Western Graduate&PostdoctoralStudies

Western University
Scholarship@Western

Electronic Thesis and Dissertation Repository

9-13-2011 12:00 AM

A model system in landscape genetics – the insect inhabitants of pitcher plants

Gordana Gojko Rasic, *University of Western Ontario*

Supervisor: Dr Nusha Keyghobadi, *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Biology
© Gordana Gojko Rasic 2011

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Ecology and Evolutionary Biology Commons

Recommended Citation
Rasic, Gordana Gojko, "A model system in landscape genetics – the insect inhabitants of pitcher plants" (2011). *Electronic Thesis and Dissertation Repository*. 271.
https://ir.lib.uwo.ca/etd/271

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlswadmin@uwo.ca.

A MODEL SYSTEM IN LANDSCAPE GENETICS – THE INSECT INHABITANTS OF PITCHER PLANTS

(Spine title: A model system in landscape genetics - pitcher plant insects)

(Thesis format: Integrated Article)

by

Gordana Rasic

Graduate Program in Biology

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

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

© Gordana Rasic 2011

THE UNIVERSITY OF WESTERN ONTARIO
School of Graduate and Postdoctoral Studies

CERTIFICATE OF EXAMINATION

Supervisor

____
Dr. Nusha Keyghobadi

Co-supervisor

____
Dr. Marc-André Lachance

Supervisory Committee

____
Dr. Irena Creed

____
Dr. Beth MacDougall-Shackleton

Examiners

____
Dr. Joanna Freeland

____
Dr. Lindi Wahl

____
Dr. Beth MacDougall-Shackleton

____
Dr. Hugh Henry

The thesis by

Gordana Rasic

entitled:

A model system in landscape genetics – the insect inhabitants of pitcher plants

is accepted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

____
Date

____
Chair of the Thesis Examination Board

Abstract

Landscape genetics is a rapidly growing field that investigates how landscape and environmental features interact with microevolutionary processes to give rise to spatial genetic variation in populations. The ability to predict landscape effects on genetic patterns has been limited by the lack of studies conducted on more than one species, over multiple spatial scales and in replicated landscapes. The insect inhabitants of the purple pitcher plant (*Sarracenia purpurea*) constitute a system that allows for such studies. The insects are the pitcher plant flesh fly (*Fletcherimyia fletcheri*), midge (*Metriocnemus knabi*) and mosquito (*Wyeomyia smithii*). In this thesis, I worked towards developing this as a potential model system in landscape genetics. I successfully developed microsatellite markers for the flesh fly and the midge. In the pitcher plant mosquito, microsatellite isolation was very problematic due to presence of microsatellite families associated with transposable elements and further aggravated by null alleles. I assessed levels of genetic differentiation across spatial scales and inferred the extent of gene flow in the flesh fly and the midge. I found that the pitcher plant flesh fly exhibits a mixture of metapopulation and patchy population attributes, with significant structuring and limited dispersal/gene flow at larger spatial scales (metapopulation characteristics), but the absence of local extinctions/recolonizations (patchy characteristics). I found that the pitcher plant midge exhibits a complex pattern of genetic differentiation across spatial scales, significantly associated with landscape variables related to habitat size, abundance and spatial arrangement. These broad-scale landscape features seem to influence the fine-scale process of female oviposition. I also found that, in this small insect, both active flight and wind-assisted dispersal mediate gene flow among bogs within a landscape. I demonstrated that the insect inhabitants of pitcher plants can be used to address general questions in landscape genetics, such as the importance of considering spatial scale in describing genetic patterns and inferring underlying processes, and the importance of replication in testing landscape genetic hypotheses. Overall, this research has laid a foundation for further studies in this system and provided insights that are of interest to the broader community of landscape genetics researchers.

Keywords: landscape genetics, *Fletcherimyia fletcheri*, *Metriocnemus knabi*, *Wyeomyia smithii*, *Sarracenia purpurea*, microsatellites, genetic structure, spatial scale, habitat structure, wind-assisted dispersal

Co-Authorship

This thesis was completed under the supervision and financial support of Dr Nusha Keyghobadi, and co-supervison of Dr Marc-André Lachance. Authorship on manuscripts arising from this thesis will be assigned as follows:

Chapter 2

Gordana Rasic: Collected samples, performed the experimental procedures, performed genetic and statistical analysis, and drafted the manuscripts.

Sheri A. Maxwell: Assisted in the experimental procedure of microsatellite isolation in *Metriocnemus knabi* Coquillett 1904, offered suggestions on the manuscript.

Nusha Keyghobadi: Offered suggestions on the experimental procedures, provided editorial comments.

Chapter 3

Gordana Rasic: Designed study, collected samples, performed genetic and statistical analysis, and drafted the manuscript.

Nusha Keyghobadi: Offered suggestions on the study design, provided editorial comments.

Chapter 4

Gordana Rasic: Designed study, collected samples, performed genetic and statistical analysis, and drafted the manuscript.

Nusha Keyghobadi: Offered suggestions on the study design, provided editorial comments.

Chapter 5

Gordana Rasic: Designed study, collected samples, performed genetic and statistical analysis, and drafted the manuscript.

Nusha Keyghobadi: Offered suggestions on the study design, provided editorial comments.

Acknowledgements

Nanos gigantium humeris insidentes.

I would like to express my gratitude to my supervisor, Nusha Keyghobadi, for all of her support and guidance throughout my PhD studies. I couldn't have hoped for a better mentor than Nusha. Her passion for science has been truly inspirational. Her knowledge, advice and kindness enabled me to pursue scientific ideas daringly and joyfully. Also, I would like to thank my co-supervisor, André Lachance, for his advice, encouragement and humor that kept me going through challenging times and made good times even better. Many thanks to my advisory committee members, Irena Creed and Beth MacDougall-Shackleton, for helpful and timely comments on my work.

Special thanks to Sheri Maxwell, for her friendship, assistance and companionship through many long hours of lab work. I am grateful to Curtis Irvine, Noah Shapiro, Jenna Donald, Kevin Schreiber and Katelyn Weaver for their invaluable assistance in the field. Many thanks to the Keyghobadi lab, especially Lindsay Crawford, Sheri Maxwell, Daria Koscinski and Liam Frape for helpful discussions and comments on my research. Thank you to all members of the Harkness Laboratory of Fisheries Research, for their generosity and support during field work.

I am deeply grateful to my husband, Igor Filipovic, for his love, patience, constant encouragement and readiness to use his exquisite computer wizardry. This has been our journey. I thank my parents, Biljana and Gojko Rasic, for their endless love and support.

Table of Contents

Abstract iii

Co-Authorship v

Acknowledgements vi

Table of Contents vii

List of Tables ix

List of Figures xi

List of Appendices xi

List of Abbreviations xii

List of Software Packages xii

Chapter 1. General Introduction 1

1.1 The purple pitcher plant and its inhabitants 4

1.2 Outline 7

1.3 References 8

Chapter 2. Isolation of microsatellite loci for the pitcher plant insects 12

2.1 Introduction 12

2.2 Materials and Methods 14

2.3 Results 23

2.4 Discussion 30

2.5 Summary 35

2.6 References 35

Chapter 3. The pitcher plant flesh fly exhibits a mixture of metapopulation and patchy characteristics 42

3.1 Introduction 42

3.2 Materials and Methods 46

3.3 Results 53

3.4 Discussion 65

3.5 Summary 69

3.6 References 70

Chapter 4. From broad scale patterns to fine scale processes: habitat structure influences genetic differentiation in the pitcher plant midge across multiple spatial scales 76

4.1 Introduction 76

4.2 Materials and Methods 79

4.3 Results 85

4.4 Discussion 95

4.5 Summary 101

4.6 References 101

Chapter 5. Wind assists gene flow among bogs in the pitcher plant midge 108

5.1 Introduction 108

5.2 Materials and Methods 110

5.3 Results 118

5.4 Discussion 122

5.5 Summary 125

5.6 References 126

Chapter 6. Summary and Conclusions 134

6.1 References 140

Appendices 142

Curriculum Vitae 146

List of Tables

Table 2.1 Microsatellite loci developed for *Fletcherimyia fletcheri* 24

Table 2.2 Microsatellite loci developed for *Metriocnemus knabi* 26

Table 2.3 Microsatellite loci developed for *Wyeomyia smithii* 28

Table 2.4 Multiplex PCR combinations for microsatellite loci of *Fletcherimyia fletcheri* 29

Table 2.5 Multiplex PCR combinations for microsatellite loci of *Metriocnemus knabi* 29

Table 2.6 BLASTn analysis for 24 *Wyeomyia smithii* sequences 33

Table 3.1 Predictions about the distribution of genetic variation and gene flow among subpopulations resulting from the three models of population structure 45

Table 3.2 Genetic diversity measures averaged over 10 microsatellite loci for *Fletcherimyia fletcheri,* from 11 bogs in Algonquin Provincial Park, Canada 55

Table 3.3 Analysis of molecular variance (AMOVA) for *Fletcherimyia fletcheri* 57

Table 3.4 Differentiation between pairs of bogs based on pairwise F_{ST} values in *Fletcherimyia fletcheri* 58

Table 3.5 STRUCTURE analysis for *Fletcherimyia fletcheri* 59

Table 3.6 Estimated effective population sizes (N_e) and 95%CI for *Fletcherimyia fletcheri* populations 64

Table 4.1 Hierarchical analysis of molecular variance in *Metriocnemus knabi* 87

Table 4.2 Matrix of hierarchical *F*-statistics computed in HIERFSTAT for *Metriocnemus knabi* 88

Table 4.3 Distance based redundancy analysis of genetic distances among *Metriocnemus knabi* samples performed at each scale (cluster, plant, leaf) 93

Table 4.4 Forward selection procedure in distance based redundancy analysis of genetic distances among *Metriocnemus knabi* samples performed at each scale (cluster, plant, leaf) 94

Table 5.1	Wind frequencies and derived wind distances for *Metriocnemus knabi*	116
Table 5.2	Partial Mantel test results for correlation of matrices of genetic distances, geographic distances, and wind distances among all pairs of clusters in SYS1 for *Metriocnemus knabi*	120
Table 5.3	Partial Mantel test results for correlation of matrices of genetic distances, geographic distances, and wind distances among all pairs of clusters in SYS2 for *Metriocnemus knabi*	120
Table 5.4	Multiple regression on distance matrices (MRM) analysis in SYS1 for *Metriocnemus knabi*	121
Table 5.5	Multiple regression on distance matrices (MRM) analysis in SYS2 for *Metriocnemus knabi*	121

List of Figures

Figure 1.1	Larvae of the pitcher plant insects, plants and clusters in a bog	5
Figure 2.1	Microsatellite isolation following the enrichment-based protocol	21
Figure 3.1	Study area and sampling locations for *Fletcherimyia fletcheri*	48
Figure 3.2	Representative display of assignment probabilities from STRUCTURE analysis for individuals of *Fletcherimyia fletcheri*	60
Figure 3.3	Correlation between pairwise genetic distances (F_{ST} / (1−F_{ST})) and log geographic distances for *Fletcherimyia fletcheri* samples	61
Figure 3.4	Spatial genetic autocorrelograms showing mean correlation coefficients between pairs of *Fletcherimyia fletcheri* individuals	63
Figure 4.1	Sampling locations and design for *Metriocnemus knabi*	81
Figure 4.2	Percentage of *Metriocnemus knabi* full-sib pairs sampled at different spatial scales	89
Figure 4.3	Spatial genetic autocorrelograms showing average correlation coefficients between pairs of *Metriocnemus knabi* individuals	90
Figure 4.4	Plots of Eigen values for the first three components of the Principal Coordinate Analysis performed on genetic distance matrix from bog and cluster samples of *Metriocnemus knabi*	91
Figure 5.1	Study area and sampling locations for *Metriocnemus knabi*	112
Figure 5.2	Wind rose used to calculate wind distances for *Metriocnemus knabi* clusters	115

List of Appendices

Appendix 1	Locations of clusters and values of landscape variables (bog size, bog and cluster plant density, cluster connectivity) for *Metriocnemus knabi* samples	142
Appendix 2	Permissions to reproduce journal articles published in *Molecular Ecology Resources*	143

List of Abbreviations

AFLP	amplified fragment length polymorphism
AMOVA	analysis of molecular variance
BLASTn	basic local alignment search tool (nucleotide)
dbRDA	distance-based redundancy analysis
GIS	geographic information system
GPS	global positioning system
IBD	isolation by distance
mtDNA	mitochondrial deoxyribonucleic acid
NCBI	National center for biotechnology information
PCoA	principal coordinate analysis
PCR	polymerase chain reaction
RAPD	randomly amplified polymorphic deoxyribonucleic acid
ReFS	repetitive flanking sequences
SNP	single nucleotide polymorphism
SNX	linkers that contain cut sites for restriction enzymes *Stu* I, *Nhe* I, *Xmn* I
SYS1	group of bogs (i.e., system) located in Algonquin Provincial Park
SYS2	group of bogs (i.e., system) located in Algonquin Provincial Park

List of Software Packages

ArcGIS 9.3	Esri (Redlands, CA)
BOTTLENECK	http://www.ensam.inra.fr/URLB/bottleneck/bottleneck.html
DISTLM *forward*	http://www.stat.auckland.ac.nz/~mja/Programs.htm
ECODIST	http://cran.r-project.org/web/packages/ecodist/index.html
FSTAT	http://www2.unil.ch/popgen/softwares/fstat.htm
GenAlEx	http://www.anu.edu.au/BoZo/GenAlEx/
Genemapper	Applied Biosystems (Forest City, CA)
GENEPOP	http://genepop.curtin.edu.au/
HIERFSTAT	http://www2.unil.ch/popgen/softwares/hierfstat.htm
IBDWS	http://ibdws.sdsu.edu/~ibdws/
Micro-Checker	http://www.microchecker.hull.ac.uk/
MicroFamily	http://gsite.univ-provence.fr/gsite/Local/egee/dir/meglecz/MicroFamily.html
Microsatellite Analyzer (MSA)	http://i122server.vu-wien.ac.at/MSA/MSA_download.html
MLNE	http://www.zsl.org/science/research/software/mlne,1151,AR.html
ML-RELATE	http://www.montana.edu/kalinowski/Software/MLRelate.htm
ONeSAMP	http://genomics.jun.alaska.edu/asp/Default.aspx
PRIMER 3	http://frodo.wi.mit.edu/primer3/
R Project	http://www.r-project.org/
STRUCTURE	http://pritch.bsd.uchicago.edu/structure.html
STRUCTURE HARVESTER	http://taylor0.biology.ucla.edu/structureHarvester/

Chapter 1. General Introduction

In natural populations, genetic diversity is not distributed uniformly over space but is spatially structured. Geographic space mediates many biological processes that shape patterns of genetic diversity, such as dispersal, gene flow, and demography (Guillot 2009). The theoretical analysis of the spatial organization of genetic diversity was pioneered by Wright (1943) and Malécot (1948). 'Isolation by distance' (IBD), a concept introduced by Wright (1943), describes a pattern of increased accumulated genetic differences with increased geographic distances among local populations, under geographically restricted dispersal. Malécot (1948) analysed how kinship between individuals is related to the distance separating them, and many authors have since used this approach to describe how genetic structure develops in different models of isolation by distance (e.g., Kimura & Weiss 1964, Maruyama 1971, 1972, 1977, Nagylaki 1978, Slatkin 1987).

The IBD pattern is essentially a consequence of the inherent limitations to species-specific dispersal and gene flow that are independent of any specific landscape features (Balkenhol et al. 2009). In most cases, however, both geographic space and landscape features simultaneously influence movement and gene flow (Trizio et al. 2005). Investigating the interaction between landscape features and gene flow, as well as other microevolutionary processes (genetic drift, selection), is the main focus of the rapidly growing field of landscape genetics. Landscape genetics integrates data and analysis methods from landscape ecology, spatial statistics, geography and population genetics to more fully understand how the spatial distribution of genetic variation arises in populations (Manel et al. 2003, Holderegger & Wagner 2006, Storfer et al. 2007, Storfer et al. 2010).

Two key steps in landscape genetic studies are (i) describing spatial patterns of genetic variability, and (ii) testing hypotheses about the effects of landscape and environmental features on these spatial patterns of genetic structure (Manel et al. 2003). Advances and innovations in several key areas have facilitated these tasks and underlie the rapid, recent growth of landscape genetics. Increased availability of hypervariable genetic markers

(microsatellites, amplified fragment length polymorphisms, single nucleotide polymorphisms) has greatly improved the spatial and temporal resolution obtainable in describing genetic structure. New applications of statistical approaches (maximum likelihood, Bayesian) are used in estimating different genetic and demographic parameters. Geographic information systems (GIS) and increased accessibility of remote sensing data have facilitated collection and quantification of numerous landscape and environmental variables. Finally, spatial statistical approaches adopted from landscape ecology and other fields allow for more sophisticated tests of correlation between genetic and landscape data.

The key distinction between landscape genetic and traditional population genetic studies is the incorporation of explicit tests of how landscape heterogeneity influences gene flow and genetic variation within and among populations (Holderegger & Wagner 2008, Storfer et al. 2010). Recent studies have demonstrated that multivariate models that include landscape variables perform significantly better than simple IBD tests in explaining variance in genetic distance among populations (Spear et al. 2005, Vignieri 2005, Foll & Gaggliotti 2006, McRae 2006, Spear & Storfer 2008, Murphy et al. 2010). Landscape genetics is also distinct from phylogeography in terms of the temporal and spatial scales typically considered (Manel et al. 2003, Wang 2010). Although both disciplines aim to understand the distribution of genetic variation across natural environments (Avise et al. 1987, Manel et al. 2003), phylogeography investigates the role of historical processes determining genetic patterns over large spatial scales, while landscape genetics focuses on more contemporary and fine spatial scale processes (Wang 2010).

Insights from landscape genetics are making significant contributions to our understanding of how natural populations function in both ‘undisturbed’ and anthropogenically altered landscapes. The integrative landscape genetic approach has addressed a variety of questions, including: identifying barriers to gene flow, identifying dispersal corridors, inferring the effects of landscape and ecological variables and landscape change on genetic diversity, identifying source-sink dynamics, predicting spread of disease and invasive species, and comparing observed genetic patterns to

historic and contemporary landscapes (reviewed in Storfer et al. 2010). For example, several landscape genetic studies have provided valuable guidelines for constructing corridors that facilitate gene flow among habitat fragments or nature reserves (Zannese et al. 2006, Epps et al. 2007, Neel 2008).

It has recently been pointed out that most studies in the field of landscape genetics (90%, Storfer et al. 2010) focus on a single species, are conducted at a single spatial scale, and lack replication at the landscape level. This limits the ability to predict the landscape effects on genetic structure across species and at more than one observational scale. Issues related to scale in particular are believed to be a critical but largely unexplored subject in landscape genetics (Balkenhol et al. 2009, Anderson et al. 2010). Spatial genetic patterns result from a potentially complex combination of evolutionary, behavioral, ecological and stochastic processes operating at different spatial and temporal scales (Balkenhol et al. 2009, Anderson et al. 2010). Furthermore, ecological processes and environmental variables can influence genetic variation differentially over varying spatial or temporal scales (Storfer et al. 2007). A limited number of studies have examined landscape effects on patterns of genetic variation at different spatial scales (e.g., Lee-Yaw et al. 2009, Murphy et al. 2010), and researchers have only recently begun to explicitly consider scale effects on landscape genetic inference (e.g., Cushman & Landguth 2010). Scale-related questions and considerations are expected to become increasingly relevant to studies trying to disentangle the complex relationships between spatial heterogeneity and genetic variability (Anderson et al. 2010).

Another key question in landscape genetics that remains largely unexplored relates to the importance of landscape configuration (i.e., spatial arrangement of habitat types) on patterns of genetic variation, relative to landscape composition (relative abundance of different habitats in the landscape). Storfer et al. (2007) defined landscape genetics as research that specifically quantifies the effects of landscape composition, configuration and/or matrix quality on gene flow and/or spatial genetic variation. Landscape composition measures the diversity and quantity of different habitat types within a landscape (i.e., what habitat is there and how much of it is there), while landscape configuration measures the spatial arrangement of habitat types (i.e., how are different

habitat patches shaped and organized in space). The ongoing challenge is to separate the effects of these two aspects of landscape structure on the spatial distribution of genetic variation, as they are often confounded, not only in landscape genetic studies, but more generally in many landscape ecological studies (McGarigal & Cushman 2002, Fahrig 2003).

Study systems in landscape genetics are thus needed that can provide a coherent framework for addressing the questions mentioned above, and that are relevant to various species and across different spatial scales. A model study system in landscape genetics should ideally allow the researcher to: (i) examine patterns and processes at various spatial scales, (ii) sample replicate ‘landscapes' at each spatial scale, (iii) potentially control landscape composition and configuration, and (iv) perform comparative studies among different species. The insect fauna associated with pitcher plants provide these key features of a candidate model system in landscape genetics (Krawchuk & Taylor 2003).

1.1 The purple pitcher plant and its inhabitants

Three insects of the order Diptera (i.e., 'flies') lay their eggs or larvae exclusively within the water-containing, pitcher-shaped leaves of the carnivorous purple pitcher plant, *Sarracenia purpurea* L. (Addicott 1974, Heard 1994, Harvey & Miller 1996). The developing larvae feed on the decomposing prey of the pitcher plant, which are primarily other insects, as well as associated microbes (Fish & Hall 1978, Heard 1994). Purple pitcher plants are found within acidic bogs through much of eastern North America (Schnell 2002). Bogs define discrete patches of habitat with the landscape, and within them pitcher plants are distributed in clusters that are readily identifiable. Multiple leaves are also found in each plant (Figure 1.1). Thus, this system provides a series of discrete habitat patches that are hierarchically nested at different spatial scales (leaf-plant-cluster-bog-group of bogs), and that are used by three different but highly specialized species. Bog habitats are widespread and display variation in their abundance and spatial distribution across their range. Therefore, a researcher has a number of potential study landscapes to choose from that differ in the abundance and spatial configuration of bog habitats.

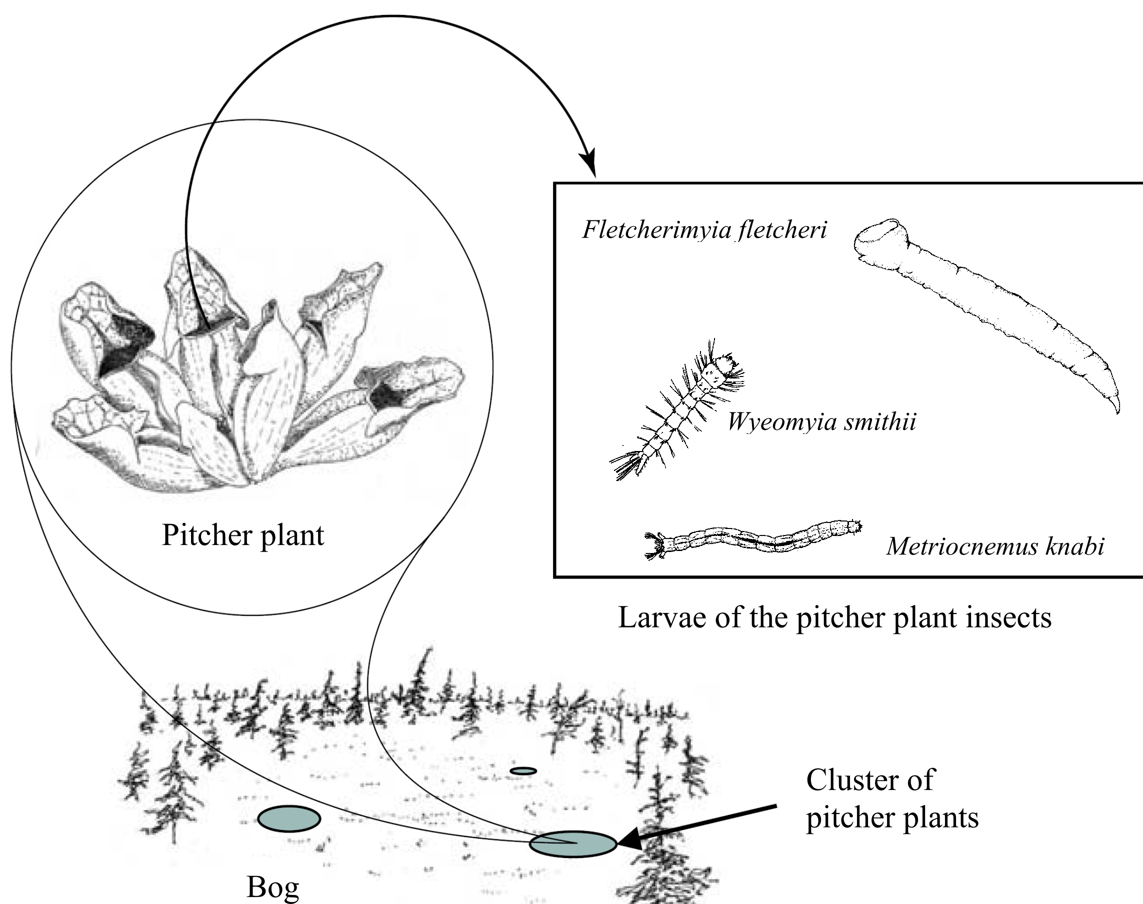


Figure 1.1 Larvae of the three insect species (*Fletcherimyia fletcheri* Aldrich 1916, *Wyeomyia smithii* Coquillett 1901, *Metriocnemus knabi* Coquillett 1904) found within leaves of the purple pitcher plant (*Sarracenia purpurea* L). Plants are distributed in clusters within a bog (adapted from Srivastava et al. 2004).

The insect inhabitants of the purple pitcher plant thus offer several important advantages for landscape genetic studies. First, easily detectable habitat patches at several nested spatial scales allow for multi-scale studies, without a need for an arbitrary delineation focal scales. Second, one can achieve replication at the landscape level at each of the different spatial scales. Also, by careful selection of study sites, one can potentially control landscape composition (how much of habitat is present) and configuration (how habitat is spatially arranged), and thereby separate the effects of these two factors.

Along with other microcosms (e.g., aquatic insects in bromeliads, micro-arthropods in moss patches, beetles in fungal sporocarps, micro-crustaceans in rockpools), the pitcher plant and its inhabitants have been proposed as a model system in ecology (Srivastava et al. 2004). This microcosm has indeed been used in community ecology and landscape ecology studies to understand community interactions and community composition (Harvey & Miller 1996, Buckley et al. 2004, 2010, Trzcinksi et al. 2005), colonization patterns (Trzcinksi et al. 2003), and species distribution and abundance (Krawchuk & Taylor 2003).

For this system to be useful in addressing questions in landscape genetics, it is first essential to obtain insight into the dispersal abilities and spatial population genetic structure of the insect species. Dispersal is a key process linking landscape structure and spatial genetic variation. Yet, data on the movement capabilities of the pitcher plant inhabiting species are very limited. Traditional methods for estimating dispersal such as mark-release-recapture have been applied only in the pitcher plant flesh fly (*Fletcherimyia fletcheri* Aldrich 1916), and indicated that adults readily move within a bog and have the potential for fluent movement among bogs (Krawchuk & Taylor 2003). For the pitcher plant midge (*Metriocnemus knabi* Coquillett 1904) and the mosquito (*Wyeomyia smithii* Coquillett 1901) such an approach is unfeasible, given their small adult body size and cryptic behaviour. Indirect dispersal estimates, based on the patterns of larval abundance, indicated that both the midge and the mosquito have very limited movement potential, even a within a bog (Krawchuk & Taylor 2003). Spatial genetic structure has only been investigated in the pitcher plant mosquito at the phylogeographic scale using high-throughput sequencing (Emerson et al. 2010), and at a smaller spatial scale using allozymes (Istock & Weisburg 1987).

1.2 Outline

My broad goal in developing this thesis was to establish a model system in landscape genetics using three insect species associated with the northern pitcher plant *Sarracenia purpurea* [the pitcher plant flesh fly (*F. fletcheri*, Sarcophagidae), midge (*M. knabi,* Chironomidae) and mosquito (*W. smithii,* Culicidae)]. To that end, I developed molecular tools to test hypotheses about the extent of gene flow in the insect species, and to test hypotheses that broad scale landscape variables influence the insects' spatial genetic patterns.

This thesis consists of four data chapters that describe research undertaken to achieve the proposed goal. The first data chapter describes *de novo* development of microsatellite markers for the three insect species. Microsatellite loci are the most commonly used genetic markers in landscape genetic research, as they provide the spatial and temporal resolution to distinguish closely related individuals and estimate contemporary dispersal and gene flow events. My second data chapter explores the pattern of spatial genetic structure and the extent of gene flow in the pitcher plant flesh fly (*F. fletcheri*). Here, I employed a hierarchical sampling design to test the theoretical predictions of different hypothesized population models (patchy populations, metapopulations or isolated populations). The third data chapter examines the importance of considering genetic patterns and ecological processes across multiple spatial scales and in multiple landscapes when investigating genetic diversity within a species. Specifically, I assessed genetic differentiation at several scales in the pitcher plant midge (*M. knabi*) and tested explicit hypotheses about the effects of several landscape variables on processes (female oviposition and dispersal) underlying spatial genetic structure across spatial scales. The fourth data chapter investigates the effect of long-term wind patterns (direction and frequency) on gene flow and genetic structure at large spatial scales in the pitcher plant midge (*M. knabi*). Although wind-assisted dispersal may be an important process in many small terrestrial arthropods, this study provides the first explicit test of a hypothesis that wind patterns influence gene flow. I conclude my thesis by discussing the overall implications of my work, and the potential future uses and benefits of this study system for landscape genetic studies.

1.3 References

Addicott JF (1974) Predation and prey community structure: an experimental study of the effect of mosquito larvae on the protozoan communities of pitcher plants. *Ecology* 55:475-492.

Anderson CD, Epperson BK, Fortin MJ, et al. (2010) Considering spatial and temporal scale in landscape-genetic studies of gene flow. *Molecular Ecology* 19:3565-3575.

Avise JC, Arnold J, Ball RM, et al. (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecological Systematics* 18:489-522.

Balkenhol N, Gugerli F, Cushman S, et al. (2009) Identifying future research needs in landscape genetics: where to from here? *Landscape Ecology* 24:455-463.

Buckley HL, Burns JH, Kneitel JM, et al. (2004) Small-scale patterns in community structure of *Sarracenia pururea* inquilines. *Community Ecology*, 5:181-188.

Buckley HL, Miller TE, Ellison AM, Gotelli NJ (2010) Local to continental-scale variation in the richness and composition of an aquatic food web. *Global Ecology and Biogeography* 19:711-723.

Cushman SA, Landguth EL (2010) Spurious correlations and inferences in landscape genetics. *Molecular Ecology* 19:3592-3602.

Emerson KJ, Merz CR, Catchen JM, et al. (2010) Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences USA* 107:16196-16200.

Epps CW, Wehausen JD, Bleich VC, et al. (2007) Optimizing dispersal and corridor models using landscape genetics. *Journal of Applied Ecology* 44:714-724.

Fahrig L (2003) Effects of habitat fragmentation on biodiversity. *Annual Review of Ecology, Evolution and Systematics* 34:487-515.

Fish D, Hall DW (1978) Succession and stratification of aquatic insects inhabiting the leaves of the insectivorous pitcher plant *Sarracenia purpurea*. *American Midland Naturalist* 99:172-183.

Foll M, Gaggiotti O (2006) Identifying the environmental factors that determine the genetic structure of populations. *Genetics* 174:875-891.

Guillot G (2009) On the inference of spatial structure from population genetics data. *Bioinformatics* 25:1796-1801.

Harvey E, Miller TE (1996) Variance in composition of inquiline communities in leaves of *Sarracenia purpurea* L. on multiple spatial scales. *Oecologia* 108:562-566.

Heard SB (1994) Pitcher-plant midges and mosquitoes: a processing chain commensalism. *Ecology* 75:1647-1660.

Holderegger R, Wagner HH (2008) Landscape genetics. *BioScience* 58:199-207.

Istock CA, Weisburg WG (1987) Strong habitat selection and the development of population structure in a mosquito. *Evolutionary Ecology* 1:348-362.

Kimura M, Weiss GH (1964) The stepping-stone model of population structure and the decrease of genetic correlation with distance. *Genetics* 49:561-576.

Krawchuk MA, Taylor PD (2003) Changing importance of habitat structure across multiple spatial scales for three species of insects. *Oikos* 103:153-161.

Lee-Yaw JA, Davidson A, McRae BH, Green DM (2009) Do landscape processes predict phylogeographic patterns in the wood frog? *Molecular Ecology* 18:1863-1874.

Malécot G (1948) *Les mathématiques sde l'hérédité*. Masson, Paris.

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

Maruyama T (1971) Analysis of population structure. II. Two-dimensional stepping stone models of finite length and other geographically structured populations. *Annals of Human Genetics* 35:179-196.

Maruyama T (1972) Rate of decrease of genetic variability in a two-dimensional continuous population of finite size. *Genetics* 70:639-651.

Maruyama T (1977) *Stochastic problems in population genetics*. Springer-Verlag, Berlin.

McGarigal K, Cushman SA (2002) Comparative evaluation of experimental approaches to the study of habitat fragmentation effects. *Ecological Applications* 12:335-345.

McRae BH (2006) Isolation by resistance. *Evolution* 60:1551-1561.

Murphy MA, Evans JS, Storfer A (2010) Quantifying *Bufo boreas* connectivity in Yellowstone National Park with landscape genetics. *Ecology* 91:252-261.

Nagylaki T (1978) The geographical structure of populations. In: Levin SA (ed.) *Studies in mathematics*. Mathematical Association of America, Washington, DC.

Neel MC (2008) Patch connectivity and genetic diversity conservation in the federally endangered and narrowly endemic plant species *Astragalus albens* (Fabaceae). *Biological Conservation* 141:938-955.

Schnell DE (2002) *Carnivorous plants of the United States and Canada*. Timber Press Inc. Portland, Oregon.

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

Spear SF, Storfer A (2008) Landscape genetic structure of coastal tailed frogs (*Ascaphus truei*) in protected vs. managed forests. *Molecular Ecology* 17:4642-4656.

Spear SF, Peterson CR, Matocq M, Storfer A (2005) Landscape genetics of the blotched tiger salamander (*Ambystoma tigrinum melanostictum*). *Molecular Ecology* 14:2553-2564.

Srivastava DS, Kolasa J, Bengtsson J, et al. (2004) Are natural microcosms useful model systems for ecology? *Trends in Ecology and Evolution* 19:379-384.

Storfer A, Murphy MA, Evans JS, 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.

Trizio I, Crestanello B, Galbusera B, et al. (2005) Geographical distance and physical barriers shape the genetic structure of Eurasian red squirrels (*Sciurus vulgaris*) in the Italian Alps. *Molecular Ecology* 14:469-481.

Trzcinksi MK, Waldea S, Taylor PD (2005) Local interactions in pitcher plant communities scale-up to regional patterns in distribution and abundance. *Environmental Entomology* 34:1464-1470.

Trzcinski MK, Walde S, Taylor PD (2003) Colonisation of pitcher plant leaves at several spatial scales. *Ecological Entomology* 28:482-489.

Vignieri SN (2005) Streams over mountains: influence of riparian connectivity on gene flow in the Pacific jumping mouse (*Zapus trinotatus*). *Molecular Ecology* 14:1925-1937.

Wang IJ (2010) Recognizing the temporal distinctions between landscape genetics and phylogeography. *Molecular Ecology* 19:2605-2608.

Wright S (1943) Isolation by distance. *Genetics* 28:114-138.

Zannese A, Morellet N, Targheta C, et al. (2006) Spatial structure of roe deer populations: towards defining management units at a landscape scale. *Journal of Applied Ecology* 43:1087-1093.

Chapter 2. Isolation of microsatellite loci for the pitcher plant insects*

2.1 Introduction

Landscape genetic studies are frequently conducted at smaller spatial scales and consequently involve sampling of individuals that are closely related, with small genetic differences among them (Holderegger & Wagner 2008). Landscape genetic questions also often revolve around contemporary rather than historical ecological processes, such as current movements of individuals or responses to recent anthropogenic landscape change (Proctor et al. 2005). Therefore, molecular analyses in this discipline rely on highly variable genetic markers that provide sufficient spatial and temporal resolution to distinguish closely related individuals and estimate contemporary dispersal events (Holderegger & Wagner 2008). A total of 18 different types of genetic markers have been used in landscape genetic studies, the most common being allozymes, mitochondrial DNA (mtDNA), microsatellites, amplified fragment length polymorphisms (AFLP), and randomly amplified polymorphic DNA (RAPD) (Storfer et al. 2010). Nuclear markers have been preferentially used for addressing questions at small spatial and temporal scales (Balkenhol et al. 2009, Anderson et al. 2010), with microsatellites being the most prevalent markers used in landscape genetic studies of animals (Holderegger & Wagner 2008, Storfer et al. 2010).

Microsatellite loci consist of 1 to 6 base-pair sequence motifs that are tandemly repeated a variable number of times (Weber & May 1989, Schlötterer 2000). The majority are likely to be selectively neutral, making them informative about the population-level processes of gene flow and genetic drift. Microsatellite markers are also highly

---

* A version of this chapter has been published: Rasic G, Maxwell SA, Keyghobadi N (2009) Characterization of microsatellite loci for the pitcher plant midge, *Metriocnemus knabi* Coq. (Diptera: Chironomidae). *Molecular Ecology Resources* 9:1388-1391.

Rasic G, Keyghobadi N (2009) Microsatellite loci characterization in the pitcher plant flesh fly, *Fletcherimyia fletcheri* Aldrich (Diptera: Sarcophagidae). *Molecular Ecology Resources* 9:1460-1466 (part of "Permanent Genetic Resources added to Molecular Ecology Resources Database 1 May 2009–31 July 2009").

polymorphic, with usually more than five alleles per locus, and are variable even in populations that have low levels of allozyme and mitochondrial DNA variation (Hedrick 1999). Therefore, they provide the power to determine relatedness among individuals, estimate contemporary gene flow and dispersal, and distinguish high levels of gene flow from panmixia (Selkoe & Toonen 2006). Microsatellites are codominant markers that follow Mendelian inheritance (Weber & May 1989). Because both alleles of a heterozygote can be distinguished, microsatellites provide direct estimates of heterozygosity and allele frequencies that are integral to various population-genetic models (e.g., Wright's F-statistics; Wright 1969), including those applied in landscape genetic hypothesis-testing (Storfer et al. 2007).

The DNA sequences surrounding a microsatellite locus are called flanking regions (Goldstein & Schlötterer 1999). Oligonucleatide primers between 17 and 25 base pairs in length can be designed to bind to the flanking regions on either side of a microsatellite and initiate polymerase chain reaction (PCR) amplification. Amplification products are typically separated and visualized via high-resolution electrophoresis, and genotypes of individuals at each microsatellite locus are assessed based on the sizes of the alleles amplified. Flanking regions are usually conserved within a species, and so primers will amplify the same target microsatellite locus in all conspecific individuals. Conversely, the same primers rarely work in all but closely related taxa, due to accumulated mutations in the flanking regions that disrupt primer binding (Glenn & Schable 2005).

Microsatellite markers have not been developed in any species that are closely related to the pitcher plant insects, precluding the development of markers by cross-species amplification for my project. My **objective** was therefore to develop microsatellite markers *de novo* for the common pitcher plant dipterans: flesh fly *Fletcherimyia fletcheri* Aldrich 1916 (Sarcophagidae), midge *Metriocnemus knabi* Coquillett 1904 (Chironomidae) and mosquito *Wyeomyia smithii* Coquillett 1901 (Culicidae).

The traditional method of microsatellite development includes screening a genomic library for microsatellite-containing clones (Rassmann et al. 1991). Several thousand clones are usually screened though colony hybridization with microsatellite repeat-

containing probes, with the number of microsatellite-containing clones typically ranging from 12% to less than 0.04% (Zane et al. 2005). Hence, traditional methods are prone to low return for a significant effort. Alternative methods involve creating DNA libraries that are highly enriched for microsatellite loci, such that a much higher proportion of the clones contain microsatellite sequences (Armour et al. 1994, Fleischer & Loew 1995, Kirkpatrick et al. 1995). I followed the enrichment-based protocol developed by Hamilton et al. (1999) which incorporates the magnetic capture of biotinylated probes bound to microsatellite-containing genomic fragments by streptavidin beads. This protocol also includes using 'SNX' linkers that serve as universal primers for the PCR recovery of the microsatellite-enriched fragments. **The rationale** for choosing the protocol by Hamilton et al. (1999) in my study was its consistently high success rate across taxa, including various insects (Zane et al. 2005, Techen et al. 2010).

2.2 Materials and Methods

2.2.1 Collection of larvae for microsatellite development

Midge (*Metriocnemus knabi*) and mosquito (*Wyeomyia smithii*) larvae were collected from a bog in the Big East River area, Huntsville, Ontario, Canada in May 2007. Flesh fly (*Fletcherimyia fletcheri*) larvae were collected from Dizzy Lake bog in Algonquin Provincial Park, Ontario, Canada in August 2007. Larvae were pipetted out from leaves of the purple pitcher plant (*Sarracenia purpurea)* and immediately stored in 100% ethanol at -20°C, preventing otherwise fast tissue and DNA degradation.

2.2.2 Collection of larvae for microsatellite variability assessment

For assessing variability of the isolated microsatellite loci, I analyzed 23 individuals of each species, which is a typical sample size in such analyses (e.g., *Molecular Ecology Resources* requires a minimum of 20 individuals, http://tomato.biol.trinity.edu/mer_faq.html). For these analyses, *M. knabi* larvae were sampled from Dizzy Lake bog, and *W. smithii and F. fletcheri* larvae were sampled from Spruce bog in Algonquin Provincial Park, Ontario, Canada in August 2007. Larvae were collected and stored as described above.

2.2.3 Processing genomic DNA for microsatellite development

Genomic DNA from larvae was extracted using the DNeasy® blood and tissue kit (Qiagen, Germantown, MC). To obtain at least 5 μg of high molecular weight genomic DNA for each species, necessary for initiating microsatellite development, I pooled extracted DNA from four larvae for the midge or mosquito, and two larvae for the flesh fly. This DNA pooling also reduces any potential ascertainment bias when designing primers based on a genome of only one individual (Glenn & Shable 2005). Whole genomic DNA in each species was completely digested to generate fragments of the desired size range (300bp - 1000bp). It is potentially helpful to combine multiple restriction endonucleases in these initial digests (Glenn & Schable 2005). For both the midge and mosquito, I used two combinations of restriction enzymes: 1) *NheI*, *MseI*, *HaeIII* and 2) *NheI*, *MseI*, *RsaI* (New England BioLabs, Pickering, CA) and combined their products in each species. The enzyme combination for the flesh fly consisted of *NheI*, *RsaI*, and *HaeIII*. Overhangs in the resulting DNA fragments were digested with exonuclease to create blunt ends. Additionally, their 5' ends were desphosphorylated, which decreased the likelihood of creating chimeric sequences (i.e., sequences created by ligation of DNA fragments from different regions in the genome).

I then followed the general protocol outlined by Hamilton et al. (1999) for constructing microsatellite-enriched libraries.

2.2.3.1 *Ligation of DNA fragments to linkers*

The genomic fragments were blunt-end ligated to SNX double stranded linkers:
SNX Forward (5' – CTAAGGCCTTGCTAGCAGAAGC – 3') and
SNX Reverse (5' – pGCTTCTGCTAGCAAGGCCTTAGAAAA – 3')
(Hamilton et al. 1999).

These linkers later served as unique primers for the PCR amplification of all microsatellite-enriched fragments that were cloned. The SNX linkers also contain necessary restriction sites for ligation into a cloning vector.

2.2.3.2 *Microsatellite enrichment with biotin-labeled oligos and streptavidin beads*

To select preferentially genomic DNA fragments that contain microsatellite repeats, the linker-ligated, digested genomic DNA was made single-stranded and then hybridized with biotin-labeled microsatellite probes: $(GT)_{15}$, $(GA)_{15}$, $(GACA)_8$ and $(GATC)_8$. Genomic DNA bound to the probes was then captured using streptavidin-coated magnetic beads (Dynabeads® M-270 Streptavidin, Dynal, Camarillo, CA), which bind to the biotin on the microsatellite probes. A magnet was used to immobilize the beads (and attached DNA) such that genomic DNA lacking the repeats and not hybridized to the probes could be washed away. I then released the microsatellite-enriched DNA from the probes by heating at 95ºC for 10 minutes, and pipetted out the single-stranded target DNA, while the probes remained bound to the beads, which were immobilized by a magnet. The single-stranded target DNA was then made double stranded by PCR, using the SNX linkers as universal primer-binding sites.

2.2.3.3 *Ligation of genomic DNA into plasmids*

To isolate individual microsatellite-containing fragments, the resulting enriched double-stranded DNA was then cloned into a bacterial plasmid vector pBluescript II SK (+) (Stratagene, Santa Clara, CA). This was accomplished by creating complementary overhangs using the restriction enzymes *Stu*I and *Nhe*I in the enriched insect DNA, and *EcoR*I in the vector. The complementary overhangs were then ligated with T4 DNA Ligase (New England BioLabs, Pickering, CA).

2.2.3.4 *Transformation of recombinant plasmids into competent E. coli*

The ligation products (i.e., plasmids containing microsatellite-enriched DNA fragments) were transformed by electroporation into *Escherichia coli* XL1 Blue MRF' electro-competent cells (Stratagene, Santa Clara, CA). I plated the cells onto Luria-Bertani (LB) medium containing ampicillin, which prevented the growth of untransformed cells. Blue/white screening was used to determine the efficiency of ligations from the previous step (Maniatis et al. 1989).

2.2.3.5 *Selection of colonies containing microsatellite inserts*

Given that the enrichment protocol is not 100% efficient (Hamilton et al. 1999), I further screened the enriched library using a standard hybridization method. The bacterial colonies were transferred onto Hybond-N^+ nylon membranes (Amersham, GE Healthcare Life Sciences) that were then air dried, UV crosslinked, and incubated with proteinase K at 55ºC (to remove bacterial debris that would cause high background). Biotinylated microsatellite repeat oligonucleotides with the same sequences as those used for the enrichment, were hybridized to the membranes at 65°C overnight. Detection of positive (i.e., microsatellite-containing) clones was carried out using a chemiluminescent Phototope®-Star Detection Kit (New England BioLabs, Pickering, CA). This protocol results in a conjugate between alkaline phosphatase and streptavidin, which becomes bound to the biotinylated probes on the membrane. When de-phosphorylated, light is emitted indicating the location of microsatellite-containing colonies. Each membrane was placed in the dark in a gel-doc with the chemiluminescence filter (Fluor Chem 8900, Alpha Innotech, Santa Clara, CA) and its digital image was recorded after two minutes of exposure. Light-emitting (i.e., microsatellite-containing) colonies appeared as dark dots in the resulting images, which were printed onto transparencies that were then matched to the original bacterial colonies in each plate. Microsatellite-containing colonies were picked from the plates using sterile pipette tips and transferred into individual micro-centrifuge tubes with 50µL of T.E. solution (10mM Tris pH = 8.0, 0.1 mM EDTA). To release the microsatellite-containing plasmids, the solutions of bacterial cells were heated to 95°C for 5 minutes and then vigorously shaken.

2.2.3.6 *Sequencing of positive clones*

Inserts in plasmids from positive colonies were amplified via PCR in a 25 µL final volume reaction (1X PCR Buffer (50 mM KCl, 10 mM Tris-HCl pH 8.3 at room temperature), 1 mM each dNTP, 1.5 U of Ampli*Taq* DNA polymerase (Applied Biosystems, Forest City, CA), 4 mM $MgCl_2$, 1.25 µM each of T3 and T7 primers, and 2µL of bacterial suspension) in the following cycling program: denaturation at 96 °C for 5 min; 35 cycles of 40 s at 94 °C, 1 min at 50 °C, 2 min at 72 °C; and a final elongation step of 5 min at 72 °C. PCR products were then run on 2% agarose gel and DNA was

extracted from bands in the gel using the QIAquick® Gel Purification kit (Qiagen, Germantown, MC). These purified inserts were then sequenced in both directions with T3 and T7 primers using BigDye cycle sequencing chemistry on a 3730 genetic analyzer (Applied Biosystems, Forest City, CA). I analyzed sequence data using the software Sequencher (Gene Codes Corp., Ann Arbor, MI).

2.2.3.7 *Designing and optimizing microsatellite primers*

I used the program PRIMER 3.0 (Rozen & Skaletsky 2000) to design primers complementary to sequences flanking the microsatellite arrays that I isolated from individual colonies. I designed primers for inserts with the following characteristics: the presence of 5 or more uninterrupted repeat units in the microsatellite, containing adequately long sequence flanking the microsatellite repeat (>30 bp on each side), and classification as *unique* in the program MicroFamily (Meglécz 2007). The latter program is designed for identifying flanking region similarities between different microsatellite sequences obtained from screening partial genomic libraries (Meglécz 2007). Non-unique microsatellites are more likely to give multiple banding patterns during PCR amplifications, which can be very difficult to interpret.

To optimize PCR amplification and test for variability, as well as Hardy-Weinberg and linkage equilibria, I amplified each locus in 23 individuals of the respective species, sampled from a single bog. Genomic DNA from these individuals was purified using the DNeasy® blood and tissue kit (Qiagen, Germantown, MC).

PCR amplifications were performed in a PTC-0200 DNA Engine Cycler (Bio-Rad, Hercules, CA) in a 20 µL final volume containing 1X PCR Buffer (50 mM KCl, 10 mM Tris-HCl pH 8.3 at room temperature), 0.15 mg/ml BSA, 0.3 mM each dNTP, 1.5 U of Ampli*Taq* DNA polymerase (Applied Biosystems, Forest City, CA), 3.75 mM $MgCl_2$, 0.25 µM of each primer, and approximately 300 ng of larval DNA. One primer of each pair was 5'-labeled with either 6FAM, NED, PET or VIC fluorescent dye (Applied Biosystems, Forest City, CA).

For each locus, one of the following PCR profiles was used (Tables 2.1-2.5):

(i) denaturation for 3 min at 96 °C; 2 cycles of 30 s at 96 °C, 30 s at 60 °C, 30 s at 72 °C; 14 touchdown cycles of 15 s at 96 °C, 15 s at 60 °C (-0.5 °C each cycle), 15 s at 72 °C; 17 cycles of 15 s at 96 °C, 15 s at 53°C, 15 s at 72 °C; and a final elongation step of 3 min at 72 °C;

(ii) denaturation for 3 min at 96 °C; 2 cycles of 30 s at 96 °C, 30 s at 56 °C, 30 s at 72 °C; 12 touchdown cycles of 15 s at 96 °C, 15 s at 56 °C (-0.5 °C each cycle), 15 s at 72 °C; 20 cycles of 15 s at 96 °C, 15 s at 50.5 °C, 15 s at 72 °C; and a final elongation step of 3 min at 72 °C;

(iii) denaturation for 3 min at 96 °C; 2 cycles of 30 s at 96 °C, 30 s at 53 °C, 30 s at 72 °C; 6 touchdown cycles of 15 s at 96 °C, 15 s at 53 °C (-0.5 °C each cycle), 15 s at 72 °C; 25 cycles of 15 s at 96 °C, 15 s at 50.5 °C, 15 s at 72 °C; and a final elongation step of 3 min at 72 °C;

(iv) only for one flesh fly locus (FF82): denaturation for 3 min at 96 °C, 30 cycles of 30 s at 96 °C, 30 s at 50 °C, 30 s at 72 °C, and a final elongation step of 3 min at 72 °C.

Negative (water) and positive controls were included in all amplifications, where the template for positive controls were 1:100 dilutions of the corresponding amplified clone inserts from the microsatellite-enriched library. The sizing of PCR products was done on a 3730 analyzer using Genemapper software (Applied Biosystems, Forest City, CA) with LIZ-500 size standard.

2.2.3.8 *Data analysis*

The number of alleles per locus, frequency of the most common allele, observed and expected heterozygosity and linkage equilibrium analyses were determined using Microsatellite Analyzer (Dieringer & Schloetterer 2003) and GENEPOP version 3.4 (Raymond & Rousset 1995). The potential presence of null alleles was checked using the program Micro-Checker (van Oosterhout et al. 2004).

2.2.3.9 *Multiplexing of microsatellite loci*

Individual amplification of large numbers of loci in many individuals of multiple species inevitably results in inefficient use of time and resources. My studies of spatial genetic variation in the pitcher plant insects demanded the analyses of hundreds of individuals, and therefore I needed to establish a more efficient genotyping protocol. This was done in two principal ways: (1) by simultaneously amplifying several loci in a single PCR reaction, also known as multiplexing (Edwards & Gibbs 1994), and (2) by combining products of different PCR reactions in a 'genotyping sample' to be analyzed in a single well of the genetic analyzer, referred to as multiloading.

To optimize multiplex reactions, I tested combinations of loci that had the same cycling profile, and either different 5' fluorescent labels or non-overlapping allele size ranges. Multiplexing amplifications were performed in a 20 µL final volume containing 1X PCR Buffer (50 mM KCl, 10 mM Tris-HCl pH 8.3 at room temperature), 0.15 mg/ml BSA, 0.3 mM each dNTP, 1.5 U of Ampli*Taq* DNA polymerase (Applied Biosystems, Forest City, CA), 3.75 mM $MgCl_2$, and 0.12-0.33µM of each primer (volumes of specific primers found in Tables 2.3-2.4) and approximately 300 ng of larval DNA. In the multiloading procedure, I combined PCR products of individual and/or multiplexed loci that had either different 5' fluorescent labels or non-overlapping allele size ranges.

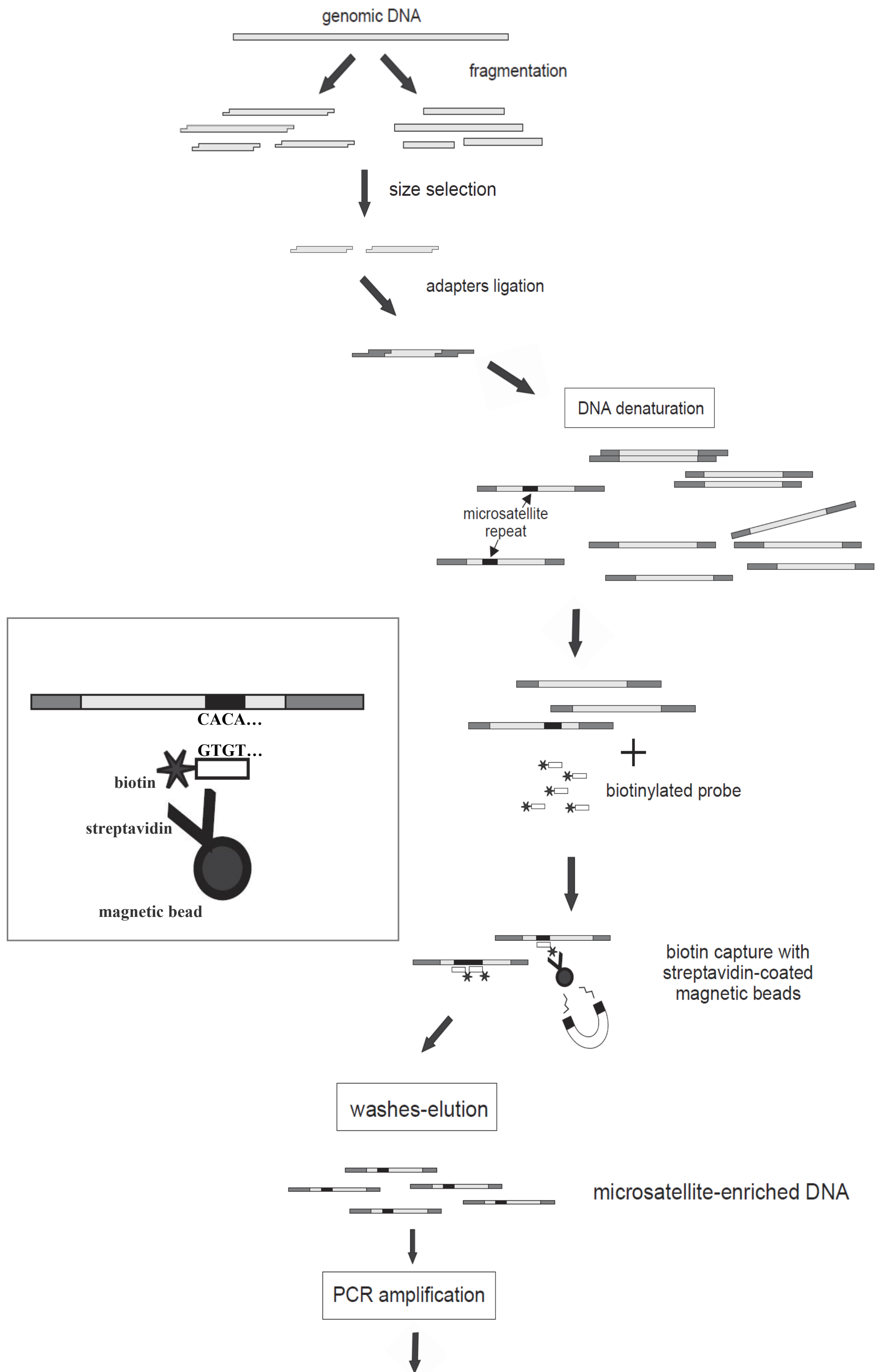


Figure 2.1 Microsatellite isolation following the enrichment-based protocol (adapted from Zane et al. 2005). Continued on next page…

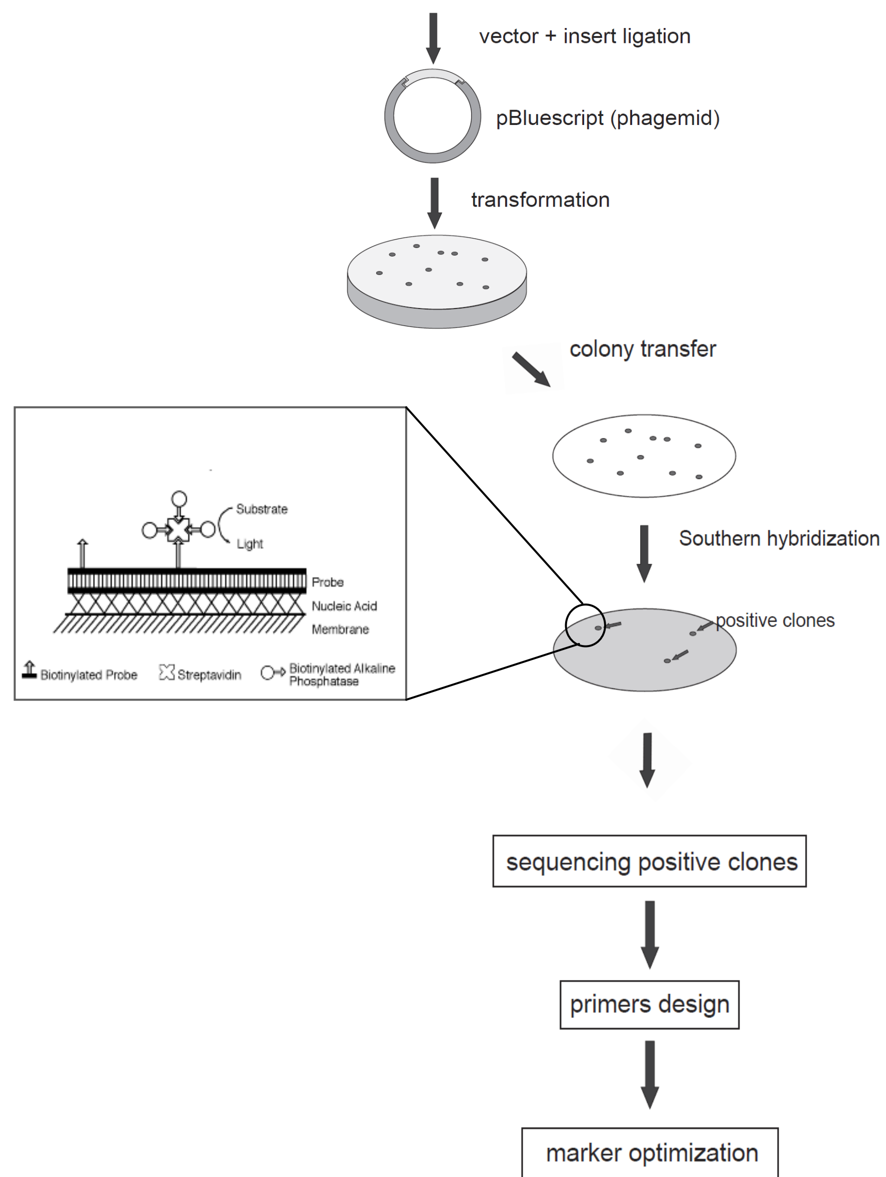


Figure 2.1 … continued from previous page

2.3 Results

2.3.1 Microsatellite development

In the flesh fly *F. fletcheri,* I detected and amplified 105 positive colonies from the microsatellite-enriched library. Of the successfully sequenced colonies, 85 contained microsatellites. Of these, I used 52 unique sequences for primer design, as they contained five or more repeats and had a sufficiently long flanking region on either side of the microsatellite. Fifteen loci amplified well with the resulting primers, but the remaining 37 loci failed to amplify even after several attempts to optimize PCR conditions. Fourteen loci amplified with a 100% success rate, whereas the locus FF217 repeatedly failed to amplify in two individuals (Table 2.1). Twelve loci were polymorphic (contained two or more alleles) in the original sample of 23 individuals. For these 12 loci, the number of alleles per locus ranged from two to eight (mean 4.58) and the observed heterozygosity per locus ranged from 0.19 to 0.91 (mean 0.49), which suggests a moderate level of genetic variability for microsatellites. Two loci, FF238 and FF217, deviated from Hardy-Weinberg equilibrium and showed a significant homozygote excess ($P < 0.001$), probably due to the presence of null alleles with the estimated frequencies of 0.216 and 0.348 respectively. The observed failure of amplification for locus FF217 in two individuals (presumably null homozygotes) was thus consistant with the estimated frequency of the null allele. No significant linkage disequilibrium was detected.

Table 2.1 Microsatellite loci developed for *Fletcherimyia fletcheri*. Number of alleles (N_a), frequency of the most common allele (*a*), observed (H_o) and expected (H_e) heterozygosity were calculated from 23 individuals. The asterix represents significant deviation from Hardy-Weinberg equilibrium (* for $P < 0.05$, *** for $P < 0.001$). *n* represents the number of individuals successfully amplified out of 23 individuals tested. Sequences are provided for forward (F) and reverse (R) primers. § denotes pig-tailed primer, where §-GTTT represents sequence added at 5' end to promote non-templated adenylation of the PCR product (as in: Brownstein et al. 1996).

Locus	Accession no.	Primer sequence (5'- 3')	Core Repeat	Size range (bp)	N_a	*a*	H_o	H_e	PCR profile	*n*/23
FF009	GQ300842	F: TGACTGCCATACGATTCACAC R: CTATACACAGCAGCGGACAAAC	$(GT)_9$	120-130	2	0.783	0.435	0.348	1	23
FF104	GQ300843	F: TGAAGAAATACCCAACATAATGAAC R: ACCGCCTAGCTTTCTAAACAC	$(GA)_{16}$	144-160	6	0.391	0.826	0.755	1	23
FF072	GQ300844	F: CGCCACTGTTTATACCAGAAATG R: AAACTGAATAGAGAAACGGCACAC	$(TG)_3CG(TG)_3GG(TG)_5$	197-201	3	0.457	0.696	0.656	1	23
FF010	GQ300845	F: CGAAAGGAATTACGTATAGCCAGAA R: GGGTGCACACTGCACAGAC	$(GACA)_2GATA(GACA)_5$	131-149	3	0.717	0.435	0.453	1	23
FF238	GQ300846	F: TGGACGGATATAGCTTTCAACAC R: GTTTGTTCGCCTACTCAGAAATG	$(AC)_{16}$	110-128	5	0.717	0.217	0.470^{*}	1	23
FF189	GQ300847	F: TCGTTCCCATGAGGTTGTATG R: CAACCATTTGCTGTTGAAGTTG	$(ACAG)_6$	165-185	4	0.630	0.478	0.558	1	23
FF217	GQ300848	F: TGTTAAGCGTCCACAAAACTAAAC R: CCCGTATAAATGAGAGCGAGAC	$(CT)_{17}$	128-188	7	0.427	0.191	0.739^{***}	1	21
FF231	GQ300849	F: VIC-CAATTTTAATCACACAAAATGGTAGG R: AGCCGACGTTCAGACTCTTC	$(GA)_6GG(GA)_2GG(GA)_{20}$	128-158	8	0.283	0.913	0.822	1	23
FF065	GQ300850	F: GATGACAATTCGATAAACAGACA R: GCTTACTGGAGTTGAAATGGT	$(GACA)_5GGCA(GACA)_4$	121-159	5	0.826	0.348	0.314	2	23
FF249	GQ300851	F: TGTTCGATAAACTTCCTCTT R: AAATCAAACACGCTACCA	$(GTCT)_6$	242-258	3	0.543	0.478	0.527	3	23
FF062	GQ300852	F: TATATGAAACGCTGTGACC R: ACGAAATAAACTAAATATTACACAA	$(TG)_{12}$	166-174	3	0.826	0.261	0.300	3	23
FF082	GQ300853	F: TTTCGTTTAAAGCTGAATAAA R: GTTTCCTATCCAAATTACGACAAC§	$(GA)_{17}$	105-117	6	0.543	0.609	0.618	4	23

In the pitcher plant midge *M. knabi,* I detected and amplified 97 positive clones of which 62 contained microsatellites. Of these, 22 sequences were suitable for primer design. Nine of the 22 loci did not give a satisfactory amplification, as they either failed to amplify or produced multiple non-specific peaks. Of the remaining 13 loci, 12 were polymorphic in a sample of 23 individuals. For these 12 loci, the number of alleles per locus ranged from two to 15 (mean 5.58) and the observed heterozygosity per locus ranged from 0.04 to 0.95 (mean 0.53), revealing a relatively high level of genetic variability (Table 2.2). The only locus that showed significant homozygote excess and deviation from Hardy-Weinberg equilibrium was MK01, probably due to the presence of a null allele with an estimated frequency of 0.123. None of the pairwise comparisons between loci for linkage disequilibrium were statistically significant after Bonferroni correction. However, the probability for the linkage disequilibrium test for MK78 and MK124 was below 0.01, suggesting that these loci deserve further scrutiny when testing across different populations.

Table 2.2 Microsatellite loci developed for *Metriocnemus knabi*. Number of alleles (N_a), frequency of the most common allele (f_a), observed (H_o) and expected (H_e) heterozygosity were calculated from 23 individuals. The asterix represents significant deviation from Hardy-Weinberg equilibrium (* for $P < 0.05$). *n* represents the number of individuals successfully amplified out of 23 individuals tested. Sequences are provided for forward (F) and reverse (R) primers. § denotes pig-tailed primers, where §-GTTTCTT; §§-GTTT, §§§-GTT represent sequences added at 5' end to promote non-templated adenylation of the PCR product (as in: Brownstein et al. 1996).

Locus	Accession no.	Primer sequence (5'-3')	Core Repeat	Size range (bp)	N_a	f_a	H_o	H_e	PCR profile	n/23
MK25	FJ665262	F:GTTTCTTTTTCAACTTTCTTTCTATGTTCTGTG§ R: CCTTCATGGCTTGGTAGAG	$(TG)_{11}$	147-151	3	0.773	0.364	0.376	2	23
MK01	FJ665263	F: GTTTGTCTCTTTTCTCAGGGTTTCAC§§§ R: CAAATTGCAGGAAGCATCAA	$(GT)_{10}$	152-194	5	0.364	0.545	0.750*	1	23
MK80	FJ665264	F: TCGCATTCCTGAATCTCGTTAG R: AGCATCGTATGAAGCCTTGTTG	$(CA)_{10}$	240-242	2	0.956	0.087	0.085	1	23
MK71	FJ665265	F: ATTACAAGGAATTATCGGAAAC R: TCTAAATTATCTTTTGTTGAGTCTG	$(GT)_{13}$	113-127	4	0.500	0.696	0.587	2	23
MK78	FJ665266	F: CGGATGACACGCAATGA R: TCATCATCATCAAGTCCTCTTTCT	$(GA)_{13}$	144-160	6	0.543	0.652	0.644	1	23
MK11	FJ665267	F: ACGTGCGATGTTTCTTG R: AATATCCAGTTTCAGTTCTTCTC	$(GA)_2A(GA)_7$	195-201	2	0.978	0.043	0.043	3	23
MK112	FJ665268	F: ACTGAAGCTCCCAAAAGTGT R: TTTTGCCTTTTCCTCTCAA	$(CA)_3TA(CA)_5$ $TA(CA)_5$	103-121	6	0.587	0.652	0.616	2	23
MK34	FJ665269	F: AATGGACAGCCTACCTCTTG R: GTTTCTATTTTAGCATTCCGCCTGTC§	$(CA)_{12}$	169-177	4	0.609	0.609	0.564	2	23
MK116	FJ665270	F: ACGGATGATTGGCGTTTTC R: GTTTCAATGCATCAACCAACACC§§§	$(AC)_{10}$	98-102	3	0.587	0.435	0.513	1	23
MK119	FJ665271	F: GGAAGATGGGGCGAGTG R: GTTTCGTATATCGTCCAGTCTGTTGTG§§	$(GA)_{17}$	88-130	11	0.435	0.696	0.766	1	23
MK124	FJ665272	F: TATGCGTGAGTGTCCGTCTC R: GTTTCCACATGCTTCTCACTGTTG§§§	$(TG)_{16}$	149-187	6	0.609	0.652	0.586	1	23
MK94	FJ665273	F: GTTTCCAATGGGTCATAATCAA§§§ R: AGCCTTCTGCGATGTAAG	$(AC)_{33}$	167-223	15	0.159	0.954	0.928	3	23

I detected and sequenced 112 positive colonies in the pitcher plant mosquito, *W. smithii*, out of which 84 contained microsatellites. Of these, 58% could not be used for primer design, as they lacked a sufficiently long flanking sequence (i.e., more than 30 nucleotides) on one or both sides of the microsatellite region. I was able to design primer pairs for 35 unique loci, but 32 of those exhibited non-specific amplification, resulting in multiple bands or smearing patterns when visualized on a gel. Repeated attempts to optimize PCR conditions (by changing the temperature profiles of PCR reactions, concentrations of different PCR ingredients, and concentration of DNA template) failed to improve these results. Only three loci amplified well (WS6, WS68 and WS92), and all three were polymorphic (Table 2.3). The number of alleles ranged from three to seven (mean 5). Average observed heterozygosity ranged between 0.05 and 0.50, but in all three loci this value was significantly lower than the expected heterozygosity ($P < 0.01$). Excess homozygosity indicated the presence of null alleles at all loci, with estimated frequencies of: 0.2 in WS6, 0.137 in WS68, and 0.163 in WS92. Last, linkage disequilibrium between pairs of loci was not detected.

In the flesh fly, six loci were multiplexed and another six loci were multiloaded, leading to only two genotyping samples per individual to analyze a total of 12 loci (Table 2.4). In the midge, I optimized three multiplexes (a 6-plex, a 4-plex and a 2-plex) (Table 2.5). The overlap of allele size ranges between loci prevented further multiloading in the midge. As each of the three usable loci in the mosquito required different PCR cycling conditions, they could not be multiplexed, but could be multiloaded.

Table 2.3. Microsatellite loci developed for *Wyeomyia smithii*. Number of alleles (N_a), frequency of the most common allele (*a*), observed (H_o) and expected (H_e) heterozygosity were calculated from 23 individuals. The asterix represents significant deviation from Hardy-Weinberg equilibrium (*** for $P < 0.001$). *n* represents the number of individuals successfully amplified out of 23 individuals tested. Sequences are provided for forward (F) and reverse (R) primers. One primer of each pair has one of the following 5' fluorescent labels: 6FAM, PET or NED (Applied Biosystems).

Locus	Primer sequence (5'- 3')	Core Repeat	Size range (bp)	N_a	*a*	H_o	H_e	PCR profile	*n*/23
WS06	F: 6-FAM-CGATCGGTTCAGTAGTTTTC R: AGGTCATATAACGCTCTTGTTC	$(CAGA)_8$	112-142	3	0.932	0.045	0.129***	2	22
WS68	F: NED-TCATAGGAGATAGAAATTAGATGAA R: GTTCCATTTGCTTGGTTAG	$(GACA)_8$	254-308	5	0.455	0.500	0.659***	3	22
WS92	F: PET-GAATCCACTCACTGCTCTCC R: TCAATCGGTTGTTGGGTTTC	$(GT)_{20}$	205-305	7	0.565	0.435	0.643***	1	23

Table 2.4 Multiplex PCR combinations for microsatellite loci of *Fletcherimyia fletcheri*. One primer of each pair has one of the following 5’ fluorescent labels: 6FAM, PET, VIC or NED (Applied Biosystems). ‘Primer conc.’ for each locus indicates the final concentration of each primer (μM) used in a final 20 μL volume PCR reaction. [&] indicates multiplex or single (-) PCR reactions that are multiloaded.

Multiplex set	Loci	Label	Primer conc. (μM)	PCR Cycling Profile
1	FF009	6-FAM	0.25	1
	FF010	NED	0.25	1
	FF072	6-FAM	0.30	1
	FF104	PET	0.25	1
	FF189	NED	0.25	1
	FF231	VIC	0.25	1
2[&]	FF217	PET	0.30	1
	FF238	NED	0.25	1
3[&]	FF062	6-FAM	0.20	3
	FF249	6-FAM	0.20	3
-[&]	FF065	VIC	0.25	2
-[&]	FF082	PET	0.25	4

Table 2.5 Multiplex PCR combinations for microsatellite loci of *Metriocnemus knabi*. One primer of each pair has one of the following 5’ fluorescent labels: 6FAM, PET, VIC or NED (Applied Biosystems). ‘Primer conc.’ for each locus indicates the final concentration of each primer (μM) used in a final 20 μL volume PCR reaction.

Multiplex set	Loci	Label	Primer conc. (μM)	PCR Cycling Profile
1	MK01	NED	0.30	1
	MK80	6-FAM	0.25	1
	MK78	6-FAM	0.25	1
	MK119	VIC	0.25	1
	MK124	PET	0.25	1
	MK116	6-FAM	0.25	1
2	MK112	PET	0.30	2
	MK34	6-FAM	0.25	2
	MK25	VIC	0.15	2
	MK71	NED	0.30	2
3	MK94	6-FAM	0.25	3
	MK11	NED	0.12	3

2.4 Discussion

I successfully developed 12 polymorphic microsatellite loci in both the pitcher plant flesh fly and midge. Between six and 15 microsatellite loci are typically used in landscape genetic studies (Holderreger & Wagner 2008). However, the degree of variability, and not simply the total number of loci, plays an important role in obtaining satisfactory power in landscape and population genetic analyses (Paetkau 2004). Using simulated and real data, Corander et al. (2003) determined that a total of 50 alleles are usually sufficient for fine-scale genetic studies. In samples of only 23 individuals, I recorded 55 alleles in the flesh fly and 67 alleles in the midge (Tables 2.1-2.2). Therefore, the microsatellite loci I developed for these two species are sufficiently numerous and variable to give a satisfactory spatial and temporal resolution even for very small-scale analyses.

I was also able to increase the efficiency of multilocus genotyping by multiplexing and multiloading. I reduced the number of genotyping reactions per individual from 12 to only two in the flesh fly (Table 2.4) and three in the midge (Table 2.5), thus achieving a significant reduction in the genotyping costs for both the fly and the midge.

Why not the mosquito?

I was not able to achieve satisfactory results with the isolation of microsatellite loci in the pitcher plant mosquito. Only three loci out of 35 tested in *W. smithii* showed clean and specific amplification, and also all three exhibited null alleles (Table 2.3). Characterization of microsatellite markers in different species of the mosquito family Culicidae has had a variable success. For example, numerous microsatellite loci have been successfully isolated in several species of *Anopheles sp*. (Zheng et al. 1993, 1996, Sinkins et al. 2000) and *Culex sp.* (Fonseca et al. 1998, Keyghobadi et al. 2004, Smith et al. 2005). However, in *Ochlerotatus* and *Aedes* species, the same procedure has proven problematic (Fagerberg et al. 2001, Widdel et al. 2005, Chambers et al. 2007).

A low abundance of microsatellite sequences in a genome has been proposed as one of the causes of difficult microsatellite isolation (Meglécz et al. 2004, 2007). For example, a low frequency of positive clones in a microsatellite-enriched DNA library was reported

for the yellow fever mosquito *Aedes aegypti* (Fagerberg et al. 2001, Huber et al. 2001) and many butterflies (e.g., Meglécz & Solignac 1998, Keyghobadi et al. 1999, 2002, Nève & Meglécz 2000, Prasad et al. 2005). However, in *W. smithii* large numbers of microsatellite-containing clones were detected, indicating that low abundance of microsatellite sequences was not a problem in this species.

Despite my success in creating the microsatellite-enriched library for *W. smithii*, only a low proportion of sequences (42%) were suitable for primer design. This was due mainly to insufficiently long flanking regions on one side of a microsatellite. Such a pattern was not pronounced in the pitcher plant midge or flesh fly whose genomes were subjected to the same screening procedure. Indeed, the microsatellite development procedure was conducted simultaneously for the mosquito and the midge. Asymmetry in the length of microsatellite flanking regions of cloned microsatellites was also reported in the yellow fewer mosquito *Aedes aegypti*, a species known to have a challenging genome for microsatellite isolation (Chambers et al. 2007). The exact mechanism responsible for this asymmetry pattern is not known.

The most common explanation for limited success in the development of microsatellite markers is the existence of duplicated microsatellite-containing regions (i.e., 'microsatellite families') throughout the genome of a species (Meglécz et al. 2007). For example, in two groups known to be problematic for microsatellite characterization, many mosquitoes of the genus *Aedes* as well as butterflies, microsatellite families are found to be at least twice as frequent as in *Anopheles* or *Culex* species (Meglécz et al. 2007). Microsatellite families contain several microsatellite loci with similar or identical flanking regions. Primers designed in the repetitive flanking regions of a given locus amplify simultaneously in several other loci with similar sequences (Zhang 2004). This results in multiple non-specifically amplified fragments, causing uninterpretable banding patterns (Van't Hof et al. 2007). I observed such a pattern of amplification in all but three tested *W. smithii* loci. Thus, the low success rate in characterizing microsatellite loci in this mosquito is most likely caused by the presence of microsatellite DNA families.

Analysis of microsatellite flanking regions using the software MicroFamily (Meglécz 2007) did not detect high sequence redundancy in cloned microsatellites from the pitcher plant mosquito, as 89% of analyzed sequences were classified as unique. However, I found clear evidence for duplication in locus WS68, as it consistently showed two identical genotyping profiles that were 200 base pairs apart in each tested individual.

The existence of microsatellite families can be at least partially explained by the association between microsatellites and transposable genetic elements (Tay et al. 2010). For example, a microsatellite could arise within a transposable element and spread across the genome during transposition. In fact, transposons (or their remnants) are significantly more associated with microsatellite families than with unique microsatellite sequences in 32 insect species (Meglécz et al. 2007). I was interested to determine whether there is any evidence of such a phenomenon in *W. smithii.* I compared flanking sequences isolated from *W. smithii* to all nucleotide sequences found in NCBI (The National Center for Biotechnology Information) database. The implemented procedure called BLASTn (nucleotide Basic Alignment Search Tool) finds regions of local similarity between nucleotide sequences. It splits the query sequence into small fragments, finds sequences that contain any of the query words (hits), extends these hits in both directions until there are no more matches, and finally, calculates the quality of the extended hit.

BLASTn analyses showed that sequences flanking microsatellites in *W. smithii* consistently had the highest match to sequences identified as either transposons or microsatellite regions in *Sabethes sp.*, *Aedes sp.* and *Ochlerotatus sp*. (Table 2.6). This result is compatible with the phylogenetic relationship among these species. Specifically, the pitcher plant mosquito and *Sabethes sp.* are members of the tribe Sabethini, whereas *Aedes sp.* and *Ochlerotatus sp*. belong to tribe Aedini, and both tribes are the members of the same subfamily Culicinae (Harbach 2007). Given that the characterization of microsatellite loci has proven problematic in all of these mosquitoes (Pedro PM *personal communication*, Widdel et al. 2005, Chambers et al. 2007), BLASTn results in my study further support the idea that the genome of *W. smithii* contains microsatellite families associated with transposable elements.

Table 2.6 BLASTn analysis for 24 *Wyeomyia smithii* sequences obtained from the microsatellite-enriched DNA library that produced highly similar matches (i.e., with Expected (E) value below 10^{-5}). Max identity indicates the maximum percentage of identical nucleotides between *W. smithii* sequence to its matched sequence.

Sequence type	Species	Max identity %	# of *W. smithii* sequence matches
transposon	*Aedes atropalpus* *Aedes epactius*	73-92 76-78	6
microsatellite DNA	*Sabethes sp.* *Aedes taeniorhynchus* *Ochlerotatus caspius*	73-93 75-78 83	11
other	*Armigeres subalbatus* *Ochlerotatus epactius,* *Aedes aegypti*	70-86 71-80 70-94	7

Null alleles were detected in all three usable *W. smithii* loci. A high incidence of null alleles at microsatellite loci has been noted in Lepidoptera and in *Anopheles* mosquitoes (Palo et al. 1995, Kamau et al. 1999, Keyghobadi et al. 1999, Meglécz et al. 2004, Sarhan 2006). It is hypothesized that many null alleles in Lepidoptera are caused by (1) mutations in primer binding sites resulting in unsuccessful PCR, or (2) insertions that produce alleles with PCR fragments sizes that fall outside the standard detection range (Van't Hof et al. 2007). For example, *W. smithii* locus WS92 contains alleles that differ in size by as much as 100 base pairs (Table 2.3), which points towards the existence of large inserts in the flanking region of this locus. Thus, in addition to the problems caused by microsatellite families associated with transposable elements, identification of usable microsatellite loci in *W. smithii* is aggravated by a relatively high flanking sequence variability that manifests itself as null alleles.

A protocol that removes highly repetitive DNA and surveys only single copy DNA for the presence of microsatellite regions has been successfully implemented in *Aedes japonicus* (Widdel et al. 2005) and could be done in *Wyeomyia smithii* as well. Another suitable alternative to the challenging isolation of microsatellite isolation in this species is the development of AFLP or SNP markers that demonstrate comparable spatial and temporal resolution in landscape genetic studies (Anderson et al. 2010).

Although my attempt to isolate microsatellite loci in *W. smithii* resulted in very few usable markers, it provided novel information about the genome of this well-studied species. My results indicate that this member of the tribe Sabethini generally exhibits more genome similarity with species from the tribe Aedini compared to the tribe Culicini (e.g., *Culex sp.*). Deeper phylogenetic relationships within the family Culicidae are largely unresolved (Harbach 2007) and the results of my study could contribute to the taxonomic reorganization of this insect group. Conserved microsatellite flanking sequences, also known as repetitive flanking sequences (ReFS), have been shown to be an effective dominant marker that can differentiate between species of Lepidoptera (Anderson et al. 2007, Molodstova et al. 2011). Unlike microsatellite markers, which are practically unusable if developed from microsatellite families, ReFS specifically utilize

the information on nucleotide variability contained within the flanking regions of these loci to determine interspecific relationships (Anderson et al. 2007). The implementation of ReFS could be a useful approach in resolving current unclear phylogenetic relationships in Culicidae, and the sequence information I obtained from microsatellite families in the pitcher plant mosquito could contribute to such an effort.

2.5 Summary

I achieved the goal of isolating microsatellite loci in the pitcher plant flesh fly, *F. fletcheri,* and midge, *M. knabi*. In each species, twelve loci were sufficiently polymorphic to provide adequate resolution for fine scale landscape genetic studies. I also optimized the protocol for highly efficient genotyping of individuals (multiplexing/multiloading). These data are now published and available to the scientific community (Rasic et al. 2009, Rasic & Keyghobadi 2009). Microsatellite isolation in the pitcher plant mosquito *W. smithii* proved to be very problematic, most likely due to existence of microsatellite families associated with transposable elements and further aggravated by the prevalence of null alleles. If microsatellite isolation in *W. smithii* were to be attempted again, I would strongly recommend the creation of a single-copy DNA library, followed by the standard enrichment protocol.

2.6 References

Anderson CD, Epperson BK, Fortin MJ, et al. (2010) Considering spatial and temporal scale in landscape-genetic studies of gene flow. *Molecular Ecology* 19:3565-3575.

Anderson SJ, Gould P, Freeland JR (2007) Repetitive flanking sequences (ReFS): novel molecular markers from microsatellite families. *Molecular Ecology Notes* 7:374-376.

Armour JAL, Neumann R, Gobert S, Jefferys AJ (1994) Isolation of human simple repeat loci by hybridization selection. *Human Molecular Genetics* 3:599-605.

Balkenhol N, Gugerli F, Cushman SA, et al. (2009) Identifying future research needs in landscape genetics: where to from here? *Landscape Ecology* 24:455-463.

Brownstein MJ, Carpten JD, Smith JR (1996) Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. *BioTechniques* 6:1004-1010.

Chambers EW, Meece JK, MGowan JA, et al. (2007) Microsatellite isolation and linkage group identification in the Yellow fever mosquito *Aedes aegypti. Journal of Heredity* 98:202-210.

Corander J, Waldmann P, Sillanpaa MJ (2003) Bayesian analysis of genetic differentiation between populations. *Genetics* 163:367-374.

Dieringer D, Schlötterer C (2003) Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes* 3:167-169.

Edwards MC, Gibbs RA (1994) Multiplex PCR: advantages, development, and applications. *Genome Research* 3:S65-S75.

Fagerberg AJ, Fulton RE, Black Iv WC (2001) Microsatellite loci are not abundant in all arthropod genomes: analyses in the hard tick, *Ixodes scapularis* and the yellow fever mosquito, *Aedes aegypti*. *Insect Molecular Biology* 10:225-236.

Fleischer RC, Loew S (1995) Construction and screening of microsatellite-enriched genomic libraries. In: Ferraris J, Palumbi S (eds.) *Molecular zoology: advances, strategies and protocols.* Wiley-Liss, New York.

Fonseca DM, Atkinson CT, Fleischer RC (1998) Microsatellite primers for *Culex pipiens quinquefasciatus*, the vector of avian malaria in Hawaii. *Molecular Ecology* 7:1617-1619.

Glenn TC, Schable NA (2005) Isolating microsatellite DNA loci. *Methods in Enzymology* 395:202-222.

Goldstein D, Schlötterer C (1999) *Microsatellites: evolution and applications*. Oxford University Press, Oxford, UK.

Hamilton MB, Pincus EL, Di-Fiore A, Fleischer RC (1999) Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. *Biotechniques* 27:500-507.

Harbach RE (2007) The Culicidae (Diptera): a review of taxonomy, classification and phylogeny. *Zootaxa* 1668:591-638.

Hedrick PW (1999) Highly variable loci and their interpretation in evolution and conservation. *Evolution* 53:313-318.

Holderegger R, Wagner HH (2008) Landscape genetics. *Bioscience* 58:199-208.

Huber K, Mousson L, Rodhain F, Failloux AB (2001) Isolation and variability of polymorphic microsatellite loci in *Aedes aegypti*, the vector of dengue viruses. *Molecular Ecology Notes* 1:219-222.

Kamau L, Mukabana WR, Hawley WA, et al. (1999) Analysis of genetic variability in *Anopheles arabiensis* and *Anopheles gambiae* using microsatellite loci. *Insect Molecular Biology* 8:287-297.

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, Matrone MA, Ebel GD, Kramer LD, Fonseca DM (2004) Microsatellite loci from the northern house mosquito (*Culex pipiens*), a principal vector of West Nile virus in North America. *Molecular Ecology Notes* 4:20-22.

Kirkpatrick BW, Bradshaw M, Barendse W, Dentine MR (1995) Development of bovine microsatellite markers from a microsatellite-enriched library. *Mammalian Genome* 6:526-528.

Maniatis T, Fritsch EF, Sambrook J (1989) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Meglécz E, Solignac M (1998) Microsatellite loci for *Parnassius mnemosyne* (Lepidoptera). *Hereditas* 128:179-180.

Meglécz E, Petenian F, Danchin E, D'Acier AC, Rasplus J-Y, Faure E (2004) High similarity between flanking regions of different microsatellites detected within each of two species of Lepidoptera: *Parnassius apollo* and *Euphydryas aurinia*. *Molecular Ecology* 13:1693-1700.

Meglécz E (2007) MicroFamily: a computer program for detecting flanking region similarities among different microsatellite loci. *Molecular Ecology Notes* 7:18-20.

Meglécz E, Anderson SJ, Bourguet D, et al. (2007) Microsatellite flanking region similarities among different loci within insect species. *Insect Molecular Biology* 16:175-185.

Molodstova D, Crowe E, Olson A, Yee J, Freeland JR (2011) Conserved flanking microsatellite sequences (ReFS) differentiate between Lepidoptera species, and provide insight into microsatellite evolution. *Systematic Entomology* 36:371-376.

Nève G, Meglécz E (2000) Microsatellite frequencies in different taxa. *Trends in Ecology and Evolution* 15:376-377.

Paetkau D (2004) The optimal number of markers in genetic capture-mark-recapture studies. *The Journal of Wildlife Management* 3:449-452.

Palo J, Varvio S, Hanski I, Vainola R (1995) Developing microsatellite markers for insect population structure: complex variation in a checkerspot butterfly. *Hereditas* 123:295-300.

Prasad MD, Muthulakshmi M, Madhu M, Archak S, Mita K, Nagaraju J (2005) Survey and analysis of microsatellites in the silkworm, *Bombyx mori*: frequency,

distribution, mutations, marker potential and their conservation in heterologous species. *Genetics* 169:197-214.

Proctor MF, McLellan BN, Strobeck C, Barclay RMR (2005) Genetic analysis reveals demographic fragmentation of grizzly bears yielding vulnerably small populations. *Procedings of the Royal Society of London, B* 272:2409-2416.

Rasic G, Maxwell SA, Keyghobadi N (2009) Characterization of microsatellite loci for the pitcher plant midge, *Metriocnemus knabi* Coq. (Diptera: Chironomidae). *Molecular Ecology Resources* 9:1388-1391.

Rasic G, Keyghobadi N (2009) Microsatellite loci characterization in the pitcher plant flesh fly, *Fletcherimyia fletcheri* Aldrich (Diptera: Sarcophagidae). *Molecular Ecology Resources* 9:1460-1466 (part of "Permanent Genetic Resources added to Molecular Ecology Resources Database 1 May 2009–31 July 2009").

Rassmann K, Schlötterer C, Tautz D (1991) Isolation of simple-sequence loci for use in polymerase chain reaction-based DNA fingerprinting. *Electrophoresis* 12:113-118.

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

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds.) *Bioinformatics methods and protocols: methods in molecular biology*. Humana Press, Totowa, NJ.

Sarhan A (2006) Isolation and characterization of five microsatellite loci in the Glanville fritillary butterfly (*Melitaea cinxia*). *Molecular Ecology Notes* 6:163-164.

Schlötterer C (2000) Evolutionary dynamics of microsatellite DNA. *Chromosoma* 109:365-371.

Selkoe KA, Toonen RJ (2006) Microsatellites for Ecologists: A practical guide to using and evaluating microsatellite markers. *Ecology Letters* 9:615-629.

Sinkins SP, Hackett BJ, Costantini C, et al. (2000) Isolation of polymorphic microsatellite loci from the malaria vector *Anopheles funestus*. *Molecular Ecology* 9:490-492.

Smith JL, Keyghobadi N, Matrone MA, Escher RL, Fonseca DM (2005) Cross-species comparison of microsatellite loci in the *Culex pipiens* complex and beyond. *Molecular Ecology Notes* 5:697-700.

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

Storfer A, Murphy MA, Evans JS, et al. (2007) Putting the 'landscape' in landscape genetics. *Heredity* 98:128-142.

Tay WT, Behere GT, Batterham P, Heckel DG (2010) Generation of microsatellite repeat families by RTE retrotransposons in Lepidopteran genomes. *BMC Evolutionary Biology* 10. doi:10.1186/1471-2148-10-144.

Techen N, Arias RS, Glynn NC, Pan Z, Khan IA, Scheffler BE (2010) Optimized construction of microsatellite-enriched libraries. *Molecular Ecology Resources* 10:508-515.

Van't Hof AE, Brakefield PM, Saccheri IJ, Zwaan BJ (2007) Evolutionary dynamics of multilocus microsatellite arrangements in the genome of the butterfly *Bicyclus anynana*, with implications for other Lepidoptera. *Heredity* 98:320-328.

van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* 4:535-538.

Weber JL, May PE (1989) Abundant class of human DNA polymorphism which can be typed using the polymerase chain reaction. *American Journal of Human Genetics* 44:388-396.

Widdel AK, McCuiston LJ, Crans WJ, Kramer LD, Fonseca DM (2005) Finding needles in the haystack: single copy microsatellite loci for *Aedes japonicus* (Diptera: Culicidae). *The American Journal of Tropical Medicine and Hygiene* 73:744-748.

Wright S (1969) *Evolution and the genetics of populations*. University of Chicago Press, Chicago.

Zane L, Bargelloni L, Patarnello T (2005) Strategies for microsatellite isolation: a review. *Molecular Ecology* 11:1-16.

Zhang DX (2004) Lepidopteran microsatellite DNA: redundant but promising. *Trends in Ecology and Evolution* 19:507-509.

Zheng L, Benedict MQ, Cornel AJ, Collins FH, Kafatos FC (1996) An integrated genetic map of the African human malaria vector mosquito, *Anopheles gambiae*. *Genetics* 143:941-952.

Zheng L, Collins FH, Kumar SV, Kafatos FC (1993) A detailed genetic map for the X chromosome of the malaria vector, *Anopheles gambiae*. *Science* 261:605-608.

Chapter 3. The pitcher plant flesh fly exhibits a mixture of metapopulation and patchy characteristics

3.1 Introduction

Phytotelmata of the carnivorous purple pitcher plant (*Sarracenia purpurea* L.) are the exclusive habitat for larval development of a group of dipteran insects that includes the pitcher plant flesh fly (*Fletcherimyia fletcheri* Aldrich 1916), midge (*Metriocnemus knabi* Coquillett 1904), and mosquito (*Wyeomyia smithii* Coquillett 1901), and also represent the habitat of specialized mites, rotifers and bacteria (Harvey & Miller 1996, Dahlem & Naczi 2006). This biological microcosm has been recognized as a strong candidate model system in ecology (Srivastava et al. 2004) and it has been used in metacommunity (e.g., Buckley et al. 2004, 2010, Holyoak et al. 2005, Mouquet et al. 2008) and landscape ecological research (Krawchuk & Taylor 2003, Trzcinski et al. 2003). This system could also be useful for looking at effects of habitat spatial structure and addressing questions of spatial scale in population and landscape genetics (Rasic et al. 2009), although it is first essential to obtain insight into the dispersal abilities and spatial population genetic structure of the resident species.

Ecological data on the insect inhabitants of the purple pitcher plant beyond the larval stage are limited, particularly with respect to dispersal characteristics. Furthermore, spatial genetic structure has only been investigated in the pitcher plant mosquito (*W. smithii*) at the phylogeographic scale using high-throughput sequencing (Emerson et al. 2010), and at a smaller spatial scale using allozymes (Istock & Weisburg 1987). Here, I investigate the spatial population structure in the largest of the pitcher plant insects, the pitcher plant flesh fly *F. fletcheri*. Using selectively neutral microsatellite genetic markers to estimate how genetic variation is distributed among and within populations, I also infer levels of gene flow and dispersal distances in this species.

F. fletcheri is exclusively associated with the purple pitcher plant that is found within peatlands throughout Eastern North America (Schnell 2002). The fly's larval development occurs within plant leaves (i.e., 'pitchers') over the summer, and the larvae

possess a uniquely large cuplike posterior spiracular pit that enables them to float at the liquid surface and feed on newly drowned prey (Johansen 1935). The peat moss surrounding the pitcher plants serves as an overwintering habitat for the diapausing pupae (Forsyth & Robertson 1975). Adult flies use the pitcher plant flower heads as overnight roosting sites and mating locations (Krawchuk & Taylor 1999). Therefore, *F. fletcheri* populations exist only where bogs with *S.purpurea* plants are found. Bogs form discrete habitat patches within a forested landscape, leading to the patchy occurrence of the flesh fly populations. Within bogs, pitcher plants also exhibit a patchy spatial distribution (Krawchuk & Taylor 2003).

Populations in nature are often spatially subdivided, due to either natural spatial heterogeneity of the habitat, as in *F. fletcheri*, or fragmentation of a previously continuous habitat. Three main models describe the organization and dynamics of spatially subdivided populations: (1) patchy population, (2) metapopulation, and (3) isolated populations (Mayer et al. 2009). Patchy populations are characterized by high connectivity among subpopulations, and essentially function as a single population with little potential for the local extinction of any given subpopulation (Harrison 1991). Metapopulations occupy partially isolated habitat patches that support local breeding populations, with extinctions and recolonizations dynamically occurring within such local populations (i.e., subpopulations). Metapopulations are thus characterized by intermediate connectivity among subpopulations and turnover within them (Levins 1970). Finally, the third model considers subpopulations that are isolated and independent from each other. In this model, habitat patches in which local extinctions occur would not be recolonized (Frankham et al. 2002).

Dispersal is an essential process underlying the conceptual framework of the three population models, and the models make contrasting predictions regarding the extent of dispersal and hence gene flow, among subpopulations with consequences for the distribution of genetic variability within and among local subpopulations (Mayer et al. 2009, summarized in Table 3.1). The patchy population model predicts absence of genetic differentiation among local subpopulations, due to the homogenizing effects of high rates of dispersal and gene flow (Harrison 1991). Therefore, isolation-by-distance

(Wright 1943) among subpopulations should be absent, and the entire patchy population (i.e., collection of all subpopulations) should be in Hardy-Weinberg equilibrium (Slatkin 1987). Bayesian clustering algorithms (Pritchard et al. 2000) that do not assume any preconceived number of subpopulations should produce a single genetic cluster. Also, subpopulations within a patchy population usually retain stable effective population sizes and do not undergo genetic bottlenecks.

The metapopulation model predicts limited dispersal, and hence and gene flow, among subpopulations, typically in a distance-dependent manner (i.e., reduced gene flow at greater distances between subpopulations) (Hanski 1994). Therefore, a metapopulation should be characterized by significant genetic differentiation among some subpopulations (Hastings & Harrison 1994) and a pattern of isolation-by-distance. The total metapopulation should exhibit a significant heterozygote deficit (i.e., deviation from Hardy-Weinberg equilibrium known as the Wahlund effect), due to the pooling of genetically differentiated samples (Wahlund 1928). Therefore, the whole metapopulation should also contain more than one genetic cluster. Local subpopulations may undergo extinction/recolonization events (Hanski 1999), and some should therefore exhibit detectable signals of recent genetic bottlenecks.

Finally, the isolated population model assumes no gene flow among subpopulations, leading to very high genetic differentiation among all of them, and a consequent high heterozygote deficit at the total population level. Due to random changes in allele frequencies (i.e., genetic drift) within subpopulations and absence of the homogenizing effect of gene flow, the level of genetic differentiation among subpopulations is not correlated with the geographic distance among them, thus isolation-by-distance should not be present (Hutchinson & Templeton 1999). The number of distinct genetic clusters should correspond to the number of subpopulations. Finally, the extent to which one might detect a signal of a bottleneck should only be a function of local dynamics only and whether there are significant local fluctuations in population size, and not related to an extinction-recolonization process.

Table 3.1 Predictions about the distribution of genetic variation and gene flow among subpopulations resulting from the three models of population structure.

	Population model predictions			
	Differentiation among subpopulations	*IBD among subpopulations*	*Number of genetic clusters (K)*	*Signatures of genetic bottleneck*
Patchy population	none	No	1	No
Metapopulation	moderate	Yes	>1, but less than # of subpopulations	Yes
Isolated populations	high	No	= # of subpopulations	Maybe - depending on local dynamics

Direct estimates of dispersal ability of *F. fletcheri,* based on a small mark-recapture experiment, showed that adults readily move within a bog and have the potential for fluent movement among bogs (Krawchuk & Taylor 2003). Thus, the available ecological data suggest that *F. fletcheri* populations should exhibit a patchy population structure within bogs, and either patchy or metapopulation structure among bogs. In this study, I used genetic data to assess the spatial population structure in *F. fletcheri* and to test theoretical predictions of the three population models. To that end, I used microsatellite markers specifically developed for this species (Rasic & Keyghobadi 2009) and both individual- and population-based analyses to explore the patterns of neutral genetic diversity and differentiation, levels of gene flow, effective population sizes and signs of bottleneck events in *F. fletcheri* samples. I employed a hierarchical sampling design and examined genetic patterns: within a bog, within a group of closely situated bogs (i.e., system of bogs), and among two such groups of bogs located in Algonquin Provincial Park (Ontario, Canada). According to the best-supported population model, I discuss implications for future landscape genetic research in this species.

3.2 Materials and Methods

3.2.1 Study area and species

Algonquin Provincial Park in Ontario, Canada (UTM: 17N 687337E 5046853N) is in an area of transition between northern coniferous forest and southern deciduous forest. These forests are dominated, respectively, by: (i) white pine (*Pinus strobus*) and red pine (*Pinus resinosa*), poplar (*Populus spp.*) and white birch (*Betula papyrifera*), and (ii) sugar maple (*Acer saccharum*), american beech (*Fagus grandifolia*), hemlock (*Tsuga canadensis*), red oak (*Quercus rubra*), and yellow birch (*Betula alleghaniensis*). Bogs are found within this forest matrix, and many of them contain *S. purpurea* and its associated commensal arthropod inhabitants. Bogs represent peat-covered wetlands in which the vegetation shows the effects of a high water table and a general lack of nutrients. Due to poor drainage and the decay of plant material, the surface water of bogs is strongly acidic. Dominated by sphagnum mosses (*Sphagnum spp.*) and heath shrubs (leather leaf *Chamaedaphne calyculata*, labrador tea *Rhododendron groenlandicum*, cranberries *Vaccinium spp.*), the bogs also contain tamarack (*Larix laricina*) and black spruce (*Picea mariana*) (Tiner 1999).

I sampled *F. fletcheri* (Diptera: Sarcophagidae) from 11 bogs located in an area of approximately 26×10 km in Algonquin Provincial Park. These included two groups of five closely spaced bogs (0.2-7.0 km apart in a mixed forest matrix), each of which constitutes a 'system'. These two systems (SYS1 and SYS2) were 26 km apart. The eleventh bog (TT) was located in between the two systems (Figure 3.1). I sampled one additional bog (Sifton) that represents a highly isolated location in the urban area London, Ontario, Canada (UTM: 17T 473541E 4757717N); the nearest known neighbouring bogs are 100 km distant, and this bog is 400 km away from the other bogs I sampled in Algonquin Park. This isolated bog was used as an outlier group in some analyses.

The focal species in this study, *F. fletcheri*, is univoltine at this latitude (Rango 1999). Adult emergence and mating occur during late spring and summer (Rango 1999, Krawchuk & Taylor 2003). Females are viviparous and deposit only one larva per leaf because larvae exhibit strong cannibalistic behaviour (Forsyth and Robertson 1975).

During early fall, larvae exit the pitchers and move to the surrounding moss, where they pupate and enter the overwintering diapause (Dahlem & Naczi 2006). Population genetic diversity and structure have never been investigated in this species.

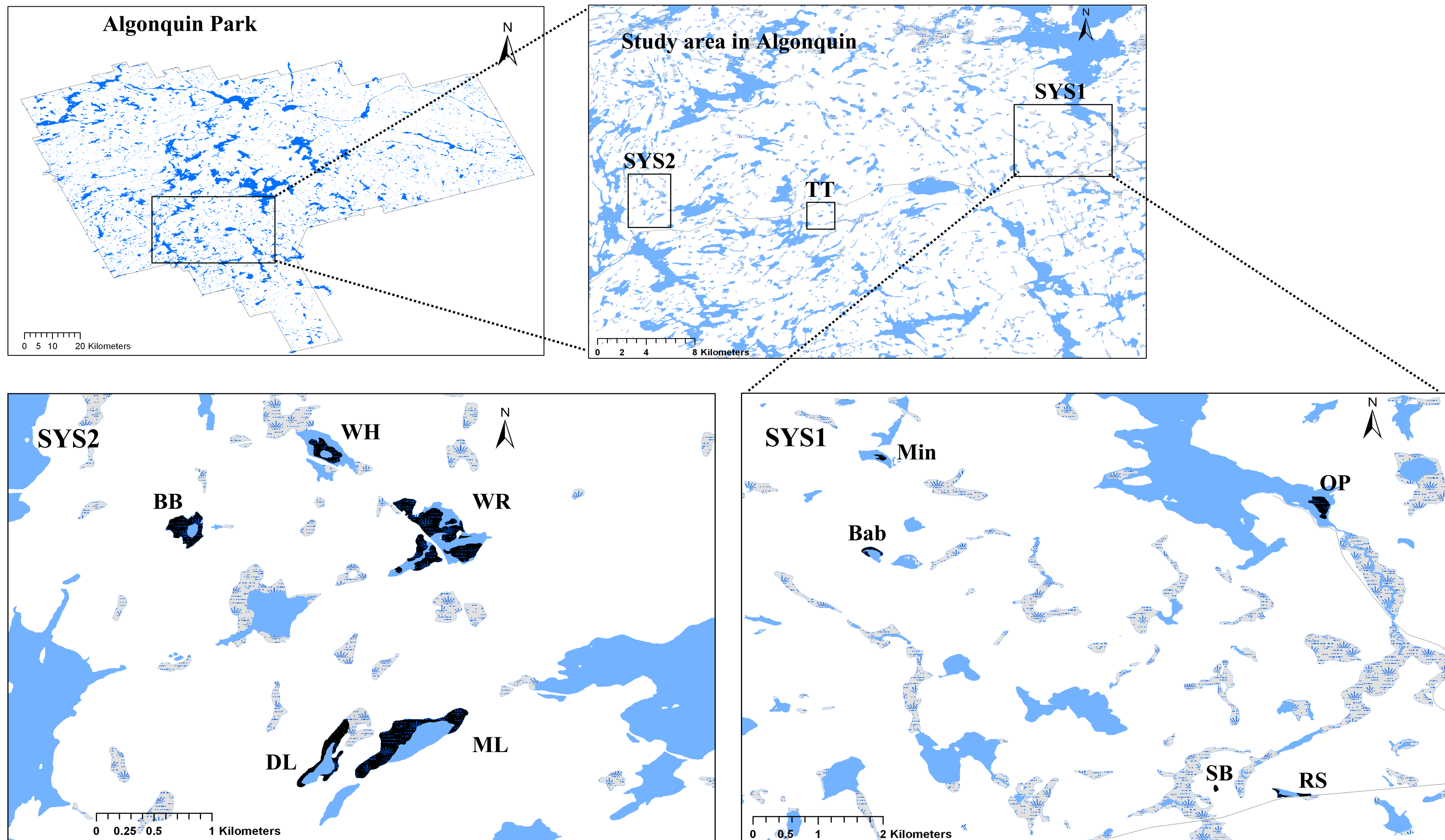


Figure 3.1 Algonquin Park (ON, CA) and study area with Track&Tower bog and two systems of bogs (SYS1 & SYS2). *F. fletcheri* larvae were sampled from Track&Tower (TT) bog and five bogs in each system (represented as dark surfaces and coded as: SB, RSB, Bab, Min, OP in SYS1; DL, ML, BB, WH, WR in SYS2) in August 2008 and August 2009.

3.2.2 Sampling and genotyping

Bogs in Algonquin Provincial Park were sampled during August 2008 and August 2009, with six of the 11 bogs being sampled in both years (Table 3.2). Larvae from the 'outgroup' Sifton bog (London, ON, Canada) were collected in August 2009. The locations of all sampled larvae were recorded to within 0.5 m using a high accuracy GPS receiver (Trimble GeoXH, Sunnyvale, CA).

Larvae were removed from the pitchers using plastic pipettes and were placed individually in absolute ethanol at -20ºC until the DNA was extracted. I used the DNeasy blood and tissue kit (QIAGEN, Germantown, MD) to extract genomic DNA from each individual and genotyped them at 12 microsatellite loci developed specifically for this species (GenBank accession numbers: GQ300842-GQ300853, Rasic & Keyghobadi 2009). I amplified these loci using the protocols detailed in Chapter 2. Sizing of PCR products was done on a 3730 genetic analyzer using Genemapper software (Applied Biosystems, Carlsbad, CA) with LIZ-500 size standard.

The total data set consisted of 12 loci scored in 670 individuals from 12 bogs sampled over two years (2008 and 2009) (Table 3.2).

3.2.3 Microsatellite variation

Genotypic data were initially tested for the presence of null alleles and other scoring errors using Micro-checker version 2.2.3 (van Oosterhout et al. 2004). Genetic variation in each bog was assessed using: allelic richness averaged over loci and corrected for the sample size (A), expected heterozygosity (H_E), and fixation index (F_{IS}; tested for significant deviation from zero using 1000 permutations) in FSTAT version 2.9.3 (Goudet 2001). This program was also used to test for the presence of linkage disequilibrium for all possible pairs of loci in each bog sample, and globally for each pair of loci across all bogs.

3.2.3.1 *Genetic differentiation*

To determine how neutral genetic variability was partitioned among and within samples collected in Algonquin Park, I employed Analysis of Molecular Variance (AMOVA) in GenAlEx (Peakal & Smouse 2006). This procedure estimated the hierarchical partitioning of genetic diversity (1) among regions (two systems of bogs and the intermediate TT bog), (2) among bogs within a system, and (3) within bogs.

Differences in allele frequencies between bogs, as well as between temporal samples from the same bog, were tested using an exact probability test in GENEPOP (Raymond & Rousset 1995). Differentiation among samples within systems (SYS1 and SYS2) was also described using F_{ST} (Wright 1951), which was calculated across all bogs (global F_{ST}) and pairs of bogs in a system (pairwise F_{ST}) (Weir & Cockerham 1984). Bootstrapping was applied over loci to produce 95% confidence intervals for F_{ST} values. Significance of F_{ST} was tested with 1000 permutations of genotypes. I also calculated F_{IT} for each system as a measure of the deviation from Hardy-Weinberg expectations for a 'total sample population' in each system and all 11 bogs combined into a single sample.

Population genetic structuring was also evaluated with the individual-based clustering method of Pritchard et al. (2000) in STRUCTURE 2.3.3, which uses a Bayesian approach to detect potential genetic structure without assuming such a pattern *a priori*. The method assigns individuals to a user-defined number of genetic clusters (K), in such a way as to minimize the departures from the Hardy-Weinberg expectations and linkage equilibrium. This method uses a Markov Chain Monte Carlo procedure to estimate the log probability of data P(X|K) for each value of K, and it also calculates the proportion of membership (Q) to each cluster (K), for each individual. I implemented the procedure without any prior information on origin or sampling location of the individuals, in order to avoid bias in estimating K under a prior that could potentially be incorrect. The results were reported for the following parameter settings: admixture model, correlated allele frequencies among populations, a burn-in period of 10000 steps, and a chain length of 10^5. Different runs with larger or smaller number of burn-in steps and chain-lengths were initially examined to establish the appropriate values that led to convergence of model results and consistency between runs. The calculations were performed for each K

between one and 11, with five runs per K. The optimal value of K was estimated: (i) based on the highest value for ln likelihood of K (lnP(K)), and (ii) following deltaK method by Evanno et al. (2005) in Structure Harvester version 0.6.6 (Earl 2011).

3.2.3.2 *Spatial genetic structure*

Spatial genetic structure was assessed both within a system of bogs and among all Algonquin Park sites. First, I wanted to determine if the observed genetic structuring among bogs can be partly explained by spatially limited gene flow. The pattern that results from such process is known as isolation-by-distance (IBD), defined as a decrease in the genetic similarity among subpopulations as the geographic distance between them increases (Wright 1943). I tested for the presence of IBD by estimating the correlation between the matrix of transformed genetic distances between pairs of bog (F_{ST} / (1−F_{ST})) and the matrix of log transformed geographic distances. Geographic distances were computed as the minimal Euclieean distances between the edges of bogs in ArcGIS 9.3 from 30-m resolution vector maps (Wetland class from Land Cover, Circa 2000, Agriculture and Agri-Food Canada) and Google Earth images. In this analysis available temporal samples for six bogs were pooled. The significance, based on the Z statistics for the correlation between the two matrices, was estimated using a Mantel test with 10000 permutations in the program IBDWS (Bohonak 2002).

I also employed a spatial autocorrelation analysis that is based on the genetic distance between pairs of individuals and is informative about the spatial extent of recent gene flow within a system of bogs. An autocorrelation coefficient (*r*) was calculated in GenAlEx (Peakall & Smouse 2006) using the matrices of pairwise geographic distances and pairwise squared genetic distances for codominant data (following Smouse & Peakall 1999). The autocorrelation coefficient was calculated for ten distance classes, providing a measure of the genetic similarity between pairs of individuals whose geographic separation falls within the specified distance class. Pairwise individual-by-individual geographic distances were calculated from the UTM coordinates of each larva. To increase the robustness of results, I defined distance classes such that they contained even sample sizes (a minimum of 1300 pairwise comparisons per class). Given that the method calculates individual-by-individual genetic distances, I performed the analysis among

individuals collected within a single year (2009), for each system separately. Spatial genetic autocorrelograms were then created by plotting the calculated autocorrelation coefficients *r* as a function of distance. The statistical significance of each autocorrelation coefficient (*r*) was tested with 999 permutations. Because individuals were sampled as larvae, the analysis ultimately reflects the dispersal behaviour of gravid females in the previous generation.

3.2.3.3 *Estimates of effective population size*

I used three different methods to calculate effective population sizes (N_e) within bogs. Method 1 employs approximate Bayesian computation to estimate N_e given summary statistics from a single sample of genotypes, in program ONeSAMP (Tallmon et al. 2008). The program generates 50000 simulated populations drawn randomly from the distribution of user-defined N_e priors. Samples are drawn from each simulated population so that they have the same size and number of loci as the actual dataset. Summary statistics are calculated and compared to the actual dataset, and simulated populations with summary values similar to the actual population are retained to generate a point estimate of N_e using weighted local regression (Tallmon et al. 2008). Lower and upper bounds on the prior for N_e were set at two and 1000, respectively.

For the six bogs that had temporal samples, I used program MLNE 2.3 (Wang & Whitlock 2003) that employs a maximum likelihood method to estimate N_e based on temporal sampling. This method either assumes that allele frequencies change randomly over time in a population isolated from the potential source of immigrants ($N_{e_closed,}$ Method 2), or allele frequencies converge to the source when immigration occurs ($N_{e_open,}$ Method 3). For the latter model, the allele frequencies from the source population are needed. I defined a source population as a pool of all samples collected in 2008 and 2009 from all bogs in the same system (analogous to Jehle et al. 2010, Fraser et al. 2007). When calculating $N_{e_open,}$ the program simultaneously calculates per-generation immigration rate (m). I applied a maximum N_e of 1000 (as in ONeSAMP, Method 1)

3.2.3.4 *Bottleneck analysis*

BOTTLENECK 1.2 (Cornuet & Luikart 1996) was used to test if samples from any bogs exhibited a signal of a recent bottleneck event. The program tests for recent population size reductions using allele frequency data, under the assumption that such an event leads to a disproportionate reduction in allelic diversity relative to heterozygosity. Thus, the software tests for the observed heterozygosity that is larger than the heterozygosity expected given the observed number of alleles at a locus at mutation-drift equilibrium. Data were tested under all three microsatellite mutation models: the Infinite Allele Model (IAM), the step-wise mutation model (SMM), and the two-phase model (TPM).

3.3 Results

3.3.1 Microsatellite variability

The initial analyses in Micro-checker (van Oosterhout et al. 2004) detected excess homozygosity at loci FF217 and FF104 consistently across samples, which indicated the presence of a null allele at each of these loci. The estimated frequencies were as high as 0.36 and null homozygote individuals were detected. Therefore, these two loci were excluded from any further analysis.

None of the six bogs in which temporal samples of *F. fletcheri* were taken showed significant changes in allele frequencies over the two years (P = 0.180-0.550 for exact probability tests). Therefore, I present results from the analyses that included pooled temporal samples, unless stated differently.

I detected a moderate level of genetic diversity at the remaining ten loci within Algonquin Park samples, with average allelic richness ranging between 3.88 and 4.42, and average gene diversity between 0.45 and 0.52 (Table 3.2). The two systems of bogs and the TT bog did not significantly differ in any of these measures. The highly isolated, outlier bog (Sifton) exhibited lower diversity however: two out of the ten tested loci were monomorphic, and mean allelic richness and gene diversity of the remaining eight loci were 3.16 and 0.43, respectively (Table 3.2). Permutation test for F_{IS} showed significant

deviation from zero (i.e., heterozygote deficit, $P < 0.001$) in three bog samples (DL, BB, TT). However, after Bonferroni correction for multiple comparisons, the corresponding F_{IS} values were not significantly different from zero at the 5% nominal level (Table 3.2). Lack of any consistent heterozygote deficit at the bog scale indicated little genetic structure within bogs and confirmed the bog as an appropriate unit of analysis. Significant linkage disequilibrium was not detected for any pairs of loci in any of the samples

Table 3.2 Genetic diversity measures averaged over 10 microsatellite loci for *Fletcherimyia fletcheri,* from 11 bogs in Algonquin Provincial Park, Canada, grouped in regions (SYS1, SYS2, TT), and one isolated bog (Sifton) in London, Canada: sample size (temporal samples from 2008, 2009 indicated in brackets) (n), allelic richness calculated from 26 individuals (A), gene diversity (H_E), inbreeding coefficient (F_{IS}), global F_{ST} and F_{IT} for systems (SYS) and all 11 bogs (Total). 95%CI were calculated using bootstrapping over loci. * designates significant heterozygote deficit before the correction for multiple tests.

Region	Bog	code	n		A	H_E	F_{IS}	SYS F_{ST} (95%CI)	SYS F_{IT} (95%CI)	Total F_{ST} (95%CI)	Total F_{IT} (95%CI)
	Spruce Bog	SB	66	(30,36)	4.42	0.511	-0.004				
	Opeongo Lake	OP	61	(43,18)	3.96	0.502	0.045	0.020	0.044		
SYS1	Roadside Bog	RSB	39		3.88	0.450	0.071	(0.016-0.025)	(0.012-0.067)		
	Bab Lake	Bab	39		3.99	0.490	-0.002				
	Minor Lake	Min	40		3.99	0.465	0.021			0.017	0.061
TT	Track & Tower	TT	27		4.27	0.509	0.137*			(0.013-0.020)	(0.027-0.094)
	Dizzy Lake	DL	85	(41,44)	4.32	0.506	0.071*				
	Mizzy Lake	ML	75	(34,41)	4.48	0.493	0.027	0.011	0.062		
SYS2	West Rose	WR	53		4.05	0.498	0.087*	(0.006-0.015)	(0.023-0.104)		
	Buggy Bog	BB	72	(30,42)	4.41	0.522	0.011				
	Wolf Howl	WH	79	(36,43)	4.31	0.506	0.068				
outgroup	Sifton bog	SIF	34		3.16	0.428	0.063				

3.3.2 Genetic differentiation

The hierarchical AMOVA revealed that, for the Algonquin Park samples, 98.1 percent of neutral genetic variation was contained within bogs and 1.5 percent among bogs within a system, leaving 0.4 percent of variation partitioned among systems of bogs and the intermediate TT bog (Table 3.3). All corresponding hierarchical fixation indices were significantly larger than zero (F_{RT} = 0.004, F_{RS} = 0.015, F_{ST} = 0.018; P = 0.01 for all), indicating overall significant differentiation among bogs and regions.

The exact probability test showed significant differences in allele frequency distributions among bogs within a system (Fisher's method χ^2 = infinity, P < 0.001 in both systems). Global F_{ST} was estimated at 0.020 (95% Confidence Interval: 0.016-0.025) in SYS1, and 0.011 (95%CI: 0.006-0.015) in SYS2 (Table 3.2). Pair-wise F_{ST} values ranged between 0.007 and 0.030 in SYS1, and between 0.004 and 0.025 in SYS2, and were significantly greater than zero for all but two pairs of bogs at the adjusted nominal level of 5% (Table 3.4). F_{IT} was estimated at 0.044 (95%CI: 0.012-0.067) and 0.062 (95%CI: 0.023-0.104) in SYS1 and SYS2 respectively. For all 11 Algonquin Park bogs together, global F_{ST} was 0.017 (95%CI: 0.013-0.020), and F_{IT} was 0.061 (95%CI: 0.027-0.093).

The cluster analysis performed in STRUCTURE indicated very weak (if any) differentiation among the 11 Algonquin Park sites. Namely, the highest value of lnP(K) was for K=1, but the highest deltaK value was for K=3 (Table 3.5), leaving the inference of true K ambiguous. When K was set to equal 3, for each individual the proportion of membership (Q) to each genetic cluster was always nearly equal (Q~1/K). The assignment of individuals was thus not dependent on their sample of origin. When the individuals from the outlier bog (Sifton) were included in the analysis, they were always assigned to a distinct genetic cluster under different K scenarios, as expected (Figure 3.2). However, Algonquin Park individuals still showed highly mixed assignment probabilities among the remaining genetic clusters.

Table 3.3 Results of analysis of molecular variance (AMOVA) for *Fletcherimyia fletcheri* from 11 bogs in Algonquin Park (ON, Canada) grouped by region (SYS1, SYS2 and TT bog). Temporal samples (taken in 2008 and 2009) are pooled.

Source	df	SS	MS	Est. Var.	%
Among Regions (SYS1, SYS2, TT)	2	18.192	9.096	0.010	0.4%
Among Bogs	8	56.349	7.044	0.038	1.5%
Within Bogs	1261	3171.320	2.515	2.515	98.1%
Total	1271	3245.862		2.562	100.0%

F Statistic	Value	P(rand ≥ data)
F_{rt}	0.004	0.010
F_{sr}	0.015	0.010
F_{st}	0.018	0.010

F_{rt} = AR / (WB + AB + AR) = AR / TOT
F_{sr} = AB / (WB + AB)
F_{st} = (AB + AR) / (WB + AB + AR) = (AB + AR) / TOT
Key: AR = Est. Var. Among Regions, AB = Est. Var. Among Bogs, WB = Est. Var. Within Bogs

Table 3.4 Differentiation between pairs of bogs based on pairwise F_{ST} values (Weir and Cockerham 1984) for 11 Algonquin Park bogs. Shaded area designates pairwise comparisons for bogs within SYS1 (upper) and SYS2 (lower). Italicized values indicate **non**-significant F_{ST} at the adjusted nominal level of 5%.

	SB	RSB	Bab	Min	OP	TT	DL	ML	WR	BB	WH
SB	*										
RSB	0.0154	*									
Bab	0.0064	0.0264	*								
Min	0.0275	0.0280	0.0197	*							
OP	0.0148	0.0197	0.0215	0.0295	*						
TT	0.0252	0.0256	0.0359	0.0428	0.0244	*					
DL	0.0086	0.0108	0.0113	0.0315	0.0206	0.0358	*				
ML	0.0227	0.0139	0.0223	0.0380	0.0130	0.0316	0.0048	*			
WR	*0.0043*	0.0127	0.0162	0.0249	0.0106	0.0159	0.0139	0.0249	*		
BB	0.0119	0.0182	0.0116	0.0245	0.0133	0.0237	*0.0035*	0.0095	0.0087	*	
WH	0.0137	0.0248	0.0125	0.0299	0.0088	0.0281	0.0143	0.0160	0.0114	0.0063	*

Table 3.5 Results of STRUCTURE analysis (Pritchard et al. 2000) and determination of number of genetic clusters in Algonquin Park samples based the statistic Delta K (Evanno et al. 2005) performed in STRUCTURE HARVESTER (Earl 2011). K is the number of clusters used in the STRUCTURE simulation. Mean and standard deviation for the probability of data given K (LnP(K)) were calculated from 5 runs. The most likely number of genetic clusters K=3 is shaded.

K	Runs	Mean LnP(K)	Stdev LnP(K)	Delta K
1	5	-12651.74	0.2701	—
2	5	-12836.96	67.8665	2.4943
3	5	-12852.90	37.1399	13.6489
4	5	-13375.76	407.2431	0.2561
5	5	-13794.34	541.4822	0.5702
6	5	-14521.70	2082.9188	0.0170
7	5	-15213.64	1194.8359	0.6115
8	5	-15174.90	1514.7263	1.8624
9	5	-17957.18	3644.3255	0.6690
10	5	-18301.32	1529.1741	0.3262
11	5	-19144.32	4351.5456	—

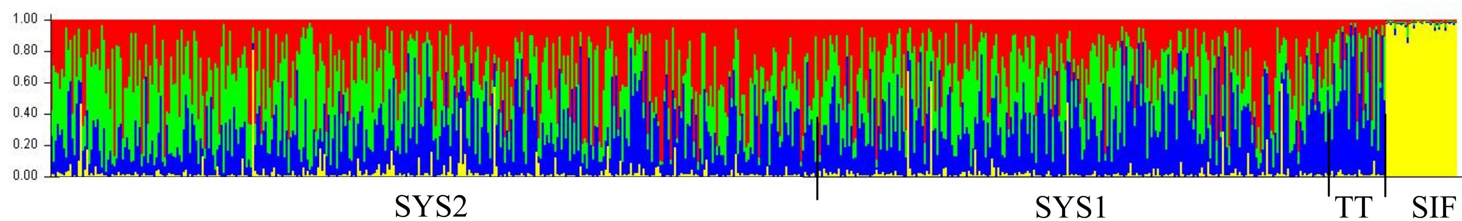


Figure 3.2 STRUCTURE analysis for *F. fletcheri* samples. Representative display of assignment probabilities for individuals to four genetic clusters for samples from Algonquin Park and Sifton bog. Each individual is represented by a thin vertical line. Each vertical line for each individual shows the proportion of ancestry in each of the four clusters. Labels on the x-axis represent sampled locations for individuals from bogs in SYS2, SYS1, TT and Sifton bog.

3.3.3 Spatial genetic structure

Isolation-by-distance (IBD) was not significant among bogs within either system (r^2 = 0.033, P = 0.247 in SYS1; r^2 = 0.165, P = 0.083 in SYS2). When TT bog was included in the analysis with SYS2, IBD was highly significant and a high proportion of genetic variation was explained by the simple straight-line distances of up to 10km (r^2 = 0.668, P = 0.011) (Figure 3.3a). The IBD pattern was marginally significant for bogs from SYS1 and TT bog, that are up to 16km apart (r^2 = 0.343, P = 0.058) (Figure 3.3b). When all 11 Algonquin Park bogs were used for the analysis, the IBD pattern was highly significant, although only a small percentage of variation in genetic differentiation was explained by the straight-line geographic distance (r^2 = 0.064, P = 0.008) (Figure 3.3c).

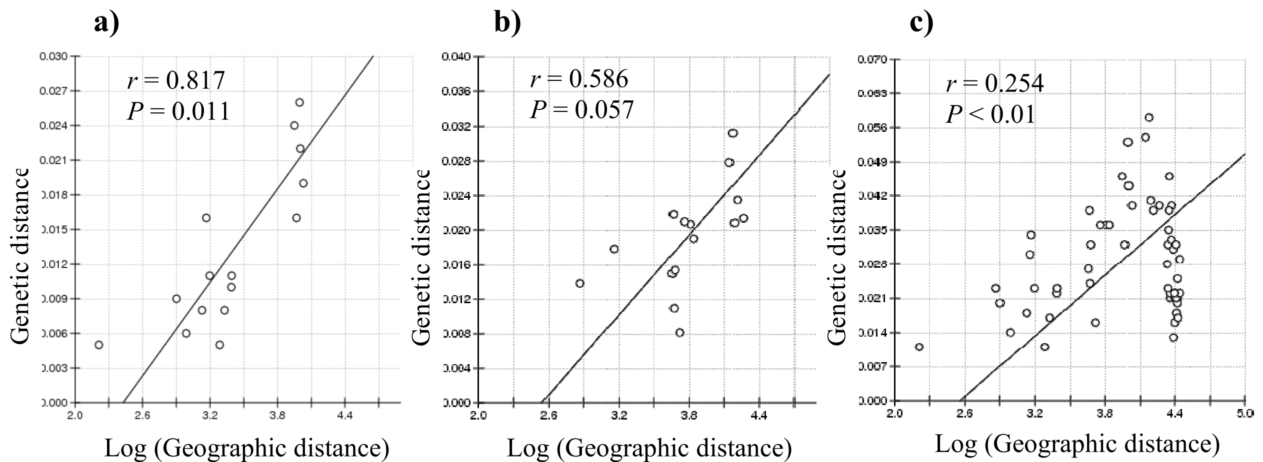
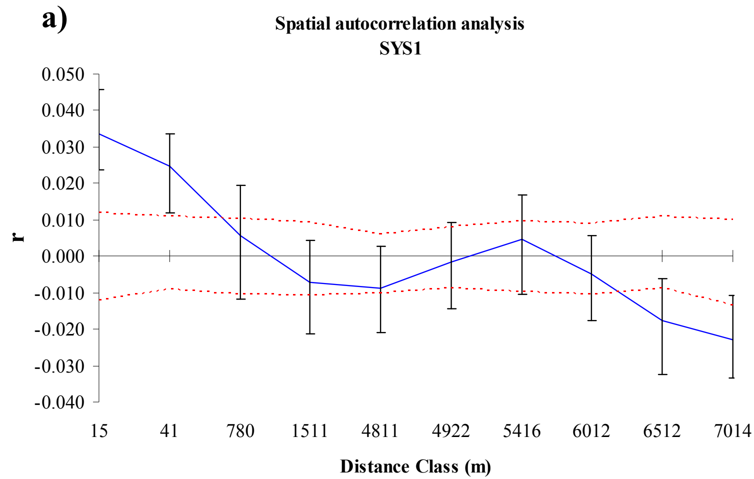


Figure 3.3 Correlation between pairwise genetic distances ($F_{ST} / (1-F_{ST})$) and log geographic distances for *F. fletcheri* samples from bogs in a) SYS2 and TT, b) SYS1 and TT, c) the entire examined area in Algonquin Park (all 11 bogs).

Spatial autocorrelation analysis, based on the genetic distances between pairs of individuals sampled within a single year (2009) within each Algonquin Park system, revealed significant genetic structure at a small spatial scale. Correlograms revealed significant positive spatial autocorrelation (P = 0.01) at distance classes within a bog (Figure 3.4a,b). Correlation coefficients (r values) values were significant and positive at some between-bog distances (1.3 km) in SYS2, while, there was no significant positive autocorrelation among samples from different bogs in SYS1. The x-axis intercept was detected at 1.1 km in SYS1 and 1.5 km in SYS2, and r values became significantly negative at distance classes beyond 6 km in SYS1 and at 1.8 km in SYS2. Significant positive r at smaller distances, accompanied by the negative r at larger distances is a pattern consistent with isolation-by-distance among individuals (Smouse & Peakall 1999).

3.3.4 Effective population size and bottleneck analysis

Methods 1 and 3 produced similar values for N_e and N_{e_open}, respectively, with larger 95% confidence intervals for N_{e_open} (Table 3.6). Estimated mean effective population sizes ranged between 18 and 135 individuals. Estimated means for N_{e_closed} (under the model of bog isolation i.e., no immigration) approached 1000 individuals (i.e., the maximum possible as determined by analysis parameters) in all but two bogs. A signal of a recent bottleneck was not detected in any of the bog samples from Algonquin Park. For comparison, this analysis was also done for the outlier Sifton bog, which might be expected to show evidence of a recent reduction in effective population size given its isolation under anthropogenic disturbance. Indeed, a significant bottleneck signal was detected in this bog, as indicated by the standardized differences test (P = 0.015) and Wilcoxon test for heterozygote excess (P = 0.01) under the IAM assumptions (Cornuet & Luikart 1996).



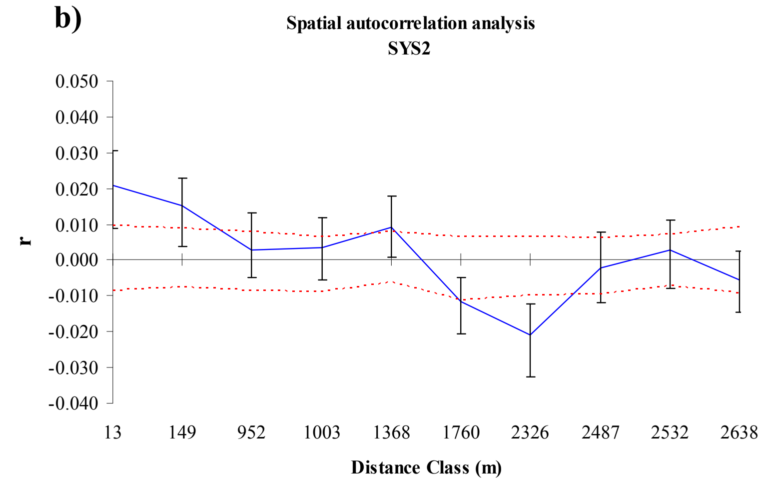


Figure 3.4 Spatial genetic autocorrelograms showing mean correlation coefficients between pairs of *F. fletcheri* individuals (r), plotted against geographic distance classes (meters) in SYS1 (a) and SYS2 (b). Horizontal dashed lines represent critical values under the null hypothesis that genotypes are randomly distributed across a landscape (α=0.05). Error bars represent 95% confidence intervals around each mean correlation coefficient. Data are for individuals sampled in 2009 only.

Table 3.6 Estimated effective population sizes (N_e) and 95%CI for *Fletcherimyia fletcheri* populations from Algonquin Park (ON, Canada), calculated using ONeSAMP (Method 1, Tallmon et al. 2008), MLNE assuming no migration (Method 2, Wang & Whitlock 2003), and MLNE assuming migration (Method 3, Wang & Whitlock 2003). **m** represents per-generation immigration rate, estimated jointly with N_{e_open} for available

	Method 1		**Method 2**		**Method 3**		
Bog	N_e	**95% CI**	N_{e_closed}	**95% CI**	N_{e_open}	**95% CI**	**m**
DL	62	41-188	182	44-1000	47	26-125	0.61
ML	111	76-371	1000	138-1000	135	48-1000	0.32
WR	48	32-124	-	-	-	-	
BB	44	32-102	1000	78-1000	72	32-394	0.67
WH	52	36-144	1000	79-1000	80	33-492	0.44
SB	42	30-110	1000	69-1000	71	30-434	0.55
OP	18	14-31	351	33-1000	42	19-346	0.43
RSB	39	28-116	-	-	-	-	
Bab	45	33-114	-	-	-	-	
Min	71	48-187	-	-	-	-	

3.4 Discussion

Analysis of genetic variation within and among bogs using ten microsatellite loci indicates that the spatial genetic structure in the pitcher plant flesh fly generally follows the patchy population model within bogs, while among bogs it generally follows the metapopulation model with some patchy population characteristics. Individual bogs contained moderately high genetic diversity and did not deviate from Hardy-Weinberg expectations. Consistent with the predictions of the metapopulation model, differentiation among bogs in Algonquin Park was small but significant (95%CI global $F_{ST} > 0$ in both systems, Table 3.2, pair-wise F_{ST} Table 3.4), leading to a significant heterozygote deficiency in the total population, even within systems (95%CI $F_{IT} > 0$ in both SYS1 ad SYS2, Table 3.2). Furthermore, limited dispersal and gene flow among closely situated bogs (i.e., within a system) was detected with the spatial autocorrelation analysis, and a highly significant pattern of isolation-by-distance was detected among bogs at larger spatial scales (Mantel test $P < 0.01$). However, the prediction of extinction/recolonization patterns was not supported in my study, as none of the bogs from Algonquin Park showed detectable signals of a genetic bottleneck.

A small mark-recapture experiment in *F. fletcheri* that did not extend beyond bog's edge estimated a maximum dispersal distance of 184 m, but suggested a potential for much larger dispersal distances (Krawchuk & Taylor 2003). However, a negative relationship between bog isolation and larval abundance in the same study also indicated that the distance among bogs could still restrict *F. fletcheri* movement (Krawchuk & Taylor 2003). Genetic results from my study are congruent with, and indeed reconcile, these ecological findings. Correlograms were significant in both systems and showed a trend consistent with isolation-by-distance between individuals in a single year of collection (Figures 3.4a,b). Positive spatial autocorrelation in SYS2 suggested a high level of adult female dispersal, and hence recent gene flow, up to 1.3 km (Figure 3.4b). This is comparable to dispersal capacity of the tsetse fly (*Glossina palpalis gambiensis*), for which both microsatellite and mark-release-recapture data produced estimates of mean dispersal distance between 1.2-2.4 km per generation (Bouyer et al. 2009).

The relationship between genetic and geographic distance among bogs was highly significant at larger spatial scales (10-26 km, Figure 3.3), but not within systems (~7 km separation). Lack of IBD among bogs at a smaller spatial scale (i.e., within systems) may be explained by a small number of samples in the analysis (five bogs per system), or absence of local drift/gene flow equilibrium. The most likely explanation however is that although dispersal of individual flies may be limited on a contemporary time scale (as seen in the spatial autocorrelation analysis), gene flow among bogs within the systems remains sufficiently high when averaged over a large number of generations to prevent the formation of IBD. In the western cherry fruit fly (*Rhagoletis indifferens)*, IBD among individuals was similarly significant at very small spatial scales (< 1 km), while IBD among samples was only detected at scales greater than 20 km (Maxwell et al. 2011). This was interpreted as reflecting substantial gene flow at scales of up to 20 km, likely maintained by large population sizes and stepping-stone gene flow, despite limited dispersal distances of individual flies.

Given the existence of IBD and small F_{ST} values, it is not surprising that the STRUCTURE algorithm gave ambiguous results: no structure (K=1) according to the maximum lnP(K) criterion, or three genetic clusters (K=3) according to the deltaK criterion (Table 3.5) in the Algonquin Park samples. The assignment probabilities for *F. fletcheri* individuals were almost equally distributed among any given K, making the interpretation of the true extent of differentiation among bogs in Algonquin Park challenging. The underlying STRUCTURE model is not well suited to data under the scenario of IBD, where it is expected that most individuals have mixed membership in multiple groups (Pritchard et al. 2010). Therefore, based on the STRUCTURE results, it is difficult to differentiate between the patchy and metapopulation models of population structure in *F. fletcheri.*

Significant gene flow among nearby bogs was also indicated by the estimates of effective population sizes. Joint short-term temporal estimates of N_{e_open} and m (Method 3) revealed high immigration rates into a bog (m = 0.32-0.67) and mean values for N_{e_open} that were comparable to the mean one-sample point estimates of N_e (Method 1) (Table 3.6). When migration was ignored in the short term temporal estimates (Method 2), mean

N_e values across populations were substantially higher (N_{e_closed} = 182-1000). Discrepancies between N_{e_open} and N_{e_closed} have been found in several different studies, where N_{e_open} is always lower than N_{e_closed} in the order of 1.4 to 87 times (Ford et al. 2004, Hoffman et al. 2004, Johnson et al. 2004, Consuegra et al. 2005, Jensen et al. 2005, Saillant & Gold 2006, Fraser et al. 2007, Watts et al. 2007). An explanation for such a consistent direction in the ratio of the two estimates across different systems can be related to the fact that temporal allele frequency changes within populations are significantly affected by immigration (Fraser et al. 2007). Specifically, higher rates of immigration can compensate the effect of drift in the short-term, leading to the overestimation of N_{e_closed} (Fraser et al. 2007). In the long term, constant migration and drift approach an equilibrium at which allele frequency changes in a population reflect such changes in the entire metapopulation, leading again to an overestimate of population N_e if migration is ignored (Wang & Whitlock 2003). Thus, the higher N_{e_closed} estimates in this and other studies imply that the migration into subpopulations (bogs in this case) is often high enough, and genetic differentiation from source populations is low enough, to lead to a substantial overestimation of N_e if migration is ignored (Fraser et al. 2007).

Populations of *F. fletcheri,* in the Algonquin Park study area at least, appear not to undergo any 'turnover' at the bog scale. Stable local dynamics of *F. fletcheri* populations is perhaps not surprising, given the distinct life history characteristics of the species. This insect has a *K* reproductive strategy, with a fecundity up to 17 times lower than in other sarcophagids (Forsyth & Robertson 1975). On average, females produce only 10 larvae that are territorial and cannibalistic, requiring that females leave a single larva per leaf (Forsyth & Robertson 1975), while choosing fresh and large leaves (Krawchuk & Taylor 2003). Bogs in Algonquin Park provide stable and abundant breeding habitat for *F. fletcheri* larvae, which occupied between 2.3-6.0% of the inspected leaves. These percentages appear very consistent in this species, as Krawchuk & Taylor (2003) detected larvae in 5% of the sampled leaves within bogs in Newfoundland (Canada). Investment into a few offspring, larval requirements well below the habitat's carrying capacity, and a relatively unvarying environment may all contribute to stable population dynamics in *F. fletcheri*. Indeed, the continued persistence of *F. fletcheri* in highly isolated Sifton bog is a testament to the stability of local populations in this species.

Overall, among bogs, the pitcher plant flesh fly exhibits metapopulation characteristics of significant spatial genetic structuring, and limited dispersal and gene flow, but populations do not experience frequent local extinction/recolonization. Few empirical studies identify examples that fit the classical view of metapopulations as groups of populations persisting in a balance between local extinction and recolonization (Harrison 1991). In practice, it is very difficult to to draw a sharp distinction between metapopulations with true local extinction, and patchy populations in which extinction is absent or unimportant (Harrison 1991). From the point of view of regional dynamics, it is not the variation in patch or population size *per se* that is significant so much as the variation in the persistence of local populations (Harrison 1991). Patchy population dynamics arise when dispersal takes place on a spatial scale greater than that of the local events causing population fluctuations and patches are thus united into a relatively persistent population in which there is little potential for local extinction (Harrison 1991). In the case of the pitcher plant flesh fly there seems to be little variation in the persistence of local populations at the bog scale, hence despite some limitation on dispersal, its regional population dynamics may be more akin to those of patchy populations. The future task remains to tease apart the importance of the 'rescue' effect of moderately high levels of gene flow among bogs from their inherent stability in the abundant and stable habitat.

It is important to consider carefully which criterion to adopt (ecological and/or genetic) when classifying a patchily distributed species, given a predominant tendency to declare all such cases as 'metapopulations' (Mayer et al. 2009). Part of the problem lies in a common focus on spatial patterns, instead of processes that shape the patterns (Sutcliffe et al. 1997). When both are considered, as with the analysis of genetic data that allows for the inference of different underlying processes that shape the spatial pattern of genetic diversity, the clear-cut distinction between different population models becomes more challenging (Mayer et al. 2009). The pitcher plant flesh fly provides another example of such a challenge, increasingly noted in empirical studies (Harrison 1991, Sutcliffe et al. 1997).

The knowledge of the spatial population structure of *F. fletcheri* obtained in this study provides a crucial initial step in designing future landscape genetic studies in this species. An understanding of spatial scales of dispersal and gene flow in particular determines the appropriate spatial scales for various analyses. Although individual-based genetic analyses in this species are likely to be informative among closely spaced bogs, analyses based on sample allele frequencies, such as using traditional genetic distances and *F*-statistics, are more appropriately applied for larger spatial scales (> ~10 km). Furthermore, given that bogs in Algonquin Park did not show notable differences in genetic diversity measures, such as allelic richness, gene diversity, and observed heterozygosity, I recommend an analysis of populations from multiple different 'landscapes' that contain bogs with much more variable habitat characteristics, as well as different levels of connectivity. A recent landscape genetic study of gene flow in the American black bear showed that only highly variable landscape features remained supported in landscape genetic models (Short Bull et al. 2011), and the same could hold true for the pitcher plant flesh fly.

3.5 Summary

This study provided novel insights into the population genetic structure of the pitcher plant flesh fly *F. fletcheri*, the largest insect inhabitant of the unique pitcher plant aquatic community. The isolated population model can clearly be ruled out in this insect. Rather, this species displays a patchy population structure within bogs, and a mixture of metapopulation and patchy population attributes among bogs. Among bogs, the majority of genetic characteristics fit the predictions of the metapopulation model, with some aspects conforming to the predictions of the patchy population model, such as stable population dynamics and a high level of gene flow at smaller spatial scales. Future landscape genetic research in *F. fletcheri* should be conducted within and among landscapes that contain bogs with highly variable habitat characteristics and connectivity.

3.6 References

Bohonak AJ (2002) IBD (Isolation By Distance): a program for analyses of isolation by distance. *Journal of Heredity* 93:153-154.

Bouyer J, Balenghien T, Ravel S, et al. (2009) Population sizes and dispersal pattern of tsetse flies: rolling on the river? *Molecular Ecology* 18:2787-2797.

Buckley HL, Miller TE, Ellison AM, Gotelli NJ (2010) Local to continental-scale variation in the richness and composition of an aquatic food web. *Global Ecology and Biogeography* 19:711-723.

Buckley H, Burns JH, Kneitel JM, et al. (2004) Small-scale patterns in community structure of *Sarracenia pururea* inquilines. *Community Ecology* 5:181-188.

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* 144:2001-2014.

Consuegra S, Nielsen EE (2007) Population size reductions in the Atlantic salmon: genetics, conservation and management. In: Verspoor E, Stradmeyer L, Nielsen J (eds.) *The Atlantic salmon: genetics, conservation and management.* Blackwell Publishing Ltd, Oxford.

Dahlem GA, Naczi RFC (2006) Flesh flies (Diptera: Sarcophagidae) associated with North American pitcher plants (Sarraceniaceae), with descriptions of three new species. *Annals of the Entomological Society of America* 99:218-240.

Earl DA (2011) Structure harvester v0.6.1. Available at: http://taylor0.biology.ucla.edu/structureHarvester/ (accessed Feb 2 2011).

Emerson KJ, Merz CR, Catchen JM, et al. (2010) Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Science USA* 107:16196-16200.

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

Ford MJ, Teel D, Van Doornik DM, Kuligowski D, Lawson PW (2004) Genetic population structure of central Oregon Coast coho salmon (*Oncorhynchus kisutch*). *Conservation Genetics* 5:797-812.

Forsyth AB, Robertson RJ (1975) K reproductive strategy and larval behavior of the pitcher plant sarcophagid fly, *Blaesoxipha fletcheri*. *Canadian Journal of Zoology* 53:174-179.

Frankham R, Ballou JD, Briscoe DA (2002) *Introduction to conservation genetics*. Cambridge University Press, Cambridge.

Fraser DJ, Hansen MM, Østergaard S, Tessier N, Legault M, Bernatchez L (2007) Comparative estimation of effective population sizes and temporal gene flow in two contrasting population systems. *Molecular Ecology* 16:3866-3889.

Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from http://www2.unil.ch/popgen/softwares/fstat.htm.

Hanski I (1999) *Metapopulation ecology*. Oxford University Press, New York.

Harrison S (1991) Local extinction in a metapopulation context: an empirical evaluation. *Biological Journal of the Linnean Society* 42:73-88.

Harvey E, Miller TE (1996) Variance in composition of inquiline communities in leaves of *Sarracenia purpurea* L. on multiple spatial scales. *Oecologia* 108:562-566.

Hastings A, Harrison S (1994) Metapopulation dynamics and genetics. *Annual Review of Ecology and Systematics* 25:167-188.

Hoffman EA, Schueler FW, Blouin MS (2004) Effective population sizes and temporal stability of genetic structure in *Rana pipiens*, the northern leopard frog. *Evolution* 58:2536-2545.

Holyoak M, Leibold MA, Holt RD (2005) *Metacommunities: spatial dynamics and ecological communities*. The University of Chicago Press, Chicago, IL.

Hutchinson DW, Templeton AR (1999) Correlation of pairwise genetic and geographic distance measures: inferring the relative influence of gene flow and drift on the distribution of genetic variability. *Evolution* 56:1898-1914.

Istock CA, Weisburg WG (1987) Strong habitat selection and the development of population structure in a mosquito. *Evolutionary Ecology* 1:348-362.

Jehle R, Wilson GA, Arntzen JW, Burke T (2005) Contemporary gene flow and the spatio-temporal genetic structure of subdivided newt populations (*Triturus cristatus, T.marmoratus*). *Journal of Evolutionary Biology* 18:619-628.

Jensen LF, Hansen MM, Carlsson J, Loeschcke V, Mensberg KLD (2005) Spatial and temporal genetic differentiation and effective population size of brown trout (*Salmo trutta,* L.) in small Danish rivers. *Conservation Genetics* 6:615-621.

Johannsen OA (1935) *Aquatic Diptera. Part II. Orthorrhapha - Brachycera and Cyclorrhapha*. Memoir, Cornell University Agricultural Experiment Station 177.

Johnson JA, Bellinger MR, Toepfer JE, Dunn P (2004) Temporal changes in allele frequencies and low effective population size in greater prairie-chickens. *Molecular Ecology* 13:2617-2630.

Krawchuk MA, Taylor PD (1999) Roosting behaviout by *Fletcherimyia fletcheri* (Diptera: Sarcophagidae) in *Sarracenia purpurea* (Sarraceniaceae). *The Canadian Entomologist* 131:829-830.

Krawchuk MA, Taylor PD (2003) Changing importance of habitat structure across multiple spatial scales for three species of insects. *Oikos* 103:153-161.

Levins RA (1970) Extinction. *Lectures on Mathematics in the Life Sciences* 2:75-107.

Maxwell SA, Thistlewood HMA, Keyghobadi N (2011) Microsatellite analysis reveals movement patterns in the western cherry fruit fly, *Rhagoletis indifferens* (Diptera: Tephritidae). Revised manuscript, re-submitted May 13, 2011 to *Molecular Ecology*.

Mayer C, Schiegg K, Pasinelli G (2009) Patchy population structure in a short-distance migrant: evidence from genetic and demographic data. *Molecular Ecology* 18:2353-2364.

Mouquet N, Daufresne T, Gray SM, Miller TE (2008) Modelling the relationship between a pitcher plant (*Sarracenia purpurea*) and its phytotelma community: mutualism or parasitism? *Functional Ecology* 22:728-737.

Peakall R, Smouse PE (2006) GenAlEx 6: Genetic Analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6:288-295.

Pritchard JK, Wen X, Falush D (2010) Documentation for structure software: version 2.3. http://pritch.bsd.uchicago.edu/structure_software/release_versions/v2.3.3/structure.

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

Rango JJ (1999) Resource dependent larviposition behavior of a pitcher plant flesh fly, *Fletcherimyia fletcheri* (Aldrich) (Diptera: Sarcophagidae). *Journal of the New York Entomological Society* 107:82-86.

Rasic G, Maxwell SA, Keyghobadi N (2009) Characterization of microsatellite loci for the pitcher plant midge, *Metriocnemus knabi* Coq. (Diptera: Chironomidae). *Molecular Ecology Resources* 9:1388-1391.

Rasic G, Keyghobadi N (2009) Microsatellite loci characterization in the pitcher plant flesh fly, *Fletcherimyia fletcheri* Aldrich (Diptera: Sarcophagidae). *Molecular*

Ecology Resources 9:1460-1466 (part of "Permanent Genetic Resources added to *Molecular Ecology Resources* Database 1 May 2009–31 July 2009").

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

Saillant E, Gold JR (2006) Population structure and variance effective size of red snapper (*Lutjanus campechanus*) in the northern Gulf of Mexico. *Fishery Bulletin U.S.* 104:136-148.

Schnell DE (2002) *Carnivorous plants of the United States and Canada*. Timber Press Inc. Portland, Oregon.

Short Bull RA, Cushman SA, Mace R, et al. (2011) Why replication is important in landscape genetics: American black bear in the Rocky Mountains. *Molecular Ecology* 20:1092-1107.

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

Smouse PE, Peakall R (1999) Spatial autocorrelation analysis of multialleleand multi-locus genetic micro-structure. *Heredity* 82:561-573.

Srivastava DS, Kolasa J, Bengtsson J, et al. (2004) Are natural microcosms useful model systems for ecology? *Trends in Ecology and Evolution* 19:379-384.

Sutcliffe OL, Thomas CD, Peggie D (1997) Area-dependent migration by ringlet butterflies generates a mixture of patchy population and metapopulation attributes. *Oecologia* 109:229-234.

Tallmon DA, Koyuk A, Luikart G, Beaumont MA (2008) ONeSAMP: a program to estimate effective population size using approximate Bayesian computation. *Molecular Ecology Resources* 8:299-301.

Tiner RW (1999) *Wetland indicators: a guide to wetland identification, delineation, classification, and mapping*. Lewis Publishers, Boca Raton, FL.

Trzcinski MK,Walde S, Taylor PD (2003) Colonisation of pitcher plant leaves at several spatial scales. *Ecological Entomology* 28:482-489.

van Oosterhout C, Weetman D, Hutchinson WF (2004) Estimation and adjustment of microsatellite null alleles in nonequilibrium populations. *Molecular Ecology Notes* 6:255-256.

Wahlund S (1928) Composition of populations and correlation appearances viewed in relation to the studies of inheritance. *Hereditas* 11:65-106.

Wang J, Whitlock MC (2003) Estimating effective population size and migration rates from genetic samples over space and time. *Genetics* 163:429-446.

Watts PC, Saccheri IL, Kemp SJ, Thompson DJ (2007) Effective population sizes and migration rates in fragmented populations of an endangered insect (*Coenagrion mercuriale*: Odonata). *Journal of Animal Ecology* 76:790-800.

Weir BS, Cockerham CC (1984) Estimating *F*-Statistics for the analysis of population structure. *Evolution* 38:1358-1370.

Wright S (1951) The genetical structure of populations. *Annual Eugenics* 15:323-354.

Wright S (1943) Isolation by distance. *Genetics* 28:114-138.

Chapter 4. From broad scale patterns to fine scale processes: habitat structure influences genetic differentiation in the pitcher plant midge across multiple spatial scales*

4.1 Introduction

Population genetic data are increasingly analyzed within an explicitly spatial framework as more and more studies, largely in the growing field of landscape genetics, relate the spatial organization of genetic variation to underlying ecological processes and associated landscape and environmental variables (Guillot 2009). Issues of scale surrounding the collection and interpretation of spatial data have been explored in ecological studies for more than two decades (Wiens 1989, Kotliar & Wiens 1990, Holling 1992, Levin 1992, Wu & Loucks 1995, Wagner & Fortin 2005). Landscape and population genetics, however, have only recently seen a strong and growing focus on spatial scale questions (Anderson et al. 2010, Cushman & Landguth 2010, Storfer et al. 2010).

The scale at which samples for genetic analysis are defined and collected is critical in determining the patterns observed and the range of processes about which inferences can be made in population genetic studies (Anderson et al. 2010). Both the extent and the grain of a study are important, where the extent represents the total area of genetic sampling and analysis, while the grain represents the smallest (elementary) sampling unit (Anderson et al. 2010, Cushman & Landguth 2010). We cannot make reliable inferences on patterns and processes beyond the extent of our study, nor detect any elements of a pattern below the grain (Wiens 1989). In gene flow analysis for example, study area size (extent) should be larger than the area occupied by the population of interest and larger than expected dispersal distances, while sampling grain should generally be smaller than the average home-range size or dispersal distance of the study organism (Fortin & Dale 2005, Anderson et al. 2010).

---

* A version of this chapter is in press: Rasic G & Keyghobadi N (2011) From broadscale patterns to fine-scale processes: habitat structure influences genetic differentiation in the pitcher plant midge across spatial scales. *Molecular Ecology.*

Population genetic patterns result from a potentially complex combination of evolutionary, behavioral, ecological and stochastic processes operating at different spatial and temporal scales (Balkenhol et al. 2009, Anderson et al. 2010). Furthermore, ecological processes and environmental variables can influence genetic variation differentially at different spatial scales (Lee-Yaw et al. 2009, Murphy et al. 2010). For example, in the boreal toad, *Bufo boreas*, the amount of precipitation during growing season, temperature and moisture affect genetic connectivity of populations across multiple spatial scales, wheareas habitat permeability is only important at a fine scale (Murphy et al. 2010). Finally, the spatial scales of dispersal and other relevant processes affecting genetic variation may not be known *a priori*, particularly in organisms that are very small or display cryptic behaviours. Thus, there is great value in conducting population and landscape genetic analyses such that multiple spatial scales of sampling are included (Diggle & Ribeiro 2007, Schwartz & McKelvey 2009).

Although there are many reports of genetic structure across more than one spatial scale, the majority of studies include only up to three scales. For example, genetic diversity was examined at: (i) fine, population and regional scale (riparian and mountain) in Manchurian ash *Fraxinus mandshurica* (Hu et al. 2010); (ii) rivers, among rivers and among regions on an island in the riparian plant *Ainsliaea faurieana* (Mitzui et al. 2010); (iii) population, watershed and drainage scales in steelhead *Oncorhynchus mykiss* (Nielsen et al. 2009). Whereas the scales of analysis in these examples reflect natural hierarchies of spatial organization, such as river-watershed-drainage, in many other studies the scales of analysis are apparently arbitrary or based primarily on an anthropocentric perception of nature. In some cases even political boundaries may be used to define scales of sampling (Blanquer & Uriz, 2010, Gonçalves da Silva et al. 2010). Here I take advantage of a unique study system associated with commensal inhabitants of the purple pitcher plant, *Sarracenia purpurea* L., to examine patterns of genetic variation across multiple, objectively defined, nested spatial scales.

The purple pitcher plant *S. purpurea* is found within acidic bogs throuout eastern North America. Like other *Sarracenia* species, it has developed carnivory as an adaptation to the poor nutrient environment. However, the plant's leaves not only are deadly traps for

different arthropods, but also represent the exclusive breeding habitat for the larvae of several insect species (Addicott 1974, Miller et al. 2002, Buckley et al. 2010). For example, larvae of the pitcher plant midge, *Metriocnemus knabi* Coquillett 1904, are usually found at the bottom of the leaf where they feed on the decomposing prey of the plant. Multiple leaves are found in each plant, the plants are distributed in easily identifiable clusters within each bog, and bogs are easily delineated in a landscape. These levels of habitat patches (leaf, plant, cluster of plants, bog, system of bogs) not only represent scales separated by a certain spatial distance ('distance scales'), but also are hierarchically nested ('nested scales'). Thus, the insects that are commensal inhabitants (i.e., 'inquilines') of the purple pitcher plant represent an excellent natural system for ecological and genetic studies across scales. The natural features of the system remove the need for an arbitrary decision on focal scales, because they offer easily detectable habitat patches that are hierarchically nested at several spatial scales. For this reason, the system has been used in landscape ecological studies to understand how local interactions in the pitcher plant communities (Trzcinksi et al. 2005), colonization patterns (Trzcinksi et al. 2003), species distribution (Krawchuk & Taylor 2003) and community composition (Harvey & Miller 1996, Buckley et al. 2010) vary across scales.

My first objective in this study was to examine population genetic structure of one of the pitcher plant's commensal inhabitants, the pitcher plant midge *M. knabi*, across several, nested scales. By considering samples of midge larvae aggregated at each scale in the spatial hierarchy (leaf, plant, cluster of plants, bog, system of bogs), I essentially changed the grain of sampling and analysis while keeping a constant extent that is large relative to the expected dispersal ability of this species (Krawchuk & Taylor 2003). My second objective was to test explicit hypotheses about the effects of landscape variables on genetic structure of the midge across scales. Specifically, results obtained under my first objective suggested that broader scale landscape variables related to habitat amount and isolation may influence spatial patterns of genetic variation observed at both fine (leaf, plant) and broad (cluster, bog) scales. I used distance based redundancy analysis to explicitly test the hypothesis that bog size, plant density, or isolation of clusters within bogs influence patterns of genetic structure at a range of scales. Although there are many potential landscape correlates of genetic structure (e.g., Murphy et al. 2010), I focused on

variables related to habitat amount, patch size and patch isolation because these factors were suggested by my initial analyses of genetic variation across scales in *M. knabi* and because they have previously been shown to influence pitcher plant midge larval densities (Krawchuk & Taylor 2003).

4.2 Materials and Methods

4.2.1 Study area and species

My study sites (bogs) were located in Algonquin Provincial Park, Ontario, Canada (UTM: 17N 687337E 5046853N). The park is in an area of transition between northern coniferous forest and southern deciduous forest. These forests are dominated, respectively, by: (i) white pine (*Pinus strobus*) and red pine (*Pinus resinosa*), poplar (*Populus spp.*) and white birch (*Betula papyrifera*), and (ii) sugar maple (*Acer saccharum*), american beech (*Fagus grandifolia*), hemlock (*Tsuga canadensis*), red oak (*Quercus rubra*), and yellow birch (*Betula alleghaniensis*). Bogs are found within this forest matrix, and many of them contain *S. purpurea* and its associated commensal arthropod inhabitants. Bogs represent peat-covered wetlands in which the vegetation shows the effects of a high water table and a general lack of nutrients. Due to poor drainage and the decay of plant material, the surface water of bogs is strongly acidic. Dominated by sphagnum mosses (*Sphagnum spp.*) and heath shrubs (leather leaf *Chamaedaphne calyculata*, labrador tea *Rhododendron groenlandicum*, cranberries *Vaccinium spp.*), the bogs also contain tamarack (*Larix laricina*) and black spruce (*Picea mariana*) (Tiner 1999).

The focal species in this study, *Metriocnemus knabi* Coq. (Diptera: Chironomidae), is expected to have one generation per year at this latitude (Rango 1999). The midge overwinters as a larva in the leaves of *S. purpurea*. Pupation, adult emergence, mating and oviposition occur during late spring and summer (Heard 1994, Rango 1999, Krawchuk & Taylor 2003). Adult midges exhibit very cryptic behaviour and have body length of only 3mm (Krawchuk & Taylor 2003), which makes it unfeasible to sample them in that life stage. Larvae on the other hand can be readily sampled from within pitchers, which represent the exact spatial locations of maternal oviposition.

4.2.2 Sampling and genotyping

I sampled second instar larvae in August 2009 at five nested spatial scales: leaf, plant, cluster, bog, and system of bogs (Figure 4.1). I balanced sampling so that: (i) each system contained the four closest bogs in a landscape (0.2-7.0 km apart in a mixed forest matrix), (ii) within every bog we randomly selected three clusters containing ten plants, (iii) I randomly chose three plants within each cluster, and (iv) I pipetted all larvae out of the bottom of three randomly selected leaves per plant. Two separate systems of bogs, located 26 km apart, were sampled in this way. The locations of all sampled plants were recorded to within 0.5 m using a high accuracy GPS receiver (Trimble GeoXH, Sunnyvale, CA).

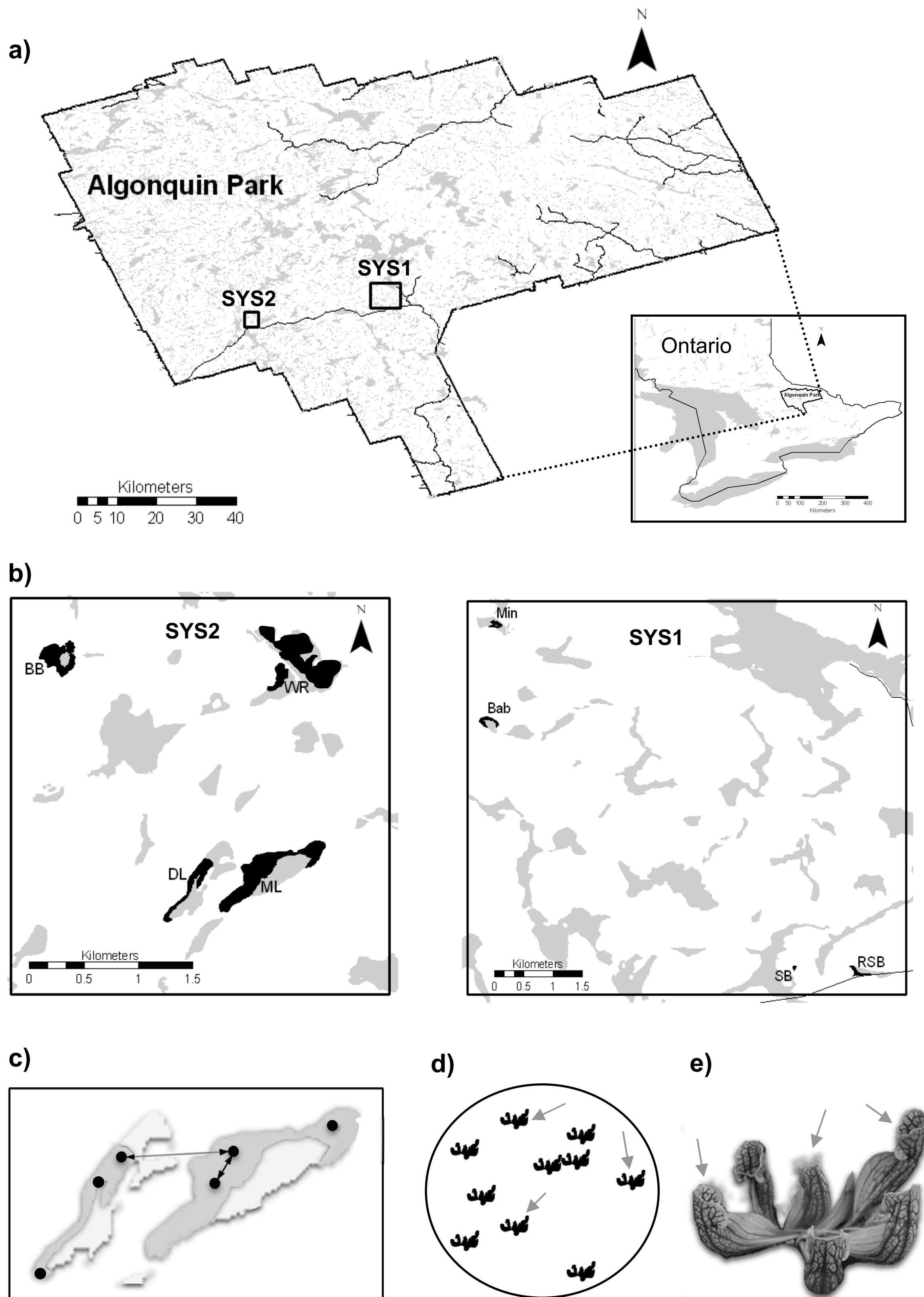


Figure 4.1 Sampling locations and design. a) The location of the two bog systems (SYS1 and SYS2) in Algonquin Park, Ontario, Canada. b) The four closest bogs were sampled in each system. c) Three clusters (shown as circles) within each bog were sampled. Arrows represent the Euclidean distances between the clusters' centres. d) Three randomly chosen plants within each cluster were sampled. e) *Sarracenia purpurea*. Three randomly chosen leaves within each plant were sampled.

Larvae were sorted with forceps in a Petri dish and placed in absolute ethanol at -20ºC until DNA extraction. I used the DNeasy tissue kit (QIAGEN, Germantown, MD) to extract genomic DNA from each individual and genotyped them at 12 microsatellite loci (GenBank accession numbers FJ665262–FJ665273) developed specifically for this species (Rasic et al. 2009). I followed amplification protocols and fragment analysis methods described in Chapter 2.

In all analyses I included only individuals with complete genotypes. My final data set consisted of 12 loci scored for 740 individuals sampled from 24 clusters (clusters 1-24) in 8 bogs (SB, RSB, Bab, Min; WR, DL, ML, BB) grouped in two systems (SYS1 & SYS2) (Appendix 1). The average number of genotyped larvae per leaf (3.5) did not significantly differ between the systems (t = 0.843, P = 0.397).

4.2.3 Habitat mapping and landscape variables

I determined the density and distribution of pitcher plants within each bog by recording the number of plants within a 2 m-radius circle positioned every 10 m along a linear transect extending from one edge of the bog to the other. The entire bog area was covered by such transects, separated from each other by 5 m. I used these point recordings and their UTM coordinates for the spherical kriging procedure performed in ArcGIS 9.3 (ESRI, Redlands, CA). From the resulting raster maps with the predicted plant distribution, I estimated landscape variables related to habitat patch density and isolation: bog plant density (average number of plants/m^2 in each bog), cluster plant density (average number of plants/m^2 in each sampled cluster), and cluster connectivity (measures the connectivity of each cluster to all other clusters in a system and is calculated as $\sum \exp(-d_{ij})$, where d_{ij} is a pairwise Euclidean distance in km between centres of clusters j and i). Bog area (m^2) was measured in ArcGIS 9.3 from 30-m resolution vector maps (Wetland class from Land Cover, Circa 2000, Agriculture and Agri-Food Canada) and Google Earth images. Bog plant density and bog size were estimated for all

eight bogs in the two systems, and cluster plant density and cluster connectivity were estimated for all 24 clusters (Appendix 1).

4.2.4 Microsatellite variability and summary statistics

Genotypic data were initially tested for the presence of null alleles and other scoring errors using MICRO-CHECKER version 2.2.3 (van Oosterhout *et al.* 2004). Standard population genetic summary statistics were generated, and tests for Hardy-Weinberg and linkage equilibrium performed, in FSTAT version 2.9.3.2 (Goudet 1995) for all scales of sampling; for simplicity we present results for the bog scale only.

4.2.5 Hierarchical AMOVA

I investigated how genetic variation was partitioned across all spatial scales using the hierarchical analysis of molecular variance (hierarchical AMOVA). The model properties of hierarchical AMOVA correspond well to the biological structural properties of this system (i.e., nested hierarchy). I employed the package HIERFSTAT for the statistical software R (Goudet 2005), which computes variance components and moment estimators of hierarchical F-statistics for any number of nested scales. I used 1000 randomizations to determine the statistical significance of genetic differentiation at a given scale (leaf, plant, cluster, bog, system), while controlling for the effects at the other scales. For example, testing for significant differences among plants (nested within clusters and above leaves) implies permutating whole units of the scale 'leaf' among plants, but keeping them within units defined by the scale 'cluster'.

4.2.6 Relationships between individuals across spatial scales

Given that we defined samples starting at a very small spatial scale (within leaves), I wanted to investigate the percentage of full-sib pairs sampled within leaves and more generally, within each of the higher scales of sampling. Although most studies try to avoid the inclusion of family groups, I was specifically interested in how this variable would change with scale of sampling. I used the data for the distributions of full-sib pairs over increasing aggregation scales to infer oviposition behaviour of midge females. Maximum likelihood estimates of pair-wise relationships between individuals were obtained in ML-RELATE (Kalinowski et al. 2006), with 10000 randomizations and 99%

confidence level. Relationships were tested between the following categories: full-sibs (FS), half-sibs (HS), unrelated (U) and parent-offspring (PO). Given that the parent-offspring relationship is not possible between larvae collected in the same year, I treated those cases as full-sibs (as in Savage et al. 2010). A confidence set for the relationship between each pair of individuals was generated with 1000 randomizations at the 99% confidence level. Every putative full-sib relationship was then tested against each of the alternative relationships indicated by the confidence sets using likelihood ratio tests with 1000 simulated random genotype pairs (Kalinowski *et al.* 2006). Only full-sib pairs that had significantly higher likelihood than the alternative relationships were further considered. The percentage of full-sib pairs thus detected was plotted for each level in the spatial hierarchy: within a leaf, between leaves within a plant, between plants within a cluster, between clusters within a bog. Note that pairwise comparisons at lower levels were thus removed as I analyzed progressively higher levels.

4.2.7 Spatial Autocorrelation and PCoA

To further examine spatial genetic structure, I employed spatial autocorrelation analysis for distance classes with equal sample sizes in GenAlEx ver. 6 (Peakall & Smouse 2006). The program calculates a matrix of mean genotypic distance values between all pairs of individuals following Smouse & Peakall (1999). A linear pair-wise geographic distance matrix was calculated as the Euclidean distance between UTM coordinates of sampled larvae. The spatial autocorrelation coefficient (r) was calculated for several distance classes that corresponded to the comparison of individuals at the following scales: within a cluster, between the clusters (within a bog), among bogs. The statistical significance of the autocorrelation coefficient (r) was tested with 999 permutations. To visualize the pattern of genetic structuring for bog and cluster samples, I also conducted Principal Coordinate Analysis (PCoA) on the same mean genotypic distance values using GenAlEx ver. 6.41 (Peakall & Smouse 2006).

4.2.8 Distance-based Redundancy Analysis (dbRDA)

I tested the effects of four landscape variables: (i) bog size, (ii) average plant density within a bog, (iii) cluster connectivity, and (iv) average plant density within a cluster, on

genetic differentiation of pitcher plant midge larvae from different leaves, plants, and clusters. To this end, I used distance-based redundancy analysis (dbRDA), a multivariate method that assesses the influence of landscape data measured at distinct points on values in a dissimilarity (in this case, mean genotypic genetic distance) matrix (Legendre & Anderson 1999). Because I had only four bogs per system, there was insufficient power to test the effects of landscape variables at the bog scale.

A matrix of mean genotypic distance values between individuals (Smouse & Peakall 1999) was calculated for leaves, plants or clusters*,* and each was used separately in DISTLM *forward* (Anderson 2003, McArdle & Anderson 2001) with all four predictor variables (bog size, plant density within bog, cluster connectivity and plant density within a cluster) entered in each analysis. The multi-locus inter-individual genetic distance measure of Smouse & Peakall (1999) is commonly used in spatial autocorrelation analysis, does not require estimation of allele frequencies from small samples, and does not assume any particular microevolutionary processes. Marginal tests (i.e., fitting of each variable individually, ignoring other variables) were followed by the forward selection procedure with conditional tests (i.e., fitting each variable one a time, conditional on the variables already included in the model). The significance of the marginal tests was done with 9999 permutations of raw data, while for the conditional test the program uses permutation of residuals under the reduced model (Anderson 2003). Tests were conducted separately for genetic distances measured at each scale (between leaves, plants or clusters), allowing me to assess which landscape variables were important at any given spatial scale. Separate analyses were conducted within each system of bogs.

4.3 Results

4.3.1 Microsatellite variability

After the initial testing for scoring errors, I excluded locus MK01 from further analyses due to the potential presence of null alleles (with estimated frequencies within bogs between 0.1-0.21). Within bogs, the number of alleles per locus ranged from two to 20 and the average allelic richness ranged from five to 7.5. Observed heterozygosity was

significantly higher than expected in two bogs in each system ($P < 0.05$) (i.e., deviations from the Hardy-Weinberg proportions were due to excess heterozygosity). On average six pairs of loci in SYS1 bogs, and 8.3 in SYS2 bogs exhibited significant linkage disequilibrium. The linkage disequilibrium results did not indicate consistent associations between any loci.

4.3.2 Hierarchical structuring

A hierarchical AMOVA revealed that the two bog systems were significantly differentiated from each other ($F_{system/Total} = 0.002$, $P = 0.014$). Within both systems variability was similarly partitioned among higher spatial scales (bog, cluster, plant). A difference between the two systems was detected at the leaf scale: in SYS1 it contributed only 0.14% to the overall genetic variation, whereas in SYS2 this scale made up 2.9% of the total variation (Table 4.1). Consequently, significant structuring was present at every hierarchical scale in SYS2, whereas leaves were not structured within plants in SYS1 (Table 4.2).

Overall, most of the genetic variation was contained within individuals (Error term), which is common for microsatellite markers (Hedrick 1999). Negative values at the individual term (Table 4.1), which is an equivalent to an individual inbreeding coefficient, imply that individuals were highly heterozygous. Consequently, hierarchical F-statistics for individuals grouped at any of the higher scales ($F_{individual/level(i)}$, F_{IS} analogues) were negative as well (the last column in Table 4.2), and indicated excess heterozygosity in groups of individuals aggregated at any scale.

4.3.3 Full-sib pairs across scales

The pattern of full-sib distribution was quite different between the two bog systems (Figure 4.2). In SYS1 the percentage of pairs of individuals that were full-sibs was around 0.2% - 4.8% across all scales, whereas in SYS2 significantly more full-sib pairs were found within a single leaf (5% - 15%) than at any higher spatial level (0.3% - 4.4%). There were two exceptions to this general pattern in SYS1: (i) Bab bog (clusters 7-9) showed higher number of full-sibs in a single plant (12%) than in higher levels (0.3% - 3%), and (ii) cluster 12 in Min bog showed extremely high percentage of full-sib pairs at

the leaf and plant scale (53% and 12%, respectively), and was omitted from the graph as an outlier. Additionally, WR bog in SYS2 had a high number of full-sibs at both the leaf and the plant scale (10.3% and 8.4%, respectively). The bogs and clusters in which we found these high levels of full-sib pairs are either characterized by low plant density (Bab and WR bogs) or are distant from the main bog area (cluster 12 in Min bog), respectively (Appendix 1).

Table 4.1 Hierarchical analysis of molecular variance in *Metriocnemus knabi*. The output from HIERFSTAT (Goudet 2005) contains: overall variance components and percentage (%) of variation at each scale. SYS indicates the variance between systems. Results for all lower scales are shown separately for the two systems (SYS1 and SYS2).

			Scale					
	SYS		**bog**	**cluster**	**plant**	**leaf**	**individual**	**error**
variance components	0.014	SYS1	0.130	0.101	0.103	0.008	-0.333	5.838
		SYS2	0.116	0.119	0.124	0.171	-0.423	5.784
% variation	0.24	SYS1	2.22	1.73	1.76	0.14	-5.70	99.84
		SYS2	1.97	2.02	2.10	2.90	-7.18	98.18

Table 4.2 Matrix of hierarchical *F*-statistics computed in HIERFSTAT (Goudet 2005). Each value in the table indicates differentiation among scales of the corresponding 'column' within scales of the corresponding 'row'. Results are shown separately for each of the two systems. For example, the *F*-statistics measuring differentiation among clusters within bogs of system1 is 0.013. The most important values are found in the last line above the empty cells and are boxed for emphasis. Values within these boxes that are significantly greater then zero are shown bold. The significance of genetic differentiation at each scale (while controlling for the effects at all other scales) was determined using 1000 permutations.

Scale	bog		cluster		plant		leaf		individual	
	SYS1	SYS2	SYS1	SYS2	SYS1	SYS2	SYS1	SYS2	SYS1	SYS 2
Total	**0.022**	**0.020**	0.035	0.041	0.053	0.062	0.050	0.075	-0.007	0.001
bog			**0.013**	**0.021**	0.031	0.043	0.026	0.052	-0.032	-0.022
cluster					**0.018**	**0.022**	0.019	0.039	-0.040	-0.037
plant							0.001	**0.031**	-0.059	-0.045
leaf									**-0.061**	**-0.079**

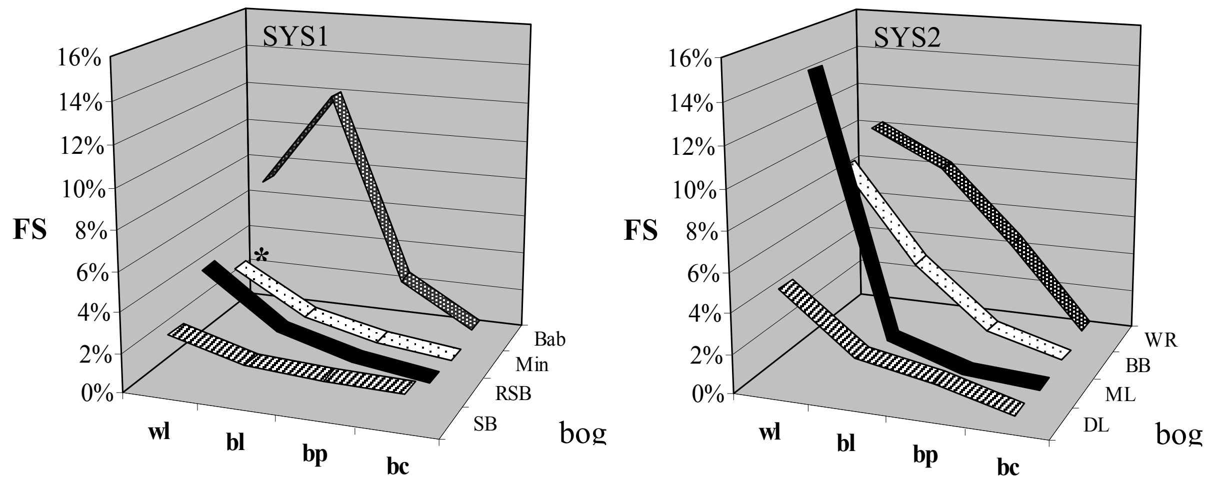


Figure 4.2 Percentage of *Metriocnemus knabi* full-sib pairs (**FS**) sampled at different spatial scales: within a leaf (**wl**), between leaves within a plant (**bl**), between plants in the same cluster (**bp**), between clusters in the same bog (**bc**). The values are averaged across 3 clusters per bog, with the exception (*) of Min bog data set that included 2 clusters, as the outlier cluster 12 contained values above 2 standard deviations for the entire data set.

4.3.4 Spatial autocorrelation

Correlograms in both systems were significant (P = 0.01; Fig. 4.3) and all distance classes contained approximately 3000 comparisons each. Significant positive spatial autocorrelation was detected at distance classes within a bog (i.e. among plants and clusters) in both systems. r values remained significant and positive at smaller between-bog distances (up to 1.5 km) in SYS2. However, there was no significant positive autocorrelation among bog samples in SYS1.

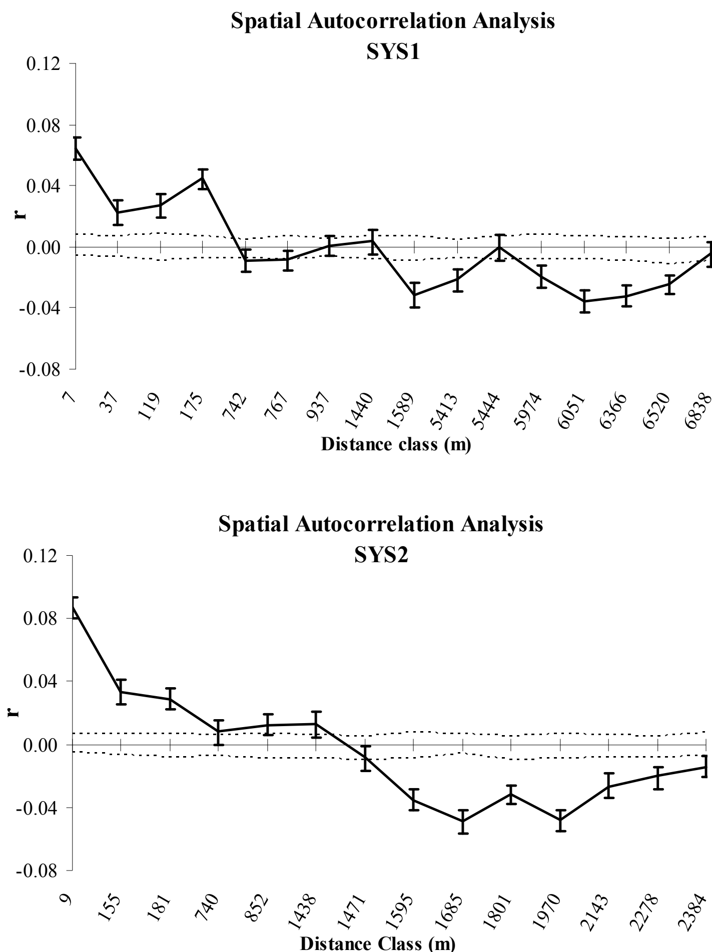


Figure 4.3 Spatial genetic autocorrelograms showing average correlation coefficients between pairs of *Metriocnemus knabi* individuals (r), plotted against geographic distance classes in SYS1 (a) and SYS2 (b). Vertical dotted lines delineate distance classes contained within scales in the spatial hierarchy, namely: within a cluster (wc), between clusters in a bog (bc) and between bogs within a system (bb). Horizontal dashed lines represent critical values under the null hypothesis that genotypes are randomly distributed across a landscape ($\alpha = 0.05$). Error bars represent 95% confidence intervals around each mean correlation coefficient.

4.3.5 Principal Coordinate Analysis (PCoA)

When individuals were grouped by bog, the first three principal coordinate axes explained 78% of the genetic variation among samples (Figure 4.4). I detected two highly differentiated bogs: Bab bog in SYS1 and WR in SYS2. When individuals were grouped according to the pitcher plant clusters from which they were sampled, 68.7% of variation was explained and cluster 12 from Min bog in SYS1 was also seen to be highly differentiated (Figure 4.4). These highly differentiated bogs and clusters are the same ones in which I detected a high proportion of full-sib pairs within leaves and plants, and they either have low abundance of pitcher plants (clusters 7-9 in Bab and 13-15 in WR) or are spatially isolated (cluster 12 in Min).

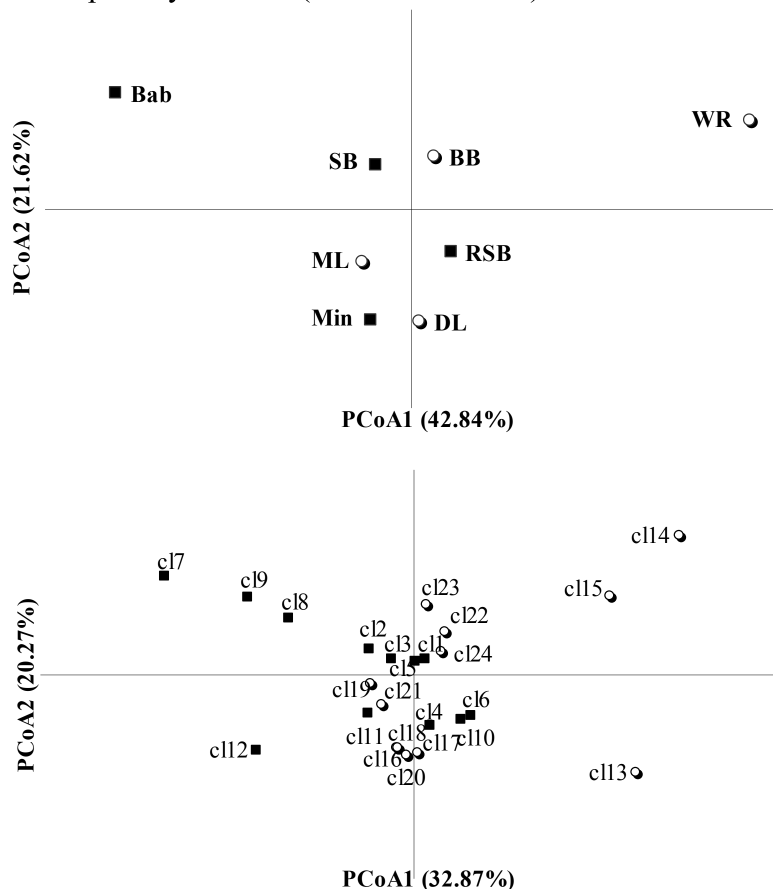


Figure 4.4 Plots of Eigen values for the first two components of the Principal Coordinate Analysis (PCoA) performed on genetic distance matrix from bog (upper) and cluster (lower) samples of *Metriocnemus knabi*. ■-SYS1 ○-SYS2.

4.3.6 Distance-based Redundancy Analysis (dbRDA)

First, at each scale (leaf, plant, cluster), the effect of each predictor variable (bog size, bog plant density, cluster plant density, cluster connectivity) was tested individually. In both systems and across all scales the strongest predictor of genetic distance was bog plant density ($P < 0.01$, Table 4.3). In SYS1 this variable explained 12.1%, 22.3% and 33.5% of variation in genetic distances at the leaf, plant and cluster scales respectively. The amount of variation explained was similar in SYS2, going from 7.9% at the leaf, 23.5% at the plant, to 29.1% at the cluster scale. Bog size and cluster connectivity were significantly associated with genetic patterns at all scales in SYS1, while cluster plant density was not significant at any scale. In SYS2, cluster connectivity and cluster plant density were significantly associated with genetic patterns at plant and leaf scales ($P < 0.05$), but not at the cluster scale. Bog size was a marginally significant explanatory variable ($P = 0.048$) only at the leaf scale in SYS2.

Sequential tests showed that bog size and bog plant density jointly explained between 18.4% and 53.4% of variation in genetic distances across scales in SYS1 (Table 4.4), and cluster connectivity was only significant in the model at the leaf scale. Two jointly significant factors in SYS2 were bog plant density and cluster connectivity, explaining between 11.2% and 32.6% of variation at the leaf and plant scales. Cluster plant density and bog size were only significant in the model at the leaf scale in SYS2.

Table 4.3 Distance based redundancy analysis of genetic distances among *Metriocnemus knabi* samples performed at each scale (cluster, plant, leaf). Each predictor variable (bog size, bog plant density, cluster plant density, cluster connectivity) was tested separately. Significant P values are bolded. '% Variation' indicates amount of variation in genetic distances explained by a particular variable.

Marginal test

Scale	Variable	SYS1 *pseudo-F*	SYS2 *pseudo-F*	SYS1 *P*	SYS2 *P*	SYS1 % Variation	SYS2 % Variation
cluster	bog size	2.75	0.44	**0.023**	0.863	21.6	4.2
plant		6.50	1.00	**0.001**	0.428	16.5	2.8
leaf		7.74	2.52	**0.001**	**0.048**	6.9	2.4
cluster	bog plant density	5.04	4.10	**0.001**	**0.005**	33.5	29.1
plant		9.48	10.47	**0.001**	**0.001**	22.3	23.5
leaf		14.40	8.89	**0.001**	**0.001**	12.1	7.9
cluster	cluster plant density	0.62	2.35	0.701	0.072	5.9	12.9
plant		1.12	5.37	0.368	**0.005**	3.3	13.6
leaf		1.45	4.46	0.261	**0.003**	1.4	4.1
cluster	cluster connectivity	2.35	1.22	**0.046**	0.301	19.1	10.9
plant		5.65	2.83	**0.001**	**0.024**	14.6	7.7
leaf		6.38	2.77	**0.001**	**0.022**	5.7	2.6

Table 4.4 Forward selection procedure in distance based redundancy analysis of genetic distances among *Metriocnemus knabi* samples performed at each scale (cluster, plant, leaf). Only significant values in a combined model are reported. Bolded Cummulative % indicates the total variation explained by combined variables in sequential tests. The top down sequence of variables corresponds to the sequence indicated by the forward selection procedure.

Sequential test

		SYS1	SYS2	SYS1	SYS2	SYS1	SYS2
Scale	Variable	*pseudo-F*		*P*		Cummulative %	
cluster	bog plant density	5.04	4.11	0.001	0.005	33.5	**29.1**
	bog size	3.84	-	0.006	-	**53.4**	-
plant	bog plant density	9.48	10.47	0.001	0.001	22.3	23.5
	bog size	8.14	-	0.001	-	**38.1**	-
	cluster connectivity	-	4.43	-	0.005	-	**32.6**
leaf	bog plant density	14.40	8.89	0.001	0.001	12.1	7.9
	bog size	8.02	5.74	0.001	0.001	18.4	15.9
	cluster connectivity	3.74	3.88	0.021	0.006	**21.2**	11.2
	cluster plant density	-	3.41	-	0.014	-	**18.7**

4.4 Discussion

Analyses of genetic variation across multiple nested scales revealed complex genetic structuring in the pitcher plant midge. Comparing two systems of bogs, the partitioning of variation was similar at the broader scales (among bogs, clusters, and plants), but different at the finest scale (among leaves within plants). The percentage of full-sib pairs across hierarchical scales was quite different between the two bog systems, and a high proportion of full sibs was found within leaves and plants that occur in isolated or low plant density patches. Positive local spatial structure extended among bogs in system 2 (SYS2), but not in system 1 (SYS1). Overall, dbRDA showed that across several scales a significant portion of genetic structure in *M. knabi* can be explained by bog size, bog plant density and cluster connectivity.

4.4.1 Genetic structure across scales in *M. knabi*

Careful consideration of the scale of sampling in population and landscape genetic studies is highly important (Anderson et al. 2010, Cushman & Langduth 2010, Storfer et al. 2010). Multiple processes operating over different spatial scales, such as natal dispersal, social and mating interactions, long-distance colonizations, etc. can influence patterns of genetic variation. These processes, in turn, may respond to landscape and environmental factors at different spatial scales (Murphy et al. 2010). Analysis of genetic structure across multiple scales of sampling can therefore be important for understanding links between genetic and landscape patterns, and the underlying ecological processes. I conducted such an analysis for the pitcher plant midge, taking advantage of a biological system where the units of sampling across various scales, defining essentially the grain of the analysis, need not be arbitrarily selected but are naturally presented by the larval habitat itself.

Overall, I detected significant structuring across multiple nested scales, going from the system to the leaf scale. The two bog systems were significantly differentiated ($F_{system/Total} = 0.002$ $P = 0.014$), which was expected as they are 26 km apart. This small F value does not mean high genetic connectivity at this distance, but simply that the vast majority of the variation is contained within lower scales in the hierarchy. In both

systems the partitioning of variability was similar at the bog, cluster and plant scales (Table 4.1), and genetic structuring was significant at all of them (Table 4.2). A difference between the systems was revealed at the leaf scale, where structuring among leaves within plants was significant in SYS2 but not in SYS1. A cruder grain in sampling would have missed this component of the overall genetic pattern in this species.

Consistent with the finding that samples from leaves within a plant were significantly different in SYS2 but not in SYS1, I observed a high proportion of full-sibs within leaves of the same plant in SYS2 but not in SYS1 (Figure 4.2). The distribution of full-sibs is a potential proxy for oviposition behaviour, given that chironomid females very rarely mate with multiple males (Armitage et al. 1995). These results suggest that females of *M. knabi* in SYS1 leave smaller number of eggs within a single leaf and tend to distribute their eggs more equally across plants, while in SYS2 they tend to leave a large number of eggs (clutches) within a single chosen leaf.

The high proportion of full-sib pairs found within leaves and plants of clusters that were either highly isolated (cluster 12 in Min bog) or occured in bogs with low plant density (Bab and WR bogs) suggest a role of habitat patch isolation and habitat amount at these higher spatial scales in influencing the fine-scale (i.e., among leaves) oviposition decisions made by females. There are a number of reports of directional flight of the chironomid females prior to oviposition, but how they are able to select the correct site is not understood (Oliver 1971). The females of the pitcher-plant midge appear to respond to leaf size (Paterson & Cameron 1982, Nastase et al. 1995), but oviposition decisions may occur at several spatial scales (Trzcinski et al. 2003), as supported by my findings. Given that this species is an extreme specialist with respect to ovipositon sites, as the pitcher plant leaves represent the exclusive habitat for the larval development, it would be highly advantageous to make active decisions about oviposition based on different characteristics of the larval habitat at several spatial scales.

4.4.2 Linking patterns and processes across scales

In PCoA analyses, I found that the same clusters that were characterized by a high proportion of full-sib pairs within leaves and plants (cluster 12 in Min bog, clusters 7-9 in

Bab bog, and clusters 13-15 in WR bog) were identified as being highly differentiated from other clusters. This result in not surprising given that the inclusion of highly related individuals within samples inflates measures of genetic differentiation (Allendorf & Phelps 1981, Anderson & Dunham 2008, Goldberg & Waits 2010). However, this result is important because it indicates that the process of female oviposition occurring at the finest scales (among leaves and plants) interacts with the sampling design (in this case, collection of juveniles at small spatial scales and before dispersal events) to affect the output from the common population genetic analyses conducted at larger scales. My results thus highlight that the scale of sampling, relative to the scales of ecological/evolutionary processes, influences the conclusions that can be drawn in population and landscape genetic studies (Anderson et al. 2010).

These findings also suggest, in this system, linkages among processes and patterns at different spatial scales. Specifically, I hypothesized that the isolation and amount of habitat at cluster and bog scales (broad-scale landscape patterns) lead females to aggregate their eggs within leaves (fine-scale ecological process), which exaggerates genetic differentiation of larvae not only among leaves but also at higher scales (fine to broad-scale genetic patterns). Based on this hypothesis, I would expect to see significant effects of broad scale landscape variables on genetic differentiation at all scales, but more so at the finer scales. This is indeed what my dbRDA analyses revealed. In the marginal tests in both systems, significant effects were seen either across all scales, or only at finer scales. The only exception was for cluster plant density in SYS1, which was not significant at any scale. In the sequential tests, going from the cluster to the plant to the leaf scale, progressively more landscape variables were included in the significant models. Pitcher plant density within bogs exhibited the strongest effect on genetic structure among leaves, plants and clusters (Table 4.3). When compared to other tested variables, it explained the largest proportion of variation in genetic distances (between 7.9% and 33.5%) and its effect was consistent in both systems and across scales. Even when the other predictor variables were accounted for in the sequential models (Table 4.4), average plant density in a bog had a pronounced effect on genetic distances among larval samples, supporting the hypothesis that females are more likely to aggregate eggs locally under conditions of low plant density at the bog scale.

Cluster connectivity measures Euclidean distance among sampling points in the entire landscape (within and among bogs), and therefore accounts for the importance of overall physical distance on the pattern of genetic distances. If lower connectivity of clusters is associated with greater genetic distances between samples, as I observed, this is analogous to isolation-by-distance and could simply indicate that spatially limited dispersal of adult midges plays a role in determining genetic patterns among larval samples. However, the strongest effects of cluster connectivity on genetic differentiation of midge larvae were observed at the finest scales, for leaves and plants, particularly in the sequential tests. Thus, it is likely that the significant influence of cluster connectivity on genetic differentiation is mediated to a large extent by effects on female oviposition, as hypothesized, which should be observable at the finest spatial scales. In contrast, if the influence of cluster connectivity was mediated simply by limited dispersal of adults, I would expect to observe stronger effects at the broader spatial scales.

Bog size was the second most important landscape factor explaining genetic distances in SYS1, but was marginally significant only at the finest scale SYS2. This difference can be explained by the characteristics of the two systems: SYS2 contains only large bogs, whereas SYS1 contains bogs more variable in size (Appendix 1). This provided more size classes for the regression analysis in SYS1, making the pattern detectable. Limited variation in a predictor variable reduces power in any analysis and Short Bull *et al.* (2011), in their gene flow analysis in American black bears found that landscape features had to be highly variable in order to be supported in landscape genetic models. My results further reinforce the conclusion that landscape genetic studies should ideally incorporate large variation in landscape attributes. The fact that the larger bogs in SYS1 exhibited a pattern of full-sib distribution similar to bogs in SYS2 (Figure 4.2), points towards a critical bog size at which female oviposition behaviour changes.

Overall, in SYS1 up to 54% of the variation in genetic distances was jointly explained by broader scale variables in the dbRDA: bog size and plant density within bogs. The joint influence of bog plant density and cluster connectivity explains up to 33% of variation across spatial scales in SYS2. The predictive power of only two habitat variables in each system is high and comparable to the results from the study by Pilot et al. (2006). Their

sequential tests in dbRDA revealed that, after accounting for geographic distance, 53% of variation in Nei's genetic distance at microsatellite loci among European wolf populations can be attributed to vegetation types.

4.4.3 Dispersal and isolation by distance

Ecological studies have made inferences about dispersal of pitcher plant midges based on spatial patterns of larval abundance (Miner & Taylor 2002, Krawchuk & Taylor 2003). These studies suggest that *M. knabi* individuals are weak fliers, aggregate around plants and clusters, and rarely move among bogs. Significant genetic structure at the plant and cluster scales in my study largely support these previous ecological inferences. However, my study also indicates that gene flow can occur among close bogs, as seen in the spatial autocorrelation analyses. Significant positive spatial genetic structure was detected among bogs in SYS2, where distances between some bogs are relatively small, but not in SYS1 where the bogs are more distant from each other. Furthermore, significant positive autocorrelation at short distances coupled with the significant negative autocorrelation at long distances is a pattern consistent with isolation-by-distance (Sokal & Oden 1991). Thus, in addition to effects of female oviposition behaviour occurring at fine scales, the balance between restricted gene flow and genetic drift must also a contributor to genetic structuring at broad spatial scales (i.e., among bogs).

4.4.4 High individual and group heterozygosity

Excess heterozygosity (i.e., negative F_{is}) at neutral loci is not often found in animal populations, and is somewhat surprising for an insect that is considered to be a weak flier, dependent on a highly specific patchy habitat for its development. Non-random mating, specifically outbreeding, is a frequent explanation for excess heterozygosity. However, looking at the individual inbreeding coefficient (calculated following Ritland 1996), I did not observe significantly different values in the individuals I sampled as compared to individuals simulated under a random-mating scenario using our observed allele frequencies (data not shown). Outbreeding is therefore not a likely explanation for the high observed heterozygosities in our study. The excess heterozygosity I observed, coupled with significant differentiation among samples, is actually very similar to the

patterns observed in social mammals as well as among communal hibernacula of the timber rattlesnake (Anderson 2010). As pointed out by Anderson (2010), these patterns counter the expectation of reduced heterozygosity within genetic "demes" as a result of restricted gene flow (Wright 1969). However such patterns may be expected to arise when there is spatial clustering of individuals at some life-history stage, in combination with either sex-biased dispersal or a limited number of breeding adults (Anderson 2010), both of which can lead to excess observed heterozygosity within samples (Prout 1981, Balloux 2004).

4.4.5 Sampling considerations

It is important to note a caveat with respect to my hierarchical sampling design: progressing from the leaf scale up to the bog and system scales, the size of each sample increases while the total number of samples decreases. This could affect the power of analyses conducted at each scale. Fewer individuals included in each sample at the leaf or plant scale would lead to more uncertainty associated with estimates of population genetic parameters (i.e., "noisier" data) and potentially more outliers. I observed significant and consistent relationships between genetic patterns and several landscape variables across spatial scales, from leaves to clusters. Thus, the decreasing sizes of samples at the finest scales did not limit my ability to detect significant effects of landscape variables on genetic differentiation between samples at these scales. A small number of samples did however prevent testing of such relationships at the bog scale. Replication at large spatial scales is a challenge in landscape and ecological studies. Even when there are sufficient resources to sample multiple larger regions or 'landscapes', each landscape may have its own history and unique features, making their true replication technically impossible (Hargrove & Pickering 1992). However, conducting research at multiple similar landscapes can nonetheless be informative (Anderson et al. 2010, Short Bull et al. 2010). My sampling of two different bog systems, which could be considered as different 'landscapes', proved to be very important given the different patterns of distribution of full-siblings between the systems (Figure 4.2). Non-identical landscape attributes may limit generalizations and strict statistical inferences, but they provide the opportunity to detect plasticity of ecological processes and patterns.

4.5 Summary

I demonstrated that encompassing a large research area (suitable extent), along with refining the resolution of sampling (gradually changing the grain), reveals links among processes and patterns across different spatial scales. Genetic differentiation at several scales in *M. knabi* is significantly associated with landscape variables related to habitat size, abundance and spatial arrangement. These broad-scale landscape features seem to influence the fine-scale process of female oviposition. This process, in turn, shapes the patterns of genetic differentiation observed at both fine and broader spatial scales (e.g., as observed through PCoA). Overall, the results of my study reinforce the value of considering patterns and processes across multiple spatial scales and in multiple landscapes when investigating genetic diversity within a species.

4.6 References

Addicott JF (1974) Predation and prey community structure: an experimental study of the effect of mosquito larvae on the protozoan communities of pitcher plants. *Ecology* 55:475-492.

Allendorf FW, Phelps SR (1981) Use of allelic frequencies to describe population structure. *Canadian Journal of Fisheries and Aquatic Sciences* 38:1507-1514.

Anderson CD, Epperson BK, Fortin MJ, et al. (2010) Considering spatial and temporal scale in landscape-genetic studies of gene flow. *Molecular Ecology* 19:3565-3575.

Anderson CD (2010) Effects of movement and mating patterns on gene flow among overwintering hibernacula of the Timber Rattlesnake (*Crotalus horridus*). *Copeia* 1:54-61.

Anderson EC, Dunham KK (2008) The influence of family groups on inferences made with the program Structure. *Molecular Ecology Resources* 8:1219-1229.

Anderson MJ (2003) DISTLM forward: a FORTRAN computer program to calculate a distance-based multivariate analysis for a linear model using forward selection. http://www.stat.auckland.ac.nz/~mja/Programs.htm

Armitage PD, Cranston PS, Pinder LCV (1995) *The Chironomidae: the biology and ecology of non-biting midges.* Chapman & Hall, London.

Balkenhol N, Gugerli F, Cushman SA, et al. (2009) Identifying future research needs in landscape genetics: where to from here? *Landscape Ecology* 24:455-463.

Balloux F (2004) Heterozygote excess in small populations and the heterozygote-excess effective population size. *Evolution* 58:1891-1900.

Blanquer A, Uriz MJ (2010) Population genetics at three spatial scales of a rare sponge living in fragmented habitats. *BMC Evolutionary Biology* doi: 10.1186/1471-2148-10-13.

Buckley HL, Miller TE, Ellison AM, Gotelli NJ (2010) Local- to continental-scale variation in the richness and composition of an aquatic food web. *Global Ecology and Biogeography* 19:711-723.

Cushman SA, Landguth EL (2010) Spurious correlations and inference in landscape genetics. *Molecular Ecology* 19:3592-3602.

Diggle PJ, Ribeiro PJ (2007) *Model-based geostatistics*. Springer Series in Statistics, Springer-Verlag, Heidelberg.

Fortin MJ, Dale MRT (2005) *Spatial analysis: a guide for ecologists*. Cambridge University Press, Cambridge.

Goldberg CS, Waits LP (2010) Quantification and reduction of bias from sampling larvae to infer population and landscape genetic structure. *Molecular Ecology Resources* 10:304-313.

Gonçalves da Silva AS, Eberhard JR, Wright TF, Avery ML, Russello MA (2010) Genetic evidence for high propagule pressure and long-distance dispersal in monk parakeet (*Myiopsitta monachus*) invasive populations. *Molecular Ecology* 19:3336-3350.

Goudet J (1995) FSTAT (version 1.2): a computer program to calculate F-statistics. *Journal of Heredity* 86:485-486.

Goudet J (2005) HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes* 5:184-186.

Guillot G (2009) On the inference of spatial structure from population genetics data. *Bioinformatics* 25:1796-1801.

Hargrove WW, Pickering J (1992) Pseudoreplication: a sine qua non for regional ecology. *Landscape Ecology* 6:251-258.

Harvey E, Miller TE (1996) Variance in composition of inquiline communities in leaves of *Sarracenia purpurea* L. on multiple spatial scales. *Oecologia* 108:562-566.

Heard SB (1994) Pitcher-plant midges and mosquitoes: a processing chain commensalism. *Ecology* 75:1647-1660.

Hedrick PW (1999) Perspective: highly variable loci and their interpretation in evolution and conservation. *Evolution* 53:313-318.

Holling CS (1992) Cross-scale morphology, geometry, and dynamics of ecosystems. *Ecological Monographs* 62:447-502.

Hu LJ, Uchiyama K, Shen HL, Ide Y (2010) Multiple-scaled spatial genetic structures of *Fraxinus mandshurica* over a riparian–mountain landscape in Northeast China. *Conservation Genetics* 11:77-87.

Kalinowski ST, Wagner AP, Taper ML (2006) ML-RELATE: a computer program for maximum likelihood estimation of relatedness and relationship. *Molecular Ecology Resources* 6:576-579.

Kotliar NB, Wiens JA (1990) Multiple scales of patchiness and patch structure: a hierarchical framework for the study of heterogeneity. *Oikos* 59:253-260.

Krawchuk MA, Taylor PD (2003) Changing importance of habitat structure across multiple spatial scales for three species of insects. *Oikos* 103:153-161.

Law B, Buckleton JS, Triggs CM, Weir BS (2003) Effects of population structure and admixture on exact tests for association between loci. *Genetics* 163:381-387.

Legendre P, Anderson MJ (1999) Distance-based redundancy analysis: testing multi-species responses in multi-factorial ecological experiments. *Ecological Monographs* 69:1-24.

Levin SA (1992) The problem of pattern and scale in ecology: the Robert H. MacArthur award lecture. *Ecology* 73:1943-1967.

Lee-Yaw JA, Davidson A, McRae BH, Green DM (2009) Do landscape processes predict phylogeographic patterns in the wood frog? *Molecular Ecology* 18:1863-1874.

Miner JA, Taylor PD (2002) Effects of peatland size and exposure on two species of Diptera inhabiting the pitcher *Sarracenia purpurea* L. *Ecoscience* 9:347-354.

Miller TE, Horth L, Reeves RH (2002) Trophic interactions in the phytotelmata communities of the pitcher plant, *Sarracenia purpurea*. *Community Ecology* 3:109-116.

Mitzui Y, Isagi Y, Setoguchi H (2010) Multiple spatial scale patterns of genetic diversity in riparian populations of *Ainsliaea faurieana* (Asteraceae) on Yakushima Island, Japan. *American Journal of Botany* 97:101-110.

McArdle BH, Anderson MJ (2001) Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology* 82:290-297.

Murphy MA, Evans JS, Storfer A (2010) Quantifying *Bufo boreas* connectivity in Yellowstone National Park with landscape genetics. *Ecology* 9:252-261.

Nastase AJ, De la Rosa C, Newell SJ (1995) Abundance of pitcher-plant mosquitoes, *Wyeomyia smithii* (Coq.) (Diptera: Culicidae) and midges, *Metriocnemus knabi* Coq. (Diptera: Chironomidae), in relation to pitcher characteristics of *Sarracenia purpurea* L. *American Midland Naturalist* 133:44-51.

Nielsen JL, Byrne A, Graziano SL, Kozfkay CC (2009) Steelhead genetic diversity at multiple spatial scales in a managed basin: Snake River, Idaho. *North American Journal of Fisheries Management* 29:680-701.

Oliver DR (1971) Life history of the Chironomidae. *Annual Review of Entomology* 16:211-230.

Paterson CG, Cameron CJ (1982) Seasonal dynamics and ecological strategies of the pitcher plant chironomid, *Metriocnemus knabi* Coq. (Diptera: Chironomidae), in southeast New Brunswick. *Canadian Journal of Zoology* 60:3075-3083.

Peakall R, Smouse PE (2006) GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6:288-295.

Pilot M, Jedrzejewski W, Branicki W, et al. (2006) Ecological factors influence population genetic structure of European grey wolves. *Molecular Ecology* 15:4533-4553.

Prout T (1981) A note on the island model with sex dependent migration. *Theoretical and Applied Genetics* 59:327-332.

Rango JJ (1999) Summer phenology of aquatic insect communities inhabiting the leaves of the northern pitcher plant, *Sarracenia purpurea* L. *Northeastern Naturalist* 6:19-30.

Rasic G, Maxwell SA, Keyghobadi N (2009) Characterization of microsatellite loci for the pitcher plant midge, *Metriocnemus knabi* Coq. (Diptera: Chironomidae). *Molecular Ecology Resources* 9:1388-1391.

Ritland K (1996) Estimator for pairwise relatedness and individual inbreeding coefficients. *Genetical Research* 67:175-185.

Savage WK, Fremier AK, Shaffer HB (2010) Landscape genetics of alpine Sierra Nevada salamanders reveal extreme population subdivision in space and time. *Molecular Ecology* 19:3301-3314.

Schwartz MK, McKelvey KS (2009) Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. *Conservation Genetics* 10:441-452.

Short Bull RA, Cushman SA, Mace R, et al. (2011) Why replication is important in landscape genetics: American black bear in the Rocky Mountains. *Molecular Ecology* 20:1092-1107.

Smouse PE, Peakall R (1999) Spatial autocorrelation analysis of multi-allele and multi-locus genetic microstructure. *Heredity* 82:561-573.

Sokal RR, Oden NL (1991) Spatial autocorrelation analysis as an inferential tool in population genetics. *The American Naturalist* 138:518-521.

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

Tiner RW (1999) *Wetland indicators: a guide to wetland identification, delineation, classification, and mapping.* Lewis Publishers, CRC Press, Boca Raton, FL.

Trzcinksi MK, Walde SJ, Taylor PD (2005) Local interactions in pitcher plant communities scale-up to regional patterns in distribution and abundance. *Environmental Entomology* 34:1464-1470.

Trzcinski MK, Walde SJ, Taylor PD (2003) Colonisation of pitcher plant leaves at several spatial scales. *Ecological Entomology* 28:482-489.

van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* 4:535-538.

Wagner HH, Fortin MJ (2005) Spatial analysis of landscapes: concepts and statistics. *Ecology* 86:1975-1987.

Wiens JA (1989) Spatial scaling in ecology. *Functional Ecology* 3:385-397.

Wright S (1969) *Evolution and the genetics of populations*. University of Chicago Press, Chicago.

Wu J, Loucks OL (1995) From balance of nature to hierarchical patch dynamics: a paradigm shift in ecology. *Quarterly Review of Biology* 70:439-466.

Chapter 5. Wind assists gene flow among bogs in the pitcher plant midge

5.1 Introduction

Landscape genetics aims to explicitly quantify the effects of landscape and environmental factors on spatial genetic variation (Manel et al. 2003, Holderegger & Wagner 2006, Storfer et al. 2007, Balkenhol et al. 2009). A common research focus in terrestrial animals revolves around habitat factors that impede or facilitate gene flow, as mediated by active dispersal, through heterogeneous landscapes (Storfer et al. 2010). Features such as rivers, mountain ridges, roads, and the extent of unsuitable habitats, are frequently tested as barriers to gene flow in landscape genetics studies (e.g., Epps et al. 2005, Funk et al. 2005, Keyghobadi et al. 2005, Coulon et al. 2006, McRae & Beier 2007). Conversely, several landscape features have been identified that aid gene flow, such as forest-regenerated shrubs and rivers in amphibians (Spear et al. 2005, Murphy et al. 2010).

In aquatic systems, patterns of water movement are recognized to affect dispersal of organisms, and related variables are often included in landscape (or ‘seascape’) genetic studies. For example, in river and spring systems, landscape genetic models that included the drainage pattern, or direction and/or speed of water flow best explained patterns of genetic structure in zooplankton (Michels et al. 2001), brook charr (Angers et al. 1999), and aquatic snails (Wilmer et al. 2008). In marine environments, landscape genetic analyses that included ocean currents best explained genetic diversity and spatial structure in blue whiting (Was et al. 2008), kelp bass (Selkoe et al. 2010), subtidal whelk (White et al. 2010) and giant kelp (Alberto et al. 2011).

Air currents, analogously to water currents, can be an important environmental factor influencing dispersal, and hence gene flow, in small terrestrial arthropods. However, any effect of air currents on genetic variation in such animals has not yet been demonstrated. Wingless arthropods, such as spiders and mites, engage in specialized behaviours (e.g., ‘ballooning’ using silk threads as parachutes) to get themselves airborne and dispersed by

the wind (Frost 1997, Bell et al. 2005). The phenomenon of wind-borne dispersal and migration has evolved independently in several insect orders and is now believed to be more prevalent than previously thought (Byrne 1999). Insects considered as weak dispersers, such as the mountain pine beetle (*Dendroctonus ponderosae*) and the carabid beetle (*Notiophilus biguttatus*), are found to undergo long-range dispersal aided by wind above the forest canopy (Chapman et al. 2005, Jackson et al. 2008). Flying aphids and fig wasps can travel distances exceeding tens of kilometres, reflecting the speed and direction of winds that carry them (Compton 2002). Wind densities were also correlated with direction and spread of a bluetongue epidemic, vectored by air-borne biting midges (Hendrickx et al. 2008).

Small winged insects (body length <10 mm) are found within the thermal atmospheric layer in high concentrations, visible as 'insect plumes' in radar signals (Reynolds & Reynolds 2009). They are generally assumed to be weak flyers that are simply passively carried by the wind currents (Drake & Farrow 1989). However, radar analysis has revealed that they oppose aerial updrafts (Geerts & Miao 2005), exhibiting active flight behaviour distinctly different from the aerial dispersal of wingless arthropods and seeds, which is passive once these organisms have launched into the air (Reynolds & Reynolds 2009). Hence, these small insects can achieve large dispersal distances by actively navigating through air-currents. This process could substantially increase the extent of gene flow to distant populations that are situated along the trajectories of frequent winds.

In midges (Chironomidae), long-distance dispersal by wind is considered an integral part of their biology (Oliver 1971, Delettre 1993). This view is indirectly supported by the observations of mass appearance of adults after a strong wind (6-7 m/s, Hirabayashi 1991), recovery of midges at significant altitude (up to 600 m, White 1970) and large distances from the nearest landmass (several hundreds of kilometres, Holzapfel & Harrell 1968). Recently, Miao et al. (2011) showed that populations of the wheat midge (*Sitodiplosis mosellana*) exhibited long-distance dispersal with air currents in a step-by-step manner over a wheat-growing area in northern China (~1000 km).

Here, I examined the effect of long-term wind pattern (direction and frequency) on gene flow and spatial genetic structure in the pitcher plant midge *Metriocnemus knabi* Coquillett 1904, a Chironomid species obligately associated with the purple pitcher plant *Sarracenia purpurea* L. (Heard 1994). Ecological studies have suggested weak flying abilities of this small insect, with adults aggregating around local groups of pitcher plants and moving rarely among bogs (Krawchuk & Taylor 2003). Genetic data, however, indicated potential gene flow among some neighbouring bogs in a landscape, as individuals were more spatially related than by chance at distances up to 1.4 km (spatial autocorrelation analysis, Chapter 4). A pattern of isolation-by-distance (Wright 1943) among individuals found within a group of closely situated bogs in a landscape is also consistent with the process of spatially limited gene flow mediated by active flight of *M. knabi* adults. Here, I was interested in whether the process of wind-assisted gene flow also occurs in *M. knabi* at larger spatial scales (i.e., among more distantly spaced bogs).

I tested the hypothesis that wind facilitates gene flow in *M. knabi*, by examining evidence for the prediction that the long-term pattern of wind direction and frequency should be correlated with genetic distances among *M. knabi* samples. Both spatially limited active flight and air-borne dispersal may be present in this insect, hence I expect the variability in genetic distances among samples to be explained respectively by their geographic separation (measured here as straight-line, Euclidean distance), as well as their orientation to the prevailing winds in the landscape. Also, the explanatory power of wind patterns is expected to be higher in a landscape in which samples are more distant from each other and are more aligned in the direction of prevailing winds.

5.2 Materials and Methods

5.2.1 Study area and species

My study area was located in Algonquin Provincial Park in Ontario, Canada (UTM: 17N 687337E 5046853N) (Figure 5.1). The park is in a transition zone between northern coniferous forest and southern deciduous forest. Bogs are found within this forest matrix, and many of them contain *S.purpurea* and its associated commensal arthropod inhabitants. Bogs represent peat-covered wetlands with high water table and generally

low nutrient status. They are dominated by sphagnum mosses (*Sphagnum spp.*) and heath shrubs (leather leaf *Chamaedaphne calyculata*, labrador tea *Rhododendron groenlandicum*, cranberries *Vaccinium spp.*), the bogs also contain tamarack (*Larix laricina*) and black spruce (*Picea mariana*) (Tiner 1999). Within bogs, *S. purpurea* is often spatially clumped. I refer to aggregations of the pitcher plants within bogs as ‘clusters’.

I sampled *Metriocnemus knabi* (Diptera: Chironomidae) from the five nearest neighbouring bogs (0.2-7.0 km apart in a mixed forest matrix) in each of two regions, referred to here as ‘systems’, in Algonquin Provincial Park. The two systems (SYS1 and SYS2) were 26 km apart. Therefore, a total of 10 bogs were sampled in August 2009.

M. knabi is expected to have one generation per year at this latitude (Rango 1999). The midge overwinters as a larva in the leaves of *S. purpurea*. Pupation, adult emergence, mating and oviposition occur during late spring and summer (Heard 1994, Rango 1999, Krawchuk & Taylor 2003). Adult midges are intractable due to their small body size (length of only 3 mm) and very cryptic behaviour (Krawchuk & Taylor 2003). Larvae on the other hand can be readily sampled from the pitcher plant leaves.

Within each bog, I randomly chose two or three clusters of plants, three plants within each cluster, and sampled all larvae from three leaves of each plant. I used the cluster as the unit of analysis because prevailing winds are unlikely to facilitate dispersal within clusters, which are only a few meters across. In total, fourteen clusters from SYS1 and fifteen clusters from SYS2 were sampled this way. The centroid of each cluster was recorded to within 0.5 m using a high accuracy GPS receiver (Trimble GeoXH, Sunnyvale, CA). Sampled clusters were at least 25 m apart.

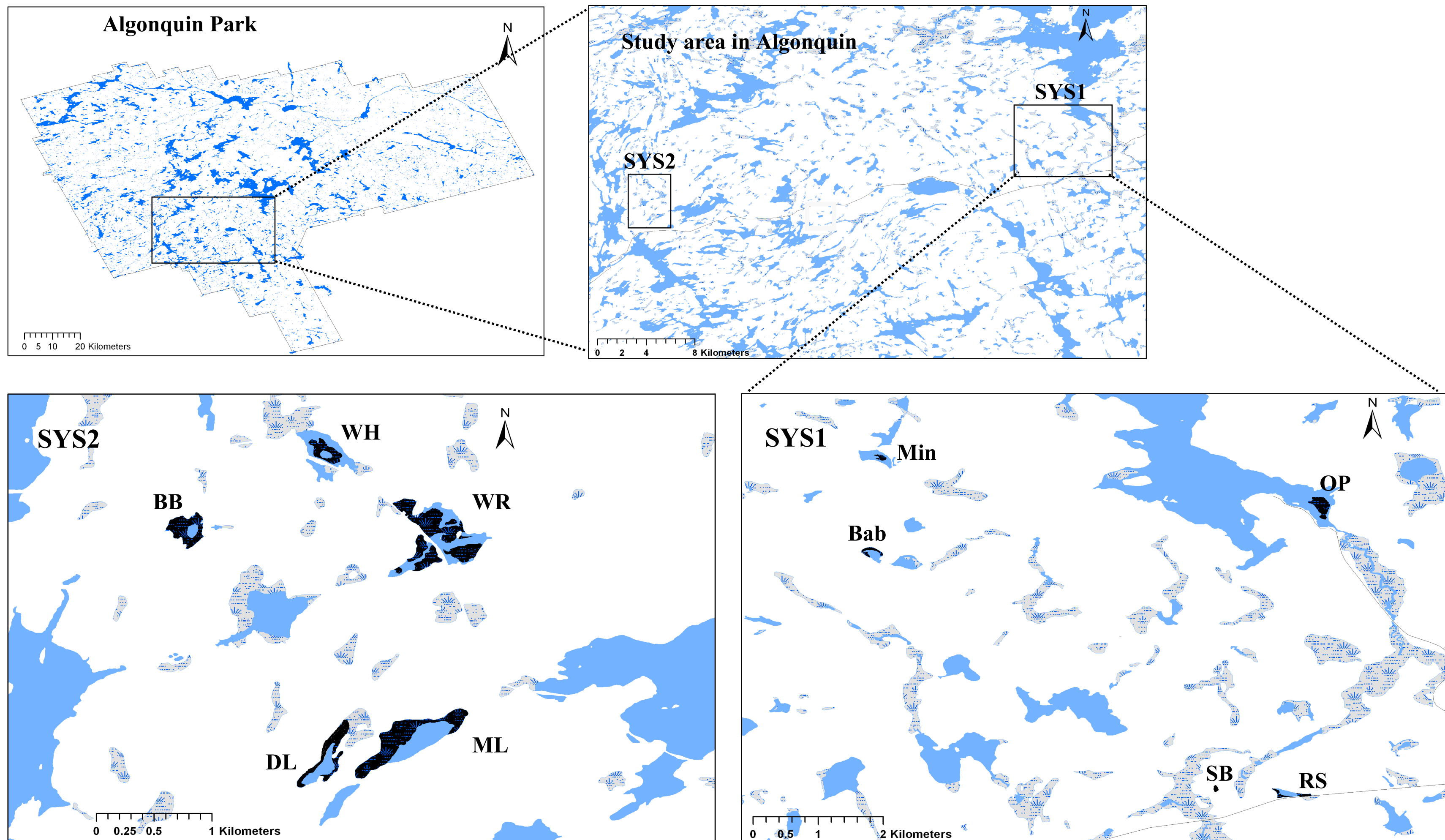


Figure 5.1 Bogs sampled for *M. knabi* larvae in Algonquin Provincial Park (Ontario, Canada) in August 2009. Sampled bogs are shown with black fill. The five nearest neighbouring bogs within each of two 'bog systems' (SYS1 and SYS2) were sampled.

5.2.2 Genetic samples and laboratory analysis

Larvae were removed from each pitcher using plastic pipettes and placed individually in absolute ethanol at -20ºC until DNA extraction. I extracted DNA using the DNeasy blood and tissue kit (QIAGEN, Germantown, MD). All individuals were analyzed at 11 microsatellite loci used in previous *M. knabi* studies (Chapter 4, Rasic & Keyghobadi 2011). Sizing of PCR products was done on a 3730 genetic analyzer using Genemapper software (Applied Biosystems, Carlsbad, CA) with LIZ-500 size standard.

5.2.3 Removal of full-siblings

Collection of larvae, especially at small spatial scales, may produce samples biased towards particular families and, thus, inflate measures of population genetic differentiation (Allendorf & Phelps 1981, Anderson & Dunham 2008, Goldberg & Waits 2010). I therefore removed full-siblings from my analyses.

I first estimated relationship for all pairs of genotyped individuals within each bog. Maximum likelihood estimates of pair-wise relationships were obtained in ML-RELATE (Kalinowski et al. 2006), with 10000 randomizations and 99% confidence level. Relationships were tested between the following categories: full-sibs (FS), half-sibs (HS), unrelated (U) and parent-offspring (PO). Given that the parent-offspring relationship is not possible between larvae collected in the same year, I treated those cases as full-sibs (as in Savage et al. 2010). I then removed all but one individual from each full-sibling family sampled from each cluster.

My final data set (after the removal of full-siblings) consisted of one-hundred-sixty-one individuals from fourteen clusters in SYS1, and one-hundred-eighty-two individuals from fifteen clusters in SYS2.

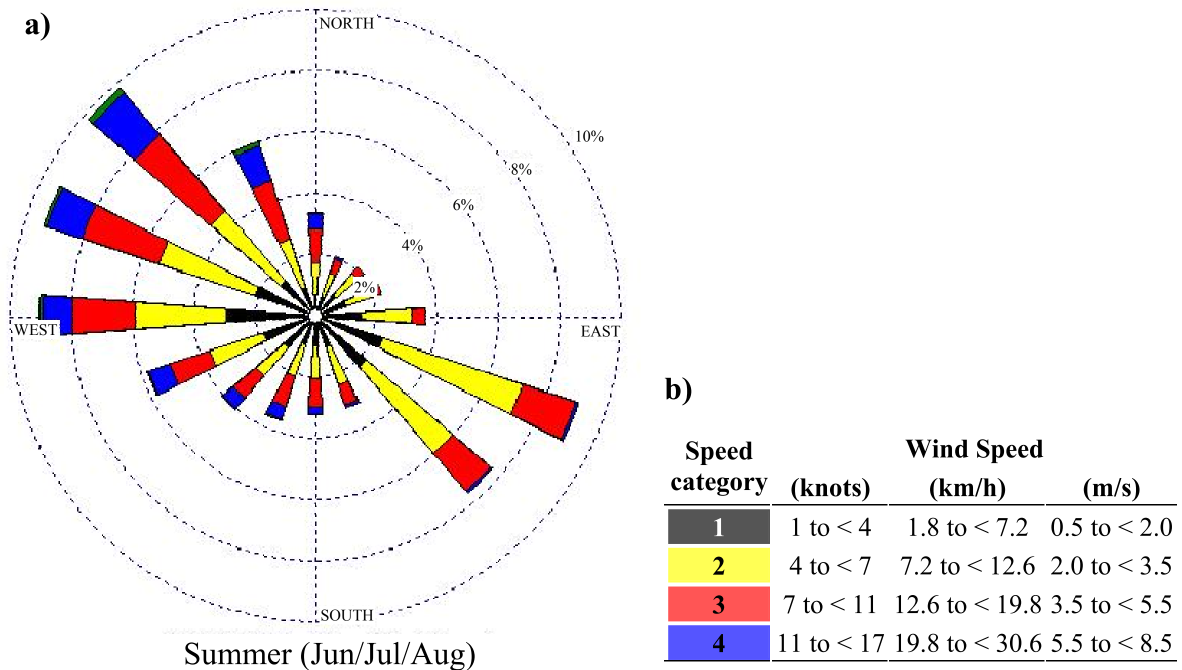
5.2.4 Descriptive statistical genetic analysis

To assess genetic variability for clusters within each system, I calculated descriptive statistics such as the number of alleles per locus, observed heterozygosity and unbiased expected heterozygosity $((2N/(2N-1))*H_e)$, using GenAlEx software (Peakall & Smouse

2006). Departure from Hardy-Weinberg equilibrium for each locus within each cluster was tested with the exact probability test in GENEPOP v.4.0.10 (Raymond & Rousset 1995).

5.2.5 Calculation of wind distances

Wind distance calculations were based on the wind rose diagram, retrieved from the Environment Canada Atmospheric Hazard database (http://ontario.hazards.ca/search/imagemap-e.html?id=1.2383) for the meteorological station nearest to the sampling locations in Algonquin Park. The wind rose used in this study (Figure 5.2) was generated based on hourly measurements of average wind speed and direction for summer months (June-August) from a 30-year period (1971-2000) at Petawawa Airport (ON, CA) (Environment Canada National Climate Data Archive). The full 360 degree range of direction is divided equally into the 16 compass points, meaning each of the compass points (e.g., N, NNE, NE, ENE, E, etc.) represents a 22.5 degree range. I recorded the percent frequency of wind occurrences (of all wind speed classes) from each of the 16 compass points and then summed two percentage values for the same direction (e.g., for N and S, NNE and SSW, NE and SW etc., Table 5.1). This way, I obtained the total percent frequency of all wind speed occurrences for all eight directions. Pairs of clusters were then categorized as follows: clusters oriented relative to each other in a direction in which the wind occurrence was ≥ 20% were given a wind distance value of 1, ≥ 10% and < 20% were given a wind distance value of 2, and < 10% were given a wind distance of 3 (Table 5.1). The ‘wind-distance’ metric should be, therefore, negatively related to the likelihood of wind-mediated movement between any two sample points.



Speed category	Wind Speed (knots)	(km/h)	(m/s)
1	1 to < 4	1.8 to < 7.2	0.5 to < 2.0
2	4 to < 7	7.2 to < 12.6	2.0 to < 3.5
3	7 to < 11	12.6 to < 19.8	3.5 to < 5.5
4	11 to < 17	19.8 to < 30.6	5.5 to < 8.5

Figure 5.2 (a) Wind rose used to calculate wind distances in this study. The wind rose is based on observed hourly measurements of average speed and direction for summer months (June-August) within a 30-year period (1971-2000) recorded in Petawawa Airport weather station (Environment Canada National Climate Data Archive). Each of the extending arms on the wind rose represents one of the 16 compass points from which wind is blowing. Concentric circles extending from the centre of the wind rose represent the percent frequency of the wind occurrences from each compass point. The length of the arm for the specific compass point corresponds to the frequency of the wind occurrences from that point (i.e., the longer the arm, the more frequent the winds). Different colours of each section of the arm represent the wind speed frequency within each speed category. (b) Wind speed categories in knots and their equivalent speeds in kilometres per hour and meters per second.

Table 5.1 Wind frequencies and derived wind distances used in the analysis. Wind frequency (%) for each of 16 compass points was obtained from the wind rose (Figure 5.2). Wind frequencies do not add up to 100%, due to the occurrence of calm periods (no wind). Corrected wind frequency was therefore calculated for wind occurrences without considering calm periods. Total wind frequency along a given direction was then calculated by summing corrected frequencies from the opposite compass points. Finally, wind distance categories based on the total wind frequency along a given direction are reported

compas point	degree	wind frequency (%)	corrected wind frequency (%)	wind direction	Total wind frequency along given direction (%)	wind distance category
N	0 (360)	3.5	4.2	N - S	7.8	3
NNE	22.5	2	2.4	NNE - SSW	6.6	3
NE	45.0	2	2.4	NE - SW	7.2	3
ENE	67.5	2.4	2.9	ENE - WSW	9.8	3
E	90.0	3.6	4.3	E - W	15.1	2
ESE	112.5	9	10.8	ESE - WNW	21.8	1
SE	135.0	7.7	9.2	SE - NW	21.0	1
SSE	157.5	3	3.6	SSE - NNW	10.8	2
S	180.0	3	3.6			
SSW	202.5	3.5	4.2			
SW	225.0	4	4.8			
WSW	247.5	5.8	6.9			
W	270.0	9	10.8			
WNW	292.5	9.2	11.0			
NW	315.0	9.8	11.7			
NNW	337.5	6	7.2			

5.2.6 Statistical analysis

To test the effect of wind on gene flow, I employed analyses that involve assessment of genetic distances between clusters, and subsequent evaluation of relationships between genetic distance and wind distance. I used two genetic distance metrics: (i) the squared genetic distance (GD) of Smouse and Peakall (1999) was calculated as the mean pair-wise individual-by-individual genotypic distance (i.e., for each pair of clusters, genotypic distances were calculated for all pairs of individuals from the two different clusters and then averaged), and (ii) linearized F_{ST} ($F_{ST}/(1-F_{ST})$), calculated from the Analysis of Molecular Variance procedure (Weir & Cockerham 1984, Peakall et al. 1995) in GenAlEx v.6 (Peakall & Smouse 2006). The former metric is based on a measure of dissimilarity of multi-locus genotypes, while the latter is based on sample allele frequencies and provides a measure of genetic variation between samples relative to the variation within.

Pair-wise geographic distances were calculated as log-transformed minimal Euclidean distances between centroids of clusters.

I employed partial Mantel tests to test for correlation between genetic distance and predictor distances (geographic and wind distance), using the ECODIST v.1.1.3 library (Goslee and Urban 2007) in the statistical software package R (R Development Core Team 2007). The partial matrix correlation coefficient (*r*) was calculated using the regression residual method of Smouse et al. (1986). This method allows for the correlation of two matrices, while partialling out the effect of a third matrix. For example, I calculated the correlation between the matrices of genetic and wind distances, while controlling (partialling out) the effect of geographic distance. The test statistic in that case was calculated by constructing two matrices of residuals for (i) regression of the genetic distance on the geographic distance, and (ii) the regression of wind distance on geographic distance. Then the two residual matrices were compared by the standard Mantel test (Mantel 1967), with 10000 permutation tests that randomize row and column order within only one of the distance matrices. Ninety-five percent confidence intervals

for *r* were calculated using a bootstrapping procedure without replacement (Goslee & Urban 2007).

To asses how much of the variation in genetic distance is explained by a model that includes geographic and wind distances, I employed multiple regression on distance matrices (MRM), following the method by Legendre et al. (1994) and Lichstein (2007). MRM allows for regression of a response (genetic distance) matrix on any number of explanatory matrices (in this case geographic distance and wind distance). MRM was also executed with the ECODIST v.1.1.3 library (Goslee and Urban 2007) in R (R Development Core Team 2007). The significance of an MRM model was tested by 10000 permutations of the response (genetic distance) matrix while holding the explanatory matrices constant. The model R^2 and regression coefficients were retained for each permutation to generate null distributions.

5.3 Results

5.3.1 Genetic variability within clusters

Over all 11 loci and clusters, the average number of alleles per locus was 11.49 (SE = 0.32) in SYS1 and 12.13 (SE = 0.21) in SYS2, observed heterozygosity was 0.542 (SE = 0.025) in SYS1 and 0.515 (SE = 0.021) in SYS2, expected heterozygosity was 0.521 (SE = 0.022) in SYS1 and 0.500 (SE = 0.020) in SYS2. For clusters in SYS1, all loci were in Hardy-Weinberg equilibrium. In SYS2, locus MK124 exhibited heterozygote deficiency in three clusters ($P < 0.01$), but this was not significant after the Bonferroni correction for multiple comparisons.

5.3.2 Genetic and wind distances

Average pair-wise genetic distance between clusters was higher in SYS1 (GD = 11.917, F_{ST} = 0.041) than in SYS2 (GD = 12.168, F_{ST} = 0.024). Average pair-wise geographic distance between clusters was higher in SYS1 (3960.3 m, log value 3.598) than in SYS2 (1618.3 m, log value 3.209). Average pair-wise wind distance between clusters was lower in SYS1 (1.56) than in SYS2 (2.04).

5.3.3 Partial Mantel tests and MRM

In SYS1, partial Mantel tests showed a significant effect of wind on genetic distance (Table 5.2), after partialling out the effect of geographic distance, and this was true for both genetic distance metrics (r = 0.347, P = 0.019 for GD; r = 0.412, P = 0.006 for F_{ST}). Conversely, genetic and geographic distances were also highly correlated after controlling for the effect of wind (r = 0.420, $P < 0.001$ for GD; r = 0.409, $P < 0.001$ for F_{ST}).

In SYS2, the effect of wind was not significant after controlling for the effect of geographic distance (r = -0.053, P = 0.706 for GD; r = -0.048, P = 0.684 for F_{ST}), nor was geographic distance significant after controlling for wind distance (r = 0.074, P = 0.063 for GD; r = 0.046, P = 0.178 for F_{ST}) (Table 5.3).

The MRM model was highly significant in SYS1 ($P < 0.001$), explaining 22.5% or 25% of variation in genetic distances between clusters (for GD and F_{ST} metric, respectively, Table 5.4). Variation in genetic distances in SYS2 could not be significantly explained by the MRM model (R^2 = 0.008, P = 0.401 for GD, R^2 = 0.004, P = 0.630 for F_{ST}; Table 5.5).

Table 5.2 Partial Mantel test results for correlation of matrices of genetic distances (mean individual-by-individual genotypic distance GD, and linearized F_{ST}), geographic distances, and wind distances among all pairs of clusters in SYS1

Correlation	Partialled out	Mantel *r*	95% CI	*P*
GD × geographic distance	wind distance	0.420	0.332-0.489	<0.001
F_{ST} × geographic distance		0.409	0.358-0.512	<0.001
GD × wind distance	geographic distance	0.347	0.123-0.472	0.019
F_{ST} × wind distance		0.412	0.277-0.529	0.006

Table 5.3 Partial Mantel test results for correlation of matrices of genetic distances (mean individual-by-individual genotypic distance GD, and linearized F_{ST}), geographic distances, and wind distances among all pairs of clusters in SYS2.

Correlation	Partialled out	Mantel *r*	95% CI	*P*
GD × geographic distance	wind distance	0.074	-0.005-0.183	0.063
F_{ST} × geographic distance		0.046	-0.004-0.133	0.178
GD × wind distance	geographic distance	-0.053	-0.169-0.102	0.706
F_{ST} × wind distance		-0.048	-0.180-0.154	0.684

Table 5.4 Multiple regression on distance matrices (MRM) analysis in SYS1, with regression coefficients and associated *P* values, regression R^2, *F*-statistic for overall *F*-test for lack of fit and associated *P* value. Each column contains model values for mean individual-by-individual genotypic distance (GD) and linearized F_{ST} as the dependent variable, separated by a semicolon (;).

MRM model	Reg. coef.	*P*	R^2	*F*	*P*
Intercept	9.732 ; -0.095	0.997 ; 0.004	0.225; 0.250	14.625; 12.803	0.0005 ; 0.0007
geographic distance	0.487 ; 0.028	0.000 ; 0.000			
wind distance	0.308 ; 0.023	0.022 ; 0.007			

Table 5.5 Multiple regression on distance matrices (MRM) analysis in SYS2, with regression coefficients and associated *P* values, regression R^2, *F*-statistic for overall *F*-test for lack of fit and associated *P* value. Each column contains model values for mean individual-by-individual genotypic distance (GD) and linearized F_{ST} as the dependent variable, separated by a semicolon (;).

MRM model	Reg. coef.	*P*	R^2	*F*	*P*
Intercept	11.778 ; 0.018	0.797 ; 0.714	0.008 ; 0.004	0.415 ; 0.216	0.401 ; 0.630
geographic distance	0.176 ; 0.004	0.101 ; 0.345			
wind distance	-0.066 ; -0.002	0.558 ; 0.605			

5.4 Discussion

Long-term patterns of wind occurrence and direction significantly contribute to gene flow in the pitcher plant midge. However, the explanatory power of wind is landscape-dependent. In SYS1, wind patterns had a significant effect on genetic distances, above any effect of geographic distance. Wind and geographic distance jointly explained 25% of the variation in genetic distances between clusters in this system. Neither wind nor geographic distance explained significant variation in genetic distances in SYS2.

The literature on dispersal in chironomid midges is exceedingly small, although overall active flight appears limited to less than a few kilometres (McLachlan 1983, 1986, Delettre & Morvan 2000). For example, dispersal distances in tanypodine midges are typically < 100 m from the site of emergence (Bohonak 1999). The ephemeral pool midges, *Chironomus imicola* and *Polypedilum vanderplanki*, have a higher propensity for extensive dispersal, with adults flying several hundred meters from the native pool (McLachlan 1983).

Adult chironomids have a short life cycle (Oliver 1971, Huryn & Wallace 2000), which considerably diminishes their potential for long-distance gene flow. To date, very few studies have investigated gene flow, and thus effective dispersal among populations of chironomid midges. Analysis of a mitochondrial COI gene in *Echinocladius martini* suggested that contemporary dispersal by females is mainly restricted to within natal stream channels (Krosch et al. 2011).

Despite long-distance dispersal limitations due to short adult life stage and poor flying abilities, chironomid midges are reported to colonize new habitats rapidly after their formation and this has been associated with the wind-borne dispersal (Oliver 1971). Strong wind (6-7 m/s) was identified as the main factor that expended ranges of adult midges of *Chironomus plumosus* and *Tokunagayusurika akamusi* (Hirabayashi 1991). Using simulated dispersal trajectories based on the air currents, along with the recordings of the wheat midge (*Sitodiplosis mosellana*) densities in balloon-supported yellow traps

located 5-75 m above the ground, Miao et al. (2011) showed that this agricultural pest can disperse with air currents over great distances (~1000 km) within a single year.

Reynolds & Reynolds (2009) have pointed out that the epithet ‘passive’ often applied to the wind-borne dispersal of small winged insects is misleading and should be abandoned. They combined a stochastic model of atmospheric dispersal with simple models of aphid behaviour, and showed that small insects actively navigate their air-borne dispersal and produce enough lift to become neutrally buoyant when they are in updraughts and cease to produce lift when they are in downdraughts. Air currents, in fact, amplify rather than ‘dampen’ the insect’s own movements (Reynolds & Reynolds 2009). Hence, winds can be considered an environmental feature that facilitates dispersal and gene flow in such insects.

A simulation study by Jaquiéry et al. (2011) indicated that landscape genetic analyses are more likely to identify variables that strongly impede dispersal and gene flow as opposed to variables that facilitate them. In this study, I was able to detect a significant effect of wind-facilitated gene flow in the pitcher plant midge, although only in one system. Within the landscapes I investigated, active flight likely also mediates gene flow in this small insect, and is more spatially limited in SYS1 where bogs are situated farther apart. This was evident from the existence of significant positive correlation between genetic and geographic distances for clusters in SYS1 (i.e., isolation-by-distance), and the absence of such correlation in SYS2 (Tables 5.2, 5.3). However, wind-assisted dispersal additionally contributes to the overall pattern of gene flow in this insect, and more so in a landscape with higher wind connectivity and spatial separation among habitat patches (SYS1).

Short Bull et al. (2011) cautioned that if landscape features are not found to influence genetic structure, researchers should not automatically conclude that the features are unimportant to the species' movement and gene flow, and they suggested studies be conducted in multiple landscapes. The importance of sampling multiple landscapes was demonstrated in my study. Namely, the influence of wind was not detected in the landscape (SYS2) where its contribution to overall gene flow would have been predicted

to be small, based on the distances among bogs and their orientation relative to the prevailing winds. Gene flow in SYS2 is most likely shaped mainly by frequent active flight among closely situated bogs. Active dispersal is far less spatially limited in SYS2 than in SYS1, leading to a lower level of differentiation between clusters, the absence of significant correlation between genetic and geographic distances, and significantly high relatedness between individuals from some neighbouring bogs (spatial autocorrelation analysis, Chapter 4). Furthermore, the spatial orientation of bogs and samples within SYS2 is such that their connectivity by frequent winds was lower than in SYS1. Therefore, it was expected that winds would contribute far less to the overall pattern of gene flow within this landscape (SYS2), making wind-assisted gene flow difficult to detect.

Factors other than geographic distance and wind could also be influencing dispersal and gene flow among bogs in the pitcher plant midge. My goal here was specifically to test the effect of prevailing winds on gene flow, a process hypothesized to occur in this and many other small arthropods. Future work that incorporates landscape characteristics such as matrix composition and configuration (i.e., amount of different forest and wetland types, as well as their spatial arrangement in the landscape) could further refine our understanding of the processes and environmental factors determining spatial genetic structure in the pitcher plant midge.

5.4.1 Implications of wind-assisted gene flow

As seen in this study, wind currents can significantly contribute to spatial genetic structure in an organism that may actively navigate via wind-borne dispersal. In cases of entirely passive wind dispersal (e.g., pollen, seeds, small non-winged arthropods), this environmental feature could also be a crucial predictor of spatial genetic structure, yet it has been completely neglected in landscape genetic models. 'Seascape genetic' studies (Hensen & Hemmer-Hensen 2007), on the other hand, have demonstrated the benefit of incorporating an environmental factor such as ocean currents into the models that explain spatial genetic structure in small marine organisms. Seemingly 'chaotic genetic patchiness' of fine-scale population structure (Johnson & Black 1984), often seen in

marine species, is now explained by ocean currents that decouple larval dispersal from Euclidean geographic distance (e.g., White et al. 2010, Alberto et al. 2011).

By influencing dispersal and gene flow, wind patterns within the landscape can have an important consequence on the metapopulation dynamics (extinction/recolonization) and broad-scale genetic structure in this, and other small air-borne organisms. It is therefore important to consider spatial arrangement of habitat patches not only in terms of their relative geographic distances, but also in terms of their orientation to prevailing winds in the landscape (i.e., relative wind distances).

Conservation efforts have greatly benefitted from landscape genetic research that focuses on examining functional connectivity among local populations and designing dispersal corridors to maintain such connectivity, but this has mainly been done for terrestrial vertebrates (Storfer et al. 2010). In small air-born organisms, wind-assisted dispersal and gene flow are important, but unexplored processes contributing to their metapopulation dynamics and functional connectivity. For them, wind-assisted dispersal corridors should be an integral part of the conservation and management plans.

5.5 Summary

In this study, I provided evidence for wind-assisted gene flow in a terrestrial arthropod. Long-term patterns of wind occurrence and direction had a significant effect on genetic distances between samples of the pitcher plant midge. This effect was, however, landscape-dependent. Specifically, wind-assisted gene flow was more pronounced in a landscape with higher wind connectivity and greater geographical separation among bogs. For this species it is therefore important, at larger spatial scales, to consider not only relative geographic separation of habitat patches, but their relative orientation to the frequent air-currents within landscapes. Wind-assisted dispersal and gene flow could be extremely important processes influencing range expansion, metapopulation dynamics and functional connectivity in many small terrestrial arthropods. Despite these important implications, there remains a scarcity of empirical data and explicit hypothesis testing on this matter. The results of my study urge consideration of these processes in future

landscape genetic models explaining spatial genetic patterns and gene flow in species that could 'sail with the wind'.

5.6 References

Alberto F, Raimondi PT, Reed DC, et al. (2011) Isolation by oceanographic distance explains genetic structure for *Macrocystis pyrifera* in the Santa Barbara Channel. *Molecular Ecology* 20:2543-2554.

Allendorf FW, Phelps SR (1981) Use of allelic frequencies to describe population structure. *Canadian Journal of Fisheries and Aquatic Sciences* 38: 1507-1514.

Anderson EC, Dunham KK (2008) The influence of family groups on inferences made with the program STRUCTURE. *Molecular Ecology Resources* 8:1219-1229.

Angers B, Magnan P, Plante M, Bernatchez L (1999) Canonical correspondence analysis for estimating spatial and environmental effects on microsatellite gene diversity in brook charr (*Salvelinus fontinalis*). *Molecular Ecology* 8:1043-1053.

Balkenhol N, Gugerli F, Cushman S, et al. (2009) Identifying future research needs in landscape genetics: where to from here? *Landscape Ecology* 24:455-463.

Bell JR, Bohan DA, Shaw EM, Weyman GS (2005) Ballooning dispersal using silk: world fauna, phylogenies, genetics and models. *Bulletin of Entomological Research* 95:69-114.

Bohonak AJ (1999) Effect of insect-mediated dispersal on the genetic structure of postglacial water mite populations. *Heredity* 82:451-461.

Byrne DN (1999) Migration and dispersal by the sweet potato whitefly, *Bemisia tabaci*. *Agricultural and Forest Meteorology* 97:309-316.

Chapman JW, Reynolds DR, Smith AS, et al. (2005) Mass aerial migration in the carabid beetle *Notiophylus biguttatus*. *Ecological Entomology* 30:264-272.

Compton SG (2002) Sailing with the wind: dispersal by small flying insects. In: Bullock JM, Kenward RE, Hails RS (eds.) *Dispersal ecology: 42nd symposium of the British ecological society*. Cambridge University Press, Cambridge.

Coulon A, Guillot G, Cosson J-F, et al. (2006) Genetic structure is influenced by landscape features: empirical evidence from a roe deer population. *Molecular Ecology* 15:1669-1679.

Cushman SA, Landguth EL (2010) Spurious correlations and inference in landscape genetics. *Molecular Ecology* 19:3592-3602.

Delettre YR, Morvan N (2000) Dispersal of adult aquatic Chironomidae (Diptera) in agricultural landscapes. *Freshwater Biology* 44:399-411.

Delettre YR (1993) Terrestrial Chironomidae: contribution of local emergence to global aerial flow in heterogeneous environment. *Netherlands Journal of Aquatic Ecology* 26:269-271.

Drake VA, Farrow RA (1989) The aerial plankton and atmospheric convergence. *Trends in Ecology and Evolution* 4:381-385.

Epps CW, Palsbøll PJ, Wehausen JD, et al. (2005) Highways block gene flow and cause a rapid decline in genetic diversity of desert bighorn sheep. *Ecology Letters* 8:1029-1038.

Frost WE (1997) Polyphenic wax production in *Abacarus hystrix* (Acari: Eriophyidae), and implications for migratory fitness. *Physiological Entomology* 22:37-46.

Funk WC, Blouin MS, Corn PS, et al. (2005) Population structure of Columbia spotted frogs (*Rana luteiventris*) is strongly affected by the landscape. *Molecular Ecology* 14:483-496.

Geerts B, Miao Q (2005) Airborne radar observations of the fight behavior of small insects in the atmospheric convective boundary layer. *Environmental Entomology* 34:369-385.

Goslee SC, Urban DL (2007) The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software* 22:1-19.

Goldberg CS, Waits LP (2010) Quantification and reduction of bias from sampling larvae to infer population and landscape genetic structure. *Molecular Ecology Resources* 10:304-313.

Hansen MM, Hemmer-Hansen J (2007) Landscape genetics goes to sea. *Journal of Biology* 6:6.

Heard SB (1994) Pitcher plant midges and mosquitoes: a processing chain commensalisms. *Ecology* 5:1647-1660.

Hendrickx G, Gilbert M, Staubach C, et al. (2008) Wind density model to quantify the airborn spread of Culicoides species during north-western Europe bluetongue epidemic. *Preventive Veterinary Medicine* 87:145-161.

Hirabayashi K (1991) Studies on massive flights of chironomid midges (Diptera: Chironomidae) as nuisance insects and plans for their control in the Lake Suwa Area, Central Japan. - Quantitative evaluations of the nuisance of chironomid midges. *Japanese Journal of Hygiene* 46:662-675.

Holderegger R, Wagner HH (2008) Landscape genetics. *BioScience* 58:199-207.

Holzapfel EP, Harrell JC (1968) Transoceanic dispersal studies of insects. *Pacific Insects* 10:115-153.

Huryn AD, Wallace JB (2000) Life history and production of stream insects. *Annual Review of Entomology* 45:83-110.

Jackson PL, Straussfogel D, Lindgren BS, Murphy B, et al. (2008) Radar observation and aerial capture of mountain beetle, *Dendroctonus ponderosae* Hopk. (Coleoptera: Scotylidae) in flight above the forest canopy. *Canadian Journal of Forest Research* 38:2313-2327.

Jaquiéry J, Broquet T, Hirzel AH, Yearsley J, Perrin N (2011) Inferring landscape effects on dispersal from genetic distances: how far can we go? *Molecular Ecology* 20:692-705.

Johnson MS, Black R (1984) Pattern beneath the chaos: the effect of recruitment on genetic patchiness in an intertidal limpet. *Evolution* 38:1371-1383.

Kalinowski ST, Wagner AP, Taper ML (2006) ML-RELATE: a computer program for maximum likelihood estimation of relatedness and relationship. *Molecular Ecology Resources* 6:576-579.

Keyghobadi N, Roland J, Matter SJ, 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: Biological Sciences* 272:553-560.

Krawchuk MA, Taylor PD (2003) Changing importance of habitat structure across multiple spatial scales for three species of insects. *Oikos* 103:153-161.

Krosch MN, Baker AM, Mather PB, Cranston PS (2011) Spatial population genetic structure reveals strong natal site fidelity in *Echinocladius martini* (Diptera: Chironomidae) in northeast Queensland, Australia. *Freshwater Biology* 56:1328-1341.

Landguth EL, Cushman SA, Schwartz MK, et al. (2010) Quantifying the lag time to detect barriers in landscape genetics. *Molecular Ecology* 19:4179-4191.

Legendre P, Lapointe F, Casgrain P (1994) Modeling brain evolution from behavior: A permutational regression approach. *Evolution* 48:1487-1499.

Lichstein J (2007) Multiple regression on distance matrices: a multivariate spatial analysis tool. *Plant Ecology* 188:117-131.

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

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

McLachlan AJ (1986) Sexual dimorphism in midges: strateges for flight in rain-pool dweller *Chironomus imicola* (Diptera: Chironomidae). *Journal of Animal Ecology* 55:261-267.

McLachlan AJ (1983) Life-history tactics of rain-pool dwellers. *Journal of Animal Ecology* 52:545-561.

McRae BH, Beier P (2007) Circuit theory predicts gene flow in plant and animal populations. *Proceedings of the National Academy of Sciences, USA* 104:19885-19890.

McRae BH (2006) Isolation by resistance. *Evolution* 60:1551-1561.

Miao J,Wu Y-Q, Zhen-Xing YU, et al. (2011) Trajectory analysis of long-distance dispersal of the wheat midge, *Sitodiplosis mosellana* (Géhin) (Diptera: Cecidomyiidae), with air current. *Acta Entomologica Sinica* 54:432-436.

Michels E, Cottenie K, Neys L, De Gelas K, Coppin P, De Meester L (2001) Geographical and genetic distances among zooplankton populations in a set of interconnected ponds: a plea for using GIS modelling of the effective geographical distance. *Molecular Ecology* 10:1929-1938.

Murphy MA, Evans JS, Storfer A (2010) Quantifying *Bufo boreas* connectivity in Yellowstone National Park with landscape genetics. *Ecology* 91:252-261.

Murphy MA, Evans JS, Cushman SA, Storfer A (2008) Representing genetic variation as continuous surfaces: an approach for identifying spatial dependency in landscape genetic studies. *Ecography* 31:685-697.

Oliver DR (1971) Life history of the Chironomidae. *Annual Review of Entomology* 16:211-230.

Peakall R, Smouse PE, Huff DR (1995) Evolutionary implications of allozyme and RAPD variation in diploid populations of dioecious buffalograss *Buchloë dactyloides*. *Molecular Ecology* 4:135-147.

Peakall R, Smouse PE (2006) GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6:288-295.

R Development Core Team (2007) http://www.r-project.org.

Rango JJ (1999) Summer phenology of aquatic insect communities inhabiting the leaves of the northern pitcher plant, *Sarracenia purpurea* L. *Northeastern Naturalist* 6:19-30.

Rasic G & Keyghobadi N (2011) From broad scale patterns to fine scale processes: habitat structure influences genetic differentiation in the pitcher plant midge across spatial scales. *Molecular Ecology (accepted with minor revisions, June 22, 2011)*

Rasic G, Maxwell SA, Keyghobadi N (2009) Characterization of microsatellite loci for the pitcher plant midge, *Metriocnemus knabi* Coq. (Diptera: Chironomidae). *Molecular Ecology Resources* 9:1388-1391.

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

Reynolds AM, Reynolds DR (2009) Aphid aerial density profiles are consistent with turbulent advection amplifying flight behaviours: abandoning the epithet ‘passive’. *Proceedings of the Royal Society B: Biological Sciences* 276:137-143.

Savage WK, Fremier AK, Shaffer HB (2010) Landscape genetics of alpine Sierra Nevada salamanders reveal extreme population subdivision in space and time. *Molecular Ecology* 19:3301-3314.

Schwartz MK, McKelvey KS (2009) Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. *Conservation Genetics* 10:441-452.

Selkoe KA, Watson JR, White C, et al. (2010) Taking the chaos out of genetic patchiness: seascape genetics reveals ecological and oceanographic drivers of genetic patterns in three temperate reef species. *Molecular Ecology* 19:3708-3726.

Short Bull RA, Cushman SA, Mace R, et al. (2011) Why replication is important in landscape genetics: American black bear in the Rocky Mountains. *Molecular Ecology* 20:1092-1107.

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

Smouse PE, Peakall R (1999) Spatial autocorrelation analysis of multialleleand multi-locus genetic micro-structure. *Heredity* 82:561-573.

Spear SF, Peterson CR, Matocq M, Storfer A (2005) Landscape genetics of the blotched tiger salamander (*Ambystoma tigrinum melanostictum*). *Molecular Ecology* 14:2553-2564.

Storfer A, Murphy MA, Evans JS, 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.

Tiner RW (1999) *Wetland indicators: a guide to wetland identification, delineation, classification, and mapping*. Lewis Publishers, CRC Press, Boca Raton, FL.

Was A, Gosling E, McCrann K, Mork J (2008) Evidence for population structuring of blue whiting (*Micromesistius poutassou*) in the Northeast Atlantic. *ICES Journal of Marine Science* 65:216-225.

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution* 38:1358-1370.

White C, Selkoe KA, Watson J, et al. (2010) Ocean currents help explain population genetic structure. *Proceedings of the Royal Society B: Biological Sciences* 277:1685-1694.

White TCR (1970) Airborn arthropods collected in South Australia with a drogue-net towed by a light aircraft. *Pacific Insects* 12:251-259.

Wilmer JW, Elkin C, Wilcox C, et al. (2008) The influence of multiple dispersal mechanisms and landscape structure on population clustering and connectivity in fragmented artesian spring snail populations. *Molecular Ecology* 17:3733-3751.

Wright S (1943) Isolation by distance. *Genetics* 28:114-138.

Chapter 6. Summary and Conclusions

I studied three insects (a flesh fly, midge, and mosquito) that use leaves of the purple pitcher plant as larval habitat and I worked towards developing this as a potential model system in landscape genetics. I successfully developed microsatellite markers for two of the insect species (Chapter 2). With these markers, I assessed levels of genetic differentiation across spatial scales and inferred the extent of gene flow in the pitcher plant flesh fly (*Fletcherimyia fletcheri* Aldrich 1916) and the midge (*Metriocnemus knabi* Coquillett 1904) (Chapters 3, 4). Furthermore, I tested explicit hypotheses about the effects of several landscape variables on processes (female oviposition and dispersal) underlying spatial genetic structure across spatial scales in the midge (Chapter 4). I also tested the effect of long-term patterns of wind occurrence and direction on gene flow in the midge (Chapter 5).

My research included sampling of individuals at very fine spatial scales (e.g., multiple leaves in the same plant). Hence, analyses in my study relied on highly variable genetic markers that provide adequate resolution to distinguish closely related individuals. Microsatellite markers offer such variability and resolution (Holderegger & Wagner 2008, Storfer et al. 2010), but have not been developed in any of the pitcher plant insects or even in closely related species. Therefore, I needed to develop microsatellite markers specific to the focal species *de novo.* I employed an enrichment-based protocol (Hamilton et al. 1999) and successfully isolated microsatellite loci for the pitcher plant flesh fly (*F. fletcheri*) and midge (*M. knabi*). For both of these species, I also optimized a protocol for highly efficient genotyping of individuals using multiplexing and multiloading. In the pitcher plant mosquito (*Wyeomyia smithii* Coquillett 1901), however, microsatellite isolation proved to be very problematic, most likely due to existence of microsatellite families associated with transposable elements and further aggravated by the prevalence of null alleles. Similar problems have been encountered in several genera of the same subfamily Culicinae (Pedro PM *personal communication*, Widdel et al. 2005, Chambers et al. 2007). For *W. smithii*, development of molecular markers such as amplified length

polymorphisms (AFLPs) and single nucleotide polymorphisms (SNPs) are a viable alternative.

An understanding of the spatial scales over which dispersal and gene flow occur is necessary to determine the appropriate spatial scales for sampling and other aspects of landscape genetic analysis (Anderson et al. 2010). Elucidating these relevant scales for each of the pitcher plant insects represented a crucial step towards developing them as a model system. Data on the ecology and behaviour of the pitcher plant flesh fly and midge beyond the larval stage, particularly dispersal abilities of adults, are very limited. Furthermore, nothing was known about their spatial genetic structure. In Chapter 3, I investigated the spatial extent of population genetic structure and gene flow in the pitcher plant flesh fly. A small mark-recapture study has shown that adults readily move within a bog and could potentially move among bogs (Krawchuk & Taylor 2003). Concordant with these results, my genetic data indicated a high level of adult dispersal and recent gene flow up to 1.3 km, and spatially limited gene flow (averaged over generations) at scales greater than 10 km. Overall, the pitcher plant flesh fly exhibits metapopulation characteristics of significant structuring, and limited dispersal and gene flow, at larger spatial scales, but populations do not experience frequent local extinctions/recolonizations. Hence, this species appears to contain a mixture of metapopulation and patchy population attributes, which is a phenomenon increasingly noted in empirical studies (Harrison 1991, Sutcliffe et al. 1997).

For the pitcher plant midge, previous ecological studies have made inferences about dispersal based on spatial patterns of larval abundance, which suggest highly limited movement potential (Miner & Taylor 2002, Krawchuk & Taylor 2003). My study, however, indicates that gene flow can occur among closely situated bogs within a landscape (Chapters 4, 5). Active flight likely mediates this process and is more spatially limited in a landscape where bogs are situated farther apart. However, another process, wind-assisted dispersal, additionally contributes to the overall pattern of gene flow in this small insect. Wind-assisted gene flow may be an important process in many small terrestrial arthropods, and my study is the first to provide explicit support for such a hypothesis (Chapter 5). My results have an important implication for landscape genetic

research. Namely, for species that could ‘sail with the wind’, landscape genetic models that explain spatial genetic structure and gene flow should consider not only the geographic separation of habitat patches, but also their relative orientation to the prevailing air-currents across landscapes. Hence, the relationship between landscape configuration and functional genetic connectivity in such species should be modeled with an additional metric of wind connectivity, such as the one developed here.

The essential first step in any landscape-level research is to define the landscape, which is a prerequisite to quantifying landscape characteristics (Pearson 2001). Landscapes can only be defined relative to an organism's perception and scaling of the environment (Wiens 1976), occupying some spatial scale intermediate between an organism's home range and its regional distribution (Pearson 2001). My analyses revealed that the pitcher plant flesh fly and the midge have quite different landscape sizes, proportional to their dispersal abilities. The two groups of bogs (SYS1 and SYS2) in Algonquin Provicial Park represent two distinct landscapes for the pitcher plant midge, and parts of a single landscape for the flesh fly. Furthermore, the two insects respond differently to their landscape characteristics. Local populations of the flesh fly do not show notable differences in genetic diversity measures (allelic richness, gene diversity) despite the variable amount and configuration of the larval habitat within pathes (i.e., bogs) (Chapter 3). For the pitcher plant midge, these variable habitat attributes have significant impact on the partitioning of genetic variability across spatial scales within a landscape (Chapter 4). Specifically, under the conditions of low plant density the midge females are more likely to aggregate eggs locally (within a single leaf or plant). My results are consistent with the finding of Trzcinski et al. (2003) that the ovipositing midge females are more ‘choosy’ when plants are sparce. This behavioural response to habitat characteristics drives the pattern of genetic differentiation at both small (leaf, plant) and broader scales (clusters, bogs) within a landscape.

The genetic implications of habitat fragmentation have received increasing attention over the past decade. The majority of studies appear to find the predicted result that increased habitat fragmentation leads to reduced genetic diversity within populations and greater genetic differentiation among local populations (Keyghobadi 2007). However, a

substantial number of studies do not show such expected responces, or even find results in the revese direction (Keyghobadi 2007). I observed such an unexpected pattern in the pitcher plant midge, where greater isolation and smaller size of habitat pathes (bogs) within a landcape (SYS1) were not accompanied by greater genetic differentiation among local populations, or a reduction in genetic diversity, when compared to a less 'fragmented' landscape (SYS2). This pattern could be explained by the balanced interplay between local dynamics (stable population sizes in an abundant habitat, high population density) and different modes of gene flow (active flight vs. wind assisted dispersal). More specifically, the different processes that I have uncovered to be operating at different spatial scales in the pitcher plant midge can explain why samples from bogs in SYS1 are not more differentiated than those in SYS2, as we might initially expect based on the higher degree fragmentation of bog habitat in SYS1. The pattern results from greater connectivity of populations at large spatial scales as a result of wind-mediated gene flow in SYS1 (Chapter 5), in combination with increased differentiation among samples in SYS2 as a result of female oviposition responses to pitcher plant abundance and distribution (Chapter 4). Swengel & Swengel (2011) found that stable populations with high abundances found in small isolated sites can be common in bog butterflies. Naturally fragmented populations, such as the populations of various bog insect species, could help in elucidating mechanisms behind 'unexpected' patterns of genetic diversity in populations found in anthropogenically fragmented landscapes.

Landscape genetics has recently seen a strong and growing focus on spatial scale questions (Anderson et al. 2010, Cushman & Landguth 2010, Storfer et al. 2010). Genetic patterns result from a potentially complex combination of biological processes operating at different spatial scales (Balkenhol et al. 2009, Anderson et al. 2010). Furthermore, ecological processes and landscape variables can influence genetic variation differentially at different spatial scales (e.g. Murphy et al. 2010). Analysis of genetic structure across multiple scales of sampling can therefore be important for understanding links between genetic and landscape patterns, and the underlying ecological processes. In Chapter 4, I demonstrated that encompassing a large research area (suitable extent), along with refining the resolution of sampling (gradually changing the grain), revealed links among processes and patterns across different spatial scales in the pitcher plant midge. Genetic

differentiation at several scales (clusters, plants, leaves) in *M. knabi* is significantly associated with landscape variables related to habitat size, abundance and spatial arrangement. These broader scale landscape features seem to influence the fine-scale process of female oviposition, which in turn shapes the patterns of genetic differentiation observed at both small and large spatial scales.

Despite recommendations from theoretical and simulation work (Schwartz & McKalvey 2009, Anderson et al. 2010), empirical researchers have not always paid adequate attention to the impact that the sampling design may have on subsequent population and landscape genetic inferences. My results have demonstrated that the scale of sampling, relative to the scales of ecological/evolutionary processes, influences the conclusions that can be drawn in population and landscape genetic studies. Specifically, the process of *M. knabi* female oviposition occurring at the finest scales (among leaves and plants) interacts with the sampling design (collection of juveniles at small spatial scales and before dispersal events) to affect the output from common population genetic analyses conducted at larger scales (Chapter 4). Hence, my work represents one of few empirical studies that explicitly highlight the impact of the spatial scale of sampling on population genetic inference.

Replication at large spatial scales is a challenge in landscape and ecological studies, because each landscape may have its own history and unique features, making true replication technically impossible (Hargrove & Pickering 1992). In Chapter 4, I demonstrated that, although non-identical landscape attributes may limit generalizations and strict statistical inferences, sampling multiple landscapes nonetheless provides the opportunity to detect plasticity of ecological processes and patterns. In *M. knabi*, the spatial pattern of distribution of full-siblings, and hence inferred female oviposition behaviour, was different between the two investigated landscapes (i.e., systems of bogs), and this was explained by differences in landscape configuration (e.g., sizes of bogs). Hence, I provided the evidence that some genetic patterns and underlying processes can be quite landscape-specific.

My studies also demonstrated that our ability to make inferences about the influence of landscape and environmental variables on genetic structure can be landscape-dependent. Landscape features have to be highly variable in order to be supported in landscape genetic models (Short Bull et al. 2011). For *M. knabi*, bog size significantly contributed to the observed pattern of genetic differentiation across spatial scales only in the landscape containing bogs that were more variable in size (Chapter 4). In Chapter 5, I found a significant effect of wind on gene flow in *M. knabi,* but only in a landscape with higher wind connectivity and greater geographical separation among bogs. My work supports the contention that if landscape features are not found to influence genetic structure, researchers should not automatically conclude that the features are unimportant to underlying processes. Landscape genetic hypothesis-testing should ideally be conducted in multiple landscapes to avoid erroneous conclusions about the importance of landscape and environmental features on gene flow and spatial genetic structure.

Model systems should be characterized by tractability, realism and generality (Srivastava et al. 2004). Ecology and evolution have only a few putative model systems, and they meet some but not all of these requirements (Srivastava et al. 2004). The pitcher plant and its inhabitants are a very tractable and natural system. A key question is the extent to which findings from this system can be readily generalized to other taxa in different landscapes. My research has demonstrated that the insect inhabitants of the pitcher plant can be used to address consequential and general questions in landscape genetics. Such questions include the importance of considering spatial scale in describing genetic patterns and inferring underlying processes, as well as the importance of replication in testing landscape genetic hypotheses. This system is therefore a viable model system for addressing specific questions in landscape genetics. The work represented in this thesis has laid a foundation for further, novel research in this system, and has also provided insights that will be of interest to the broader community of landscape genetics researchers.

6.1 References

Anderson CD, Epperson BK, Fortin MJ, et al. (2010) Considering spatial and temporal scale in landscape-genetic studies of gene flow. *Molecular Ecology* 19:3565-3575.

Balkenhol N, Gugerli F, Cushman S, et al. (2009) Identifying future research needs in landscape genetics: where to from here? *Landscape Ecology* 24:455-463.

Chambers EW, Meece JK, MGowan JA, et al. (2007) Microsatellite isolation and linkage group identification in the Yellow fever mosquito *Aedes aegypti. Journal of Heredity* 98:202-210.

Cushman SA, Landguth EL (2010) Spurious correlations and inference in landscape genetics. *Molecular Ecology* 19:3592-3602.

Hamilton MB, Pincus EL, Di-Fiore A, Fleischer RC (1999) Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. *Biotechniques* 27:500-507.

Hargrove WW, Pickering J (1992) Pseudoreplication: a sine qua non for regional ecology. *Landscape Ecology* 6:251-258.

Harrison S (1991) Local extinction in a metapopulation context: an empirical evaluation. *Biological Journal of the Linnean Society* 42:73-88.

Holderegger R, Wagner HH (2008) Landscape genetics. *Bioscience* 58:199-208.

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

Krawchuk MA, Taylor PD (2003) Changing importance of habitat structure across multiple spatial scales for three species of insects. *Oikos* 103:153-161.

Miner JA, Taylor PD (2002) Effects of peatland size and exposure on two species of Diptera inhabiting the pitcher *Sarracenia purpurea* L. *Ecoscience* 9:347-354.

Pearson SM (2001) Interpreting landscape patterns from organism-based perspective. In: Gergel SE, Turner MG (eds.) *Learning landscape ecology: a practical guide to concepts and techniqes.* Springer-Verlag, Heidelberg.

Schwartz KM, McKalvey KS (2009) Why sampling scheme matters? The effect of sampling scheme on landscape genetic results. *Conservation Genetics* 10:441-452.

Short Bull RA, Cushman SA, Mace R, et al. (2011) Why replication is important in landscape genetics: American black bear in the Rocky Mountains. *Molecular Ecology* 20:1092-1107.

Srivastava DS, Kolasa J, Bengtsson J, et al. (2004) Are natural microcosms useful model systems for ecology? *Trends in Ecology and Evolution* 19:379-384.

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

Sutcliffe OL, Thomas CD, Peggie D (1997) Area-dependent migration by ringlet butterflies generates a mixture of patchy population and metapopulation attributes. *Oecologia* 109:229-234.

Swengel AB, Swengel SR (2011) High and dry or sunk and dunked: lessons from tallgrass prairies to quaking bogs. *Journal of Insect Conservation* 15:165-178.

Widdel AK, McCuiston LJ, Crans WJ, Kramer LD, Fonseca DM (2005) Finding needles in the haystack: single copy microsatellite loci for *Aedes japonicus* (Diptera: Culicidae). *The American Journal of Tropical Medicine and Hygiene* 73:744-748.

Wiens JA (1976) Population responces to patchy environments. *Annual Review of Ecology and Systematics* 7:81-120.

Appendix 1 Locations of clusters (cl 1-24) where the pitcher plant midge (*Metriocnemus knabi*) was sampled, in bogs from Algonquin Provincial Park (Ontario, Canada). UTM coordinates (Zone 17N) represent clusters' centroids. *Bog size* is presented as a peatland area (m^2). *Bog* and *cluster plant density* were calculated as the average number of plants per $1m^2$ of a bog or cluster area, respectively. *Cluster connectivity* was calculated as: $\sum \exp(-d_{ij})$, where d_{ij} is a pairwise Euclidean distance in km between centroids of clusters j and i.

Sampling location	**Bog code**	**Cluster code**	**Coordinates**		*bog size* (m^2)	*bog plant density* (#plants/m^2)	*cluster plant density* (#plants/m^2)	*cluster connectivity*
			E	**N**				
Spruce bog	**SB**	**cl1**	705195.92	5052036.35	1895	1.48	2	3.31
		cl2	705176.03	5052055.59			1	3.28
		cl3	705194.11	5052060.97			2	3.31
'Roadside'	**RSB**	**cl4**	705957.61	5051991.23	3395	1.75	6	3.19
		cl5	705934.40	5052032.54			2.5	3.21
		cl6	706130.41	5052008.97			0.5	2.84
Bab Lake	**Bab**	**cl7**	701497.16	5055855.84	7766	0.42	0.5	2.42
		cl8	701396.43	5055919.78			4	2.56
		cl9	701351.19	5055915.38			0.5	2.52
Minor Lake	**Min**	**cl10**	701500.03	5057446.78	8294	2.00	2.5	2.46
		cl11	701466.32	5057401.92			3	2.48
		cl12	701556.67	5057290.41			1.5	2.43
West Rose	**WR**	**cl13**	680915.39	5049168.81	58900	0.15	1	3.21
		cl14	680865.45	5048958.36			0.5	3.43
		cl15	680898.52	5049050.99			1	3.42
Dizzy Lake	**DL**	**cl16**	680389.01	5047151.81	36143	1.00	2	3.76
		cl17	680343.70	5046982.95			2	3.64
		cl18	680244.01	5046847.36			3.5	3.25
Mizzy Lake	**ML**	**cl19**	681128.13	5047302.45	80772	1.27	5	3.79
		cl20	681124.71	5047382.00			7	3.83
		cl21	680968.16	5047380.73			3	3.96
'Buggy' bog	**BB**	**cl22**	679459.58	5049083.45	34532	1.57	1	3.00
		cl23	679438.99	5049235.05			1	2.97
		cl24	679536.95	5049251.28			2	3.03

Appendix 2

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS
Jun 10, 2011

This is a License Agreement between Gordana Rasic ("You") and John Wiley and Sons ("John Wiley and Sons") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by John Wiley and Sons, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	2685321365215
License date	Jun 10, 2011
Licensed content publisher	John Wiley and Sons
Licensed content publication	Molecular Ecology Resources
Licensed content title	Characterization of microsatellite loci for the pitcher plant midge, Metriocnemus knabi Coq. (Diptera: Chironomidae)
Licensed content author	GORDANA RASIC,SHERI A. MAXWELL,NUSHA KEYGHOBADI
Licensed content date	Sep 1, 2009
Start page	1388
End page	1391
Type of use	Dissertation/Thesis
Requestor type	Author of this Wiley article
Format	Print and electronic
Portion	Full article
Will you be translating?	No
Order reference number	
Total	0.00 USD

John Wiley & Sons Limited, England, registered number 641132. Registered office address: The Atrium, Southern Gate, Chichester, West Sussex, United Kingdom. PO19 8SQ.

"Butler, Verity - Chichester" <vbutler@wiley.com>,

Dear Gordana Gojko Rasic,

Thank you for your email request.
Permission is granted for you to use the material requested for your thesis/dissertation subject to the usual acknowledgements and on the understanding that you will reapply for permission if you wish to distribute or publish your thesis/dissertation commercially.

Permission is granted solely for use in conjunction with the thesis, and the article may not be posted online separately.

Any third party material is expressly excluded from this permission. If any material appears within the article with credit to another source, authorisation from that source must be obtained.

Best Wishes
Verity Butler
Permissions Assistant
John Wiley & Sons Ltd.

Reply to:
From: Gordana Gojko Rasic [mailto:grasic@uwo.ca]
Sent: 10 June 2011 13:18
To: Permission Requests - UK; Permissions - US; editorial.office@molecolres.com
Subject: Author unable to obtain the licence!

John Wiley & Sons Limited

Permission granted for:

Molecular Ecology Resources Primer Development Consortium, ALMANY, G. R., DE ARRUDA, M. P., ARTHOFER, W., ATALLAH, Z. K., BEISSINGER, S. R., BERUMEN, M. L., BOGDANOWICZ, S. M., BROWN, S. D., BRUFORD, M. W., BURDINE, C., BUSCH, J. W., CAMPBELL, N. R., CAREY, D., CARSTENS, B. C., CHU, K. H., CUBETA, M. A., CUDA, J. P., CUI, Z., DATNOFF, L. E., DÁVILA, J. A., DAVIS, E. S., DAVIS, R. M., DIEKMANN, O. E., EIZIRIK, E., FARGALLO, J. A., FERNANDES,F., FUKUDA, H., GALE, L. R., GALLAGHER, E., GAO, Y., GIRARD, P., GODHE, A., GONALVES, E. C., GOUVEIA, L., GRAJCZYK, A. M., GROSE, M. J., GU, Z., HALLDÉN, C., HÄRNSTRÖM, K., HEMMINGSEN, A. H., HOLMES, G., HUANG, C. H., HUANG, C.-C., HUDMAN, S. P., JONES, G. P., KANETIS, L., KARUNASAGAR, I., KARUNASAGAR, I., **KEYGHOBADI, N**., KLOSTERMAN, S. J., KLUG, P. E., KOCH, J., KOOPMAN, M. M., KÖPPLER, K., KOSHIMIZU, E., KRUMBÖCK, S., KUBISIAK, T., LANDIS, J. B., LASTA, M. L., LEE, C.-Y., LI, Q., LI, S.-H., LIN, R.-C., LIU, M., LIU, N., LIU, W. C., LIU, Y., LOISEAU, A., LUAN, W., MARUTHACHALAM, K. K., McCORMICK, H. M., MELLICK, R., MONNAHAN, P. J., MORIELLE-VERSUTE, E., MURRAY, T. E., NARUM, S. R., NEUFELD, K., De NOVA, P. J. G., OJIAMBO, P. S., OKAMOTO, N., OTHMAN, A. S., OVERHOLT, W. A., PARDINI, R., PATERSON, I. G., PATTY, O. A., PAXTON, R. J., PLANES, S., PORTER, C., PRATCHETT, M. S., PÜTTKER, T., **RASIC, G**., RASOOL, B., REY, O., RIEGLER, M., RIEHL, C., ROBERTS, J. M. K., ROBERTS, P. D., ROCHEL, E., ROE, K. J., ROSSETTO, M., RUZZANTE, D. E., SAKAMOTO, T., SARAVANAN, V., SARTURI, C. R., SCHMIDT, A., SCHNEIDER, M. P. C., SCHULER, H., SERB, J. M., SERRÃO, E. T. A., SHI, Y., SILVA, A., SIN, Y. W., SOMMER, S., STAUFFER, C., STRÜSSMANN, C. A., SUBBARAO, K. V., SYMS, C., TAN, F., TEJEDOR, E. D., THORROLD, S. R., TRIGIANO, R. N., TRUCCO, M. I., TSUCHIYA-JEREP, M. T. N., VERGARA, P., Van De VLIET, M. S., WADL, P. A., WANG, A., WANG, H., WANG, R. X., WANG, X., WANG, Y., WEEKS, A. R., WEI, F., WERNER, W. J., WILEY, E. O., WILLIAMS, D. A., WILKINS, R. J., WISELY, S. M., WITH, K. A., WU, D., YAO, C.-T., YAU, C., YEAP, B.-K., ZHAI, B.-P., ZHAN, X., ZHANG, G.-Y., ZHANG, S. Y., ZHAO, R. and ZHU, L.

(2009), Permanent Genetic Resources added to Molecular Ecology Resources Database 1 May 2009–31 July 2009. Molecular Ecology Resources, 9: 1460–1466. doi: 10.1111/j.1755-0998.2009.02759.x

Curriculum Vitae for Gordana Rasic

Education

2006-2011	PhD (Biology), University of Western Ontario
2002-2005	MSc (Biology), Faculty of Biology, University of Belgrade
1996-2002	BSc (Biology), Faculty of Biology, University of Belgrade

Awards and Scholarships

2010	UWO Graduate Thesis Research Award
2010	UWO Award for Excellence in Scientific Communication
2002-2004	Graduate Studies Stipend
1996	Undergraduate Studies Stipend

Refereed Contributions

Rasic G, Keyghobadi N (2011) From broad scale patterns to fine scale processes: habitat structure influences genetic differentiation in the pitcher plant midge across multiple spatial scales. *Molecular Ecology* (In Press).

Rasic G, Keyghobadi N (Molecular Ecology Resources Primer Development Consortium et al.) (2009) Microsatellite loci characterization in the pitcher plant flesh fly *Fletcherimyia fletcheri* (Diptera: Sarcophagidae). *Molecular Ecology Resources* 9:1460-1466.

Rasic G, Maxwell SA, Keyghobadi N (2009) Characterization of microsatellite loci for the pitcher plant midge, *Metriocnemus knabi* Coq. (Diptera: Chironomidae). *Molecular Ecology Resources* 9:1388-1391.

Maxwell SA, **Rasic G**, Keyghobadi N (2009) Characterization of microsatellite loci for the western cherry fruit fly, *Rhagoletis indifferens* (Diptera: Tephritidae). *Molecular Ecology Resources* 9:1025-1028.

Rašić G, Stamenković-Radak M, Savić T, Andjelković M (2008) Inbreeding reveals interpopulation differences in inversion polymorphism of *Drosophila subobscura*. *Journal of Zoological Systematics and Evolutionary Research* 46:31-37.

Stamenkovic-Radak M, **Rasic G**, Savic T, Kalajdzic P, Kurbalija Z, Kenig B, Andjelkovic M (2008) Monitoring of the genetic structure of natural populations: change of the effective population size and inversion polymorphism in *Drosophila subobscura*. *Genetica* 133:57-63.

Stamenković-Radak M, **Rasić G**, Cucković M, Savić T, Andjelković M (2008) Effect of microhabitat variability on the body size in *Drosophila subobscura*. *Folia Biologica Krakow* 56:51-56.

Stamenkovic-Radak M, Kalajdzic P, Savic T, Savic M, Kurbalija Z, **Rasic G**, Andjelkovic M (2008) The effect of lead on fitness components and developmental stability in *Drosophila subobscura*. *Acta Biologica Hungarica* 59:47-56.

Conference Contributions (selected list)

October 2010	Entomological Society of Ontario Annual Meeting (oral presentation), Grand Bend, ON
April 2010	Canadian Society for Ecology and Evolution Annual Meeting (oral presentation), Laval University, QC
May 2007	Canadian Society for Ecology and Evolution Annual Meeting (poster presentation), University of Toronto, ON
November 2004	III Congress of Serbian Geneticists (poster presentation), Subotica, Serbia
September 2003	Symposium of Entomologists of Serbia (oral presentation), Ivanjica, Serbia