D, Lindquist S (1993). The Function of heat-shock proteins in stress tolerance: Degradation and reactivation of damaged proteins. *Annual Review of Genetics* 27: 437–496.

Pearson G, Robinson F, Beers Gibson T, Xu B, Karandikar M, Berman K, et al. (2001). Mitogen-activated protein (MAP) kinase pathways: regulation and physiological functions. *Endocrine Reviews* 22: 153–183.

Prange HD (1995). Evaporative cooling in insects. *Journal of Insect Physiology* 42: 493–499.

Ramenofsky M, Wingfield JC (2007). Regulation of migration. *BioScience* 57: 135–143.

Robinson MD, McCarthy D, Smyth GK (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139–149.

Roff DA (1986). The evolution of wing dimorphism in insects. *Evolution* 40, 1009–1020.

Roland J, Keyghobadi N, Fownes S (2000). Alpine *Parnassius* butterfly dispersal: Effects of landscape and population size. *Ecology* 8: 1642–1653.

Roland J, Matter SF (2007). Encroaching forests decouple alpine butterfly population dynamics. *Proceedings of the National Academy of Sciences of the United States of America* 104: 13702–13704.

Ronce O (2007). How does it feel to be like a rolling stone? Ten questions about dispersal evolution. *Annual Review of Ecology, Evolution, and Systematics* 38: 231–253.

Saastamoinen M, Bocedi G, Cote J, Legrand D, Guillaume F, Wheat CW, et al. (2018). Genetics of dispersal: Genetic of dispersal. *Biological reviews* 93: 574–599.

Saastamoinen M, Brakefield PM, Ovaskainen O (2012). Environmentally induced dispersal-related life-history syndrome in the tropical butterfly, *Bicyclus anynana*. *Journal of Evolutionary Biology* 25: 2264–2275.

Saastamoinen M, Rantala MJ (2013). Influence of developmental conditions on immune function and dispersal-related traits in the Glanville Fritillary (*Melitaea cinxia*) butterfly. *PLoS ONE* 8: e81289.

Schulz MH, Zerbino DR, Vingron M, Birney E (2012). Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 28: 1086–1092.

Schwander T, Leimar O (2011). Genes as leaders and followers in evolution. *Trends in Ecology & Evolution* 26: 143–151.

Sekar S (2012). A meta-analysis of the traits affecting dispersal ability in butterflies: can wingspan be used as a proxy? *Journal of Animal Ecology* 81: 174–184.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31: 3210–3212.

Somervuo P, Kvist J, Ikonen S, Auvinen P, Paulin L, Koskinen P, et al. (2014). Transcriptome Analysis Reveals Signature of Adaptation to Landscape Fragmentation (W Salzburger, Ed.). *PLoS ONE* 9: e101467.

Srygley R, Chai P (1990). Predation and the elevation of thoracic temperature in brightly colored Neotropical butterfly. *The American Naturalist* 135: 766–787.

Štětina T, Koštál V, Korbelová J (2015). The role of inducible Hsp70, and other heat shock proteins, in adaptive complex of cold tolerance of the fruit fly (*Drosophila melanogaster*). *PLOS ONE* 10: e0128976.

Travis JMJ, Delgado M, Bocedi G, Baguette M, Bartoń K, Bonte D, et al. (2013). Dispersal and species' responses to climate change. *Oikos* 122: 1532–1540.

Tsuji JS, Kingsolver JG, Watt WB (1985). Thermal physiological ecology of Colias butterflies in flight. *Oecologia* 69: 161–170.

UniProt (2014). Activities at the universal protein Resource (UniProt). *Nucleic Acids Research* 14: D191-198.

Van Dyck H, Matthysen E (1998). Thermoregulatory differences between phenotypes in the speckled wood butterfly: Hot perchers and cold patrollers? *Oecologia* 114: 326–334.

Vogt FD, Heinrich B (1983). Thoracic temperature variations in the onset of flight in dragonflies (Odonata: Anisoptera). *Physiological Zoology* 56: 236–241.

Wang XH, Kang L (2005). Differences in egg thermotolerance between tropical and temperate populations of the migratory locust *Locusta migratoria* (Orthoptera: Acridiidae). *Journal of Insect Physiology* 51: 1277–1285.

Wang Z, Lindquist S (1998). Developmentally regulated nuclear transport of transcription factors in *Drosophila* embryos enable the heat shock response. *Development* 125: 4841–4850.

Wickman P (2009). Thermoregulation and habitat use in butterflies. Pp. 55–61 in Settele J, Shreeve T, Konvicka M, Van Dyck H, eds. Ecology of butterflies in Europe. Cambridge University. Press, Cambridge.

Wong SC, Oksanen A, Mattila ALK, Lehtonen R, Niitepõld K, Hanski I (2016). Effects of ambient and preceding temperatures and metabolic genes on flight metabolism in the Glanville fritillary butterfly. *Journal of Insect Physiology* 85: 23–31.

Young MD, Wakefield MJ, Smyth GK, Oshlack A (2010). GO ontology analysis for RNA-seq: Accounting for selection bias. *Genome Biology* 11: R14

Zarubin T, Han J (2005). Activation and signaling of the p38 MAP kinase pathway. *Cell Research* 15: 11–18.

Chapter 5

5 Genetic variation at the *Pgi* locus is associated with dispersal in the alpine butterfly, *Parnassius smintheus*

5.1 Introduction

The phosphoglucose isomerase gene (*Pgi)* has been proposed as a candidate gene for a number of ecologically important traits, including movement and dispersal, in arthropods (Haag et al. 2005; Wheat et al. 2006; Kallioniemi and Hanski 2011). This gene encodes the enzyme phosphoglucose isomerase (PGI), which catalyzes the second step in glycolysis, converting glucose-6-phosphate (G6P) into fructose-6-phosphate (F6P). Polymorphism at *Pgi* has been described in a wide range of taxa, ranging from plants to insects to barnacles (Hoffman 1981; Filatov and Charlesworth 1999; Wheat 2010). Diverse lines of evidence, including DNA sequence variation, allelic clines, and allele frequency changes over time indicate selection on this locus in many different species (Hoffman 1981; Rank and Dahlhoff 2002; Wheat et al. 2006; Orsini et al. 2009). Furthermore, specific *Pgi* genotypes have been associated with higher components of fitness (survival, mating success and fecundity) and with performance in traits such as peak metabolic rate, running speed, and flight (Filatov and Charlesworth 1999; Dahlhoff and Rank 2000; Wheat et al. 2006; Orsini et al. 2009).

PGI is a dimeric enzyme involved in the early steps of glycolysis and therefore in a pathway that ultimately releases ATP and NADH to provide energy for cellular activity. PGI activity is considered to occur at a metabolic ‘branch point’ since the reaction it catalyzes is reversible (G6P ⇔ F6P) and G6P can enter alternative pathways

for the pentose phosphate shunt or glycogen biosynthesis (Wheat and Hill 2014). In mammals, additional potential activities for PGI (i.e., moonlighting activities) involving nerve cell growth and differentiation have also been identified (Chaput et al. 1988). Only a single copy of the *Pgi* gene is present in the large majority of organisms in which it has been studied, with some exceptions including certain plant species, fish, and stick insects (Thomas et al. 1993; Sato and Nishida 2007; Dunning et al. 2013).

In arthropods, researchers have demonstrated that PGI variants can differ in their biochemical performance (Watt et al. 1983; Zera 1987; Li and Andersson 2016), leading to differential physiological performance and fitness at the organismal level (Watt et al. 1983; Watt 1992). The association between *Pgi* variation and fitness and performance traits have been studied most extensively in the clouded yellow butterflies, *Colias* spp. and the Glanville fritillary, *Melitaea cinxia*. The first studies showing selection and fitness effects in this locus were the now classic allozyme surveys of *Colias eurytheme* by Watt (1977). In this and other *Colias* species, *Pgi* genotype affects survival, male mating success, female fecundity and flight activity (reviewed in Watt 2003). Similarly, in the Glanville fritillary, *Pgi* genotype predicts lifespan, female fecundity, larval growth and survival, and peak metabolic rate (Haag et al. 2005; Niitepõld et al. 2009; Orsini et al. 2009; Kallioniemi and Hanski 2011; Saastamoinen et al. 2012). Population-level effects of *Pgi* have also been documented in the Glanville fritillary, such that allele frequencies are associated with population growth rate, and a higher frequency of some genotypes is found in newly colonized and isolated populations (Haag et al. 2005; Hanski and Saccheri 2006). In both of these butterfly systems, there is evidence for balancing selection and heterozygote advantage at *Pgi* (Watt 1977; Wheat et al. 2006; Niitepõld et al. 2009).

However, in many other organisms selection on *Pgi* is primarily directional (Rank and Dahlhoff 2002; Karl et al. 2008). Even within *Colias*, molecular and field data indicate historical directional selection among species, but contemporary balancing selection within populations and species (Watt 2003; Wheat et al. 2006).

Two main hypotheses have been proposed to explain the association of *Pgi* variation with performance traits. The primary working hypothesis is the 'flux hypotheses' that suggests that PGI, as a key branch-point enzyme, regulates rates of glycolytic flux (Watt 1986, 2003). According to this hypothesis, genetic variation at *Pgi* is correlated with performance because it affects the supply of ATP via glycolysis to support peak physiological demands (Watt 1977, 1983; Watt and Dean 2000). The other hypothesis is the 'moonlighting hypothesis' that suggests that additional functions of PGI, aside from energy metabolism, result in epistatic effects of the *Pgi* gene. PGI is known for diverse moonlighting functions in other taxa, that are separate from its role in energy metabolism, such as acting as an autocrine motility factor or as a neuroleukin (Marden 2013). The moonlighting hypothesis proposes that genetic variation at *Pgi* affects these additional, moonlighting functions. Regardless of the underlying mechanism, differences in enzyme function determined by variation at the *Pgi* gene can translate into differences in metabolic performance and ultimately, dispersal ability or tendency (Storz and Wheat 2010).

Thermal environment appears to play a key role in driving selection associated with *Pgi*. *Pgi* allele frequencies vary along thermal clines in diverse taxa including mussels (*Mytilus edulis*, Hall 1985), sea anemones (*Metridium senile*, Hoffman 1981), butterflies (*Lycaena tityrus*, Karl et al. 2008; *Colia meadii*, Watt et al. 2003), water

striders (*Limnoporus canaliculatus*, Zera 1987) and willow leaf beetles (*Chrysomela aeneicollis*, Dahlhoff and Rank 2000). Several studies in insects have established that temperature is an important selective agent favoring different *Pgi* genotypes (Dahlhoff and Rank 2000; Orsini et al. 2009; Kallioniemi and Hanski 2011). For example, in the Glanville fritillary, individuals heterozygous at a PGI amino acid are able to fly longer distances than homozygotes in low to moderate ambient temperatures, while at high ambient temperature, homozygotes move longer distances (Niitepõld et al. 2009). Similarly, in *C. eurytheme*, heterozygotes can maintain flight at lower ambient temperatures than homozygotes. This allows heterozygotes to remain active across a wider daily time window, and leads to an advantage in mating for males and oviposition for females (Watt 1977; Wheat et al. 2006). In willow leaf beetles, *Pgi* genotypes differ in heat shock protein expression (Dahlhoff and Rank 2000; Rank et al. 2007). These temperature-related effects of *Pgi* may derive, at least partially, from a trade-off between kinetic performance and thermal stability among isoforms of the PGI enzyme (Watt 1983; Watt and Dean 2000).

In butterflies, morphological traits such as the color and surface structure of wings affect heat gain and loss from the environment (Guppy 1986b; Brakefield and Reitsma 1991; Van Dyck and Matthysen 1998). For example, individuals with darker bodies and wings can absorb more solar radiation and therefore heat their bodies, particularly the flight muscles in the thorax, more rapidly than paler individuals (Guppy 1986a). Given the importance of thermal environment in driving selection on the *Pgi* locus, body and wing coloration could potentially interact with *Pgi* genotype to determine

fitness and performance traits of individuals. Such interactions have not, to our knowledge, been assessed in insect species.

Pgi is among a relatively small number of genes that have been clearly identified as influencing animal movement with large effect (Niitepõld et al. 2009; Orsini et al. 2009). In insects, there is considerable evidence that *Pgi* affects movement ability. In several butterflies, such as the Glanville fritillary, *Colias* sp., and European map butterfly (*Araschnia levana*), *Pgi* variation is associated with flight activity and flight metabolic rate (Watt 2003; Haag et al. 2005; Niitepõld et al. 2009; Mitikka and Hanski 2010). In willow leaf beetles the effect of thermal stress on running speed depends on *Pgi* genotype (Rank et al. 2007). The most comprehensive studies of *Pgi* effects on movement have arguably been in the Glanville fritillary, where PGI amino acid variation has been linked to flight performance using a variety of approaches (Niitepõld et al. 2009).

Since dispersal requires successful movement of individuals between habitat patches (Clobert et al. 2009; Bonte et al. 2012) and *Pgi* can influence insect movement, this gene is a potential candidate for determining variation in dispersal ability (Haag et al. 2005). In the Glanville fritillary *Pgi* variation has been linked directly to dispersal rate in the field (Niitepõld et al. 2009). In both Glanville fritillary and the European map butterfly, differences in *Pgi* allele and genotype frequencies between old established populations versus those in newly colonized habitat patches provide further, indirect evidence for an effect of *Pgi* on dispersal (Haag et al. 2005; Hanski and Saccheri 2006; Mitikka and Hanski 2010). Despite considerable evidence for heritability of movement and dispersal traits in animals, *Pgi* is still only one of a few genes that have been clearly shown to influence animal dispersal (Clobert et al. 2012; Wheat 2012; Zera and Brisson

2012). Movement and dispersal are not limited by distance alone, but are strongly influenced by land cover and geographic features (Roland et al. 2000; Ricketts 2001). To date, a genetic basis for reduced ability to traverse a specific dispersal barrier (a landscape feature that limits dispersal) has not yet been documented.

Here, I assess nucleotide and amino acid polymorphism in the coding sequence the *Pgi* locus, for the first time, in the Rocky Mountain Apollo butterfly, *Parnassius smintheus*. This species occupies naturally patchy, high-altitude alpine meadows in western North America. Given the thermal constraints placed on high altitude species, and the links between thermal environment and *Pgi* variation found in other taxa, I hypothesized that *Pgi* may affect fitness and performance traits in this species. Specifically, I examine potential associations between movement and dispersal of this species and *Pgi* variation.

Movement and dispersal in *P.smintehus* have been studied in a network of populations in Alberta, Canada since 1995 using mark-recapture methods (Matter et al. 2014). In this system, dispersal among habitat patches has been shown to have important population dynamic and genetic consequences (Roland et al. 2000; Keyghobadi et al. 2005; Jangjoo et al. 2016). The majority of movements, as inferred by mark-recapture data, occur within habitat patches and mean movement distances are on the order of ~ 150 m (Roland et al. 2000). Maximum recapture distances are ~ 2 km (Roland et al. 2000). Movement and dispersal of *P. smintheus* are strongly affected by land cover; dispersal declines exponentially with distance, but typically at a much higher rate over forest than over open meadows (Roland et al. 2000). Forest therefore is an important barrier to dispersal in this species. Movement and dispersal behaviours differ between males and

females. Females are generally less active and more cryptic than males. Although the sexes display similar mean dispersal distances (Roland et al. 2000), females appear less sensitive to intervening forest (Goff et al. 2018).

A previous study using comparative transcriptomics (Chapter 4) revealed that there was no difference in levels of *Pgi* expression between individuals of *P. smintheus* that had dispersed between habitat patches and individuals that had not dispersed. Here, I explore whether differences among sequence variants at this candidate locus may be a source of inter-individual variation in movement or dispersal behaviour. Hence, I assess whether specific alleles or genotypes at *Pgi* are associated with various aspects of dispersal and movement, including dispersal over forest barriers.

5.2 Material and Methods

My study took place in a network of populations occupying patches of alpine meadow habitat above treeline (~2100 m) along Jumpingpound Ridge in Alberta, Canada (50° 57'N, 114° 54'W). The meadows are separated primarily by intervening forest (Figure 1.1; Roland and Matter 2007).

There were two parts to my study. First, I sampled a smaller number (n = 49) of adult butterflies to obtain RNA from which I determined the coding sequence of *Pgi*. I began with RNA at this stage, rather than sequencing genomic DNA, because the *Pgi* locus contains large intron sequences and, in its entirety, spans over 10,000 bp of the genome (Wheat at al. 2006). I then identified single nucleotide polymorphisms (SNPs)

within the *Pgi* coding sequence and designed assays to genotype individuals at those SNPs using samples of genomic DNA. For the 49 individuals, I also measured thoracic and ambient temperature at the time of capture, and quantified darkness of the wings.

In the second part of my study, I genotyped a larger number of individuals (n = ~500) at the previously identified *Pgi* SNPs. Small samples of wing tissue (i.e., wing-clips) had been collected from these individuals through the course of yearly mark-recapture studies in my study system (Matter et al. 2014), and genomic DNA was extracted from these wing-clips. The mark-recapture history for each of these genotyped individuals was available, including locations and times of all capture events, allowing us to classify them as non-dispersers or dispersers (defined below) and estimate their movement distances.

5.2.1 Part I: *Pgi* coding sequence, thoracic temperature and wing darkness

5.2.1.1 Sampling and field measurements.

I sampled 49 individuals (36 male and 13 female) from ten different patches across the study area, and at different times of day (in the morning and around mid-day), in 2015 (Figure 1.1; Table 5.1).

Individuals were captured with a hand-net while in flight. I measured thoracic temperature for each individual within five seconds of capture with a digital thermometer (OMEGA HH91) attached to a copper thermocouple (OMEGA type T) housed within a disposable hypodermic needle (precisionGlide 19G1). I inserted the needle into the thorax through the net to avoid any direct contact between the butterfly's body and our hands. The same thermometer was used to measure ambient air temperature at the time and location of capture. In my subsequent analyses, I used the difference between thoracic and ambient temperature for each individual as a variable of interest.

Immediately after recording thoracic temperature, I carefully separated the thorax from the head, abdomen and wings using a clean scalpel and submerged the entire thorax into RNA later solution (Qiagen, Germantown MD). Samples were taken to the laboratory the same day and placed at 4 °C overnight, before being transferred to -20 °C where they were stored until RNA extraction was conducted. I also placed the wings individually in glassine envelopes and stored at -20 °C for wing color analysis.

Table 5.1 Information about individuals of *Parnassius smintheus* sampled in 2015 and used for determining *Pgi* coding sequence, and relationships with thoracic temperature and wing darkness (part I). Every individual was either classified as a disperser or non-disperser based on their mark-recapture history. For dispersers, the first patch in which the individual was captured is indicated.

Patch/ Population	Sex	Disperser	Non-disperser
F	male	0	2
	female	0	1
G1	male	0	2
	female	0	2
H	male	0	1
	female	0	0
I	male	0	2
	female	0	1
J	male	1	3
	female	0	0
K	male	1	1
	female	0	0
L	male	0	5
	female	0	1
M	male	1	7
	female	0	4
Q	male	2	2
	female	0	2
P	male	2	1
	female	0	1
Z	male	0	3
	female	0	1
Total		7	42

Sampling for this part of my study was concurrent with the annual mark-recapture study in the population network. For the mark-recapture study, butterflies are marked on the lower hind-wing with a unique three-letter code, and the capture locations for each individual are recorded using an x-y coordinate grid overlaid on aerial photos of the study area (Roland et al. 2000; Matter et al. 2014). All 49 individuals sampled for RNA in this study had been previously marked, and could therefore be sorted, based on their recapture history, into two groups: dispersers were those individuals that I captured in a patch different than the one in which they were originally marked, and non-dispersers were individuals that I sampled in the same patch in which they were originally marked.

5.2.1.2 Wing colour analysis

In the laboratory, the dorsal surfaces of the wings were photographed against a standard black background under consistent lighting using a Nikon D1 camera at a fixed focal length of 58 cm. I analyzed the wing images in ImageJ software (NIH; http://rsb.info.nih.gov/ij/) to determine the darkness of the wing dorsal surfaces by measuring grey values (from 0 = black to 255 = white), and extracted a single value for both fore- and hindwings of each individual.

5.2.1.3 *Pgi* coding region sequencing

Total RNA was extracted from each of the 49 thorax samples using the RNeasy Mini Kit (Qiagen) according to the manufacturer's protocol. To determine the coding sequence of *Pgi*, which had not previously been described for this species, I first conducted

RNA−sequencing (RNA-seq) for a subset of 12 individuals who were also used as part of a gene expression study, as described in Chapter 4. I assembled a reference transcriptome using the RNA-seq data for those individuals and generated a comprehensive annotation report from which the coding sequence of *Pgi* was extracted (see Chapter 4 for details). Nucleotide sequences of the entire *Pgi* gene from *C. eurytheme* and *M. cinxia* were used in a BLAST (Basic Local Alignment Search Tool) search against the *Pgi* coding sequence of *P. smintheus* to determine the exon-intron boundaries.

Based on the *Pgi* coding sequences determined by RNA-seq for the 12 individuals, I designed conserved primers, using Primer3 v. 0.4.0 (Koressaar and Remm 2007), to amplify the entire coding sequence of *Pgi* in five overlapping fragments. For the remaining 37 individuals from which RNA had been extracted (i.e., not used in the RNA-seq study), I synthesized first-strand cDNA from 3 µg of total RNA per individual, using oligo$(dT)_{20}$ primer and Superscript III reverse transcriptase (Invitrogen, Waltham MA) according to the manufacturer's protocol. This cDNA was then used as template to amplify and sequence the five overlapping fragments that comprise the *Pgi* coding region. I added common sequence 1 (CS1) and common sequence 2 (CS2) universal sequence tags to the 5'end of all forward and reverse primers, respectively, to enable fragments to be prepared for next generation sequencing using the Fluidigm Access Array (Fluidigm Corporation, San Francisco CA). The primer and tag sequences, and PCR chemistry, are provided in the supplementary material (Table D1 and D2). I used the following cycling conditions for all five PCR reactions: denaturation for 120 s at 94 °C; followed by 39 cycles of 18 s at 94 °C, 24 s at 56 °C, 60 s at 72 °C; and final elongation at 72 °C for 420 s. PCR amplifications were performed in a PTC 0200 DNA Engine Cycler (Bio-Rad,

Hercules CA). Quality and size of PCR products were checked by 1.5% agarose gel electrophoresis using SYBR Green (Bio-Rad) on a UV transilluminator.

The five amplified fragments for each individual were pooled and sent to the McGill University and Génome Québec Innovation Centre (McGill University, Montréal, Québec; MGU-GQ), where the samples from each individual were given unique barcodes, and libraries were prepared and sequenced on the Illumina MiSeq PE300 bp platform, using a total of ~100 000 reads. I sorted reads by individual, and identified and called SNPs (including both synonymous and non-synonymous SNPs) using the SAMtools software package (Li et al. 2009). I used VCFtools software (Danecek et al. 2011) to extract SNPs that were called only for positions with a minimal mapping quality (-Q) and coverage (-d) of 25. The maximum read depth (-D) was set at 200. A genotyped SNP was excluded if the minor allele was observed less than three times across all individuals. I computed genotype likelihoods using the SAMtools utilities and determined variable positions in the aligned reads compared to the reference with the BCFtools utilities (Li 2011).

The *Pgi* reference coding sequence was translated to amino acid sequence using MEGA 6.06. I predicted the effect of amino acid changes (i.e. changes in charge, polarity and size of amino acids) resulting from non-synonymous substitutions in the online SIFT platform (http://sift.jcvi.org/) using the SIFT Sequence option (Sim et al. 2012).

5.2.1.4 Linkage Disequilibrium (LD)

Across all 49 individuals for which the full *Pgi* coding sequence was determined, I tested for linkage disequilibrium between all pairs of SNPs using Fisher's exact test with Bonferroni correction (Weir 1996) implemented in DnaSP5 software (Librado and Rozas 2009).

5.2.1.5 Association of wing darkness, thoracic temperature, and *Pgi* variation with dispersal history

With the dataset of 49 individuals, I explored the association of dispersal history with wing color, thoracic temperature, and *Pgi* variation using generalized linear models (GLM). Specifically, I modeled dispersal history (response variable) as a function of wing darkness, thoracic temperature relative to ambient, and *Pgi* SNP genotype. I considered genotypes at both non-synonymous SNPs (nsSNPs) and synonymous SNPs (ssSNPs) and ran separate sets of models for each SNP. I coded genotypes in two alternative ways. First, I coded genotype at a given SNP as the number of copies of the major allele possessed by the individual (0 = homozygous for minor allele, 1 = heterozygous and 2 = homozygous for major allele). This coding assumes that heterozygotes are intermediate in phenotype to the two homozygotes and essentially measures an additive effect of the number of major alleles. Second, I treated genotype as a categorical variable (AA = homozygous for major allele, AB = heterozygous and BB = homozygous for minor allele). This coding does not assume that the heterozygote is intermediate in phenotype, and allows for the possibility of heterozygote advantage (i.e., overdominance) or disadvantage (i.e., underdominance). Since all females in this dataset were non-

dispersers, and females of the species are also considerably darker than males, I removed females from these analyses and ran the models only on male individuals (n = 36). I used the glm function of the Stats package in R v.3.4.4 (R Core Team 2017) to run GLMs, with the binomial family, and confirmed that the predictors used in each model were not collinear based on pairwise Pearson's correlation coefficient values ($r^2 < 0.5$).

5.2.2 Part II: *Pgi* SNP variation and dispersal in a larger dataset

5.2.2.1 Genotyping DNA samples at *Pgi* SNPs

I used genomic DNA previously extracted from wing-clips of 491 adult individuals from 14 different patches collected in the years 1995, 2005, 2008 and 2013 (Table 5.2; Keyghobadi et al. 1999; Caplins et al. 2014; Jangjoo et al. 2016).

After determining the position of all synonymous (ss) and non-synonymous (ns) SNPs in the coding sequence of *Pgi* in my initial sample of 49 sequenced individuals (Part I), I designed primers to assay all nsSNPs (n = 16) and an approximately equal number of ssSNPs (n = 14) from the samples of genomic DNA using iPLEX Gold (Agena Bioscience, San Diego CA). For two SNPs (nsSNP550 and nsSNP891) that were located near an inferred intron/exon boundary (Figure D4), I used intron sequences from *C. eurytheme* to design the primer on the intron side. An assay panel for the *Pgi* SNPs was designed, and samples were genotyped, by the MGU-GQ facility using iPLEX Gold (Agena Bioscience).

Table 5.2 Information about the samples used for genotyping *Pgi* SNPs from genomic DNA in *Parnassius smintheus* (Part II). Every individual was either classified as a disperser or non-disperser based on their mark-recapture history. 'No. genotyped individuals' is the number of individuals that were genotyped for the *Pgi* SNPs.

		No. genotyped individuals	
Patch/ Population	**Sex**	**Disperser**	**Non-disperser**
F	male	3	0
	female	0	0
G1	male	10	32
	female	0	9
H	male	2	8
	female	0	7
g2	male	3	37
	female	0	11
I	male	3	0
	female	0	0
J	male	7	44
	female	0	10
K	male	5	49
	female	0	11
L	male	19	31
	female	3	11
M	male	10	50
	female	0	10
O	male	4	26
	female	0	7
P	male	0	4
	female	0	0
R	male	2	0
	female	1	0
S	male	2	0
	female	0	0
Z	male	0	44
	female	1	15
Total		75	416

I combined the genotypic data derived from the 491 wing clip samples with the genotypes of the 49 fully sequenced individuals (from Part I), and used GLMs to examine the effect of *Pgi* variation on dispersal and movement in this larger dataset. I ran a separate model for each *Pgi* SNP, using SNP genotype as a predictor of dispersal or movement. As with the smaller dataset, I coded SNP genotype in two alternative ways, either numerically or categorically. Movement and dispersal behaviours differ between male and female *P. smintheus* (Goff et al. 2018). In the larger dataset, I had very few female dispersers (n = 5), which did not allow for tests of interaction between SNP genotypes and sex. Therefore, I ran all models for the larger dataset with only males.

With the larger dataset, I was able to quantify and examine four different aspects of dispersal or movement as response variables, which were tested independently. First, as for the smaller dataset of 49 individuals, I used binomial response of disperser (individual that at some point was re-captured in a patch different from the one in which it was originally marked) or non-disperser (individual that was only ever re-captured in the same patch in which it was originally marked). This variable indicates whether an individual emigrated from one patch and successfully moved to a different patch, irrespective of the distance or landscape between the patches.

Second, for all dispersers (n = 77; males only), I used the distance between the patch in which it was first marked and the patch in which it was last re-captured as a response. Distances were measured along ridge-tops, between the centroids of butterfly capture in each patch (Roland et al. 2000). The total distance between any two patches, along the ridge-top, can be divided into a portion that occurs over forest and a portion that occurs over meadow (Roland et al. 2000). As a third response variable, for each disperser,

I used the distance over forest between the patch in which the individual was first marked and the patch in which it was last re-captured. This variable measures the distance moved over forest in dispersing between the patches. These models also included the distance over meadow between the patches as a controlling variable, to account for the fact that patches that are further apart are also more likely to be separated by more intervening forest. Finally, for all individuals from one patch (M), I calculated a measure of within-patch movement. Specifically, I measured the largest linear distance between any two capture events within the patch based on their location on the x-y coordinate system. Because within-patch movement distance will be limited by patch size, for this analysis I only looked at individuals from the largest patch, M (n = 34; males only).

I used the glm function of the Stats package in R v.3.4.4 to run and evaluate all GLMs. Across all models that were analyzed for a given response variable, *P*-values were adjusted for multiple comparisons using the Benjamini-Hochberg False Discovery Rate (FDR) procedure (Benjamini and Hochberg 1995).

5.3 Results

5.3.1 Part I: *Pgi* coding sequence, thoracic temperature and wing darkness

5.3.1.1 *Pgi* coding sequence of *Parnassius smintheus*

The full-length coding sequence of *Pgi* in *Parnassius smintheus* is 1671 bp long, corresponding to 557 amino acids, similar to *Colias eurytheme* (1,668 bp, 556 aa; Wheat

et al. 2006) and *Melitaea cinxia* (1671 bp, 557 aa; Orsini et al. 2009). Percent identity to the coding sequences from *M. cinxia*, *C. eurytheme*, and *Bombyx mori* were 77%, 77%, and 76%, respectively, in nucleotide sequence, and 86%, 87% and 86%, respectively, in amino acid sequence (Figure D4 and D5). The *Pgi* intron-exon boundaries appear to be highly conserved among butterfly species, as my coding sequence aligned perfectly to identified exon sequences of *M. cinxia* and *C. eurytheme* (Figure D4). Therefore, like these other butterfly species, the *P. smintheus* coding sequence appears to be divided among 12 exons (Table 5.3).

Overall, I identified 50 variable sites (16 non-synonymous and 34 synonymous) in the coding sequence of *Pgi* across 49 individuals (Table 5.4 and 5.5). All polymorphic sites were segregating for only two alleles. The overall nucleotide diversity was 0.0068 and the estimated nucleotide diversity for ssSNPs (π_{ss} = 0.0165) was higher than nsSNPs (π_{ns} = 0.0039). Among the 49 individuals, five pairs of sites were in significant linkage disequilibrium (Table D3).

5.3.1.2 Amino acid variation in *P. smintheus Pgi*

Of the 16 nsSNPs I observed, three represent a charge change at the target amino acids: at codon 10, there is polymorphism for a negatively charged aspartic acid (Asp; 133 Da) and polar but uncharged tyrosine (Tyr; 181 Da), while at codons 26 and 330, there is polymorphism between a positively-charged lysine (Lys; 146 Da) and a polar asparagine (Asn; 132Da).

I detected four additional nsSNPs that represent a change in polarity of the target amino acids: at codons 24 and 340, a non-polar alanine (Ala: 89Da) switches with polar

serine (Ser;105Da) or threonine (Thr; 119 Da), respectively, while codon 209 is polymorphic for a non-polar phenylalanine (Phe ;165Da) and polar Ser, and codon 414 is polymorphic for a polar glutamine (Gln; 146 Da) and a non-polar leucine (Leu;131Da).

Table 5.3 The inferred positions of exons in the coding sequence of the *Pgi* gene in *Parnassius smintheus*, based on alignment to coding sequences of the butterflies, *Colia eurytheme* and *Melitaea cinxia*.

	Exon 1	Exon 2	Exon 3	Exon 4	Exon 5	Exon 6	Exon 7	Exon 8	Exon 9	Exon 10	Exon 11	Exon 12
Start (bp)	1	132	272	441	550	715	828	955	1050	1209	1342	1484
Stop (bp)	131	271	440	549	714	827	954	1049	1208	1341	1483	1674

Table 5.4 Non-synonymous SNPs detected in the coding region of *Pgi* in *Parnassius smintheus*. For each non-synonymous SNP, the exon in which it occurs, position within the coding sequence ('nsSNP Site'), position within the PGI amino acid sequence ('Codon/AA Site'), codon triplet ('Triplet'), identity of common and alternate alleles, and identity of the common and alternate amino acids (AA), along with the polarity (P = polar, NP = non-polar), charge (Pos = positively charged, Neg = negatively charged, U = uncharged, H = hydrophobic), and molecular weight (Daltons) of each amino acid, are provided.

Exon	nsSNP Site (bp)	Codon/ AA Site	Triplet	Common Allele	Alternate Allele	Common AA	Alternate AA
1	28*	10	GAT	G	T	Asp; P, Neg (133)	Tyr; P, U (181)
1	70*	24	GCT	G	T	Ala; NP, H (89)	Ser; P, U (105)
1	78*	26	AAA	A	C	Lys; P, Pos (146)	Asn; P; U (132)
1	99	33	TTT	T	A	Phe; NP, H (165)	Leu; NP, H (131)
3	346	116	ATG	A	T	Met; NP, H (149)	Leu; NP, H (131)
4	469*	157	ATC	A	G	Ile; NP, H (131)	Val; NP, H (117)
5	550	184	GTG	G	A	Val; NP, H (117)	Met; NP, H (149)
5	626*	209	TTC	T	C	Phe: NP, H (165)	Ser; P, U (105)
5	664*	222	CTT	C	A	Leu; NP, H (131)	Ile; NP, H (131)
5	690*	230	AAC	C	A	Asn; P, U (132)	Lys; P, Pos (146)
7	891	297	GAG	G	C	Glu; P; Neg (147)	Asp; P, Neg (133)
8	1018*	340	GCC	G	A	Ala; NP, H (89)	Thr; P, U (119)
9	1129*	377	TCC	T	A	Ser; P, U (105)	Thr; P, U (119)
10	1241*	414	CAG	A	T	Gln; P, U (146)	Leu; NP, H (131)
12	1610*	537	CCA	C	T	Pro; NP, H (115)	Leu; NP, H (131)
12	1612*	538	GTA	G	A	Val; NP, H (117)	Ile; NP, H (131)

*SNPs successfully genotyped from wing clips using iPlexGold

Table 5.5 All synonymous SNPs detected in the coding region of *Pgi* in *Parnassius smintheus*. For each synonymous SNP detected, the exon in which it occurs, position within the coding sequence ('ssSNP Site'), and the identity of the common and rare allele are provided.

Exon	ssSNP Site (bp)	Common allele	Alternate allele	Exon	sSNP Site (bp)	Common allele	Alternate allele
1	108*	C	T	9	1029	A	G
2	169	C	T	9	1050	G	A
3	294*	G	A	9	1092*	C	A
3	321*	G	A	9	1107*	G	A
3	360	G	T	10	1245*	G	A
3	369*	A	T	10	1317*	A	G
3	402*	G	A	10	1323	G	A
5	606*	G	C	11	1368	A	G
5	622	C	T	11	1419	C	A
5	627	C	T	11	1446	G	A
5	654*	G	A	11	1458	G	A
6	756	C	T	12	1497	C	T
6	768	G	A	12	1503	C	T
6	795	C	T	12	1512*	A	G
7	849	T	C	12	1570*	C	T
7	945	G	A	12	1623	C	T
8	1008*	G	A	12	1635	T	C

* SNPs successfully genotyped from wing clips using iPlexGold

I compared the positions of amino acid polymorphisms in *P. smintheus* to those described for *M. cinxia* and *C. eurytheme* (Wheat et al. 2006; Figure D5). Both *P. smintheus* and *C. eurytheme* have a polymorphism at codon 538 for the amino acids valine (Val) and isoleucine (Ile) (Wheat et al. 2006). This amino acid change does not result in a change in charge or polarity. Interestingly, Val is the more common variant in *P. smintheus* while Ile is more common in *C. eurytheme* (Wheat et al. 2006).

One nsSNP, at codon 377, was close to the location of a SNP potentially experiencing selection in *M. cinxia* and *C. eurytheme*, at codon 375 (Wheat et al. 2006, 2010). While variation at codon 375 in these other species leads to a change in amino acid charge, the variation I observed in *P. smintheus* causes no change in polarity or charge of the target amino acid (Ser → Thr). The amino acid polymorphisms at codons 24 and 26 in *P. smintheus*, representing a change in polarity and charge, respectively, are very close to the site of a codon in *C. eurytheme* (codon 21) containing a charge-changing amino acid polymorphism (i.e. between polar Asn and negatively charged Asp; Wheat et al. 2006, 2010).

5.3.1.3 Wing color, thoracic temperature and dispersal history

As expected for this species, females were significantly darker than males ($t_{16.7} = 7.15$, $P < 0.001$), however, I found no significant difference in thoracic temperature between males and females ($t_{22.03} = 0.8$, $P = 0.42$). In the sample of 49 individuals used to determine the full *Pgi* coding sequence, there were seven dispersers, all of which were male. Given that all females sampled in this part of the study (n = 13) were non-dispersers, that they were darker than males, and that wing darkness was correlated with

thoracic temperature (below), I removed females from the subsequent analyses with this smaller dataset examining the joint effects of wing darkness, thoracic temperature and *Pgi* variation on dispersal history.

Among males (n = 36), thoracic temperature was positively correlated with ambient temperature (r = 0.61, P <0.001; Figure 5.1). Male dispersers were caught flying at a narrower range of ambient temperatures (20 °C -28 °C) than non-dispersers (11 °C - 32 °C). Also, across the range of ambient temperatures at which male dispersers were captured flying (20 °C - 28 °C), the dispersers displayed thoracic temperatures within a lower and narrower range (19 °C -25°C) than non-dispersers (20 °C -32°C) (Figure 5.1). Indeed, male dispersers tended to have thoracic temperatures lower than ambient (mean ± SD = -2.01 ± 2.5) while male non-dispersers tended to have thoracic temperatures higher than ambient (mean ± SD 2.9 ± 4.8).

Among males, thoracic temperature relative to ambient was correlated with wing darkness, such that darker individuals had higher relative thorax temperatures (r= -0.60, P < 0.001; Figure 5.2). Male dispersers tended to have lighter wing color (mean grey value ± SD: 157.1 ± 4.05; Figure 5.2), while non-dispersers tended to have darker wing color (mean grey value ± SD: 148 ±8.08; Figure 5.2).

Despite the trends described above, in a model including wing darkness and relative thoracic temperature (and their interaction) as predictors of dispersal, none of the terms were significant (P < 0.05). Models with just one of these two predictors were significant for relative thorax temperature (P = 0.02, β ±SE= 0.27± 0.12) and marginally non-significant for wing darkness (P = 0.08, β ±SE= -0.10± 0.06). Because of the small size of the dataset, the correlation between wing darkness and relative thoracic

temperature, and the fact that thoracic temperature was a slightly better predictor of dispersal history, in models evaluating the influence of individual *Pgi* SNPs on dispersal, I chose to include only relative thoracic temperature as an additional variable. Thus, for each SNP I evaluated a model with genotype at a given SNP, relative thoracic temperature, and their interaction as predictors. Among these models, I did not detect significant effects of any of the SNPs on dispersal history.

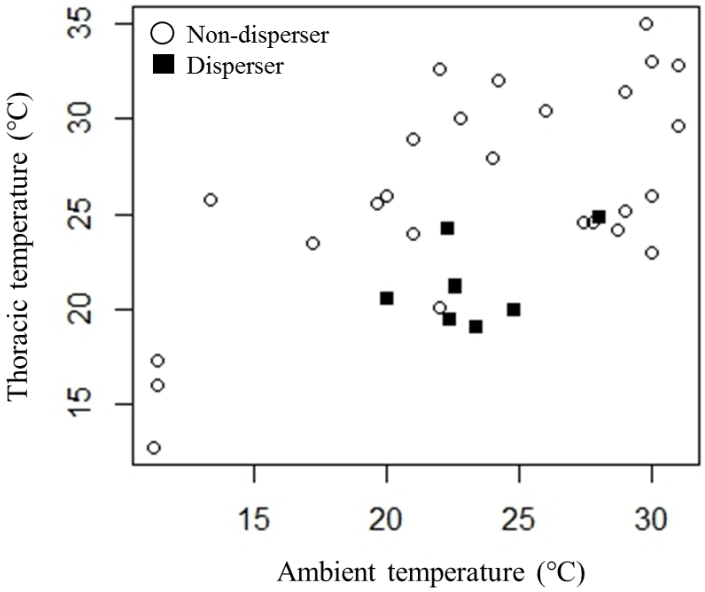


Figure 5.1 Relationship between thoracic temperature and ambient temperature in male dispersers and non-dispersers of *Parnassius smintheus*.

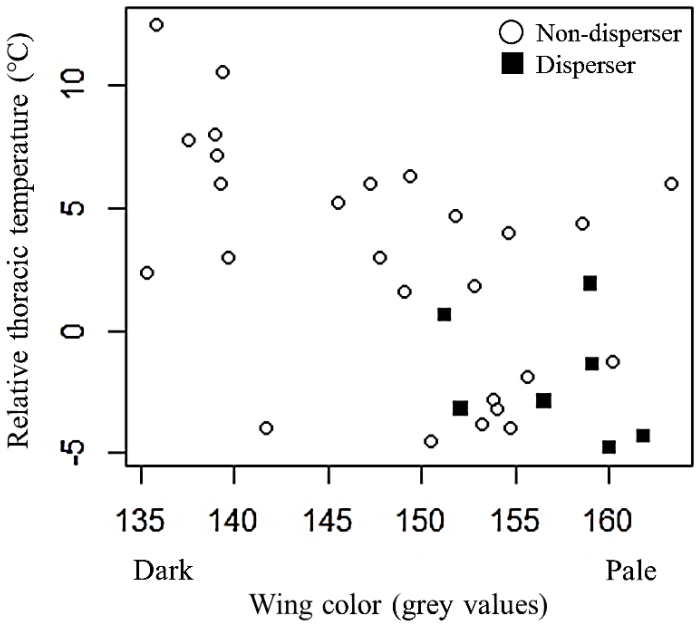


Figure 5.2 Relationship between relative thoracic temperature (thoracic temperature minus ambient temperature) and wing color in male dispersers and non-dispersers of *Parnassius smintheus*.

5.3.2 Part II: *Pgi* SNP variation and dispersal in a larger dataset

The iPlex Gold assays for four of the 16 nsSNPs failed because of either proximity to an intron-exon boundary (n = 2) or high variability in the surrounding sequences (n = 2). Both assays designed using *Colias* intron sequences failed. Assays for all 14 ssSNPs were successful. Thus, I obtained genotypes at 12 nsSNPs and 14 ssSNPs within *Pgi* for 491 individuals, which I combined with the genotype data for the same SNPs from the individuals whose full coding sequence I determined in Part I.

For the binary response of dispersal history (disperser vs. non-disperser), I did not detect significant effects of any of the SNPs when genotypes were treated as ordered numerical variables. Treating genotypes as categorical predictors however, I found significant effects of nsSNP1018 and nsSNP1241 on dispersal history. For both SNPs, heterozygotes were significantly different from homozygotes for the major (i.e., common) allele (nsSNP1018: adjusted $P = 0.011$; nsSNP1241: adjusted $P = 0.018$). The odds ratios for the effects of heterozygotes versus major allele homozygotes for nsNSPs 1081 and 1241 were, respectively 4.71 (95% CI: 2.06–10.77) and 3.03 (95% CI: 1.89–4.88). At nsSNP1018, heterozygotes were not significantly different from homozygotes for the minor (i.e, rare) allele (adjusted $P = 0.20$), while at nsSNP1241 no homozygotes for the minor allele were observed. At these two SNP sites, the frequency of the rare allele was higher in dispersers (nsSNP1018 = 0.087 and nsSNP1241 =0.1) than in non-dispersers (nsSNP1018 = 0.033 and nsSNP1241 = 0.041; Figure 5.3). Furthermore, at both nsSNP sites, the frequency of heterozygotes was higher in dispersers (nsSNP1018 = 0.128 and nsSNP1241 = 0.20) compared to non-dispersers (nsSNP1018 = 0.030 and nsSNP1241 =

0.079; Figure 5.3), with the effect being stronger for SNP1018. I did not find significant effects of any ssSNPs on dispersal history.

Among dispersers, nsSNP1018 was a significant predictor of total dispersal distance (i.e., between the centroids of the patch of origin and the final patch of capture), but only when genotype was coded as a categorical variable. The mean total distance dispersed by heterozygotes at nsSNP1018 was higher, but not significantly different than, the total distance dispersed by individuals homozygous for the minor allele (adjusted P = 0.29). The total dispersal distance for heterozygotes at nsSNP1018 was significantly higher than that for individuals homozygous for the major allele (adjusted P = 0.003; $\beta \pm SE$ = -1.48±0.35 for effect of homozygote relative to heterozygote; Figure 5.4). I did not detect significant effects of any other SNPs, treated either as ordered numerical or categorial predictors, on dispersal distance (all adjusted $P > 0.05$).

Finally, no SNPs were significant predictors of either distance dispersed over forest (controlling for distance over meadow; all adjusted $P > 0.05$) or the maximum estimated displacement distance within the largest patch, M (all adjusted $P > 0.05$).

Because of the small number of female dispersers (n = 5), I could not test for an interaction of sex with SNP genotype on any of the dispersal or movement responses. However, I observed that all those females that dispersed were homozygotes for the major allele at nsSNP1018, three were homozygotes for the major allele at nsSNP1241, and two were heterozygotes at nsSNP1241.

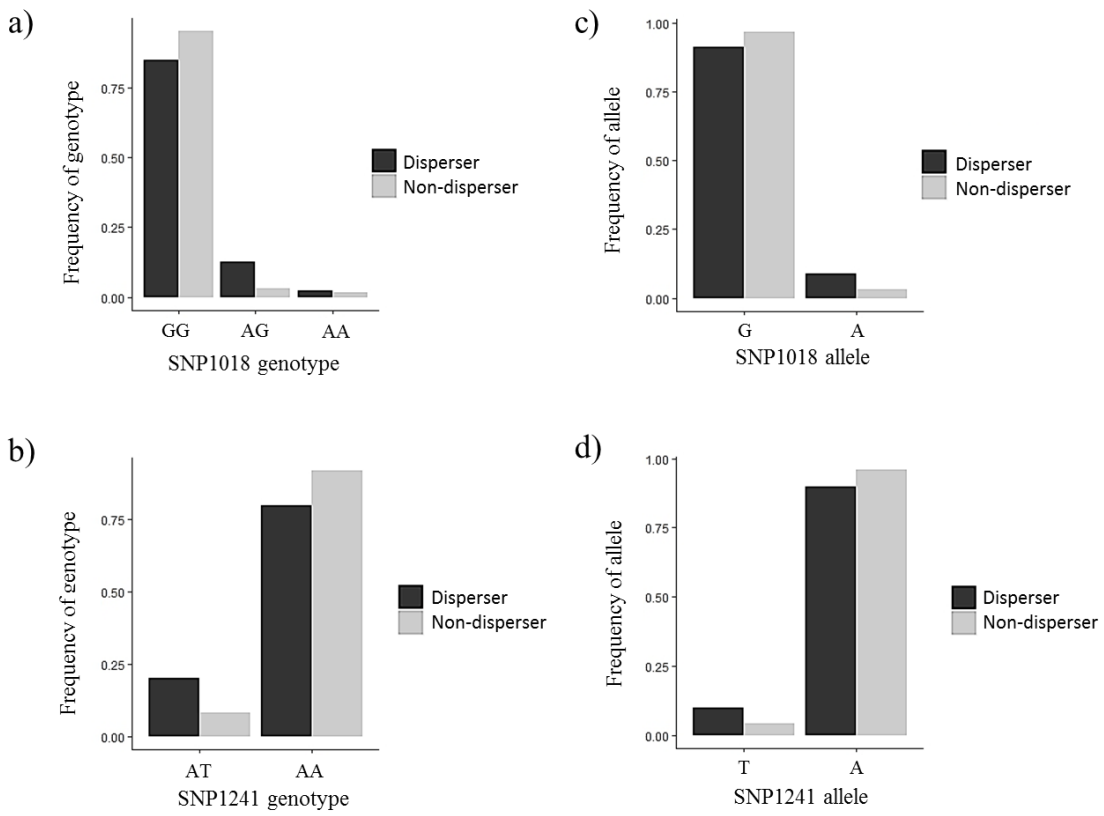


Figure 5.3 The frequency of genotypes (a, b) and alleles (c, d) for dispersal-related SNPs nsSNP1018 and nsSNP1241, from the coding region of *Pgi* in *Parnassius smintheus*. Both SNPs are non-synonymous.

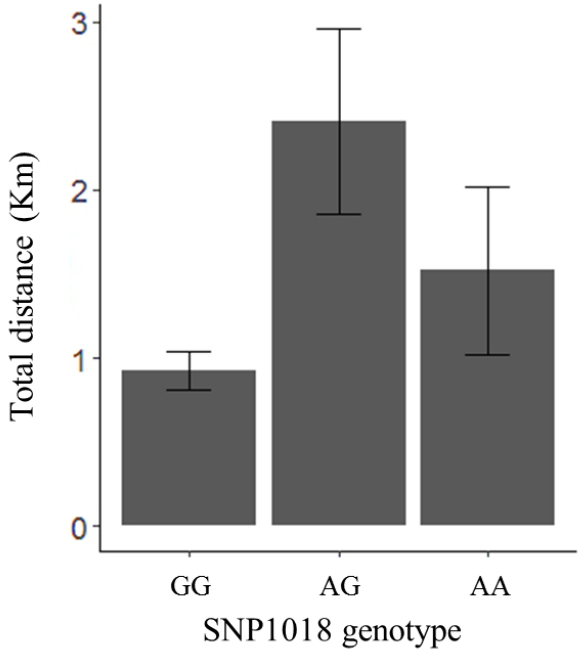


Figure 5.4 The total distance moved between patches for dispersing males with differing genotypes at the non-synonynous *Pgi* SNP, nsSNP1018, in *Parnassius smintheus* (GG= homozygous for major allele, AG= heterozygous and AA= homozygous for minor allele). I have also shown means with standard error bars.

5.4 Discussion

The *Pgi* coding sequence of *P. smintheus* aligned to that of other butterflies, including the well-studied *C. eurytheme* and *M. cinxia* (Wheat et al. 2006; Orsini et al. 2009) with 77% nucleotide sequence and 86-87% amino acid sequence identity. The levels of nucleotide and amino acid sequence variability I detected are also comparable to those in *C. eurytheme* and *M. cinxia* (Wheat et al. 2006; Orsini et al. 2009). I observed 34 synonymous and 16 non-synonymous substitutions in *P. smintheus*, and estimated nucleotide diversity of 0.017 for synonymous sites and 0.0039 for synonymous sites. Seven of the 16 non-synonymous substitutions I observed result in a difference in amino acid charge or polarity. In comparison, in *C. eurytheme* Wheat et al. (2006) detected a much larger total number of variable sites (130), but a similar number of sites with non-synonymous substitutions (17), of which five resulted in altered amino acid charge. They observed nucleotide diversity of 0.073 for synonymous sites and 0.0024 for non-synonymous sites. In *M. cinxia* Wheat et al. (2009) observed 45 synonymous and 10 non-synonymous substitutions, and nucleotide diversity of 0.013 for synonymous sites and 0.0076 for non-synonymous sites. Detection of amino acid variation at PGI in *P. smintheus* is significant because although many species display high variability in *Pgi*, some species including bumble bees and the butterfly *Bicyclus anynana* display little to no amino acid variation at this locus (Ellis et al. 2013; Wheat and Hill 2014).

Although the *Pgi* intron-exon boundaries in *P. smintheus* appear to align perfectly with those of *C. eurytheme* and *M. cinxia*, the intron sequences are likely quite different based on the failure of all assays for SNPs located near exon/intron boundaries where I

used intron sequences from *C. eurytheme* to design primers. These intron/exon boundaries seem generally conserved among Lepidoptera as they also align with those of the silk moth *Bombyx mori* (Orsini et al. 2009).

I detected several nsSNPs in the coding sequence of Pgi that cause a broad range of changes in polarity and size of amino acids (Table 5.4). These changes might affect conformation, kinetics or stability of the PGI enzyme depending on whether they occur on the surface or interior of the enzyme (Watt and Dean 2000). Amino acid changes at the catalytic center of an enzyme can directly affect substrate affinity or reaction mechanism (Dean and Golding 1997; Newcomb et al. 1997), while changes at the surface can affect an enzyme's kinetics by altering its geometry (Gerstein and Chothia 1991; Watt and Dean 2000). As with many other enzymes, naturally occurring variation in PGI in other species is concentrated at the surface, where the amino acids would be exposed to the surrounding solvent (Wheat et al. 2006). I do not yet know the exact three-dimensional structure of PGI in *P. smintheus*, but based on extrapolation from the inferred structure in *C. eurytheme* and *M. cinxia* (Wheat et al. 2006; Wheat et al. 2009), the charge or polarity changing substitutions I observed are also most likely at the enzyme surface. I observed only one non-synonymous substitution that, based on comparison to the amino acid sequence and structure of PGI in *C. eurytheme,* is likely to occur near the enzyme's center. This was nsSNP1612, which corresponds to an Ile/Val polymorphism at amino acid 538 (Table 5.4). Interestingly, *C. eurytheme* has the same polymorphism in exactly the same amino acid position, although the identity of the minor and major allele is reversed relative to *P. smintheus*: Val is the common allele in *P. smintheus* while Ile is the common allele in *C. eurytheme.* This particular polymorphism is likely to be of limited

significance for enzyme function however, as the Ile/Val change is a conservative one that does not result in any change in charge or polarity.

Consistent with a previous, smaller study in *P. smintheus* (Chapter 3), I observed that thoracic temperature during flight, relative to ambient temperature, differed between dispersers and non-dispersers. Dispersers had lower relative thoracic temperatures than non-dispersers and also displayed a narrower range of thoracic temperature (Figure 5.1). I also found that dispersers tended to have lighter wing color than non-dispersers, and were captured flying at a narrower range of ambient temperature (Figure 5.2). Long-distance flight is a very energetically demanding activity, requiring use of stored energy reserves (Arrese and Soulages 2010). Lipid reserves are a particularly important fuel source for long-distance insect flight, and individuals with greater lipid reserves may be more likely to be successful in dispersal (Chaplin and Wells 1982). High body temperature speeds up metabolic rate, which will accelerate consumption of lipid and energy reserves. For this reason, migrating monarch butterflies maintain lower body temperatures during flight to avoid rapid depletion of reserves (Calvert and Brower 1986; Masters et al. 1988). My data suggest that maintenance of lower body temperature is also an important feature of dispersal in *P. smintheus*. Lower body temperatures, particularly in the energetically demanding flight muscles of the thorax, could allow dispersers to reduce either the risk of overheating (Masters et al. 1988; Neve and Hall 2016) or reduce the rate of energy (particularly lipid) consumption (Calvert and Brower 1986; Masters et al. 1988), or both. Furthermore, lighter wing color appears to be at least one factor that contributes to reduced thoracic temperatures in dispersing *P. smintheus*.

Although thoracic temperature relative to ambient is associated with dispersal history in *P. smintheus* (Figure 5.1), I found no interaction effect of thoracic temperature and *Pgi* variation on dispersal. A number of studies have demonstrated potential adaptation at *Pgi* to the local thermal environment (Hoffman 1981; Zera 1987; Dahlhoff and Rank 2000; Watt et al. 2003; Karl et al. 2008). These studies also indicate differential performance of *Pgi* variants at different temperatures, suggesting that thermal context can be an important selective factor on *Pgi* variation. For example, Niitepõld (2010) found that *M. cinxia Pgi* homozygotes displayed better flight performance at higher temperatures, while heterozygotes are able to fly at lower ambient temperatures. Watt (1983) also demonstrated that *C. eurytheme* heterozygotes are able to fly at a broader range of temperatures. Another example comes from male willow leaf beetles in which individuals with different *Pgi* genotypes show differential physiological stress and running speed in response to elevated air temperatures (Dick et al. 2013). My inability to detect any interaction of *Pgi* variation with thoracic temperature, or indeed any direct effect of *Pgi* variation, on dispersal history in this part of my study could reflect the small sample size (n=49), particularly the small number of dispersers sampled (n=7). Similar sampling over additional years may be required to obtain sufficient statistical power to test for potential interactions between *Pgi* variation and aspects of thermal biology.

In a much larger dataset of ~500 individuals genotyped at *Pgi* using DNA from non-lethal wing tissue samples, I found a significant association between dispersal history and two non-synonymous *Pgi* SNPs (nsSNP1018 and 1241) corresponding to amino acids 340 and 414, respectively (Figure 5.3). Genotype at nsSNP1018 was also a significant predictor of the total distance moved by dispersing individuals (Figure 5.4). I found no

evidence that the two nsSNPs at sites 1018 and 1241 are linked based on an analysis of linkage disequilibrium (Table D3). Indeed, these nsSNPs are located in different exons (exons 8 and 10, respectively; Wheat et al. 2006), suggesting that these SNPs are likely to be independently affecting PGI function.

I detected significant effects of nsSNP1018 and 1241 only when genotypes were coded categorically. In all cases, the heterozygote either was more likely to disperse, or dispersed a greater distance, compared to homozygotes of the major (i.e, common) allele. No homozygotes of the minor allele were observed at nsSNP1241. At nsSNP1018, dispersal traits of the heterozygotes did not differ from those of homozygotes for the minor (i.e, rare) allele. These results indicate strongly that the heterozygote at nsSNP1018 is not intermediate in phenotype between the two homozygotes. However, there is also no clear evidence for any heterozygote advantage; although trends in dispersal distance for genotypes at nsSNP1018 are consistent with a pattern of heterozygote advantage (Figure 5.4), the difference between heterozygotes and minor allele homozygotes is not statistically significant. Therefore, based on the data currently available, it appears that simply possessing a single copy of the rare allele at either nsSNP1018 or 1241 contributes to a higher likelihood of dispersal or greater dispersal distance.

The polymorphism at nsSNP1241 represents a change in polarity and size of the amino acid at codon 414, by switching a polar glutamine to a non-polar leucine (Table 5.4). At nsSNP1018, the polymorphism also represents a change in amino acid polarity and size, switching a non-polar alanine with polar threonine at codon 340. Based on comparison to the three-dimensional structure of PGI in *Colias* (Wheat et al. 2006), the amino acids encoded by nsSNP1018 and 1241 are likely near the surface of the enzyme

rather than at its center. It is possible that the observed variation at each of these nsSNPs results in changes in the geometry or flexibility of the enzyme, which alters either catalytic efficiency, thermal stability, or both.

Interestingly, variation at an amino acid site near the location of codon 340 (represented in *P. smintheus* by nsSNP1018) is also implicated in *M. cinxia* (at codon 361, specifically) as being important in explaining flight metabolic rate, population growth rate, and dispersal (Orsini et al. 2009; Niitepold et al. 2011). Furthermore, Wheat et al. (2006) demonstrated that several amino acid sites in *Colias* butterflies that are also nearby (although slightly further away), at codons 317, 369 and 375, are under balancing selection. Thus, it is possible that nsSNP1018 lies in a potential 'hot zone' where variation has strong effects on PGI performance.

In summary, I have detected potential associations between dispersal and each of genetic variation at *Pgi*, wing color, and thoracic temperature. In combination with previous work (Chapter 4) showing gene expression differences between dispersing and non-dispersing individuals, this points to a potential suite of traits characterizing dispersing individuals, and hence a possible 'dispersal syndrome' in *P. smintheus*. Further study is required to determine whether the *Pgi* genotypes associated with dispersal in *P. smintheus* have different enzyme-kinetic properties. We also require estimates of flight capacity and flight metabolic rate, under controlled conditions, in individuals of *P. smintheus* with different *Pgi* genotypes. Given the importance of dispersal to the ecology and evolution of spatially structured populations and communities, unraveling the genetic variation underlying dispersal is critical for a full understanding of the dynamics of spatially structured systems.

5.5 Literature cited

Arrese EL, Soulages JL (2010). Insect fat body: Energy, metabolism, and regulation. *Annual Review of Entomology* 55: 207–225.

Benjamini Y, Hochberg Y (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological* 57: 289–300.

Bonte D, Van Dyck H, Bullock JM, Coulon A, Delgado M, Gibbs M, *et al.* (2012). Costs of dispersal. *Biological Reviews* 87: 290–312.

Brakefield PM, Reitsma N (1991). Phenotypic plasticity, seasonal climate and the population biology of Bicyclus butterflies (Satyridae) in Malawi. *Ecological Entomology* 16: 291–303.

Calvert WH, Brower LP (1986). The location of monarch butterfly (*Danaus plexippus l.*) Overwintering colonies in Mexico in relation to topography and climate. *Journal of the Lepidopterists' Society* 19: 164–187.

Caplins SA, Gilbert KJ, Ciotir C, Roland J, Matter SF, Keyghobadi N (2014). Landscape structure and the genetic effects of a population collapse. *Proceedings of the Royal Society B* 281: 20141798.

Chaplin SB, Wells PH (1982). Energy reserves and metabolic expenditures of monarch butterflies overwintering in southern California. *Ecological Entomology* 7: 249–256.

Chaput M, Claes, V, Portetelle D, Cludts I, Cravador A, Burny A, *et al.* (1988). The neurotrophic factor neuroleukin is 90% homologous with phosphohexose isomerase. *Nature* 332: 454–455.

Clobert J, Baguette M, Benton TG, Bullock JM (2012). Dispersal ecology and evolution. Oxford University Press. Oxford, UK.

Clobert J, Le Galliard J-F, Cote J, Meylan S, Massot M (2009). Informed dispersal, heterogeneity in animal dispersal syndromes and the dynamics of spatially structured populations. *Ecology Letters* 12: 197–209.

Dahlhoff EP, Rank NE (2000). Functional and physiological consequences of genetic variation at phosphoglucose isomerase: Heat shock protein expression is related to enzyme genotype in a montane beetle. *Proceedings of the National Academy of Sciences of the United States of America* 97: 10056–10061.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, *et al.* (2011). The variant call format and VCFtools. *Bioinformatics* 27: 2156–2158.

Dean AM, Golding GB (1997). Protein engineering reveals ancient adaptive replacements in isocitrate dehydrogenase. *Proceedings of the National Academy of Sciences of the United States of America* 94: 3104–3109.

Dick CA, Rank NE, McCarthy M, McWeeney S, Hollis D, Dahlhoff EP (2013). Effects of temperature variation on male behavior and mating success in a montane beetle. *Physiological and Biochemical Zoology* 86: 432–440.

Dunning LT, Dennis AB, Thomson G, Sinclair BJ, Newcomb RD, Buckley TR (2013). Positive selection in glycolysis among Australasian stick insects. *BMC Evolutionary Biology* 13: 215.

Ellis JS, Turner LM, Knight ME (2013). Lack of Variation at phosphoglucose isomerase (Pgi) in bumblebees: Implications for Conservation Genetics Studies. *PLoS ONE* 8: e65600.

Filatov DA, Charlesworth D (1999). DNA polymorphism, haplotype structure and balancing selection in the Leavenworthia PgiC Locus. *Genetics* 153: 1423–1434.

Gerstein M, Chothia C (1991). Analysis of protein loop closure. *Journal of Molecular Biology* 220: 133–149.

Goff J, Yerke C, Keyghobadi N, Matter SF (2018). Dispersing male *Parnassius smintheus* butterflies are more strongly affected by forest matrix than are females. *Insect Science*. Early view, doi: 10.1111/1744-7917.12592.

Guppy CS (1986a). Geographic variation in wing melanism of the butterfly *Parnassius phoebus* F. (Lepidoptera: Papilionidae). *Canadian Journal of Zoology* 64: 956–962.

Guppy CS (1986b). The Adaptive significance of alpine melanism in the butterfly *Parnassius phoebus* F. (Lepidoptera: Papilionidae). *Oecologia* 70: 205–213.

Haag CR, Saastamoinen M, Marden JH, Hanski I (2005). A candidate locus for variation in dispersal rate in a butterfly metapopulation. *Proceedings of the Royal Society B* 272: 2449–2456.

Hall JG (1985). The adaptation of enzymes to temperature: catalytic characterization of glucosephosphate isomerase homologues isolated from *Mytilus edul*is and *Isognomon alatus*, Bivalve Molluscs inhabiting different thermal environments. *Molecular Biology and Evolution* 2: 25 l – 269.

Hanski I, Saccheri I (2006). Molecular-level variation affects population growth in a butterfly metapopulation. *PLoS Biology* 4: e129.

Hoffman R. (1981). Evolutionary genetics of Metridium senile. I. Kinetic differences in phosphoglucose isomerase allozymes. *Biochemical Genetics* 19: 129–144.

Jangjoo M, Matter SF, Roland J, Keyghobadi N (2016). Connectivity rescues genetic diversity after a demographic bottleneck in a butterfly population network. *Proceedings of the National Academy of Sciences of the United States of America* 113: 10914–10919.

Kallioniemi E, Hanski I (2011). Interactive effects of *Pgi* genotype and temperature on larval growth and survival in the Glanville fritillary butterfly: *Pgi* genotype and temperature effects on larval growth. *Functional Ecology* 25: 1032–1039.

Karl I, Schmitt T, Fischer K (2008). Phosphoglucose isomerase genotype affects life-history traits and cold stress resistance in a Copper butterfly. *Functional Ecology* 22: 887–894.

Keyghobadi N, Roland J, Strobeck C (1999). Influence of landscape on the population genetic structure of the alpine butterfly *Parnassius smintheus* (Papilionidae). *Molecular Ecology* 8: 1481–1495.

Keyghobadi N, Roland J, Strobeck C (2005). Genetic differentiation and gene flow among populations of the alpine butterfly, *Parnassius smintheus*, vary with landscape connectivity: Landscape connectivity and gene flow. *Molecular Ecology* 14: 1897–1909.

Koressaar T, Remm M (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23: 1289–1291.

Li H (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27: 2987–2993.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, *et al.* (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–2079.

Li Y, Andersson S (2016). The 3-D Structural Basis for the Pgi Genotypic Differences in the Performance of the Butterfly *Melitaea cinxia* at Different Temperatures (CJ Breuker, Ed.). *PLOS ONE* 11: e0160191.

Librado P, Rozas J (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451-1452.

Marden JH (2013). Nature's inordinate fondness for metabolic enzymes: why metabolic enzyme loci are so frequently targets of selection. *Molecular Ecology* 22: 5743–5764.

Masters AR, Malcolm SB, Brower LP (1988). Monarch butterfly (*Danaus Plexippus*) thermoregulatory behavior and adaptations for overwintering in Mexico. *Ecology* 69: 458–467.

Matter SF, Keyghobadi N, Roland J (2014). Ten years of abundance data within a spatial population network of the alpine butterfly, *Parnassius smintheus*. *Ecology* 95: 2985–2985.

Mitikka V, Hanski I (2010). *Pgi* genotype influences flight metabolism at the expanding range margin of the European map butterfly. *Annales Zoologici Fennici* 47: 1–14.

Neve G, Hall C (2016). Variation of thorax flight temperature among twenty Australian butterflies (Lepidoptera: Papilionidae, Nymphalidae, Pieridae, Hesperiidae, Lycaenidae). *European Journal of Entomology* 113: 571–578.

Newcomb RD, Campbell PM, Ollis DL, Cheah E, Russell RJ, Oakeshott JG (1997). A single amino acid substitution converts a carboxylesterase to an organophosphorus

hydrolase and confers insecticide resistance on a blowfly. *Proceedings of the National Academy of Sciences of the United States of America* 94: 7464–7468.

Niitepõld K (2010). Genotype by temperature interactions in the metabolic rate of the Glanville fritillary butterfly. *Journal of Experimental Biology* 213: 1042–1048.

Niitepõld K, Smith AD, Osborne JL, Reynolds DR, Carreck N, Martin AP, *et al.* (2009). Flight metabolic rate and Pgi genotype influence butterfly dispersal rate in the field. *Ecology* 90: 2223–2232.

Orsini L, Wheat CW, Haag CR, Kvist J, Frilander MJ, Hanski I (2009). Fitness differences associated with Pgi SNP genotypes in the Glanville fritillary butterfly (Melitaea cinxia): Reduced fitness in a butterfly metapopulation. *Journal of Evolutionary Biology* 22: 367–375.

Rank NE, Bruce DA, McMillan DM, Barclay C, Dahlhoff EP (2007). Phosphoglucose isomerase genotype affects running speed and heat shock protein expression after exposure to extreme temperatures in a montane willow beetle. *Journal of Experimental Biology* 210: 750–764.

Rank NE, Dahlhoff EP (2002). Allele frequency shifts in response to climate change and physiological consequences of allozyme variation in a montane insect. *Evolution* 56: 2278–2289.

Ricketts TH (2001). The matrix matters: Effective isolation in fragmented landscapes. *The American Naturalist* 158: 87–99.

Roland J, Keyghobadi N, Fownes S (2000). Alpine *Parnassius* butterfly dispersal: Effects of landscape and population size. *Ecology* 81: 1642–1653.

Roland J, Matter SF (2007). Encroaching forests decouple alpine butterfly population dynamics. *Proceedings of the National Academy of Sciences of the United States of America* 104: 13702–13704.

Saastamoinen M, Brakefield PM, Ovaskainen O (2012). Environmentally induced dispersal-related life-history syndrome in the tropical butterfly, *Bicyclus anynana*. *Journal of Evolutionary Biology* 25: 2264–2275.

Sato Y, Nishida M (2007). Post-duplication charge evolution of phosphoglucose isomerases in teleost fishes through weak selection on many amino acid sites. *BMC Evolutionary Biology* 7: 204.

Sim N-L, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC (2012). SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Research* 40: W452–W457.

Storz JF, Wheat CW (2010). Integrating evolutionary and functional approaches to infer adaptation at specific loci. *Evolution* 64: 2489–2509.

Thomas R, Pichersky E, Gottlieb LD (1993). Molecular characterization of duplicate cytosolic phosphoglucose isomerase genes in clarkia and comparison to the single Gene in Arabidopsis. *Genetics* 135: 895–905.

Van Dyck H, Matthysen E (1998). Thermoregulatory differences between phenotypes in the speckled wood butterfly: Hot perchers and cold patrollers? *Oecologia* 114: 326–334.

Watt WB (1983). Adaptation at specific loci. II. Demographic and biochemical elements in the maintenance of the *Colias* PGI polymorphism. *Genetics* 103: 691–724.

Watt WB (1986). Power and Efficiency as Indexes of Fitness in Metabolic Organization. *The American Naturalist* 127: 629–653.

Watt WB (1992). Eggs, enzymes, and evolution: natural genetic variants change insect fecundity. *Proceedings of the National Academy of Sciences of the United States of America* 89: 10608–10612.

Watt WB (1977). Adaptation at specific loci. I. Natural selection on phosphoglucose isomerase of *Colias* butterflies: biochemical and population aspects. *Genetics* 87: 177-194.

Watt WB (2003). Mechanistic studies of butterfly adaptation. In Butterflies: Ecology and evolution taking flight. University of Chicago, Chicago.

Watt WB, Cassin RC, Swan (1983). Adaptation at specific loci. III. Field behavior and survivorship differences among *Colias Pgi* genotypes are predictable from in vitro biochemistry. *Genetics* 103: 725–739.

Watt WB, Dean AM (2000). Molecular-functional studies of adaptive genetic variation in prokaryotes and eukaryotes. *Annual Review of Genetics* 34: 593–622.

Watt WB, Wheat CW, Meyer EH, Martin J-F (2003). Adaptation at specific loci. VII. Natural selection, dispersal and the diversity of molecular–functional variation patterns among butterfly species complexes (*Colias*: Lepidoptera, Pieridae). *Molecular Ecology*: 12: 1265-75.

Weir B. (1996). Genetic data analysis II. Sinauer Associates. Sunderland, MA.

Wheat CW (2010). Phosphoglucose isomerase (*Pgi*) performance and fitness effects among Arthropods and its potential role as an adaptive marker in conservation genetics. *Conservation Genetics* 11: 387–397.

Wheat CW (2012). Dispersal genetics: emerging insights from fruit flies, butterflies, and beyond. In Dispersal Ecology and Evolution. Oxford University Press, USA, pp 95–107.

Wheat CW, Haag CR, Marden JH, Hanski I, Frilander MJ (2010). Nucleotide Polymorphism at a Gene (*Pgi*) under Balancing Selection in a Butterfly Metapopulation. *Molecular Biology and Evolution* 27: 267–281.

Wheat CW, Hill J (2014). *Pgi*: the ongoing saga of a candidate gene. *Current Opinion in Insect Science* 4: 42–47.

Wheat CW, Watt WB, Pollock DD, Schulte PM (2006). From DNA to fitness differences: sequences and structures of adaptive variants of *Colias* phosphoglucose isomerase (PGI). *Molecular Biology and Evolution* 23: 499–512.

Zera AJ (1987). Temperature-dependent kinetic variation among phosphoglucose isomerase allozymes from the wing-polymorphic water strider. *Molecular Biology and Evolution* 4: 226–285.

Zera AJ, Brisson JA (2012). Quantitative, physiological, and molecular genetics of dispersal/migration. In Dispersal Ecology and Evolution. Oxford University Press, USA, pp 63–82.

Chapter 6

6 General Discussion

6.1 Overview of dissertation

In my thesis, I built upon the rich body of work that describes patterns of genetic variation in natural populations, and attempts to determine the ecological and evolutionary processes that determine those patterns. I integrated data on neutral and adaptive genetic variation with data on dispersal, population dynamics and landscape effects. I did this in a naturally occurring spatial population network of the Rocky Mountain Apollo butterfly, *P. smintheus*, which is arguably emerging as a model system for the integrated study of population dynamics and population genetics. The first goal of my dissertation was to dissect the neutral processes (i.e., genetic drift and gene flow) that affect genetic variation in populations inhabiting heterogeneous landscapes during repeated demographic bottlenecks, and determine how those processes are mediated by landscape structure and patch connectivity (Chapters 2 and 3). My second goal was to evaluate genetic variation that might underlie the ecologically important process of dispersal, using gene expression and candidate locus approaches (Chapters 4 and 5).

Taken together, the chapters of my dissertation provide valuable evidence of how dispersal and population size fluctuations affect drift, gene flow and potentially selection, to ultimately shape genetic diversity and patterns of genetic differentiation in dynamic populations occupying heterogeneous landscapes.

6.2 Main contributions of dissertation

6.2.1 Integration of genetic variation with population dynamics

All natural populations will vary in size over time due to a wide potential range of density dependent and density independent factors (Hansen et al. 1999; Bjørnstad and Grenfell 2001). If the fluctuations in population size are pronounced enough, they can drive loss of genetic diversity, and divergence of allele frequencies among populations, even in the absence of any selection (Nei et al. 1975; Garza and Williamson 2001). There has been considerable scientific interest in the genetic effects of population size fluctuations. Much of the research has focused on effects of demographic bottleneck events, with or without subsequent demographic recovery. Both theoretical and empirical studies have explored the genetic consequences of such events (Hoelzel et al. 2002; Spielman et al. 2004; Caplins et al. 2014). A large portion of this work has been conducted in the area of conservation genetics, due to concern about loss of genetic diversity in populations that have experienced significant bottlenecks, and the resulting possibility of inbreeding depression and erosion of evolutionary potential (Frankham et al. 2010). Despite considerable interest in the genetic effects of demographic bottlenecks, most studies to date have assessed single bottlenecks in single populations (Whitehouse and Harley 2001; Groombridge et al. 2000; Hoelzel et al. 2002; Spielman et al. 2004).

I took advantage of a long-term dataset derived from multi-year sampling of genetic and demographic data, obtained by mark-recapture, to investigate the effects of repeated bottlenecks in a population network. I examined, in detail, how the distribution of neutral genetic variation within and among several populations changed over repeated

demographic bottlenecks. In Chapter 2, through comparison of the effects of two bottlenecks of differing intensity and duration, I found that the severity of a bottleneck event, which is reflected by the extent to which population size is reduced, the duration at low size, and the extent of population recovery (Williamson-Natesan 2005), can determine the amount of genetic diversity lost from within populations. I also showed how the recovery of genetic variation following a bottleneck depends on immigration, which is mediated by habitat patch connectivity. Finally, in both Chapters 2 and 3, I showed how the dynamics of genetic variation across the population network are driven by a continual shifting of the relative dominance of genetic drift and gene flow as populations fluctuate in size. Bottlenecks appear to consistently erase spatial patterns of genetic structure from this system through drift effects, but with rebounding population sizes and through the action of gene flow, spatial patterns of genetic variation are quickly re-established among the populations.

6.2.2 Integration of genetic variation with dispersal data

Dispersal is a critical process affecting the ecology and evolution of populations, communities and ecosystems (Clobert et al. 2012). Dispersal is necessary for metapopulation persistence, metacommunity structure, nutrient flows, and gene flow (Clobert et al. 2012; Travis et al. 2013). Traditionally, the estimation of dispersal rates and patterns has been divided into ‘direct’ and ‘indirect’ approaches (Balkenhol et al. 2015). Direct approaches rely on techniques such as mark-recapture or radio-telemetry to track movement of individuals across the landscape (Růžičková and Veselý 2016). Indirect methods typically use spatial patterns of genetic variation to infer historical or

contemporary dispersal rates and patterns (Slatkin 1987). A large proportion of studies of genetic variation in natural populations have attempted to indirectly estimate or make inferences about dispersal (Shipham et al. 2013; Lowe et al. 2006).

Despite the clear link between dispersal, which underlies gene flow, and patterns of genetic variation, as well as widespread interest in making indirect inferences about dispersal based on genetic data, a relatively small proportion of all population genetic studies have combined genetic data with direct estimates of dispersal (Bohonak 1999; Keyghobadi et al. 2005, b; Fedy et al. 2008; Sigaard et al. 2008). Among studies that have done so, some have shown directly how rates of movement between sites can be related to the degree of genetic differentiation between those sites (Keyghobadi et al. 1999; Sigaard et al. 2008). In the Glanville fritillary butterfly, direct movement and dispersal data have been combined with genotyping at the *Pgi* locus to show the influence of this gene on dispersal behavior (Niitepõld et al. 2009).

Working in a system where mark–recapture studies have been conducted in a population network since 1995, I was able to combine analyses of genetic variation with data on dispersal rates and dispersal history of individuals. In Chapter 3, I examined the correlation between genetic differentiation and movement patterns inferred from mark-recapture data. Importantly, I showed that the degree of correlation can be highly dynamic over short time periods in response to population size fluctuations. Using estimates of dispersal, as well as measures of connectivity calibrated with mark-recapture data, I showed that dispersal accompanied by gene flow is a key process that not only maintains genetic variation within population (Chapter 2) but also rapidly redistributes genetic variation among populations after a demographic bottleneck (Chapter 3).

In Chapter 4, I used data on dispersal history of individuals in combination with RNA-seq data to determine gene expression differences that are likely to be a consequence of long-distance flight associated with dispersal. My results suggest a high potential cost of dispersal, as individuals that had dispersed between habitat patches showed elevated expression of genes related to re-establishment of energy reserves and stress responses, compared to non-dispersers. I also used data on dispersal history of individuals to explore potential adaptive variation underlying dispersal at the well-endorsed candidate locus *Pgi*. In Chapter 5, I found non-synonymous *Pgi* polymorphisms associated with dispersal and movement distance. My study provides another example of the functional consequences of variation at *Pgi* in insects (Wheat and Hill 2014), and represents one of few systems in which a specific gene of large effect underlying dispersal has been identified.

6.2.3 Evaluation of landscape effects

A large and growing body of work in 'landscape genetics' has, over the past two decades, explored the influence of landscape composition and configuration on patterns of genetic variation in natural populations (Manel et al. 2003; Storfer et al. 2007; Manel and Holderegger 2013). Although landscape genetic studies can be very diverse in scope and approach, the large majority of studies focus on evaluating the influence of intervening landcover on patterns of genetic differentiation among populations (Storfer et al. 2010; DiLeo and Wagner 2016), and on how reduced connectivity among populations affects genetic differentiation (Keyghobadi et al. 2005; Vandergast et al. 2009).

My work provides at least three important contributions to this field. First, Chapter 2 represents one of the first studies to document the direct effect of patch connectivity on recovery of genetic diversity after a demographic bottleneck, highlighting the importance of connectivity for maintaining genetic diversity in natural populations. Second, in Chapter 3, I showed that the relationship between landscape variables (in this case, intervening forest) and genetic differentiation can change very rapidly as populations experience repeated fluctuations in size over time. This is an important result because it indicates how temporary the associations between landscape variables and genetic differentiation, which form the basis of most landscape genetic studies, can potentially be. My work therefore suggests more caution in the interpretation of landscape genetic studies conducted at a single point in time (which represent the large majority of landscape genetic studies) and suggests more multi-time point studies are needed. Finally, in Chapter 5 I found evidence that variation at a gene, *Pgi*, is potentially influencing dispersal in *P. smintheus*. This works opens the door to future studies in this system that can explore how genetic variation at *Pgi* is distributed across space and in relation to landscape features and patch connectivity.

6.3 General summary

I conducted an analysis of neutral and adaptive genetic variation in a spatial population network. My analyses included samples collected over multiple years and were informed by a unique long-term mark-recapture dataset. I demonstrated how landscape structure

and population size fluctuations can interact to shape neutral patterns of genetic variation within and among populations (Chapters 2 and 3). I also demonstrated how patterns of spatial genetic structure can change rapidly over time (Chapter 3). I showed that the dispersal history of individuals is reflected in differences in overall gene expression profiles (Chapter 4), as well as in the DNA sequence of the candidate gene *Pgi* (Chapter 5).

6.4 Future directions

My work suggests several worthwhile lines of inquiry in this study system. Further study is required to determine how adaptive genetic variation changes across the population network of *P. smintheus* in response to the demographic bottlenecks. Population size fluctuations can affect adaptive genetic variation by providing an opportunity for drift and selection to increase the frequency of some potentially adaptive genetic variants during and immediately after bottlenecks, while gene flow in the periods between bottlenecks spreads those variants across the network. Patterns of change over time at the candidate locus *Pgi* should be examined using data that I have collected on *Pgi* variation from four different time periods (Chapter 5).

Analysis of genome-wide SNPs could also yield important insights into the dynamics of adaptive genetic variation, over both space and time, in this system. The RNA-seq dataset and reference transcriptome that I developed for *P. smintheus* in Chapter 4 represent a significant resource for such studies. These resources could be used

to develop assays for SNPs that are located in transcribed regions, and therefore more likely to be of adaptive significance. The transcriptome can also serve as a reference to potentially identify the function of any SNPs that are assayed using methods such as RADSeq and GBS and are found to display signatures of selection.

Measuring flight performance and its relation to variation at the *Pgi locus* is also an important area of further inquiry. Controlled flight trial studies could dissect the short- and long-term effects of sustained flight on gene expression. Flight trials under controlled conditions as well as in the field would provide more insight into performance differences among individuals with different *Pgi* genotypes. If coupled with studies of gene expression and flight metabolic rates, such studies could yield important insights into the functional significance of *Pgi* variation and the factors potentially contributing to a 'dispersal phenotype' in *P. smintheus*

6.5 Implications for conservation of alpine species

My findings have implications for the conservation of *P. smintheus* and other alpine species. Because they occupy colder, high-altitude areas, alpine species may be especially vulnerable to the effects of climate change (Peters and Darling 1985; Parmesan 1996; Roland et al. 2000; Roland and Matter 2016). One prediction for high-altitude species under a generally warming climate is that the areas providing suitable habitat and climatic conditions for these species are likely to be reduced considerably as they are 'pushed' upwards along elevational gradients (Taylor 1995; Woodward et al. 1995). In this way,

climate change can contribute indirectly to habitat loss and fragmentation for such species. The meadow habitats occupied by *P. smintheus* for example are likely to be reduced in area and fragmented as tree-line encroaches upward along ridges and mountain slopes (Gehrig-Fasel et al. 2007). Climatic instability introduced by climate change may also affect populations of alpine species. In the case of *P. smintheus*, more extreme winter weather, both warm and cold, is predicted to lead to more frequent years of very low abundance and thereby, more frequent and severe demographic bottlenecks (Roland and Matter 2016).

My results suggest there will be an important interaction between the two factors of habitat fragmentation, which reduces connectivity among populations, and increasing climate variability, which may increase demographic stochasticity, on genetic diversity of *P. smintheus* populations. Specifically, my work shows that connectivity among populations is necessary to counteract the loss of genetic diversity that occurs as a result of demographic bottlenecks (Chapter 2). Increased population isolation and lower levels of gene flow resulting from habitat fragmentation will hinder recovery of genetic diversity and spatial genetic patterns after demographic bottlenecks. My works also suggests that as a result of the combined effects of habitat fragmentation and increased climatic and demographic variability, the relative influence of drift in populations of *P. smintheus* will increase (Chapter 3). Overall therefore, genetic diversity of populations is likely to decrease.

To the extent that other alpine species are also exposed to habitat loss and demographic stochasticity as a result of a changing climate, my results suggest these species are vulnerable to a loss of genetic diversity. Erosion of genetic diversity, in turn,

can result in inbreeding depression and reduce the potential to adapt to future environmental change, thereby increasing extinction risk (Saccheri et al. 1998; Spielman et al. 2004). My work highlights the need to maintain connectivity across landscapes and among populations in the face of climate change.

6.6 Literature cited

Balkenhol N, Cushman S, Storfer A, Waits L (2015). Landscape genetics: concepts, methods, applications. Wiley-Blackwell, Chichester, UK.

Bjørnstad ON, Grenfell BT (2001). Time series analysis of population fluctuations in animals. *Science* 293: 638–643.

Bohonak AJ (1999). Dispersal, gene flow, population structure. *The Quarterly Review of Biology* 74: 21–45.

Caplins SA, Gilbert KJ, Ciotir C, Roland J, Matter SF, Keyghobadi N (2014). Landscape structure and the genetic effects of a population collapse. *Proceedings of the Royal Society B* 281: 20141798.

Clobert J, Baguette M, Benton TG, Bullock JM (2012). Dispersal ecology and evolution. Oxford University Press, Oxford, UK.

DiLeo MF, Wagner HH (2016). A landscape ecologist's agenda for landscape genetics. *Current Landscape Ecology Reports* 1: 115–126.

Fedy BC, Martin K, Ritland C, Young J (2008). Genetic and ecological data provide incongruent interpretations of population structure and dispersal in naturally subdivided populations of white-tailed ptarmigan (*Lagopus leucura*): white-tailed ptarmigan dispersal. *Molecular Ecology* 17: 1905–1917.

Frankham R, Briscoe DA, Ballou JD (2010). Introduction to conservation genetics. Cambridge University Press, Cambridge.

Garza JC, Williamson EG (2001). Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology* 10: 305–318.

Gehrig-Fasel J, Guisan A, Zimmermann NE (2007). Tree line shifts in the Swiss Alps: Climate change or land abandonment? *Journal of Vegetation Science* 18: 571–582.

Groombridge JJ, Jones CG, Bruford MW, Nichols RA (2000). Conservation biology: 'Ghost' alleles of the Mauritius kestrel. *Nature* 403, 616.

Hansen TF, Stenseth NC, Henttonen H, Tast J (1999). Interspecific and intraspecific competition as causes of direct and delayed density dependence in a fluctuating vole population. *Proceedings of the National Academy of Sciences* 96: 986–991.

Hoelzel AR, Fleischer RC, Campagna C, Le Boeuf BJ, Alvord G (2002). Impact of a population bottleneck on symmetry and genetic diversity in the northern elephant seal. *Journal of Evolutionary Biology* 15: 567–575.

Keyghobadi N, Roland J, Strobeck C (1999). Influence of landscape on the population genetic structure of the alpine butterfly *Parnassius smintheus* (Papilionidae). *Molecular Ecology* 8: 1481–1495.

Keyghobadi N, Roland J, Strobeck C (2005). Genetic differentiation and gene flow among populations of the alpine butterfly, *Parnassius smintheus*, vary with landscape connectivity. *Molecular Ecology* 14: 1897–1909.

Lowe WH, Likens GE, McPeek MA, Buso DC (2006). Linking direct and indirect data on dispersal: isolation by slope in a headwater stream salamander. *Ecology* 87: 334–339.

Manel S, Holderegger R (2013). Ten years of landscape genetics. *Trends in Ecology & Evolution* 28: 614–621.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003). Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution* 18: 189–197.

Nei M, Maruyama T, Chakraborty R (1975). Bottleneck effect and genetic-variability in populations. *Evolution* 29: 1–10.

Niitepõld K, Smith AD, Osborne JL, Reynolds DR, Carreck N, Martin AP, Marden JH (2009). Flight metabolic rate and *Pgi* genotype influence butterfly dispersal rate in the field. *Ecology* 90: 2223–2232.

Orsini L, Wheat CW, Haag CR, Kvist J, Frilander MJ, Hanski I (2009). Fitness differences associated with *Pgi* SNP genotypes in the Glanville fritillary butterfly (*Melitaea cinxia*): Reduced fitness in a butterfly metapopulation. *Journal of Evolutionary Biology* 22: 367–375.

Parmesan C (1996). Climate and species'range. *Nature* 382: 765–766.

Peters RL, Darling DS (1985). The greenhouse effect and nature reserves. *BioScience* 35: 707–717.

Roland J, Keyghobadi N, Fownes S (2000). Alpine *Parnassius* butterfly dispersal: Effects of landscape and population size. *Ecology* 81: 1642–1653.

Roland J, Matter SF (2016). Pivotal effect of early-winter temperatures and snowfall on population growth of alpine *Parnassius smintheus* butterflies. *Ecological Monographs* 86: 412–428.

Růžičková J, Veselý M (2016). Using radio telemetry to track ground beetles: Movement of *Carabus ullrichii*. *Biologia* 71: 924–930.

Saccheri I, Kuussaari M, Kankare M, Vikman P, Fortelius W, Hanski I (1998). Inbreeding and extinction in a butterfly metapopulation. *Nature* 392: 491–494.

Shipham A, Schmidt DJ, Hughes JM (2013). Indirect estimates of natal dispersal distance from genetic data in a stream-dwelling fish (*Mogurnda adspersa*). *Journal of Heredity* 104: 779–790.

Sigaard P, Pertoldi C, Madsen AB, SøGaard B, Loeschcke V (2008). Patterns of genetic variation in isolated Danish populations of the endangered butterfly *Euphydryas aurinia*: Genetic variability in *Euphydryas aurinia*. *Biological Journal of the Linnean Society* 95: 677–687.

Slatkin M (1987). Gene flow and the geographic structure of natural populations. *Science* 236: 787–792.

Spielman D, Brook DW, Frankham R (2004). Most species are not driven to extinction before genetic factors impact them. *Proceedings of the National Academy of Sciences* 101: 15261–15264.

Storfer A, Murphy MA, Evans JS, Goldberg CS, Robinson S, Spear SF, et al. (2007). Putting the 'landscape' in landscape genetics. *Heredity* 98: 128–142.

Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP (2010). Landscape genetics: where are we now? *Molecular Ecology* 19: 3496–3514.

Taylor AH (1995). Forest expansion and climate change in the mountain hemlock (*Tsuga mertensiana*) zone, Lassen Volcanic National Park, California, U.S.A. *Arctic and Alpine Research* 27: 207-216.

Travis JMJ, Delgado M, Bocedi G, Baguette M, Bartoń K, Bonte D, et al. (2013). Dispersal and species' responses to climate change. *Oikos* 122: 1532–1540.

Vandergast AG, Lewallen EA, Deas J, Bohonak AJ, Weissman DB, Fisher RN (2009). Loss of genetic connectivity and diversity in urban microreserves in a southern California endemic Jerusalem cricket (Orthoptera: Stenopelmatidae: *Stenopelmatus* n. sp. "santa monica"). *Journal of Insect Conservation* 13: 329–345.

Wheat CW, Hill J (2014). *Pgi*: the ongoing saga of a candidate gene. *Current Opinion in Insect Science* 4: 42–47.

Whitehouse AM, Harley EH (2001). Post-bottleneck genetic diversity of elephant populations in South Africa, revealed using microsatellite analysis. *Molecular Ecology* 10: 2139–2149.

Williamson-Natesan EG (2005). Comparison of methods for detecting bottlenecks from microsatellite loci. *Conservation Genetics* 6: 551–562.

Woodward A, Schreiner EG, Silsbee DG (1995). Climate, geography, tree establishment in subalpine meadows of the Olympic Mountains, Washington, U.S.A. *Arctic and Alpine Research* 27: 217-225.

Appendices

Appendix A: Chapter 2

Table A1 Metrics of genetic diversity for populations of *Parnassius smintheus* before and after the 2010 demographic bottleneck.

Patch/population	Year	G1	g2	H	J	K	L	M	O	Z
No. genotyped individuals	2008	39	17	12	46	20	72	41	30	31
	2013	10	9	6	17	16	12	37	6	5
A_R	2008	4.71	4.89	4.84	4.57	4.88	4.75	4.69	4.79	4.70
	2013	4.15	4.44	3.39	4.70	4.38	4.13	4.64	3.30	3.57
H_E	2008	0.69	0.74	0.66	0.69	0.73	0.70	0.69	0.70	0.70
	2013	0.64	0.71	0.55	0.71	0.68	0.67	0.69	0.64	0.66
Null freq.	2008	0.12	0.07	0.14	0.13	0.08	0.10	0.10	0.11	0.12
	2013	0.06	0.10	0.02	0.09	0.08	0.11	0.12	0.10	0.03
N	N_{2010}	4.4	6.8	4.9	30.1	22.3	42.3	84.6	6.6	5.6

No. genotyped individuals is the number of individuals that amplified at seven microsatellite loci. Allelic richness (A_R; rarefaction to 10 gene copies), unbiased expected heterozygosity (H_E) and frequency of null allele (Null Freq.) are averaged across loci. N2010 is an index of 2010 adult abundance estimated using Craig's method for mark–recapture data.

Table A2 Ratio of sample size to the index of population size for populations of *Parnassius smintheus* in the years 2008, 2010, 2011 and 2013.

	2008			2010			2011			2013		
Patch/Population	**n**	**Index**	**Ratio**	**n**	**Index**	**Ratio**	**n**	**Index**	**Ratio**	**n**	**Index**	**Ratio**
F	-	-	-	2	6.8	0.29	5	5.7	0.88	-	-	-
G1	39	201.8	0.19	2	4.4	0.45	-	-	-	10	27.8	0.36
g2	17	139.4	0.12	3	6.8	0.44	2	6.6	0.30	9	14.7	0.61
H	12	104	0.11	2	4.9	0.41	-	-	-	6	8.4	0.71
J	46	297.2	0.15	-	-	-	-	-	-	17	56.6	0.30
K	20	227.8	0.09	5	22.3	0.22	11	22.5	0.49	16	42.6	0.38
L	72	429.8	0.17	3	42.3	0.07	-	-	-	12	29	0.41
M	41	803.7	0.05	13	84.6	0.15	11	41.2	0.27	37	112	0.33
O	30	79.1	0.38	3	6.6	0.45	-	-	-	6	4.4	1.36
S	-	-	-	2	4.4	0.45	-	-	-	-	-	-
Z	31	39.1	0.79	-	-	-	3	4.4	0.68	5	11.1	0.45
Mean	34.2	258.0	0.23	3.9	20.3	0.40	6.4	16.1	0.52	13.1	34.1	0.55

'n' is the number of individuals genotyped (sample size) for each population in each year, 'Index' is an index of population size calculated as the maximum daily estimate of population size from Craig's method applied to mark-recapture data. 'Ratio' is 'n' divided by 'Index'.

Table A3. Allelic richness rarefied to four gene copies in the years 2008, 2010, 2011 and 2013.

Patch/population	Year	F	G1	g2	H	J	K	L	M	O	S	Z
No. genotyped individuals	2010	2	2	3	2	-	5	3	13	3	2	-
	2011	5	-	2	-	-	11	-	11	-	-	3
A_R	2008	2.75	2.68	2.82	2.69	2.65	2.77	2.71	2.70	2.68	2.66	2.72
	2010	2.71	2.43	2.44	2.29	-	2.61	2.65	2.75	2.41	2.29	-
	2011	2.66	-	1.86	-	-	2.24	-	1.87	-	-	2.58
	2013	-	2.56	2.66	2.31	2.73	2.61	2.55	2.67	2.35	-	2.52

No. genotyped individuals is the number of individuals that amplified at seven microsatellite loci. Allelic richness (A_R; rarefaction to four gene copies) is averaged across loci. Sample sizes for all sites in 2008 and 2013 are provided in Table A2. Dashes indicate that tissue samples were not available.

Table A4 Summary of linear regression models with $\Delta AIC_c < 12$ explaining the proportional loss of allelic richness (A_R; rarefaction to 10 gene copies) between 2008 and 2013. Models are sorted in increasing order of AICc values.

Models	**AICc**	**ΔAICc**	***r²***	**Adjusted *r²***	***P* value**
C2012	-17.77	0	0.81	0.78	0.0009
C2011	-13.85	3.93	0.71	0.66	0.004
N2010+C2012	-13.64	4.13	0.86	0.82	0.002
C2009	-12.53	5.24	0.66	0.61	0.007
C2010	-11.44	6.34	0.62	0.56	0.011
N2010+C2009	-11.15	6.62	0.82	0.76	0.005
N2011+C2012	-10.71	7.06	0.81	0.75	0.006
N2010+C2010	-10.31	7.46	0.8	0.74	0.007
N2010+C2011	-9.93	7.84	0.79	0.73	0.008
N2011+C2009	-7.55	10.22	0.73	0.65	0.017
N2010	-6.89	10.89	0.37	0.28	0.081
N2011+C2011	-6.69	11.08	0.71	0.61	0.023

Abbreviations used for variables: C, connectivity for given year; N, population size for given year.'+' indicates additive term.

Table A5 Summary of linear regression models explaining the proportional change in allelic richness (A_R; rarefaction to four gene copies) during distinct phases of demographic decline and recovery. For models predicting loss of A_R between 2008 and 2010, those with $\Delta AIC_c < 12$ are shown and are sorted in increasing order of AIC_c values. For the other time periods, small samples sizes precluded calculation of AIC_c for some models due to overfitting. For these periods, models are ranked based on adjusted r^2 and those with adjusted $r^2 > 0.6$ are shown.

Dependent Variable	**Models**	**AICc**	**ΔAICc**	***r²***	**Adjusted *r²***	***P* value**
Loss of A_R between 2008 and 2010	N2010	-22.83	0	0.57	0.51	0.018
	N2011	-20.06	2.77	0.41	0.33	0.059
	C2012	-18.07	4.76	0.27	0.17	0.14
	N2010+C2009	-17.72	5.10	0.66	0.54	0.038
	C2009	-17.51	5.32	0.22	0.11	0.19
	C2011	-17.23	5.59	0.2	0.09	0.22
	N2010+C2012	-16.63	6.20	0.49	0.49	0.05
	N2010+C2010	-16.37	6.45	0.6	0.47	0.06
	N2010+C2011	-16.28	6.54	0.6	0.47	0.06
	C2010	-16.26	6.56	0.11	-0.01	0.37
	N2011+C2009	-14.18	8.64	0.49	0.33	0.12
	N2011+C2012	-13.80	9.02	0.47	0.3	0.14
	N2011+C2011	-13.47	9.35	0.45	0.27	0.16
	N2011+C2010	-13.36	9.47	0.45	0.26	0.16
Loss of A_R between 2010 and 2011	N2011+C2010	-51.88	0.00	0.9	0.96	0.17
	N2010+C2010	-47.38	4.50	0.71	0.90	0.3
	C2009	-	-	0.86	0.79	0.07
	N2011+C2009	-48.18	3.71	0.92	0.76	0.28
	N2010+C2009	-46.75	5.13	0.88	0.66	0.33
Gain of A_R between 2011 and 2013	C2012	-	-	0.9	0.86	0.047
	N2011+C2012	-45.52	-45.52	0.94	0.84	0.22
	N2010+C2012	-43.71	-43.71	0.91	0.75	0.28

Abbreviations used for variables: C, connectivity for given year; N, population size for given year.'+' indicates additive term.

Appendix B: Chapter 3

Table B1 Summary of Mantel tests and partial Mantel tests examining the relationship between pairwise genetic distance (F_{ST}) and total distance, landscape distances, and estimated contemporaneous movement, for subsampled datasets. For years preceding bottleneck events (1995 and 2008), 25 datasets were randomly subsampled (without replacement) with within-patch sample sizes equivalent to those in the smaller, 2005 dataset. Mantel and partial Mantel tests were performed for each subsampled dataset: Mantel tests were used to examine the relationship between F_{ST} and each of total distance and contemporaneous movement, while partial Mantel tests were used to examine the effect of forest distance on F_{ST} controlling for meadow distance (Forest effect) and vice-versa (Meadow effect). The median correlation coefficient (*r*) from all subsampled datasets is reported along with the result of Wilcoxon signed rank (WSR) tests for a significant difference from zero. Data for 1995 are from Keyghobadi et al. (1999, 2005) and data for 2005 are from Caplins et al. (2014).

Models	**1995**	**2008**
	r (*P*)	*r* (*P*)
Total distance	0.44 (<**0.001**)	0.36 (<**0.001**)
Forest effect	0.34 (<**0.001**)	0.22 (<**0.001)**
Meadow effect	0.05 (0.06)	0.025 (0.10)
Movement	-0.22 (<**0.001**)	-0.25 (<**0.001**)

Significant values are in bold typeface.

Table B2 Summary of maximum likelihood population effects (MLPE) models explaining pairwise genetic differentiation (F_{ST}) between populations of *Parnassius smintheus* as a function of year (at four different time points) and either total geographical distance, forest distance, or estimated contemporaneous movement. Interaction effects are included in all models. Year is treated as a categorical factor and contrasts are against 1995. Data for 1995 are from Keyghobadi et al. (1999, 2005) and data for 2005 are from Caplins et al. (2014).

Models	**Predictors**		
		(*P*)	β ±SE
Year+ total distance + year* total distance	Intercept (year1995)	0.12	0.007±0.004
	Year 2005	**0.001**	0.015±0.004
	Year 2008	0.22	0.005±0.004
	Year 2013	**0.00**	0.025±0.005
	Total distance	**0.003**	0.002±0.001
	Year 2005* total distance	**0.03**	0.002±0.001
	Year 2008* total distance	0.16	0.001±0.001
	Year 2013* total distance	**0.03**	0.004±0.002
Year+ forest distance + meadow distance + year* forest distance	Intercept (year1995)	0.10	0.006±0.004
	Year 2005	**0.0001**	0.013±0.003
	Year 2008	0.26	0.004±0.003
	Year 2013	**0.00**	0.034±0.004
	Forest distance	**0.02**	0.005±0.002
	Meadow distance	0.34	0.001±0.001
	Year 2005* forest distance	**0.01**	-0.004±0.002
	Year 2008* forest distance	0.12	-0.003±0.002
	Year 2013* forest distance	0.57	0.003±0.005
Year+ movement + year* movement	Intercept (year1995)	0.10	0.01±0.005
	Year 2005	**0.01**	0.012±0.005
	Year 2008	0.38	0.004±0.004
	Year 2013	**0.001**	0.02±0.006
	Movement	0.30	-0.0004±0.0003
	Year 2005* movement	0.51	0.0003±0.0005
	Year 2008* movement	0.93	0.00004±0.0005
	Year 2013* movement	0.55	-0.002±0.0004

β: MLPE regression coefficient ± SE. Significant values are in bold typeface.

Literature Cited

Caplins SA, Gilbert KJ, Ciotir C, Roland J, Matter SF, Keyghobadi N (2014). Landscape structure and the genetic effects of a population collapse. *Proceedings of the Royal Society B* 281: 20141798.

Keyghobadi N, Roland J, Matter SF, Strobeck C (2005). Among- and within-patch components of genetic diversity respond at different rates to habitat fragmentation: an empirical demonstration. *Proceedings of the Royal Society B* 272: 553–560.

Keyghobadi N, Roland J, Strobeck C (1999). Influence of landscape on the population genetic structure of the alpine butterfly *Parnassius smintheus* (Papilionidae). *Molecular Ecology* 8: 1481–1495.

Appendix C: Chapter 4

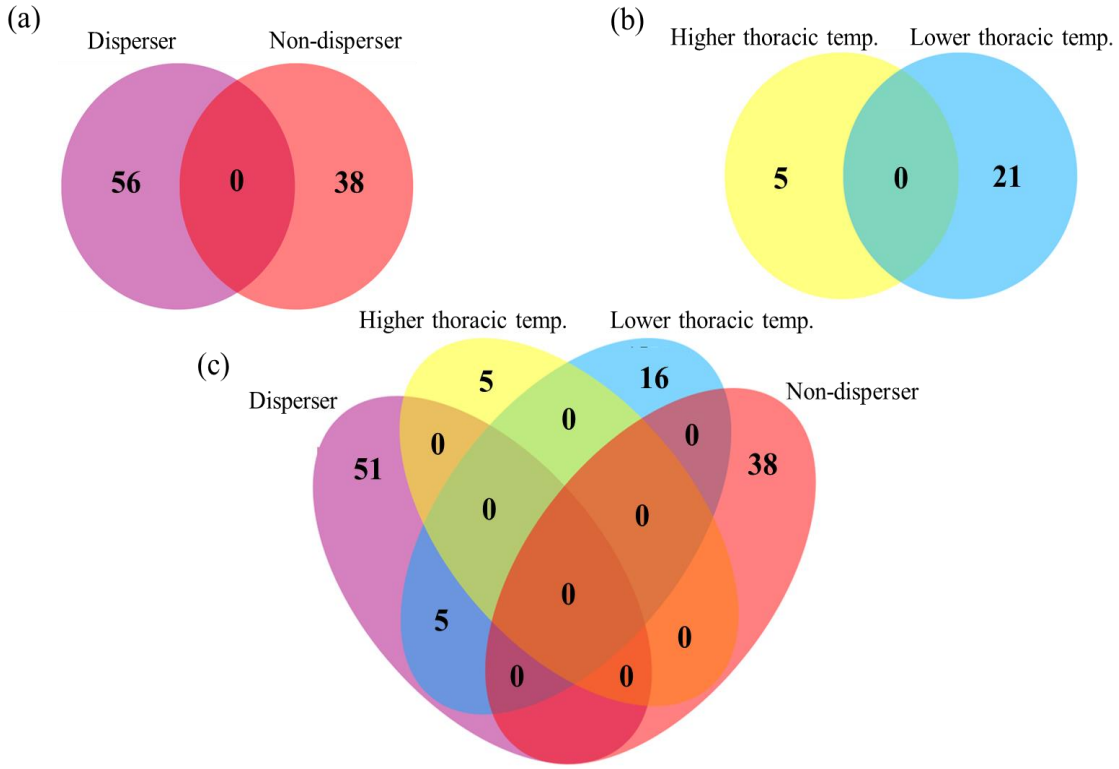


Figure C1 Venn diagrams of upregulated differentially expressed genes (DEGs) for: a) disperser and non-disperser individuals b) individuals with thoracic temperature higher and lower than ambient, and c) all four categories. The numbers in each large circle indicate the total number of DEGs unique to each comparison group, and numbers in overlapping sections indicate the number of DEGs shared among the comparison groups.

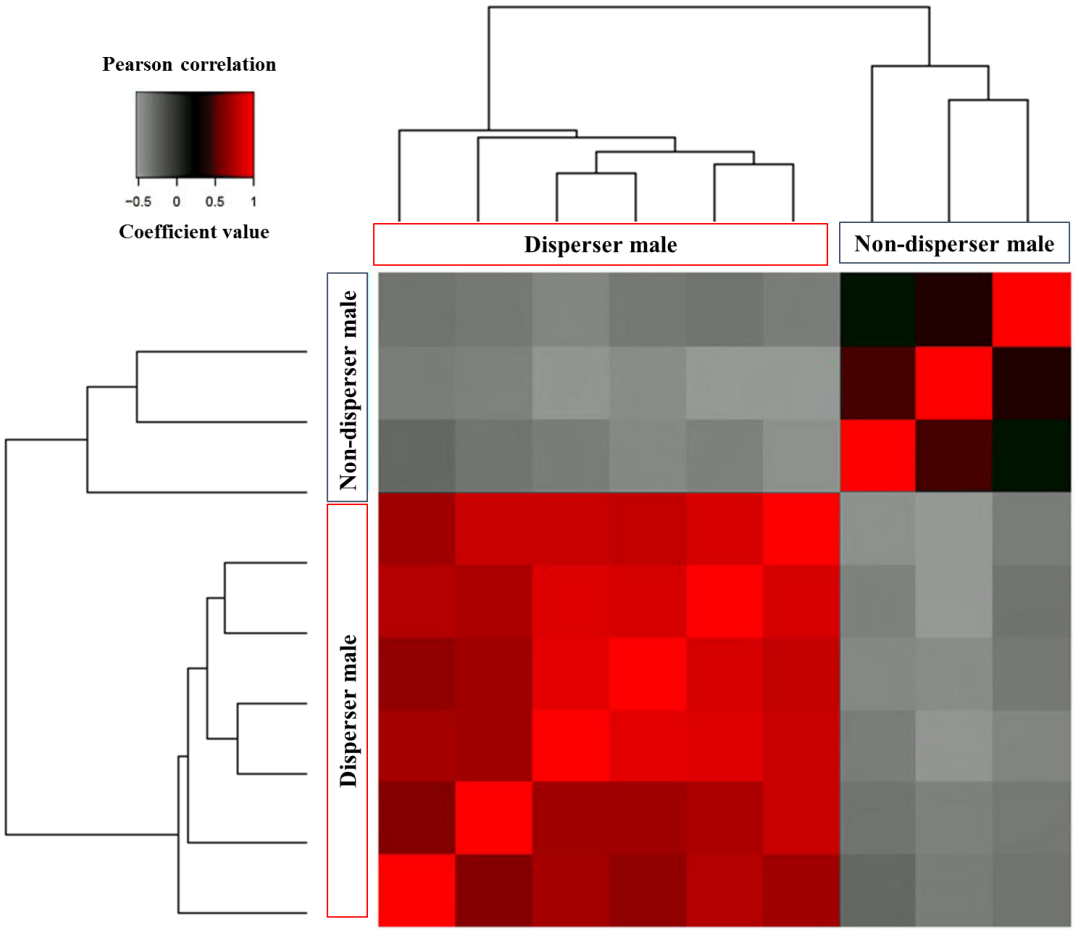


Figure C2 Cluster analysis based on Pearson correlation coefficients showing similarity of gene expression profiles in male *Parnassius smintheus* samples. Samples are classified based on their dispersal history as inferred by mark-recapture data. Red is indicative of similarity, while grey is indicative of dissimilarity in the level of gene expression. Each cell represents the average correlation coefficient of a set of n = 90 dispersal-related genes. The diagram is symmetric across the red-cell diagonal.

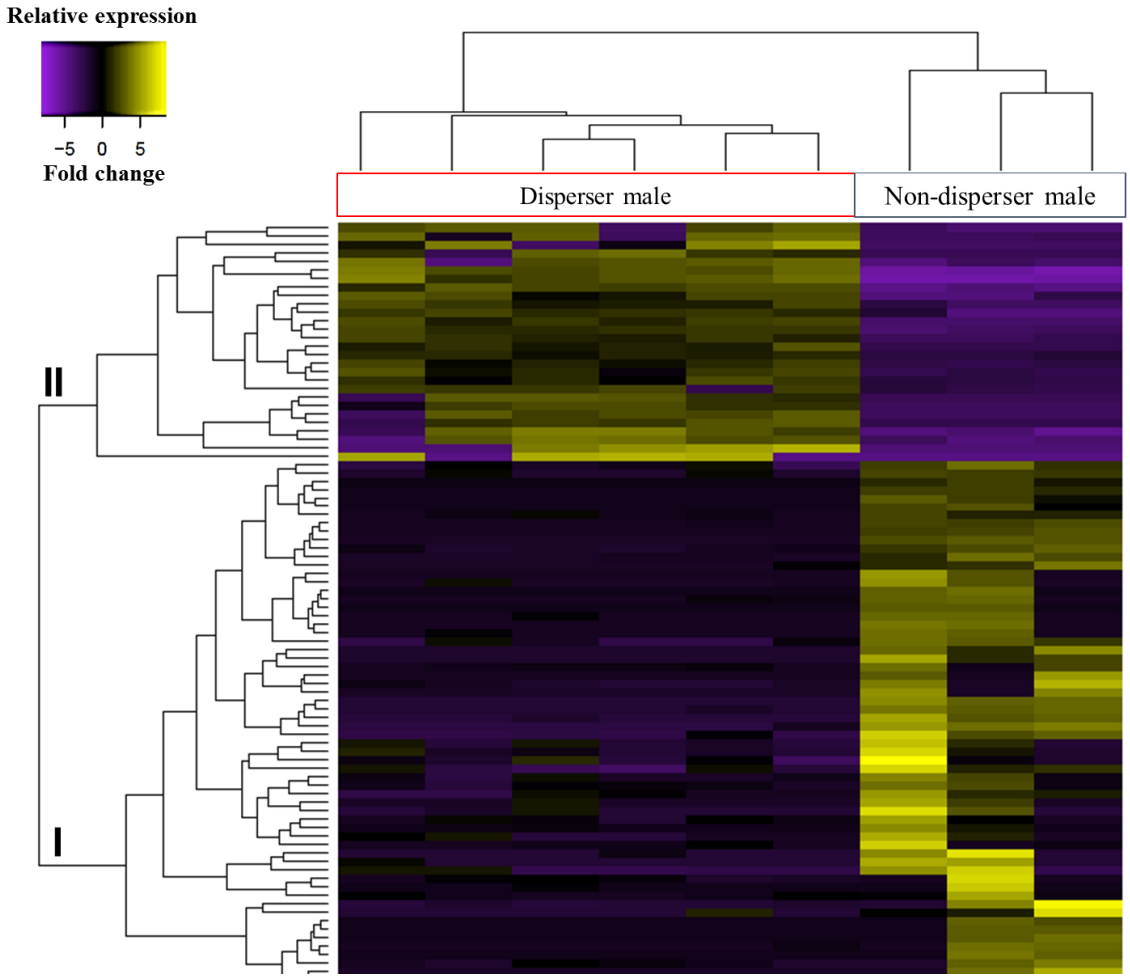


Figure C3 Heat map matrix of 90 genes differentially expressed between male *Parnassius smintheus* that are dispersers and non-dispersers (FDR < 0.05 and minimum four-fold change). The colour code represents the relative expression, where yellow represents upregulation, purple represents down-regulation, and black represents no change in expression. Genes were clustered by means of a hierarchical clustering algorithm presenting two gene sets, I and II (vertical axis).

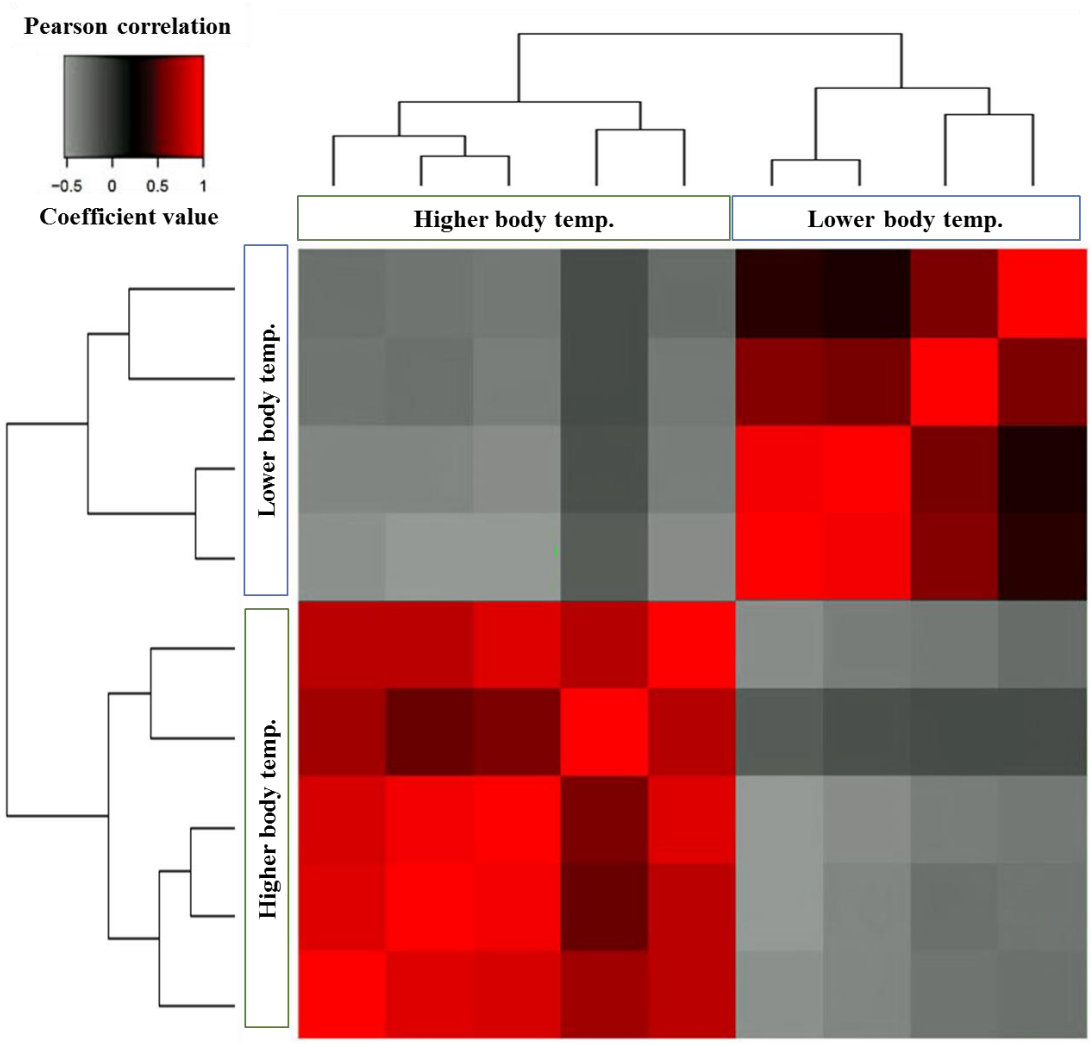


Figure C4 Cluster analysis based on Pearson correlation coefficients showing similarity of gene expression profiles in male *Parnassius smintheus* samples. Samples are classified based on whether thoracic temperature was higher or lower than ambient temperature. Red is indicative of similarity, while grey is indicative of dissimilarity in the level of gene expression. Each cell represents the average correlation coefficient of a set of n = 24 genes. The diagram is symmetric across the red-cell diagonal.

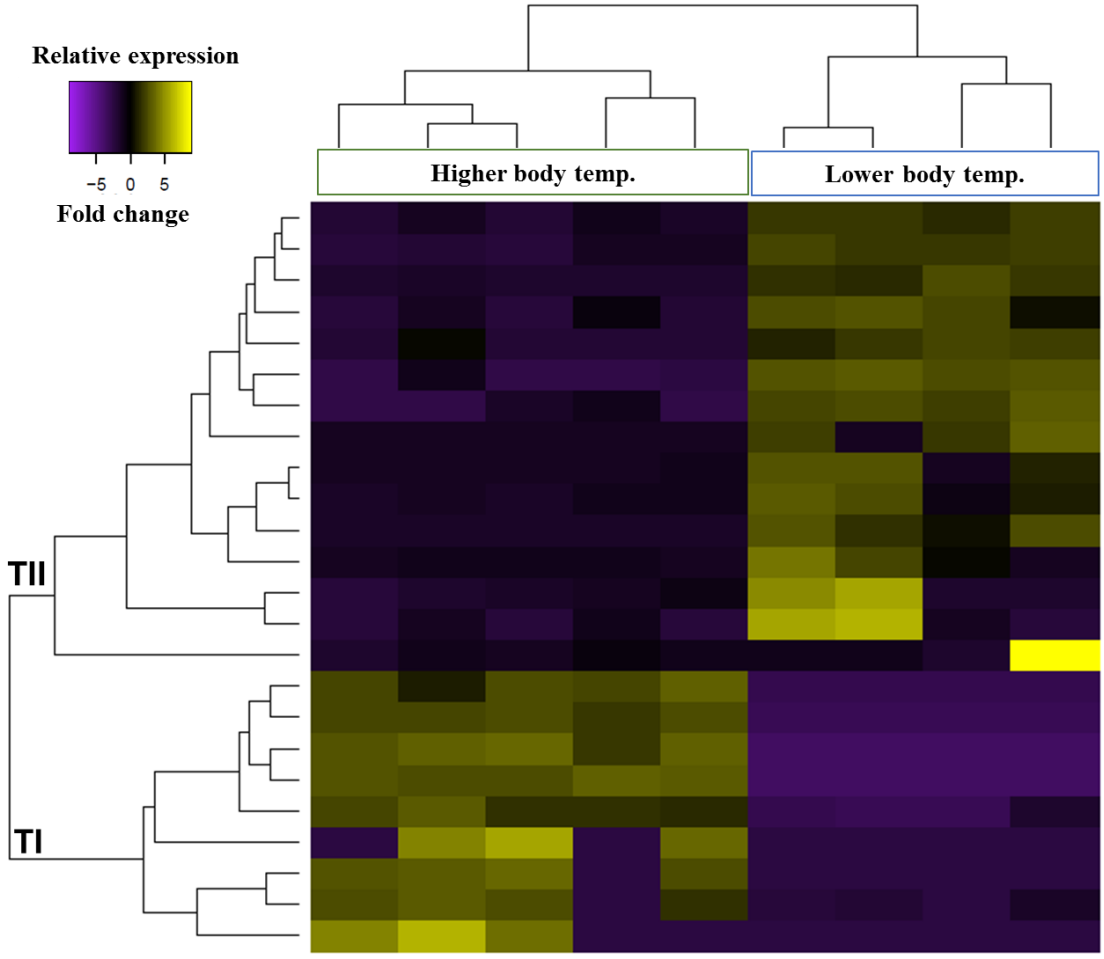


Figure C5 Heat map matrix of 24 genes differentially expressed between male *Parnassius smintheus* individuals with thoracic temperature higher and lower than ambient (FDR < 0.05 and minimum four-fold change). The colour code represents the relative expression, where yellow represents upregulation, purple represents down-regulation, and black represents no change in expression. Genes were clustered by means of a hierarchical clustering algorithm presenting two gene sets, TI and II (vertical axis).

Appendix D: Chapter 5

Table D1 Information on the primers used in this study to amplify five overlapping fragments of coding sequence of *Pgi* in *Parnassius smintheus*.

	Length of fragment (bp)	Name of primer	Sequence from 5' to 3'
Fragment1	531	Ps_Pgi_frag1_F	[ACACTGACGACATGGTTCTACA][1]CGAAGACTTAACCATATAAATTACGAG
		Ps_Pgi_frag1_R	[TACGGTAGCAGAGACTTGGTCT] [2]GGCATCAACACGGTTCTTT
Fragment2	546	Ps_Pgi_frag2_F	[ACACTGACGACATGGTTCTACA] [1]TTCAAACACCCAATGATGGA
		Ps_Pgi_frag2_R	[TACGGTAGCAGAGACTTGGTCT] [2]TTGGCAGAAGTAGCGTTGGT
Fragment3	515	Ps_Pgi_frag3_F	[ACACTGACGACATGGTTCTACA] [1]CGCTGTTCATCATAGCCTCA
		Ps_Pgi_frag3_R	[TACGGTAGCAGAGACTTGGTCT] [2]CGGTGGAGTAGTCTGCCTGT
Fragment4	466	Ps_Pgi_frag4_F	[ACACTGACGACATGGTTCTACA] [1]CAGCAGGGAGACATGGAGAG
		Ps_Pgi_frag4_R	[TACGGTAGCAGAGACTTGGTCT] [2]TTGATGTCCCAGATCACACC
Fragment5	314	Ps_Pgi_frag5_F	[ACACTGACGACATGGTTCTACA] [1]CCATCGCAAAGATTCTACCTC
		Ps_Pgi_frag5_R	[TACGGTAGCAGAGACTTGGTCT][2]TTAACTGCAGACGGCTTACAAA

[1]CS1 universal tag (Fluidigm Corporation, South San Francisco, California, United States)
[2]CS2 universal tag (Fluidigm Corporation, South San Francisco, California, United States)

Table D2 PCR reaction conditions used to amplify the coding sequence of *Pgi* from cDNA, in five overlapping fragments (primers provided in Table S1), in *Parnassius smintheus*. Amplitaq enzyme and and Amplitaq buffer (Applied Biosystems) were used.

PCR reagents	**per reaction (µL) Fragment 1**	**per reaction (µL) Fragment 2,3 and 4**	**per reaction (µL) Fragment 5**
10X Buffer	2.5	2.5	2.5
25mM $MgCl_2$	2	1.625	1.5
10 mM dNTP Mix	0.5	0.5	0.5
Sense primer (10 µM)	0.75	0.5	0. 5
Antisense primer (10 µM)	0.75	0.5	0. 5
Taq DNA polymerase (5 U/µl)	0.2	0.2	0.2
cDNA (from first-strand reaction)	1	1	1
Milli-Q water	17.3	18.175	18.3
Total	25	25	25

Table D3 All pairs of *Pgi* SNPs in *Parnassius smintheus* showing significant linkage disequilibrium (LD), based on Fisher's exact test with Bonferroni correction (based on 49 sequenced individuals; Part I). The location of each SNP within the coding sequence (SNP1 and SNP2; bp) is shown.

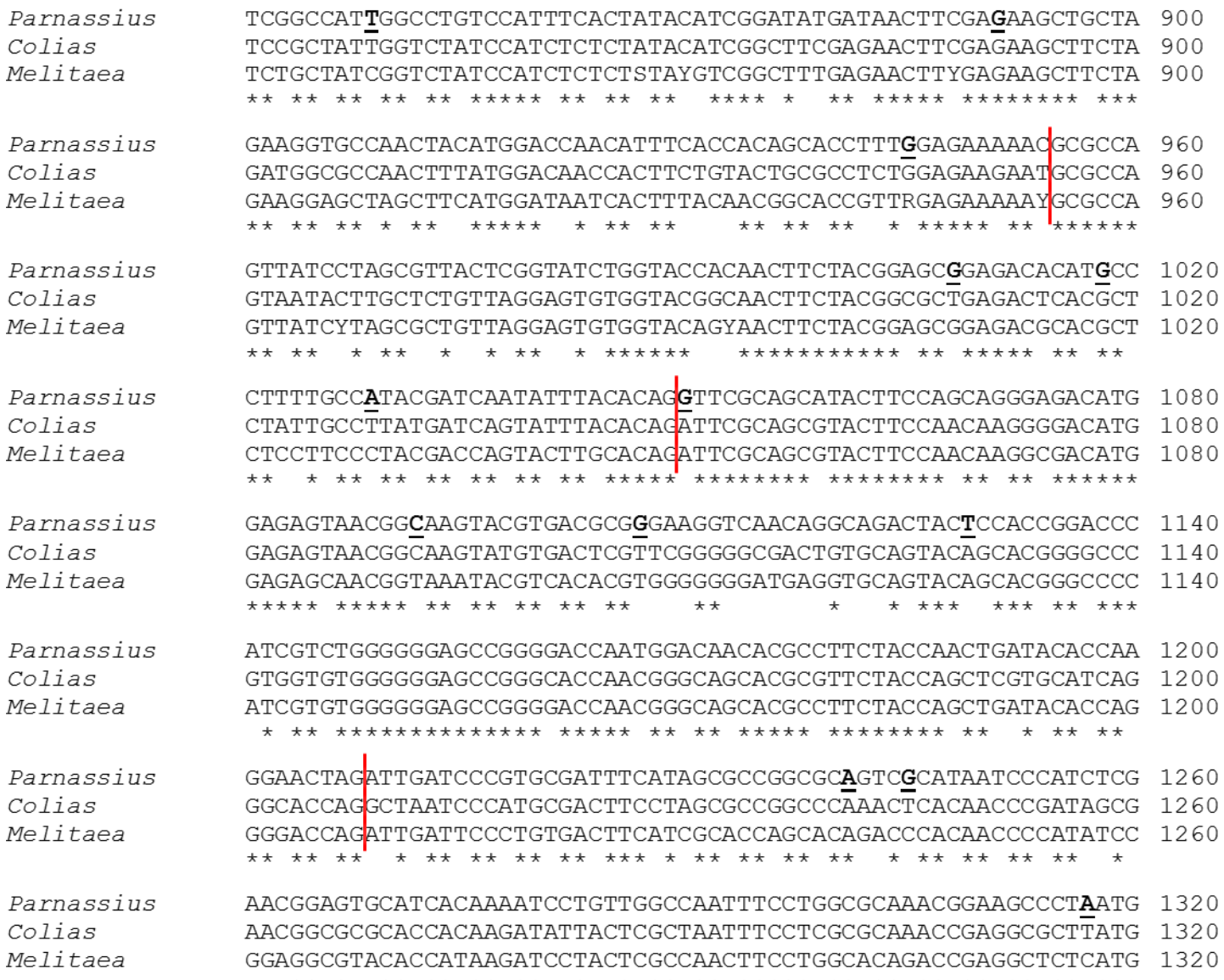
SNP1 (bp)	**SNP2 (bp)**	**D'**	**R**	***P* (Fisher exact test)**
294	1245	1	1	< 0.001
849	945	0.745	0.666	< 0.001
1008	1241	1	0.767	< 0.001
1317	1512	1	0.886	< 0.001
1503	1623	1	0.886	< 0.001

D-prime: proportion of the possible LD that was present between the SNPs; R: correlation coefficient.

Figure D4 Alignment of the consensus coding sequence of *Pgi in Parnassius smintheus* to coding sequences of *Colias eurytheme* and *Melitaea cinxia.* The coding sequences of *Pgi* for *Colias eurytheme and Melitaea cinxia* were extracted from GenBank (accession no. ACS27508.1 and ADB11194.1, respectively). Locations of *Pgi* SNPs in *P. smintheus* are underlined and in bold typeface. Asterisks (*) show similarity among all three sequences. Red vertical lines indicate the location of intron/exon boundaries.

```
Parnassius      ATGGAGCCAAAAATTAACTTAAAACAGGATCCAGCTTACAAAAAACTGCAAGACTTTTTT  60
Colias          ATGGAACCAAAAGTGAATTTAAAACAAGACCCGGCGTATCAGAAGTTACAAGAATATTAT  60
Melitaea        ATGGAGCCTAAAGTGAATTTGAAACAAGACCCGGCTTATAAAAAACTACAAGAATTCTAC  60
                ***** ** *** * ** ** ***** ** ** ** **  * **  * ***** *  *

Parnassius      AACACAAATGCTGAAAAAATCAATATTCTTCAACTTTTTCAACAAGACGCCGACCGCTTC  120
Colias          AACAACAATGCTGATAAGATAAATATCCTACAACTATTCCAACAAGATGCAGACCGCTTT  120
Melitaea        GATGTGAATGGTGGAAAAATTAATATTCTACAACTATTTCAACAAGATCCGGAAAGGTTT  120
                 *    **** **  ** ** ***** ** ***** ** ********  * **  * **

Parnassius      AAAAAGTTCAGTCTCAGACTTCAAACACCCAATGATGGAGACATCTTGCTGGATTACTCA  180
Colias          ATTAAATACAGTCTTCGTATTCCCACACCAAATGATGGTGAGATCCTCCTGGATTACTCC  180
Melitaea        AATAAGTTCAGTTTACGCATCCCCACACCAAATGATGGTGACATCCTTCTGGATTACTCT  180
                *  ** * **** *  *  * *  ***** ******** ** *** * ***********

Parnassius      AAGAACCGTGTTGATGCCACAGCTTTCGACCTTCTCCTCAATCTGGCTAAAAGTCGTGGC  240
Colias          AAAAACCGCATCGATGACACCACCTTTTCATTGCTGCTCAACTTAGCTAAGAGCCGCAAT  240
Melitaea        AAGAACAGGGTAGATGCTGCTGCCTTGTCATTACTGCTGGAGCTGGCCAAGAGTCGTGGA  240
                ** *** *  * ****   *  * **     * ** **  *  * ** ** ** **

Parnassius      GTCGAGCAAGCTAGAGATGCTATGTTTTCAGGTGAAAAGATAAACTTCACGGAGAACCGT  300
Colias          GTTGAGAAAGCTAGAGATGCTATGTTTGCTGGTGAAAAAATCAACTTCACTGAAGACCGA  300
Melitaea        GTGGAACAGGCTCGAGATGCTATGTTTTCTGGTCAAAAGATCAACTTCACTGAGGATCGC  300
                ** **  * *** ************** * *** **** ** ******** **  * **

Parnassius      GCGGTGTTGCACGTAGCTCTGCGCAACCGTCAAAATAAACCGATTATGGTAAATGGGGCG  360
Colias          GCAGTGCTCCACGTAGCCCTCCGCAACCGTCAGAACCGGCCAATAATGGTCAACGGTAAA  360
Melitaea        GCAGTACTACACATAGCACTTCGTAATAGAAAAAATAGACCTATCTTAGTCAACGGCAAG  360
                ** **  * *** **** ** ** **  *  * **    ** **  * ** ** **
```


Parnassius GATGTCACACCTGACGTGAATGCTGTCCTTGCTCATATGAAGGAGTTTTCTCAACAAGTA 420
Colias GATGTCACACCTGATGTGAACGGAGTGCTCGCACATATGAAGGAGTTCTCCACGCAAGTT 420
Melitaea GATGTTACTCCTGATGTCAATGCTGTACTCGCTCATATGAAGGAATTTTCAGAACAAGTT 420
Parnassius ATAGGCGGTCAGTGGAAGGGATACACTGGAAAACCAATCACTGACGTCATCAATATCGGT 480
Colias ATTAGTGGAGCGTGGAAGGGGTACACTGGCAAACCAATCACTGACGTCATCAACATTGGA 480
Melitaea ATCAGTGGCAAGTGGAAGGGGTACACAGGCAAAGCAATAACAGATGTTATCAACATCGGC 480
Parnassius ATTGGGGGCTCTGATCTCGGCCCTTTGATGGTGACGGAGGCTTTGAAACCTTACGCTAAT 540
Colias ATCGGTGGTTCCGATCTCGGACCTCTGATGGTCACAGAGGCTTTGAAACCTTATGCCAAT 540
Melitaea ATCGGGGGCTCAGATCTCGGACCACTCATGGTCACTGAGGCGCTCAAACCCTACGCTAAT 540
Parnassius CACCTTAAGGTGCATTTCGTGTCGAATATTGACGGCACTCATCTCGCAGAAGTACTGAAA 600
Colias CATCTTAAGGTACATTTCGTATCAAACATCGACGGTACCCACCTAGCCGAGGTGCTAAAG 600
Melitaea CATCTTAAGGTACATTTCGTATCCAACATCGACGGAACCCATTTAGCTGAAGTCCTGAAG 600
Parnassius CGTCTGGAATCCAGAAACGGCGCTGTTCATCATAGCCTCAAAGACCTTCACTACGCAGGAA 660
Colias CGATTGAACCCAGAAACAGCATTATTTATAATCGCCTCCAAAACTTTCACCACACAAGAG 660
Melitaea AAATTGAACCCTGAGACAGCTCTCTTCATCATAGCGTCCAAAACMTTTACAACGCAAGAG 660
Parnassius ACCCTTACCAACGCTACTTCTGCCAAGAACTGGTTCTTAGATGTGGCTAAAGATCCATCG 720
Colias ACGATCACCAACGCTACGTCAGCGAAGACATGGTTCTTGGAGGCTGCTAAGGACCCAGCA 720
Melitaea ACCATCACCAACGCGACYTCAGCCAAGAACTGGTTCTTGGATGTCGCYAAAGACCCGTCC 720
Parnassius GCGGTAGCGAAGCATTTTGTAGCACTCTCCACTAACGCTGAGAAGGTGACCGCCTTCGGT 780
Colias GCAGTATCAAAGCACTTTGTAGCGCTATCCACTAATGGCGAAAAGGTGACAGCTTTCGGT 780
Melitaea KCAGTATCMAAGCACTTCGTAGCGCTGTCTACAAACGGGGAAAAGGTTTCCGCATTCGGT 780
Parnassius ATAGACTCTAAGAACATGTTCGGCTTCTGGGATTGGGTTGGTGGAAGATATTCCTTGTGG 840
Colias ATCGACCCCAAGAATATGTTTGGTTTCTGGGACTGGGTTGGAGGAAGGTATTCTCTCTGG 840
Melitaea ATCGACCCCAAGAACATGTTTGGTTTCTGGGACTGGGTGGGAGGCAGATACTCGCTCTGG 840


Parnassius TCGGCCATTGGCCTGTCCATTTCACTATACATCGGATATGATAACTTCGAGAAGCTGCTA 900
Colias TCCGCTATTGGTCTATCCATCTCTCTATACATCGGCTTCGAGAACTTCGAGAAGCTTCTA 900
Melitaea TCTGCTATCGGTCTATCCATCTCTCTSTAYGTCGGCTTTGAGAACTTYGAGAAGCTTCTA 900
Parnassius GAAGGTGCCAACTACATGGACCAACATTTCACCACAGCACCTTTGGAGAAAAACGCGCCA 960
Colias GATGGCGCCAACTTTATGGACAACCACTTCTGTACTGCGCCTCTGGAGAAGAATGCGCCA 960
Melitaea GAAGGAGCTAGCTTCATGGATAATCACTTTACAACGGCACCGTTRGAGAAAAAYGCGCCA 960
Parnassius GTTATCCTAGCGTTACTCGGTATCTGGTACCACAACTTCTACGGAGCGGAGACACATGCC 1020
Colias GTAATACTTGCTCTGTTAGGAGTGTGGTACGGCAACTTCTACGGCGCTGAGACTCACGCT 1020
Melitaea GTTATCYTAGCGCTGTTAGGAGTGTGGTACAGYAACTTCTACGGAGCGGAGACGCACGCT 1020
Parnassius CTTTTGCCATACGATCAATATTTACACAGGTTCGCAGCATACTTCCAGCAGGGAGACATG 1080
Colias CTATTGCCTTATGATCAGTATTTACACAGATTCGCAGCGTACTTCCAACAAGGGGACATG 1080
Melitaea CTCCTTCCCTACGACCAGTACTTGCACAGATTCGCAGCGTACTTCCAACAAGGCGACATG 1080
Parnassius GAGAGTAACGGCAAGTACGTGACGCGGGAAGGTCAACAGGCAGACTACTCCACCGGACCC 1140
Colias GAGAGTAACGGCAAGTATGTGACTCGTTCGGGGGCGACTGTGCAGTACAGCACGGGGCCC 1140
Melitaea GAGAGCAACGGTAAATACGTCACACGTGGGGGGGATGAGGTGCAGTACAGCACGGGCCCC 1140
Parnassius ATCGTCTGGGGGGAGCCGGGGACCAATGGACAACACGCCTTCTACCAACTGATACACCAA 1200
Colias GTGGTGTGGGGGGAGCCGGGCACCAACGGGCAGCACGCGTTCTACCAGCTCGTGCATCAG 1200
Melitaea ATCGTGTGGGGGGAGCCGGGGACCAACGGGCAGCACGCCTTCTACCAGCTGATACACCAG 1200
Parnassius GGAACTAGATTGATCCCGTGCGATTTCATAGCGCCGGCGCAGTCGCATAATCCCATCTCG 1260
Colias GGCACCAGGCTAATCCCATGCGACTTCCTAGCGCCGGCCCAAACTCACAACCCGATAGCG 1260
Melitaea GGGACCAGATTGATTCCCTGTGACTTCATCGCACCAGCACAGACCCACAACCCCATATCC 1260
Parnassius AACGGAGTGCATCACAAAATCCTGTTGGCCAATTTCCTGGCGCAAACGGAAGCCCTAATG 1320
Colias AACGGCGCGCACCACAAGATATTACTCGCTAATTTCCTCGCGCAAACCGAGGCGCTTATG 1320
Melitaea GGAGGCGTACACCATAAGATCCTACTCGCCAACTTCCTGGCACAGACCGAGGCTCTCATG 1320

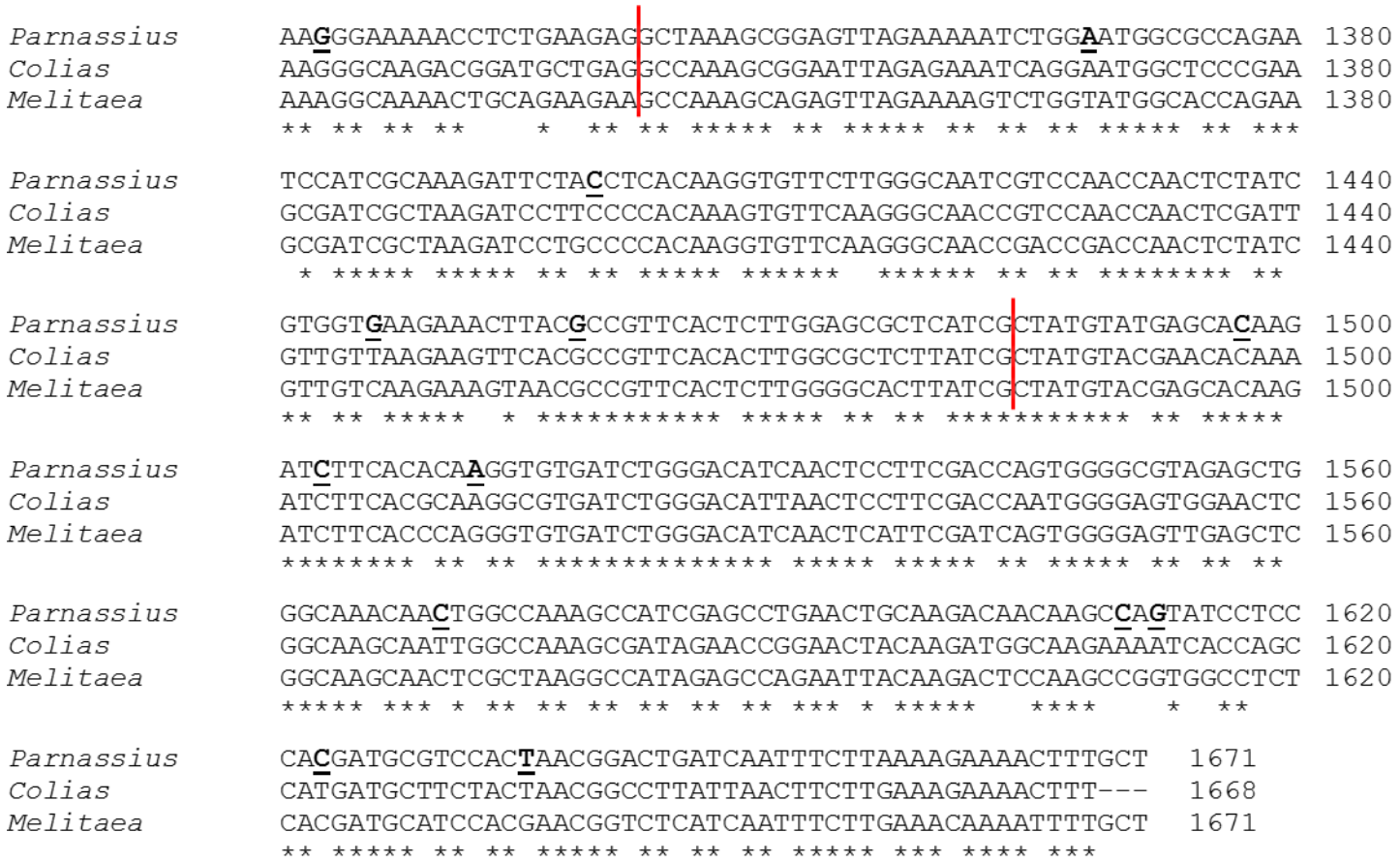

Parnassius AAGGGAAAAACCTCTGAAGAGGCTAAAGCGGAGTTAGAAAAATCTGGAATGGCGCCAGAA 1380
Colias AAGGGCAAGACGGATGCTGAGGCCAAAGCGGAATTAGAGAAATCAGGAATGGCTCCCGAA 1380
Melitaea AAAGGCAAAACTGCAGAAGAAGCCAAAGCAGAGTTAGAAAAGTCTGGTATGGCACCAGAA 1380
Parnassius TCCATCGCAAAGATTCTACCTCACAAGGTGTTCTTGGGCAATCGTCCAACCAACTCTATC 1440
Colias GCGATCGCTAAGATCCTTCCCCACAAAGTGTTCAAGGGCAACCGTCCAACCAACTCGATT 1440
Melitaea GCGATCGCTAAGATCCTGCCCCACAAGGTGTTCAAGGGCAACCGACCGACCAACTCTATC 1440
Parnassius GTGGTGAAGAAACTTACGCCGTTCACTCTTGGAGCGCTCATCGCTATGTATGAGCACAAG 1500
Colias GTTGTTAAGAAGTTCACGCCGTTCACACTTGGCGCTCTTATCGCTATGTACGAACACAAA 1500
Melitaea GTTGTCAAGAAAGTAACGCCGTTCACTCTTGGGGCACTTATCGCTATGTACGAGCACAAG 1500
Parnassius ATCTTCACACAAGGTGTGATCTGGGACATCAACTCCTTCGACCAGTGGGGCGTAGAGCTG 1560
Colias ATCTTCACGCAAGGCGTGATCTGGGACATTAACTCCTTCGACCAATGGGGAGTGGAACTC 1560
Melitaea ATCTTCACCCAGGGTGTGATCTGGGACATCAACTCATTCGATCAGTGGGGAGTTGAGCTC 1560
Parnassius GGCAAACAACTGGCCAAAGCCATCGAGCCTGAACTGCAAGACAACAAGCCAGTATCCTCC 1620
Colias GGCAAGCAATTGGCCAAAGCGATAGAACCGGAACTACAAGATGGCAAGAAAATCACCAGC 1620
Melitaea GGCAAGCAACTCGCTAAGGCCATAGAGCCAGAATTACAAGACTCCAAGCCGGTGGCCTCT 1620
Parnassius CACGATGCGTCCACTAACGGACTGATCAATTTCTTAAAAGAAAACTTTGCT 1671
Colias CATGATGCTTCTACTAACGGCCTTATTAACTTCTTGAAAGAAAACTTT--- 1668
Melitaea CACGATGCATCCACGAACGGTCTCATCAATTTCTTGAAACAAAATTTTGCT 1671

Figure D5 Alignment of the consensus amino acid sequence of PGI *in Parnassius smintheus* to PGI amino acid sequences of *Colias eurytheme* and *Melitaea cinxia.* The sequences of PGI amino acids for *Colias eurytheme and Melitaea cinxia* were extracted from GenBank (accession no. ACS27508.1 and ADB11194.1, respectively). Locations of amino acid variation in *P. smintheus* are underlined and in bold typeface. Asterisks (*) shows similarity among all three sequences.

```
Parnassius      MEPKINLKQDPAYKKLQDFFNTNAEKINILQLFQQDADRFKKFSLRLQTPNDGDILLDYS  60
Colias          MEPKVNLKQDPAYQKLQEYYNNNADKINILQLFQQDADRFIKYSLRIPTPNDGEILLDYS  60
Melitaea        MEPKVNLKQDPAYKKLQEFYDVNGGKINILQLFQQDPERFNKFSLRIPTPNDGDILLDYS  60
                **** ******** ***     *  ***********  ** * ***  ***** ******

Parnassius      KNRVDATAFDLLLNLAKSRGVEQARDAMFSGEKINFTENRAVLHVALRNRQNKPIMVNGA  120
Colias          KNRIDDTTFSLLLNLAKSRNVEKARDAMFAGEKINFTEDRAVLHVALRNRQNRPIMVNGK  120
Melitaea        KNRVDAAALSLLLELAKSRGVEQARDAMFSGQKINFTEDRAVLHIALRNRKNRPILVNGK  120
                *** *     *** ***** ** ****** * ****** ***** ***** * ** ***

Parnassius      DVTPDVNAVLAHMKEFSQQVIGGQWKGYTGKPITDVINIGIGGSDLGPLMVTEALKPYAN  180
Colias          DVTPDVNGVLAHMKEFSTQVISGAWKGYTGKPITDVINIGIGGSDLGPLMVTEALKPYAN  180
Melitaea        DVTPDVNAVLAHMKEFSEQVISGKWKGYTGKAITDVINIGIGGSDLGPLMVTEALKPYAN  180
                ******* ********* *** * ******* ****************************

Parnassius      HLKVHFVSNIDGTHLAEVLKRLNPETALFIIASKTFTTQETLTNATSAKNWFLDVAKDPS  240
Colias          HLKVHFVSNIDGTHLAEVLKRLNPETALFIIASKTFTTQETITNATSAKTWFLEAAKDPA  240
Melitaea        HLKVHFVSNIDGTHLAEVLKKLNPETALFIIASKXFTTQETITNAXSAKNWFLDVXKDPS  240
                ******************** ************* ****** *** *** ***   ***

Parnassius      AVAKHFVALSTNAEKVTAFGIDSKNMFGFWDWVGGRYSLWSAIGLSISLYIGYDNFEKLL  300
Colias          AVSKHFVALSTNGEKVTAFGIDPKNMFGFWDWVGGRYSLWSAIGLSISLYIGFENFEKLL  300
Melitaea        XVXKHFVALSTNGEKVSAFGIDPKNMFGFWDWVGGRYSLWSAIGLSISXXVGFENXEKLL  300
                 * ********* *** ***** *************************   *  * ****

Parnassius      EGANYMDQHFTTAPLEKNAPVILALLGIWYHNFYGAETHALLPYDQYLHRFAAYFQQGDM  360
Colias          DGANFMDNHFCTAPLEKNAPVILALLGVWYGNFYGAETHALLPYDQYLHRFAAYFQQGDM  360
Melitaea        EGASFMDNHFTTAPXEKXAPVIXALLGVWYXNFYGAETHALLPYDQYLHRFAAYFQQGDM  360
                 **  ** ** *** ** **** **** ** ****************************
```

```
Parnassius      ESNGKYVTREGQQADYSTGPIVWGEPGTNGQHAFYQLIHQGTRLIPCDFIAPAQSHNPIS 420
Colias          ESNGKYVTRSGATVQYSTGPVVWGEPGTNGQHAFYQLVHQGTRLIPCDFLAPAQTHNPIA 420
Melitaea        ESNGKYVTRGGDEVQYSTGPIVWGEPGTNGQHAFYQLIHQGTRLIPCDFIAPAQTHNPIS 420
                ********* *    ***** *************** *********** **** ****

Parnassius      NGVHHKILLANFLAQTEALMKGKTSEEAKAELEKSGMAPESIAKILPHKVFLGNRPTNSI 480
Colias          NGAHHKILLANFLAQTEALMKGKTDAEAKAELEKSGMAPEAIAKILPHKVFKGNRPTNSI 480
Melitaea        GGVHHKILLANFLAQTEALMKGKTAEEAKAELEKSGMAPEAIAKILPHKVFKGNRPTNSI 480
                 * ********************  ************** ********** ********

Parnassius      VVKKLTPFTLGALIAMYEHKIFTQGVIWDINSFDQWGVELGKQLAKAIEPELQDNKPVSS 540
Colias          VVKKFTPFTLGALIAMYEHKIFTQGVIWDINSFDQWGVELGKQLAKAIEPELQDGKKITS 540
Melitaea        VVKKVTPFTLGALIAMYEHKIFTQGVIWDINSFDQWGVELGKQLAKAIEPELQDSKPVAS 540
                **** ************************************************ *   *

Parnassius      HDASTNGLINFLKENFA  557
Colias          HDASTNGLINFLKENF-  556
Melitaea        HDASTNGLINFLKQNFA  557
                ************* **
```

VITA

Name:	Maryam Jangjoo
Post-secondary Education and Degrees:	Bahonar University Kerman, Iran 2002-2006 B.Sc. Ferdowsi University of Mashhad Mashhad, Iran 2007-2010 M.Sc. University of Western Ontario London, Ontario, Canada 2013-2015, Intent to transfer to PhD University of Western Ontario 2015-2018 Ph.D.
Awards	Biology Graduate Travel Award-2017
Related Work Experience	Graduate Teaching Assistant University of Western Ontario 2013-2018 Research Assistant The University of Western Ontario, London, Canada 2013-2018

Publications:

Jangjoo M, Matter SF, Roland J, Keyghobadi N (2016). Connectivity rescues genetic diversity after a demographic bottleneck in a butterfly population network. *Proceedings of the National Academy of Sciences of the United States of America* 113: 10914–10919.

Jangjoo M, Matter SF, Roland J, Keyghobadi N. Demographic fluctuations lead to rapid and cyclic shifts in genetic structure among populations of an alpine butterfly, (*Parnassius smintheus*), (submitted to *Heredity*).

Jangjoo M, Darvish J, Hashemi N (2011). Application of outline analysis on fossil and

modern specimens of *Apodemus*. *Iranian Journal of Animal Biosystematics*. 7: 143-154.

Jangjoo M (2010). Geometric morphometric analysis of the second upper molar of the genus *Apodemus* (Muridae, Rodentia) in Northern Iran. *Iranian Journal of Animal Biosystematics* 6: 33-44.

Tarahomi SM, Malek M, Karami M, Darvish J, Jangjoo M (2010). Geometric morphometric comparison of mandible and skull of five species of genus *Allactaga* (Rodentia: Dipodidae) from Iran. *Iranian Journal of Animal Biosystematics* 6: 61-69.

Conference Presentations:

Jangjoo M, Matter SF, Roland J, Keyghobadi N (2018) Demographic fluctuations lead to rapid and cyclic shifts in genetic structure among populations of an alpine butterfly, (*Parnassius smintheus*). *The Biology of Butterflies Conference.* Bengaluru, India (Poster presentation).

Jangjoo M, Matter SF,Benoit JB, Keyghobadi N (2017) Differential gene expression patterns associated with dispersal ability in the alpine butterfly, *Parnassius smintheus. Entomological Society of Ontario*. University of Guelph, Guelph, Canada (Oral presentation).

Jangjoo M, Matter SF,Benoit JB, Keyghobadi N (2017). Gene expression associated with dispersal ability under different temperature conditions in the alpine butterfly, *Parnassius smintheus. Canadian Society for Ecology & Evolution*. Victoria, Canada (Oral presentation).

Jangjoo M, Matter SF, Roland J, Keyghobadi N (2016). Connectivity rescues genetic diversity after a population collapse. *Canadian Society of Zoologists.* Western University, London, Canada (Poster presentation).

Jangjoo M, Matter SF, Roland J, Keyghobadi N (2015). Connectivity rescues genetic diversity after a population collapse: empirical evidence from a butterfly population network. *Biology Graduate Research Forum*. Western University, London, Canada (Poster presentation).

Jangjoo M, Matter SF, Roland J, Keyghobadi N (2015) Connectivity rescues genetic diversity after a population collapse: empirical evidence from a butterfly population network. *The Fallona Family Interdisciplinary Showcase*. Western University, London, Canada (Poster presentation).

Jangjoo M, Darvish J, Hashemi N, Mashkour M (2010) Geometric morphometric analysis of steppe field mouse *Apodemus witherbyi* (Rodentia: Muridae) from Northern part of Iran, *International Conference of Archaeozoology*. Paris, France (Oral presentation).