University of North Dakota
UND Scholarly Commons

Theses and Dissertations | Theses, Dissertations, and Senior Projects

12-2007

The Population Genetics Of The Wood Frog, Rana Sylvatica, Across Its Geographic Range

Tina M. Squire

Follow this and additional works at: https://commons.und.edu/theses

Part of the Biology Commons

Recommended Citation

Squire, Tina M., "The Population Genetics Of The Wood Frog, Rana Sylvatica, Across Its Geographic Range" (2007). *Theses and Dissertations*. 745.
https://commons.und.edu/theses/745

This Dissertation is brought to you for free and open access by the Theses, Dissertations, and Senior Projects at UND Scholarly Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of UND Scholarly Commons. For more information, please contact zeineb.yousif@library.und.edu.

THE POPULATION GENETICS OF THE WOOD FROG, *RANA SYLVATICA*, ACROSS ITS GEOGRAPHIC RANGE

by

Tina M. Squire
Bachelor of Science, Montana State University-Billings, 1994
Master of Science, University of North Dakota, 2000

A Dissertation

Submitted to the Graduate Faculty

of the

University of North Dakota

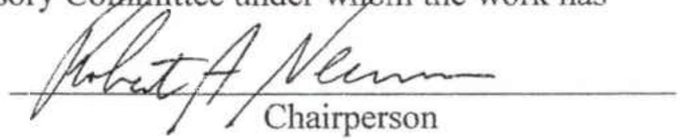
in partial fulfillment of the requirements

for the degree of

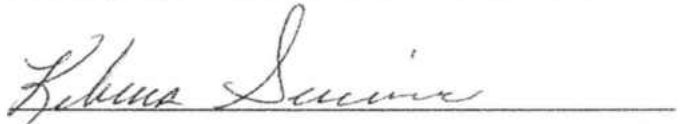
Doctor of Philosophy

Grand Forks, North Dakota
December
2007

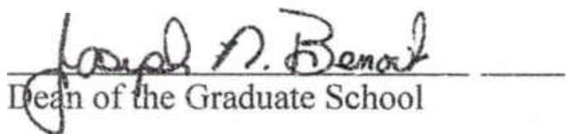
This dissertation, submitted by Tina M. Squire in partial fulfillment of the requirements for the Degree of Doctor of Philosophy from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.



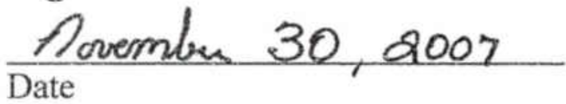
Chairperson



This dissertation meets the standards for appearance, conforms to the style and format requirements of the Graduate School of the University of North Dakota, and is hereby approved.



Dean of the Graduate School



Date

PERMISSION

Title The Population Genetics of the Wood Frog, *Rana sylvatica*

Department Biology

Degree Doctor of Philosophy

In presenting this dissertation in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my dissertation work or, in his absence, by the chairperson of the department or the dean of the Graduate School. It is understood that any copying or publication or other use of this dissertation or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my dissertation.

Signature ______________________

Date 5 Dec 2007

TABLE OF CONTENTS

LIST OF FIGURES v

LIST OF TABLES vi

ACKNOWLEDGMENTS vii

ABSTRACT viii

CHAPTER

I. GLACIATION, MICROSATELLITES AND PATTERNS 1

II. RANGEWIDE ANALYSIS 15

III. REGIONAL PATTERNS 37

IV. CONCLUSIONS 53

APPENDICES 56

REFERENCES 64

LIST OF FIGURES

Figure Page

1. Estimate of the Laurentide Ice sheet at its maximum. 2

2. Geographic range of *Rana sylvatica* outlined in the bold line. 8

3. Sample locations and number of individuals. 21

4. Allelic diversity across all loci. 24

5. F_{IS} values across the geographic range. 25

6. Number of private alleles per population. 29

7. Pairwise F_{ST} values plotted against geographic distance in kilometers 30

8. The log likelihood values for each value of K for all populations.....................30

9. In STRUCTURE output each individual represented by a vertical bar. 31

10. Sample locations and number of samples from each. 40

11. Alleles found in each population for all loci (red) compared to the total number of alleles found in for all populations for all loci (blue). 42

12. Pairwise F_{ST} values plotted against geographic distance in kilometers. 45

13. The log likelihood values for each value of K for all populations...................48

14. Visual representation of STRUCTURE results. 49

LIST OF TABLES

Table Page

1. Loci and primer sequences....19

2. Populations, with the number of different alleles (A) at each locus and the number of individuals genotyped at each locus (N). 23

3. Linkage disequilibrium significance values across all population pairs for each pair of loci. Significant p-values are bolded....26

4. Hardy-Weinberg p-values by population over all loci....26

5. Non-significant genotypic differentiation in population pairs....27

6. Frequency estimations for null alleles by population (for populations estimated to have null alleles) for each locus. 28

7. Combined areas with large numbers of private alleles....34

8. The inbreeding value (F_{IS}), within population variation (1-Q intra) and between population variation (1-Q inter) for each population....42

9. Populations, with the number of different alleles (A) at each locus and the number of individuals genotyped at each locus (N). 43

10. Population pairs with non-significant genotypic differentiation. 43

11. Global Hardy-Weinberg p-values by population....44

12. Linkage disequilibrium across all population pairs for each pair of loci....44

13. Pairwise F_{ST} values for each population pair....46

14. Frequency estimations for null alleles by population (for populations estimated to have null alleles) for each locus. 47

ACKNOWLEDGEMENTS

Funding for this research was provided by North Dakota EPSCoR (OSR-9452892), the National Science Foundation (DEB-9733276), a University of North Dakota Graduate Student Fellowship, an EPSCoR doctoral fellowship, a Wheeler Award, and an APSAC grant.

I would to thank my committee members, past and present, including Dr. Brett Goodwin, Dr. Joseph Hartman, Dr. Colin Hughes, Dr. John La Duke, and Dr. Rebecca Simmons. I also owe special thanks to my advisor, Dr. Robert Newman.

My sincere appreciation goes out to Nikki Squire, daughter and field assistant, as well as friends and family without whose encouragement and support this would not have happened. Thank you Chase for your support.

ABSTRACT

This study aimed to determine the level of genetic variation across the continental-wide range of the wood frog, *Rana sylvatica*. Levels of genetic differentiation between sampled populations were investigated as was the possible locations of glacial refugia for this species. DNA microsatellites were used as the genetic marker. This study found significant genetic differentiation across the geographic range of *Rana sylvatica* that increased with geographic distance. In addition three likely glacial refugia, Alaska, New York and the southern Appalachians, were identified.

A subset of the populations used in the geographic range study was used to investigate the patterns at a regional scale including North Dakota, Minnesota and Manitoba. While glaciation and recolonization would be expected to play a major role in the patterns seen at the geographic range it was unclear if these forces would play such an important role at a smaller scale. Microsatellite DNA showed that while glaciation and recolonization were likely important in the establishment of populations it appears current geographical barriers, such as the Red River of the North, are keeping populations on either side genetically divergent.

CHAPTER I

GLACIATION, MICROSATELLITES AND PATTERNS

Introduction

One of the most important events in recent geologic history affecting both the distribution and population genetics of northern species is glaciation (Alsos et al. 2005). The transition from the Pleistocene to the Holocene (10,000 years ago) marks the end of the most recent glaciation, the Wisconsin glaciation. During the last glacial maximum a large ice sheet, called the Laurentide ice sheet (Figure 1), covered most of present day Canada and extended south of the Great Lakes (Schwalb and Dean 1998). The Laurentide ice sheet lay just east of the Cordilleran ice sheet or ice field that covered western Canada and possibly part of Alaska. Because of the amount of water sequestered in these and other ice sheets sea level was 85 to 130 meters below current level. Lower sea levels exposed more areas of continental shelves. A land bridge connecting Eurasia and North America lasted until approximately 15,500 years ago, following a period when the volume of ice in the world's ice sheets decreased rapidly and sea levels rose (Pielou 1991).

Melting did not occur evenly over a period of time subsequent to glaciation. Periods of stagnation and reverses in the melting may have lasted centuries (Levesque et al. 1997). The "prairie potholes" of the northern Great Plains are the result of small areas

of ice left after the uneven retreat of the main ice sheet leading edge. Fossil evidence indicates many of the small lakes of this region contained fish and bivalves ten thousand years ago, most likely brought in by migrating waterfowl. In other parts of North America large proglacial freshwater lakes formed as the ice melted and the water was trapped (Fenton et al. 1983, Ashworth and Cvancara 1983). Although the exact dates of these lakes is disputed, it is clear that these large bodies of water continually changed over time, impacted the local climate at the time, and the landscape that exists in those areas today (Pielou 1991).

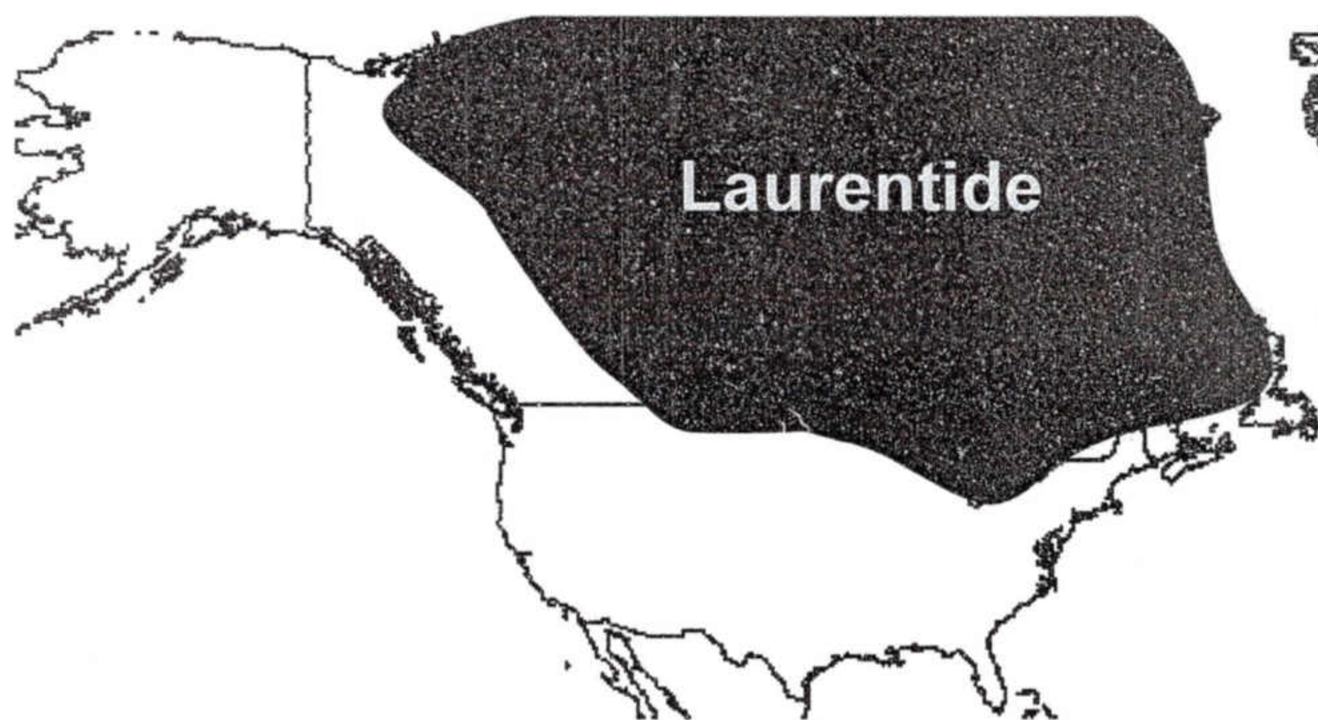


Figure 1. Estimate of the Laurentide Ice sheet at its maximum. Redrawn from Pielou (1991).

Around 12,000 years ago the Laurentide ice sheet had receded to roughly the Canada/United States border. Vegetation quickly colonized exposed land, though the type of vegetation varied by region. In some areas tundra developed, likely due to remaining permafrost. The rate of thawing of permafrost not only affected the vegetation

in the area, but also the animal community. Because the permafrost thawed more gradually than the ice sheets there were associated gradual changes to temperate plant and animal communities. It was not until about 6500 years ago that the Laurentide ice sheet finally disappeared (Pielou 1991).

Following the disappearance of the ice sheets intermittent droughts occurred in the Rocky Mountains and the northern Great Plains that rivaled the intensity of the 1930's dustbowl (Laird et al. 1996). In contrast, the climate of the last 750 years is wetter and cooler than the preceding 1500 years (Laird et al. 1996, Valero-Garcés et al. 1997). These changes in climate greatly affected the population genetics of the regional fauna due to local extinctions and recolonizations (Pielou 1991).

At the end of the Wisconsin glaciation populations could again expand into the newly available terrain. The nature of the range expansion and how quickly it occurred would affect the population structure of local species. Rapid northward expansion would lead to a reduction in allelic diversity in northern populations because the relatively few founders of new populations will likely not carry all of the alleles present in the original population (Nichols and Hewitt, 1994; Ibrahim et al, 1996; Hewitt 2001, 2004). Ibrahim et al.'s (1996) modeling study found rapid leptokurtic dispersal of individuals (i.e. long distance dispersal occurring more often that would occur under a normal distribution of dispersal distances) produced patches of homogeneity. Essentially long distance founders would be able to colonize an area and expand the population to fill that suitable habitat and nearby habitats. Because these few, subsequent founders would again carry fewer alleles than the original population or the secondary source population, these areas would experience genetic homogeneity.

A more normal distribution of colonization distances, more recently called Hewitt's "phalanx" model (Nichols and Hewitt, 1994), expects recently founded populations to retain more genetic diversity and produce less population structure. In this type of recolonization founding individuals tend to colonize nearby suitable habitats more often and further away habitats less often. In this case, the main front of colonists rather than a few long distance colonizers expands to fill by nearby areas. This pattern of colonization allows more gene flow between the source and the newly founded populations and potentially larger numbers of colonists. Large numbers of colonists and increased gene flow would allow for retention of genetic variation seen in the refugial population. Populations in refugial areas would have been able to both diverge genetically and become geographically subdivided as genetic drift, mutation, nonrandom mating and lack of gene flow would all potentially have had time to act (Ibrahim et al, 1996; Hewitt 2001, 2004).

These modes of recolonization have been supported by many studies of species occupying previously glaciated areas in North America. In the Stellar's jay, *Cyanocitta stelleri,* and chestnut-backed chickadee, *Poecile rufescen,* genetic similarity measured by microsatellites between populations in previously glaciated regions is not found in geographically proximal populations (Burg et al. 2005, Burg et al. 2006, respectively) as would be expected under the phalanx model of colonization. Chestnut-backed chickadees show significant genetic structure in the more recently colonized areas of their range, though the northwest song sparrow, *Melospiza melodia,* showed a different pattern. In the northwest song sparrow geographically close populations were the most genetically similar, suggesting a stepping-stone model of recolonization in which

A number of amphibians show a similar east-west split to song birds. In leopard frogs, *Rana pipiens,* there are two main genetic lineages seen in mitochondrial DNA, a western and an eastern lineage (Hoffman & Blouin 2004). The western refugia appears to be located in Nebraska, while the eastern refugia occurs from Illinios to New York. The zone of overlap between these lineages is in central Ontario. A study that included microsatellites also found that west and east leopard frogs were genetically dissimilar (Hoffman & Blouin 2004). A study of spring peepers revealed that northern population are closely related and probably originated from a single refuge, or from closely related refugial populations (Austin et al. 2002). Austin et al. (2002) concluded that dispersing spring peepers went north into eastern Ontario, then westward into Manitoba. The authors hypothesize that historical barriers to dispersal may have prevented western (Arkansas to Kansas to southwest Illinois) haplotypes from expanding north into Ontario and Manitoba. Diverse haplotypes in Virginia and Ontario suggested a possible recent colonization event from multiple refugia. The possibility that a single refugial population containing very high levels of genetic diversity or that gene flow is higher than previously thought in anurans cannot be ruled out (Austin et al. 2002). The evidence is less clear in the toad *Bufo fowleri* (Smith and Green 2004). In this species, populations on the Canadian shore of Lake Erie possess similar mitochondrial DNA haplotypes to populations on the opposite shore. Another population on the Canadian shore of Lake Erie is most similar to a population on Lake Michigan, geographically further away than other populations along the Canadian shore of Lake Erie. These patterns obscure postglacial routes of recolonization (Smith and Green 2004).

There is limited evidence on which to base any inference of wood frog phylogeographic history. Currently, they are found throughout the Appalachian Mountains as far south as northern Georgia, as well as west of the Mississippi in Missouri (Stebbins 1985), and disjunct locations in Arkansas and northern Colorado/southern Wyoming (Figure 2) although there is some evidence that the Rocky Mountain populations may represent a separate species (Porter 1969). Wood frogs probably persisted during the Pleistocene in the southeastern (including southern areas of what is referred to as the Midwest) and south central plains regions of the United States, based on their current and historical distributions and may have survived glaciation northwest of the continental ice sheets as well (Stewart and Lindsay 1983). Colonization of the formerly glaciated parts of the range, therefore, may have come from multiple sources. One hypothesis is that populations in the glaciated northeast would have come north from the southern Appalachians or mid-Atlantic (an east of the Mississippi clade), whereas populations in the north central (northern plains north into Canada) and northwest (western Canadian provinces and Alaska) would have come from the south central plains and/or Rocky Mountains (a west of the Mississippi clade). Alternatively, the westernmost portion of the range (Alaska, northwestern Canada) may have been colonized from a northwestern refuge. The logic of the hypothesis of an east-west split stems from both the current or historical distribution and that the Mississippi River and glacial Lake Agassiz represented major barriers to dispersal. Habitat changes subsequent to the retreat of glaciers permitted the wood frog to persist in the southeast, but largely eliminated presumptive populations from the south-central plains.

Supporting the east to west recolonization hypothesis, several morphological characters exhibit clines following an east-west pattern (Martof and Humphries 1959). It is possible; however, these characters may be under selection driven by climate or habitat, and may not simply reflect colonization history (Fishbeck and Underhill 1971). Based on museum specimens, Martof and Humphries (1959) found that body size decreases gradually going north from the southern Appalachians, with a weaker gradient in the western portion of Canada. Relative leg length shows the same overall pattern

Figure 2. Geographic range of *Rana sylvatica* outlined in the bold line. Redrawn from Stebbins (1985).

(Martof and Humphries 1959). The middorsal white stripe present in most Alaskan and northern Canadian individuals is absent from the southeastern portion of the geographic range of wood frogs (Martof and Humphries 1959). Dorsal stripe is thought to be inherited in a simple Mendelian fashion (1 locus, with 2 alleles, Browder et al. 1966), and

may, therefore, be treated as a single-locus genetic marker. The pattern (albeit based on limited sampling) appears to exhibit a relatively abrupt transition in the northern plains and in northeastern Canada, with striped individuals absent in the east and southeast, infrequent at midlatitudes in Canada, and common in the northwestern portion of the range. The Rocky Mountain populations also exhibit a high frequency of striped individuals, with Colorado and Wyoming populations apparently fixed for the striped allele (Martof and Humphries 1959).

Reconstruction of the biogeographic history of a species may be based on a variety of sources. Fossil evidence is available for some taxa (e.g., Holman 1995, Stewart and Lindsay 1983), but is generally limited both in spatial coverage and time period. Wood frog fossils dating to the Pleistocene are known from the central plains (Kansas, Nebraska), the Midwest (southern Indiana), as well as numerous sites in the Appalachian Mountains (Holman 1995), suggesting these areas as possible refugia. Genetic data have become the tool of choice for phylogeographic studies (Avise 2000) in the absence of or to complement fossil evidence.

The field of population genetics and in particular molecular population genetics has grown as a field of study in the last few decades due to the increasing understanding and use of molecular markers (Avise 2004). Molecular population genetics allows insights into the development of a species' range and the impact of habitat and climatic change on the population genetics within a species. Understanding the history of a species' response to historical environmental changes can also allow predictions of the potential impact of future environmental change (Davis and Shaw 2001).

Microsatellites

Microsatellites are tandem repeats found throughout the genome. Microsatellite loci most often contain 5 to 40 repeats of 2-4 nucleotides. Microsatellites are highly variable (Selkoe and Toonen 2006 & refs therein, Eisen 1999). In many eukaryotes microsatellites occur almost entirely in noncoding regions (see Li et. al. 2002 for review). Triplet repeats, because they do not cause frameshift mutations, are more likely than dinucleotide or tetranucleotide repeats to be found in protein coding regions of the genome (Selkoe and Toonen 2006 & refs therein & Hancock 1999).

Microsatellites have generally been thought to mutate through one of two processes. The first is slippage during DNA replication and the second is genetic recombination (Selkoe and Toonen 2006, Eisen 1999, Hancock 1999). Studies in yeast and *E. coli* support replication slippage as the primary mechanism for microsatellite length change. Allele lengths often cluster, further supporting these findings (Hancock 1999).

Microsatellites have desirable traits and advantages over other genetic markers. Microsatellites are preferable to allozymes and mitochondrial DNA for many applications because of their high mutation rate, which varies greatly between species, resulting in greater variation (Avise 2004, Zhang and Hewitt 2003) and more precise estimates of genetic distance between populations (Kalinowski 2002). Mitochondrial DNA is also a single locus while microsatellites offer more loci (Selkoe and Toonen 2006). More loci also allow more precise estimates of genetic distance between populations (Kalinowski 2002).

There are logistic advantages to microsatellites as well. Sample preparation is easier than for other genetic markers due to only small pieces of DNA being required to be amplified. Also, because of the small fragment size, DNA degradation is less of an issue (Selkoe and Toonen 2006). Because the primers for each locus are designed based on the DNA sequence flanking the microsatellite the chances of primers amplifying a region of DNA in other unrelated taxa during PCR is very low. Cross-contamination is therefore less of an issue as well (Selkoe and Toonen 2006). Additionally, microsatellites are inherited in a Mendelian manner and are generally presumed to be selectively neutral as they tend to occur in noncoding regions (Selkoe and Toonen 2006).

Like all genetic markers microsatellites are not without their drawbacks. Primer site mutations may cause alleles to be unamplifiable (null alleles), though primers can often be redesigned to overcome this (Selkoe and Toonen 2006). There is always the chance of homoplasy, alleles that are identical in size but are not identical by descent (Avise 2004, Selkoe and Toonen 2006). Homoplasy is detectable if it can be detected when alleles are sequenced. Selkoe and Toonen (2006) state such detectable homoplasy is uncommon unless the repeat region of the microsatellite is either compound or interrupted. Undetectable homoplasy is a potential problem for all types of genetic markers (Selkoe and Toonen 2006).

Technical and human factors can both introduce errors to any study (Bonin et al. 2004). Allele dropout and false alleles (stuttering) are additional problems that have recently been given more attention. Allele dropouts are alleles that do not amplify in heterozygotes, causing those individuals to appear as homozygotes (Broquet and Petit 2004, Bonin et al. 2004). This type of error is locus specific and might be mitigated with

rigorous screening of loci before their use (DeWoody et al. 2006). False alleles are alleles that appear due to slippage during the initial cycles of the polymerase chain reaction (PCR) (Broquet and Petit 2004, Bonin et al. 2004). Bonin et al. (2004) found ten errors in genotyping in 1209 alleles in Scandinavian brown bears, *Ursus arctos*. Six of these errors were allelic dropouts while false alleles or contaminants accounted for the other four. These ten errors equate to a 0.8% error rate or 17.6% of 34 samples' multilocus genotypes with at least one error. Error rates were higher for DNA extracted from fecal samples (Bonin et al. 2004). Stuttering can be especially troublesome with dinucleotide repeats. Extensive screening and elimination of loci with this problem is the easiest way to eliminate this issue (DeWoody et al. 2006).

Microsatellites were chosen as the genetic marker of choice in this study for a variety of reasons. These microsatellite markers were used in previous studies (Newman and Squire 2001, Squire and Newman 2002) and therefore were already available and determined to be potentially informative for this study. Mitochondrial DNA sequencing was not chosen because it is generally not as polymorphic as microsatellites (Avise 2004, Zhang and Hewitt 2003). Once this study was begun it was learned that a similar study using mitochondrial DNA was underway. The results from this study confirmed that mitochondrial DNA was too conserved to bring to bear on this study (Irwin, pers. comm.). At the time this study was begun other markers such as AFLPs and SNPs were not as commonly used as they are today; if this study were continued I would likely include these markers as well, but that is beyond the scope of this dissertation.

Objectives

This study was undertaken to investigate the patterns of genetic variability and differentiation in *R. sylvatica* across its range. The patterns investigated were both continent wide (Chapter 2) and in the upper Great Plains region (Chapter 3). A corollary to the main goals of the project was a look at the usefulness of microsatellites, a marker generally used to investigate fine-scale structure (Newman and Squire 2001, Squire and Newman 2002), on species range-wide scale (Chapter 4). The specific hypotheses tested are as follows:

1. Genetic variation in *R. sylvatica* will be less in previously glaciated and more recently recolonized areas than in putative glacial refugial areas.

Morphological evidence shows clinal variation in a number of phenotypic traits. Body size decreases gradually going north from the southern Appalachians (a putative glacial refugia) with a weaker gradient in the western portion of Canada (Martof and Humphries 1959). Relative leg length shows the same overall pattern (Martof and Humphries 1959). The middorsal white stripe present in most Alaskan and northern Canada individuals is not found in individuals from the southeastern portion of the wood frog's geographic range (Martof and Humphries 1959). The morphological evidence suggests significant differences in genetic variation across the range of this species. Based on the pattern of morphological variation a pattern of significant genetic differentiation across the geographic range of the species is expected. Additionally, populations exhibiting the same phenotype in the morphological traits described would be expected to be more closely genetically related than those not exhibiting the same phenotype.

2. ***R. sylvatica*** **found glacial refugia in the southern Appalachian region of the southeast United States.**

Wood frogs are cold tolerant, typically found in wooded areas and are currently found at high elevations in the Appalachians. The southern Appalachians region also forms the southern boundary of the wood frog's geographic range (Stebbins 1985). Martof and Humphries (1959) identified an Appalachian phenotype (among other phenotypes) and proposed possible refugia in this area. This glacial refugium should show a greater genetic diversity than more recently colonized regions, and possibly unique genotypes due to the longer time during which they accumulated mutations (Taberlet and Cheddadi 2002).

3. A genetic pattern should be apparent if Manitoba, Canada was a point of convergence for dispersing ***R. sylvatica*** **during range expansion.**

An area of convergence would show an increase in genetic variation over geographically proximal areas. In a previous study of *Rana sylvatica,* Dr. Jason Irwin found strong support for an eastern clade that included samples from the southeastern United States north through the Atlantic states into Canada, that was divergent from a western clade that included samples from the upper Midwest (Minnesota and North Dakota), western Canada, Alaska, and Colorado and Wyoming based on a 562 base pair sequence of the mtDNA cytochrome *b* gene. The zone of overlap appeared to be in Manitoba (Irwin, unpublished data). If the zone of overlap is in Manitoba then we should see a mixture between eastern and western alleles and higher allelic diversity than populations either east or west of this zone.

CHAPTER II

RANGEWIDE ANALYSIS

Abstract

Wood frogs, *Rana sylvatica,* are freeze tolerant. This trait has allowed this species one of the northernmost geographic ranges of all amphibians, providing an exceptional biological system in which to examine the effects of the retreat of the glaciers. Six microsatellite loci were used to identify large scale patterns of genetic variability and differentiation in the wood frog. These patterns have been affected by the pattern of glaciation, the location of glacial refugia, modes of recolonization, and areas of secondary contact.

Introduction

During the last glacial maximum the Laurentide ice sheet covered most of present day Canada and extended further south than the Great Lakes. The Laurentide ice sheet lay just east of the Cordilleran ice sheet or ice field that covered western Canada and possibly part of Alaska. By around 12,000 years ago the Laurentide ice sheet had receded to roughly the Canada/United States border but it was not until about 6500 years ago that this ice sheet finally disappeared (Pielou 1991). In the northern Great Plains the glaciers were followed by intermittent severe droughts lasting over a hundred years. It has not been until the last 750 or so years that a cooler and wetter climate has come about

in this region (Laird et al. 1996). One of the most important events in recent geologic history affecting northern species range and genetic makeup is glaciation (Alsos et al. 2005.

The recolonization histories of a variety of North American species have been inferred from molecular markers (e.g., Good and Sullivan 2001, Milot, et al. 2000, Nesbø et at. 1999, Stone et al. 2002). In addition to spatial clines in allele or haplotype frequencies resulting from expansion of colonists from refugial populations, populations in glaciated regions tend to exhibit lower differentiation among populations than among populations persisting in isolated glacial refugia, as well as reduced levels of genetic variability (reviewed in Hewitt 1999, 2000). Rapid northward expansion would lead to a reduction in genetic diversity in a species most northern populations. A slower recolonization would allow populations to retain more of their genetic diversity throughout their range and allow refugial areas time to both diverge genetically and become geographically subdivided (Hewitt 2001, 2004). Patterns of variation may be complex, however, depending on the number and location of refugial populations, routes and timing of colonization of formerly glaciated regions, and locations (if any) of zones of contact between colonists derived from different refugia (Ibrahim et al. 1996).

This study was undertaken to look specifically at three questions listed below.

1. Is genetic variation in *R. sylvatica* less in previously glaciated and more recently recolonized areas than in putative glacial refugial areas?

Morphological evidence shows clinal variation in a number of phenotypic traits in *R. sylvatica* (Martof & Humphries 1959). The morphological evidence suggests significant differences in genetic variation across the range of this species with those

populations exhibiting the same clinal traits will have similar alleles to those of other populations.

2. Does *R. sylvatica* have a glacial refugia in the southern Appalachian region of the southeast United States?

The southern Appalachians region forms the southern boundary of the wood frog's geographic range and has been suggested as refugia for a number of species, including the Northern short-tailed shrew *Blarina brevicauda* (Brant and Orti 2003), the pygmy salamander *Desmognathus wrighti* (Crespi et. al. 2003) and the spring peeper *Pseudacris crucifer* (Austin et. al. 2002). Martof and Humphries (1959) identified an Appalachian phenotype (among others) and proposed possible refugia for *Rana sylvatica* in this area. This glacial refugia should show a greater genetic diversity than more recently colonized regions, and possibly unique genotypes (Taberlet and Cheddadi 2002).

3. Is there a point of convergence for dispersing *R. sylvatica* during range expansion in Manitoba, Canada?

An area of convergence would show in increase in genetic variation over geographically proximal areas. Strong support for an eastern clade that included samples from the southeastern United States north through the Atlantic states into Canada, that was divergent from a western clade that included samples from the upper Midwest (Minnesota and North Dakota), western Canada, Alaska, and Colorado and Wyoming was found in a previous study with a possible zone of overlap in Manitoba (Irwin, unpublished data). If the zone of overlap is in Manitoba then we should see a mixture between eastern and western alleles and higher allelic diversity than populations either east or west of this zone.

Methods

Sample Handling

A total of 547 samples were collected from 25 populations (individual or neighboring ponds) across North America (see Figure 3 locations and the number of samples of each and Appendix 1A for collectors). Tissue samples (toe tips or hatchlings (one per egg mass)) were stored and transported in 95% ethanol. DNA was extracted by a standard phenol-chloroform extraction and an ethanol precipitation (Hitchings & Beebee1998). Six microsatellite loci were used (Table 1). PCR reactions were conducted in 10ul volumes and contained 1X buffer, 0.5 uM dNTPs, 2.5 mM $MgCl_2$, 1uM unlabeled primer, 1uM fluorescently labeled primer, and 0.5 units taq DNA polymerase (Promega). The PCR reactions were carried out in a BIORAD thermocycler. For Rs23 (Newman and Squire 2001) the cycles consisted of 3 minutes at 94°C, and 3 minutes at 55°C for one cycle, 1 minute at 72°C, 1 minute at 94°C, 1 minute at 55°C for 30 cycles and a 3 minute extension at 72°C. For all other primers (see Julian and King 2003) the PCR reaction cycles were 3 minutes at 94°C for one cycle, 1 minute at 94°C, 30 seconds at 54°C, and a final 1 minute at 72°C for 30 cycles, and a 2 minute extension step at 72°C. PCR products were run on a 2% agarose gel and stained with ethidium bromide (2ug/ml). Products in the correct size range were then run on an ABI model 3100 (Applied Biosystems) using 0.5ul 500ROX as a size standard in each sample. ABI output was using ABI GeneScan software (version 3.7, 2000) and allele sizes determined by ABI Genotyper software.

Table 1. Loci and primer sequences.

Locus	Primer sequence
Rs11	F: 5′-TTACTTTCAGTTTCAAAAGGCAG-3′ R: 5′-TACACAGTGCTTCACAAGTTCC-3′
Rs23	F: 5'-CACCGATTTTTTTTAATACAG-3' R: 5'-AACCCGGTTAGTGTATTGTC-3'
Rs32	F: 5′-GGACACACAATTCCTTGGTTC-3′ R: 5′-GAGGAGATTTCCAAAACAATCC-3′
Rs55	F: 5′-GAGTTGGGACTCCTGAATAGAG-3′ R: 5′-AGTCTTTGCTTTGTAAATTGGC-3′
Rs70	F: 5′-CAAAGTGCACAGTCTATTTGTC-3′ R: 5′-TGGACCATTTGGTTTATTTGTC-3′
Rs88	F: 5′-TCAATCCATCAGTCTGTCTGTC-3′ R: 5′-GGATTTTGTAAAGAATGCTCCTC-3′

Data Analysis

Characterization of genetic diversity (allele frequencies, genic diversity, tests for Hardy-Weinberg equilibrium) was conducted using GENEPOP version 4 (Rousset 2007). Levels of inbreeding (F_{IS}), intrapopulation and interpopulation genetic variation were calculated and linkage disequilibrium between pairs of loci was tested for in GENEPOP using a log likelihood ratio statistic (Rousset 2007). GENEPOP was also used to test for differentiation in allele frequencies among pairs of populations using a Markov Chain Monte Carlo approximation of Fisher's exact test (Raymond and Rousset 1995). Because of the large num r of pairwise tests involved, the significance of pairwise comparisons was determined using a sequential Bonferroni adjustment (Rice 1989). Pairwise F_{ST} (averaged over loci) were estimated in GENEPOP and used in analyses of isolation by distance, with a Mantel test of significance in IBDWS 3.14 (Jensen et al. 2005). Presence and estimation of null allele frequencies was examined in GENEPOP (Rousset 2007) and

MICRO-CHECKER (Van Oosterhout 2004) was used to investigate the possibility of large allele dropout, evidence of scoring errors and presence of null alleles and estimation of null allele frequencies.

Bayesian methods implemented via STRUCTURE 2.2 (Pritchard *et al.* 2000), were used to delineate groups of individuals or clusters within the data set. STRUCTURE does not assume a mutational model or process but uses individuals' genotypes to determine the population(s) to which they should be assigned. The model accounts for Hardy–Weinberg or linkage disequilibrium by introducing population structure and attempts to find population groupings that are not in Hardy-Weinberg disequilibrium (Pritchard et al. 2000). The number of clusters or groups (K) was set from K=2 to K=25. For each setting for K the log transformed likelihood of the data matching that value of K was determined by STRUCTURE. All of the log likelihood values were plotted and the smallest value determined the most likely K value. After the most likely number of groups was identified each group was subsequently separated and the most likely number of clusters within this grouping was determined. STRUCTURE 2.2 analyses were based on genotypes at 5 loci (Rs11, Rs23, Rs32, Rs55, Rs88). Locus Rs70 was removed from the STRUCTURE analysis. Removal of this locus was done to avoid introduction of the appearance of a departure from Hardy-Weinberg in the absence of true population structure as recommended by Pritchard et al. (2000).

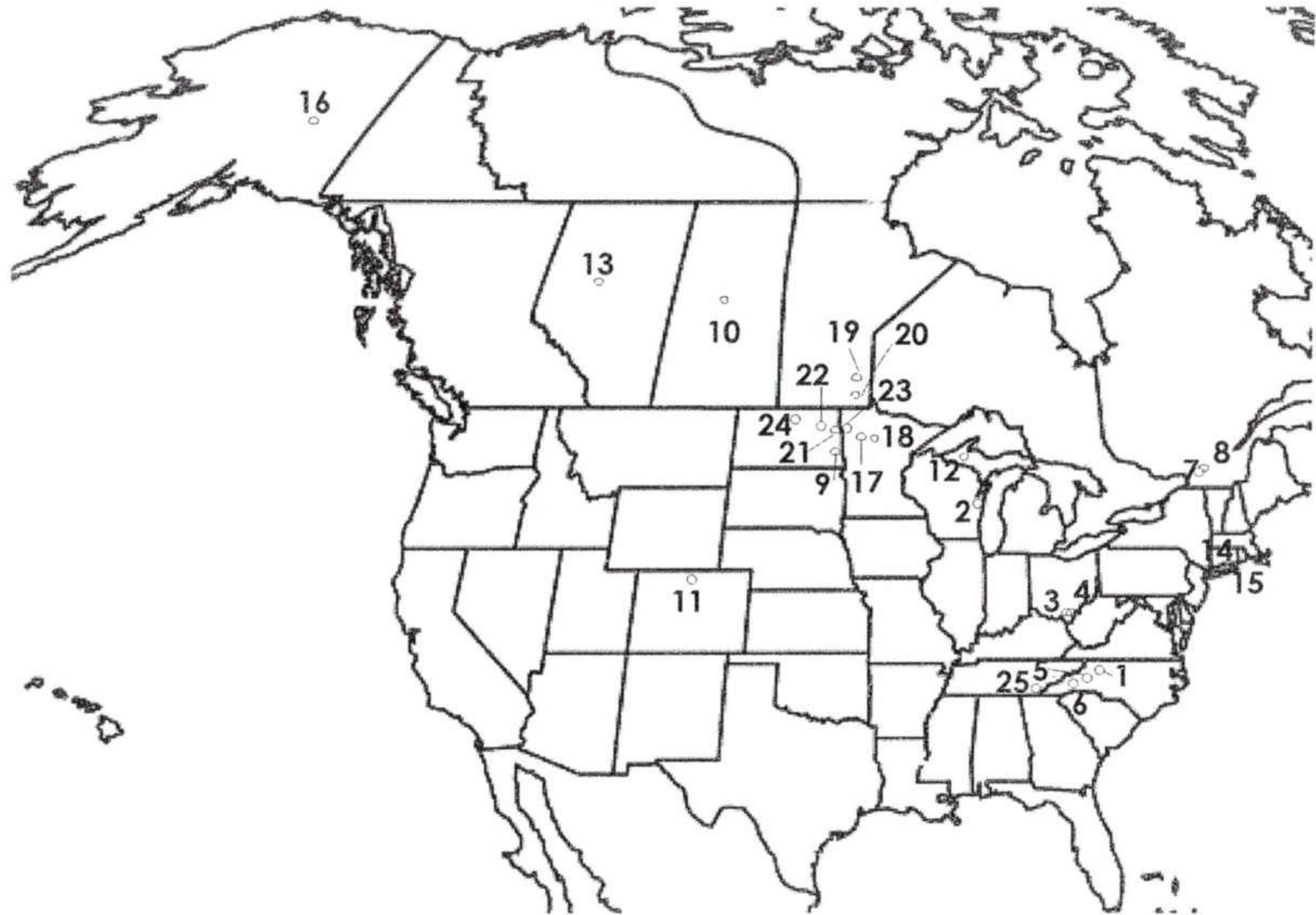


Figure 3. Sample locations and number of individuals. 1. North Carolina - 17; 2. Wisconsin - 25; 3. Ohio1 - 36; 4. Ohio2 - 20; 5. South Appalachia1 - 20; 6. South Appalachia2 - 20; 7. Quebec1 - 20; 8. Quebec2 - 20; 9. Sheyenne National Grasslands - 16; 10. Saskatchewan - 25; 11. Colorado - 16; 12. Michigan - 15; 13. Alberta - 21; 14. New York - 13; 15. SSP New York - 13; 16. Alaska - 16; 17. Crookston - 20; 18. Rydell - 24; 19. Manitoba1 - 29; 20. Manitoba2 - 25; 21. Oakville - 41; 22. Hill Pond East - 47; 23. East Grand Forks - 23; 24. Turtle Mountains – 25; 25. Tennessee - 15

Results

One population, Tennessee, was excluded from the analysis because these individuals could be successfully genotyped at only one locus, Rs23. No bands were obtained for 5 of the loci in this population. Eighty individuals from with genotypes at fewer than three loci were also excluded. Populations ranged from zero individuals that could not be genotyped at two or more loci to a high of 19 individuals in Saskatchewan. Poor DNA extraction appeared to be the cause of the large number of failed attempts to genotype individuals from the Saskatchewan population. Numbers of individuals genotyped and number of alleles for each population and locus are in Table 2. The number of alleles in each population compared to the total number of alleles found across the range of this species is seen in Figure 4.

The Alaska population showed the lowest level of inbreeding (F_{IS} = -0.1604) while Alberta and Saskatchewan showed the highest (F_{IS} = 0.3885 and 0.3368, respectively). The degree of inbreeding for each population is shown in Figure 5. Please see Appendix B for all results.

Within population variation was lowest in the Alaska and Colorado populations at 0.659 and 0.5453 respectively and highest in the second Manitoba population at 0.9223. Within individual population was lowest in the Colorado population at 0.5068 and highest in the second Manitoba population at 0.8702.

Table 2. Populations, with the number of different alleles (A) at each locus and the number of individuals genotyped at each locus (N).

	Rs11		**Rs23**		**Rs32**		**Rs55**		**Rs70**		**Rs88**	
	A	N	A	N	A	N	A	N	A	N	A	N
Alaska	5	11	2	11	5	5	6	13	0	0	5	11
Alberta	6	6	3	15	7	8	4	11	2	1	10	15
Colorado	4	10	3	12	3	10	3	11	7	11	3	12
Crookston	13	15	6	19	12	18	13	20	0	0	11	19
E Grand Forks	13	12	6	14	12	10	6	6	10	7	8	13
Hill Pond E	15	18	7	37	20	30	12	34	12	19	13	26
Manitoba 1	13	10	4	19	16	15	13	19	12	15	10	11
Manitoba 2	15	16	3	2	12	15	16	14	15	12	11	13
Michigan	20	14	4	15	10	13	11	14	11	14	11	14
North Carolina	11	12	7	13	10	14	9	10	14	13	9	14
New York	12	12	2	9	8	9	8	12	6	8	8	9
Oakville	15	33	5	14	16	38	13	35	11	22	10	32
Ohio 1	31	34	11	27	15	35	16	21	19	23	13	21
Ohio 2	20	20	7	20	9	20	10	13	9	12	12	13
Quebec 1	17	19	4	20	11	20	11	15	17	17	17	20
Quebec 2	15	20	7	18	13	20	11	18	13	13	14	19
Rydell	15	19	5	17	9	18	11	16	8	5	11	19
Saskatchewan	6	6	6	4	4	3	6	5	3	4	8	5
Sheyenne	11	16	4	16	9	16	10	14	8	15	7	16
S Appalachia 1	17	18	5	20	10	20	9	13	15	19	9	9
S Appalachia 2	19	16	9	16	11	15	3	2	8	8	5	4
SSP New York	6	7	1	5	8	9	10	10	6	5	11	9
Turtle Mountains	12	11	5	21	0	0	12	16	9	14	9	18
Wisconsin	16	12	7	16	13	14	12	11	15	16	11	16

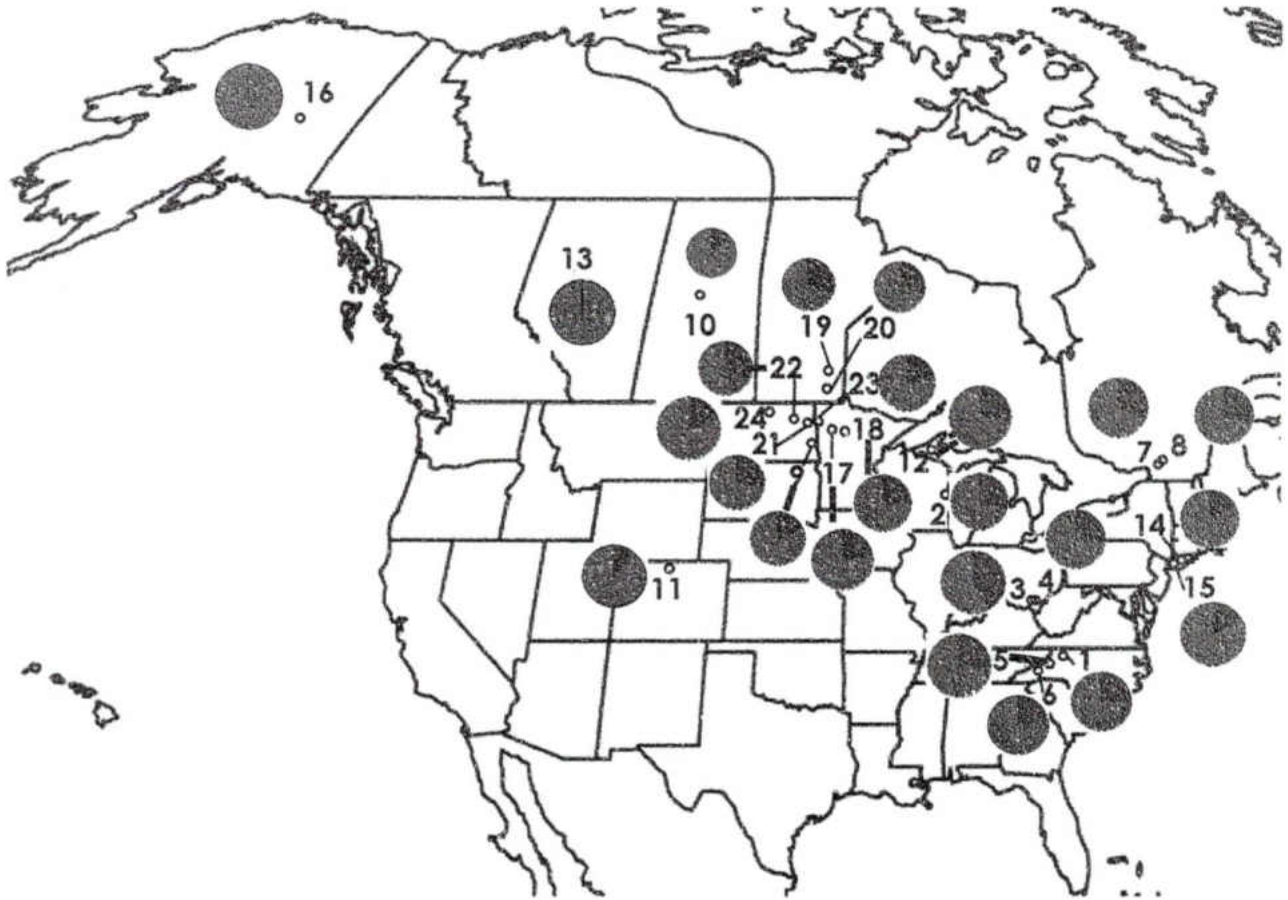


Figure 4. Allelic diversity across all loci. Red is the percentage of alleles present in the population compared to the total number of alleles found across the geographic range (blue).

Two pairs of loci, Rs11 and Rs70, and Rs11 and Rs88 were found to be in linkage disequilibrium. Each pair of loci in linkage disequilibrium did not have all populations in significant linkage disequilibrium (Table 3). Rs11 and Rs70 were in disequilibrium in South Appalachia 1 and Oakville while Rs11 and Rs88 were in disequilibrium only in Oakville (see Appendix C). Rs11 and Rs70 also had one other population that was nearly significant (North Carolina's unbiased estimate of the p-value of p=0.0804). In both cases at least one of the loci involved had a large number of missing genotypes in enough populations that a number of those populations could not be used to test for linkage disequilibrium. This smaller number of populations used in the calculations placed

undue weight on a very low number of significant finding for that loci pair. For this reason all of the loci were used in most of the further data analysis.

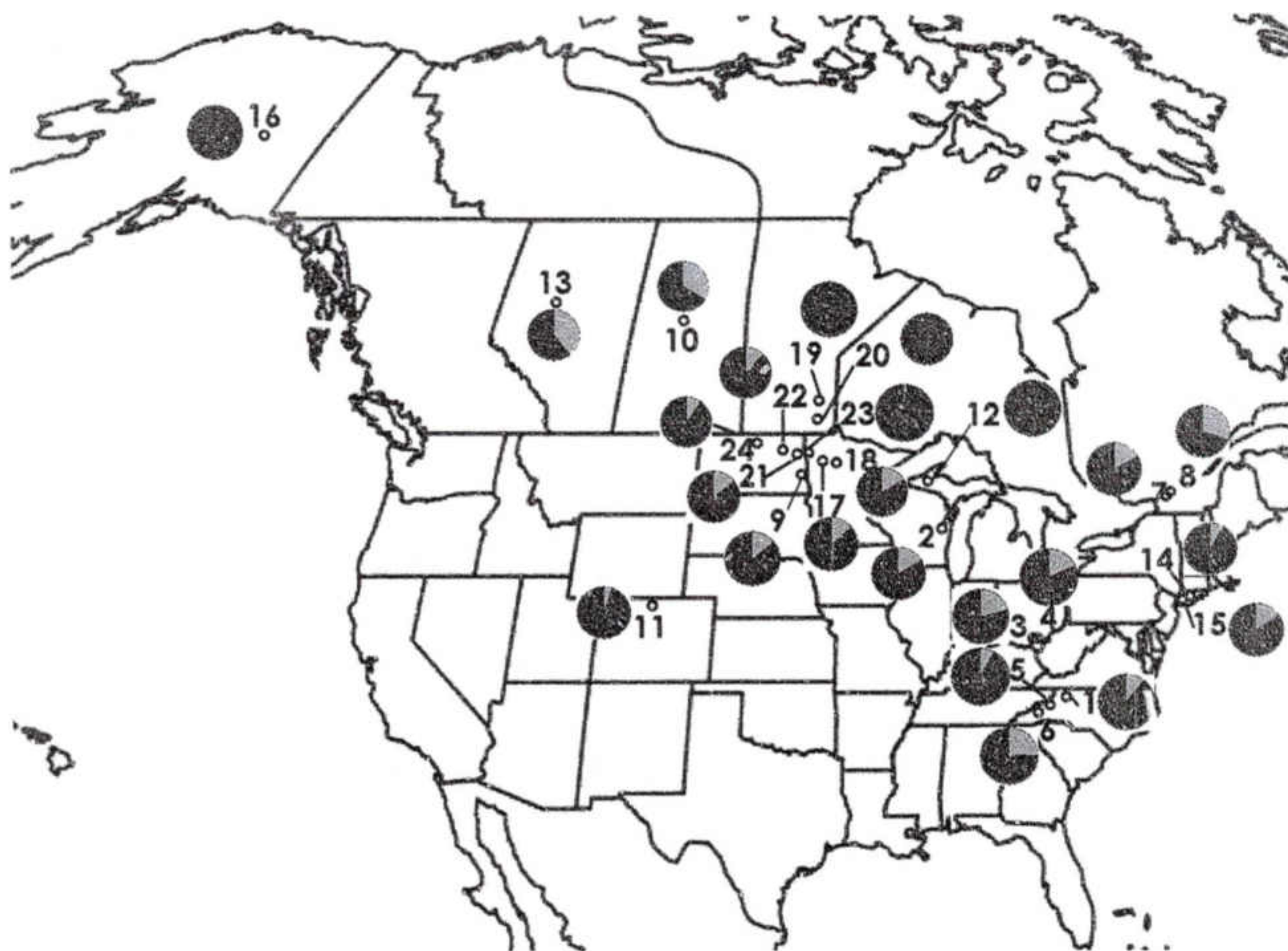


Figure 5. F_{IS} values across the geographic range. The yellow is the degree of inbreeding in each population.

Only populations from Colorado, Long Island, New York, Alaska, Manitoba (both populations) and East Grand Forks, Minnesota were in Hardy-Weinberg equilibrium. All other populations showed a significant heterozygote deficiency (Table 4). All populations were significantly differentiated from all other populations in pairwise genotypic differentiation tests except for the six population pairs listed in Table 5. Two populations in close geographic proximity, southern Appalachia 1 and 2, were significantly differentiated ($p<0.001$).

Table 3. Linkage disequilibrium significance values across all population pairs for each pair of loci. Significant p-values are bolded.

	Rs23	Rs32	Rs55	Rs70	Rs88
Rs11	0.993	0.888	1.000	**Highly sign.**	**Highly sign.**
Rs23		1.000	0.729	0.556	0.980
Rs32			0.777	1.000	1.000
Rs55				0.212	1.000
Rs70					0.993

Table 4. Hardy-Weinberg estimation of p-values by Markov chain method by population over all loci. Non-significant values are bolded.

Population	p-value	Population	p-value
Alaska	**0.8616**	Ohio 1	0.0000
Alberta	0.0000	Ohio 2	0.0000
Colorado	**0.1583**	Quebec 1	0.0000
Crookston	0.0002	Quebec 2	0.0000
E Grand Forks	**0.3708**	Rydell	0.0001
Hill Pond E	0.0000	Saskatchewan	0.0002
Manitoba 1	**0.3808**	Sheyenne	0.0034
Manitoba 2	**0.7196**	S Appalachia 1	0.0003
Michigan	0.0185	S Appalachia 2	0.0000
North Carolina	0.0005	SSP New York	0.0177
New York	**0.1368**	Turtle Mtns	0.0022
Oakville	0.0000	Wisconsin	0.0000

The number of private alleles found in each population over all loci was relatively small, with the second Manitoba population having no private alleles and the first Quebec and one of the New York populations having seven private alleles (Figure 6).

F_{ST} values ranged from 0 to 0.34 (see Appendix D). Of 276 pairwise F_{ST} values, 35 of these were greater than 0.20. Of these 35 pairwise combinations all but one involved Colorado or Alaska. The pairwise F_{ST} values under 0.03 included geographically proximal populations gathered from the same state or province (Ohio (2), Quebec (2), Minnesota (3) and Manitoba (2)) except in the cases of North Carolina and

New York. The pairwise F_{ST} values for the three North Carolina populations range from 0.0419 to 0.0264. The two New York populations had an F_{ST} value of 0.11.

Table 5. Non-significant genotypic differentiation in population pairs.

Population pair	Unbiased estimate of p-value
East Grand Forks, MN and Manitoba 2	0.085
Quebec 1 and Quebec 2	0.106
Crookston, MN and Manitoba 2	0.295
Crookston, MN and Rydell, MN	0.316
Manitoba 1 and Manitoba 2	0.338
Ohio 1 and Ohio 2	0.509

Due to small sample sizes in some populations and at some loci, not every population was able to be analyzed by MICRO-CHECKER to determine presence and frequency of null alleles. The GENEPOP estimation method also uses genotype failure rate information and was able to estimate null allele frequencies in more populations. Estimation of null allele frequencies performed in MICRO-CHECKER (Brookfield 1 method from Brookfield et al. 1996) and GENEPOP (EM algorithm from Dempster et al. 1977) gave similar results in all but two cases (Table 6). In each of these two exceptions the null allele frequency was under 0.2 in both estimations (Table 6). Rs70 was the locus with largest number of populations determined to have null alleles. This high frequency of null alleles at the Rs70 locus is likey the result of a large number of samples that did not produce a PCR product at this locus and a lack of heterozygotes in sampled populations. The Mantel test performed in IBDWS 3.14 showed a significant association between genetic and geographic distance (r=0.47, p=0.013). Pairwise F_{ST} values for each pair of populations were plotted against geographic distance in kilometers in Figure 7.

Table 6. Frequency estimations for null alleles by population (for populations estimated to have null alleles) for each locus. GP is the GENEPOP value and MC is the MICROCHECKER Brookfield et al. (1996) value. Boxed values are those with greater than 10% difference between the two estimations.

	Rs 11		Rs 23		Rs 32		Rs 55		Rs 70		Rs 88	
Population	GP	MC	GP	MC	GP	MC	GP	MC	GP	MC	GP	MC
North Carolina			0.1525	0.1522								
Wisconsin			0.1196						0.1869	0.1862		
Ohio1	0.1196	0.1125	0.1666	0.1483					0.2569	0.2675		
Ohio2							0.1854	0.1824	0.3171	0.3169	0.1113	0.102
S Appalachia1					0.1136				0.1517	0.1476		
S Appalachia2									0.3284			
Quebec1							0.3248	0.329	0.1171	0.1226		
Quebec2			0.2963	0.2941	0.0787	0.0891	0.1795	0.1846	0.3168	0.3137		
Sheyenne							0.2102					
Saskatchewan					0.4		0.4683					
Colorado	0.1133		0.1164									
Alberta	0.1855				0.3825		0.3114					
New York			0.2206	0.207								
SSP New York							0.1931	0.1957				
Crookston					0.119							
Rydell	0.1596	0.1744			0.1115	0.1212						
Oakville					0.1615	0.1462			0.1079	0.1027		
Hill Pond E			0.1226	0.118	0.0975	0.097			0.0835	0.1325		
Turtle Mtns			0.1287				0.1012					

Figure 6. Number of private alleles per population.

STRUCTURE 2.2 analyses were based on genotypes at 5 loci. The least admixture was found when K=2 (K=2, -9672.9; K=3, -9593.9, K=4, -9469.4; Figure 8). The data set was then divided into two groups, called "East" and "West" based on the proportion of membership in that group (Figure 9). In the East group 3 clusters were the best fit (K=2, -4387.1; K=3, -4702.1, K=4, -4363.1) while in the West group 6 clusters were the best fit (K=5, -4356.2; K=6 -4654.4; K=7 -4360.3). The East group was analyzed again without Alaska, Colorado and Alberta but this did not change the best fit of K=3. When the West group was reanalyzed with Alaska, Colorado and Alberta, however, the best fit was K=2 (K=2, -5078.4; K=3, -4982.1; K=4, -4886).

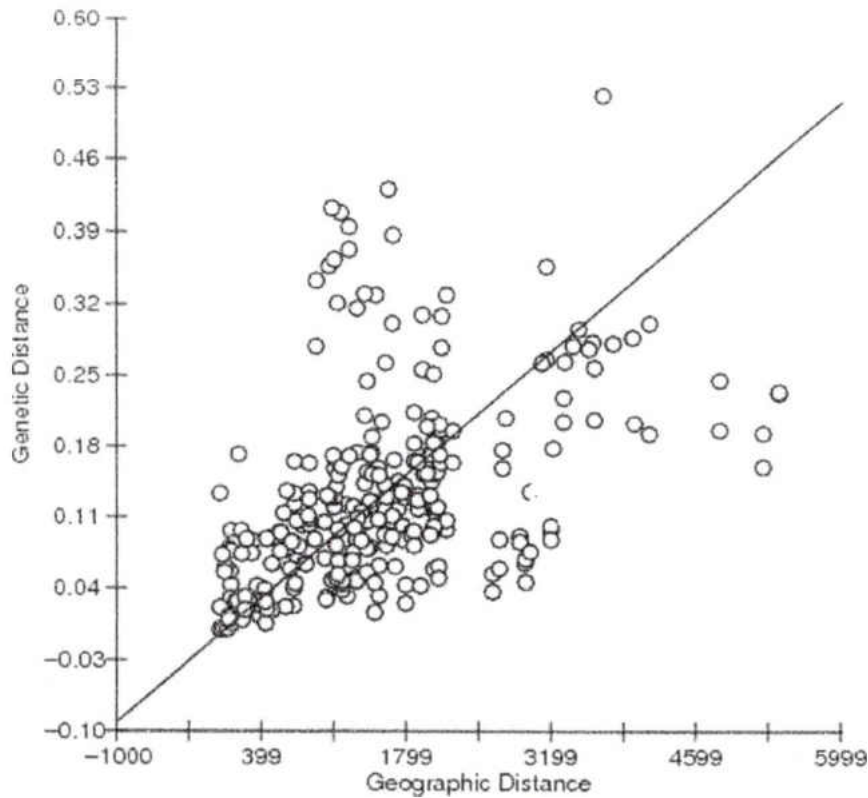


Figure 7. Pairwise F_{ST} values plotted against geographic distance in kilometers. Mantel test r= r=0.47, p=0.013. Line represents the reduced major axis regression line used to calculate slope and intercept.

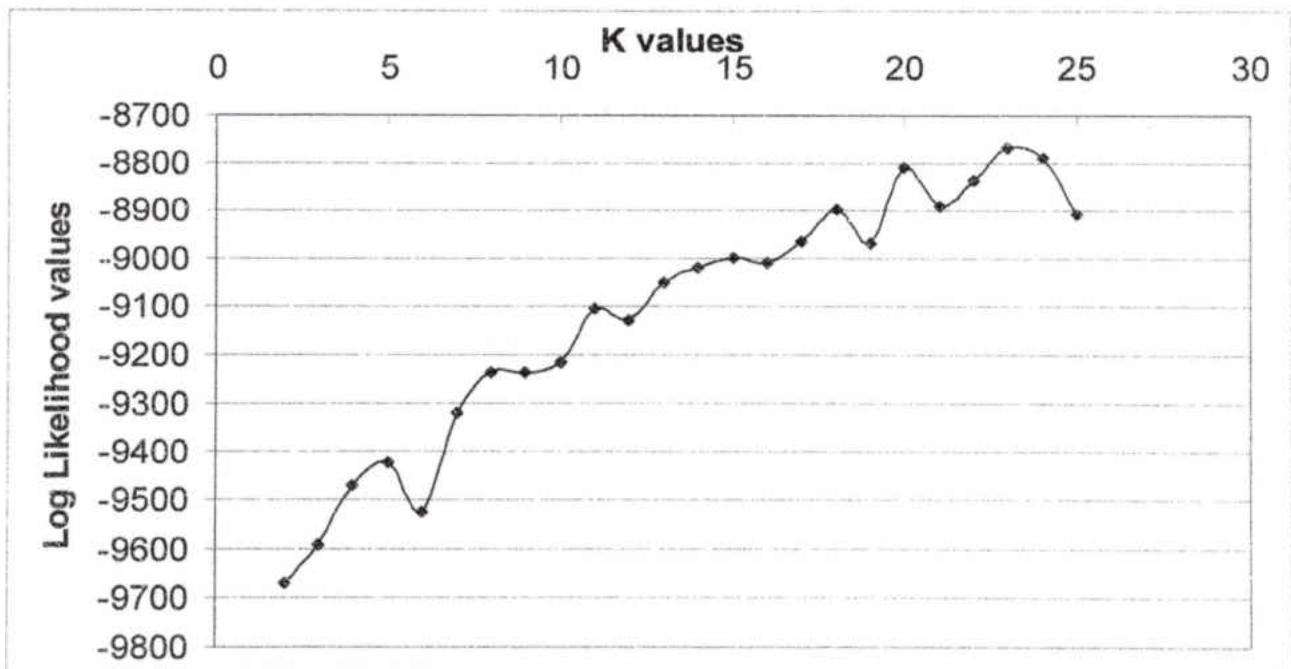


Figure 8. The log likelihood values for each value of K for all populations. The highest likelihood (and therefore lowest log likelihood) is K=2.

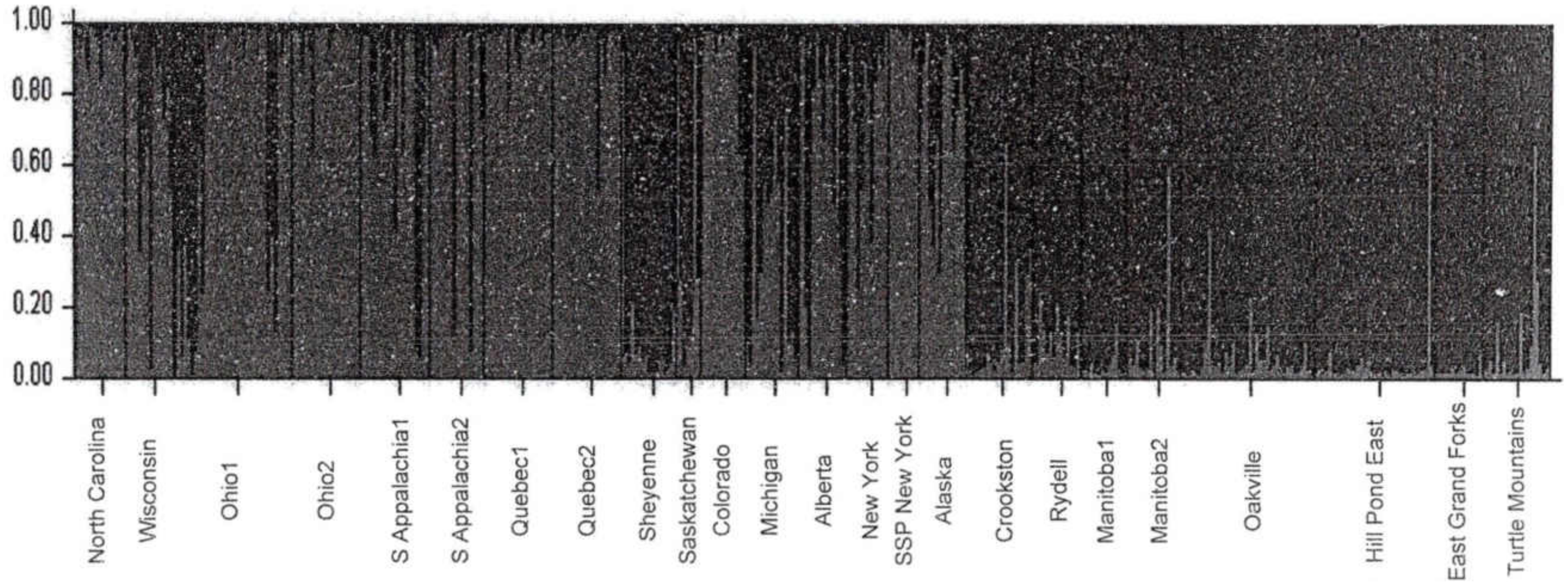


Figure 9. In STRUCTURE output each individual represented by a vertical bar. The proportion of each individual's 5 locus genotype (see text) that is most similar to each group is colored (see the y-axis for proportion). The population names are on the x-axis and the name is colored according to the group to which that population as a whole was assigned. The "East" group is green, the "West" group is red.

Discussion

Significant genotypic population differentiation between most pairs of populations (exceptions listed in Table 5), the pairwise values of F_{ST}, and the STRUCTURE result all indicate significant population structure in *Rana sylvatica*. The Colorado and Alaska populations are of special interest because of their very high pairwise F_{ST} values and the lowest within population genetic variation of any sampled population found in the Colorado population. The Colorado population is one of a few remnant populations that are not continuous with the rest of the geographic range of the species (Figure 2). It may be that as little as 100 years ago those remnant populations were part of a continuous range. According to Martof and Humphries (1959) Rocky Mountains populations connected Colorado populations to the rest of the range until this time based on two poorly preserved museum specimens from Montana. A study done by Porter (1969) showed a high degree of interpopulation incompatibility between Colorado and Manitoba populations, based on interpopulational breeding crosses, that might be an indication of a more temporally distant split. Since the range's contraction it is possible that a genetic drift, lack of gene flow or both have caused the Colorado population to differentiate from other populations.

Colorado was likely a peripheral population, occurring near the outer boundary of the species range, before it became disjunct (Martof and Humphries 1959). Peripheral populations are subject to lower gene flow and often different natural selection pressures compared to more central populations. Both of these will have the effect of causing peripheral populations to diverge from the central populations (Lessica and Allendorf 1995).

Since wood frog generation time is as little as 2 years (Berven 1990), it is conceivable that a significant loss of genetic variation due to genetic drift in the Colorado population has caused differentiation in as little as 100 years. A simulation study by Lacy (1987) showed that 25 identical populations each containing 120 randomly mating, hermaphroditic individuals lost an average 58% of their heterozygosity in 100 generations (Lacy 1987). In fewer than 5 generations the New Zealand snapper, (*Pagrus auratus*), showed significant loss of genetic diversity and heterozygosity when effective population sized dramatically decreased due to overharvesting (Hauser et al. 2002). These studies lend support to the hypothesis that a combination of isolation and drift caused differentiation of the Colorado population.

The Alaska populations were likely isolated during the last glacial maximum and became more differentiated with time. White spruce (Anderson et al. 2006), root voles (Brunhoff et al. 2003), tundra voles (Galbreath and Cook 2004) and arctic hares (Waltari and Cook 2005) all show evidence of an Alaskan refugium as well, often as part of a greater Beringean refugium. The Alaska populations show the lowest level of inbreeding in this study. The sample size from the Alaska population was low. More samples from more populations in Alaska have recently been gathered by R. Newman: future work will shed light on the genetic variability and pattern within Alaska.

Glacial refugia might also be expected to contain unique (private) alleles and high allelic richness. The Alaska population does not fit this profile with only 3 private alleles; however this may due to a small sample size. A number of other populations do have large numbers of unique alleles (Figure 6). It is likely that there was not a single refugial population, but that a refugial area contained multiple populations. When

populations are combined with those in closest geographic proximity, the two New York populations have the highest combined number of private alleles and intermediate levels of allelic diversity (Table 7). New York may have been glacial refugium for wood frogs. This observation is consistent with the findings of Hoffman and Blouin (2004) in their study of *Rana pipiens* who also suggest a New York refugium for leopard frogs. The Southern Appalachian populations (North Carolina, S Appalachia1 and S Appalachia2) when combined also show a relatively high number of private alleles (6) and a moderate level of differentiation from one another (F_{ST} = 0.0419 to 0.0264) based on levels of differentiation found in previous studies of *R. sylvatica* (Newman and Squire 2001, Squire and Newman 2002). Both the number of private alleles and the F_{ST} values are most likely an indication of a third, Southern Appalachian glacial refugia. The number of private alleles in North Dakota is also high. This might simply reflect the largest samples sizes in the study and the largest number of sampled populations. Alternatively, this may be due to a lack of gene flow between the populations that has caused new alleles to remain within a single population.

Table 7. Combined areas with large numbers of private alleles.

Area (number populations combined)	Number private alleles
Manitoba (2)	5
Ohio (2)	5
Southern Appalachia (3)	6
Quebec (2)	9
North Dakota (4)	10
New York (2)	11

Generally, the populations nearest geographically are also those genetically most similar with a clear isolation by distance pattern (see Figure 7). This is consistent with a slow phalanx model of recolonization proposed by Hewitt (2001, 2004). Under this

model we expect to see maintenance of genetic variation and no large geographic split and no corresponding genetic split. The pattern of genetic diversity and latitude is less clear. Many of the northern populations are no less diverse than the southern populations. The Manitoba populations in particular show greater diversity than many of the more southern populations. This increase in diversity could be an indication of a zone of secondary contact; however, our sampling is insufficient to determine this. Ontario, north of the Great Lakes, appears to be a zone of secondary contact in *Rana pipiens* (Hoffman and Blouin 2004). Austin et al. 2002 found that dispersing spring peepers went north into eastern Ontario, then westward into Manitoba and that barriers to dispersal may have prevented western haplotypes from expanding north into Ontario and Manitoba. With more sampling, particularly in Ontario, it might be possible to more accurately determine the existence and possible location of a zone of secondary contact.

STRUCTURE results showed that in the East group Alaska formed its own group with fair support for the inclusion of one of the New York populations. The inclusion of Alaska, Colorado and Alberta is unexpected. The inclusion of these three in the East group is most likely due in part to the fact that these populations are so different from all other populations, though Alaska is most closely related to the New York population. This relationship may reflect a chance sharing of alleles and the small number of individuals sampled in the Alaska population. Because Alaska, Colorado and Alberta added more support to the formation of further groups when in the East group when K=2 the clustering placed them in the East group. The removal of Alaska, Alberta and Colorado from the East group decreased the support for the remaining populations in that group. When forced to join the West group, Alaska, Alberta and Colorado formed a very

well supported group with Sheyenne and Saskatchewan, a more geographically logical grouping. Additional samples from the New York populations in conjunction with more samples from Alaska might shed more light on these unexpected groups.

Like many other species, the recolonization history of wood frogs is complex. This study has found indications of three glacial refugia (Alaska, North Carolina, and New York) and at least one possible area of secondary contact in Manitoba. With more intensive sampling in future studies the pattern could become clearer.

CHAPTER III

REGIONAL PATTERNS

Abstract

Local geography had a major role in shaping the genetic structure of *R. sylvatica* in the upper Great Plains region of North America. Populations in Minnesota and Manitoba are most closely related to each other and less closely related to those of North Dakota, indicating a geographical barrier. No pattern of isolation by distance was seen at this scale.

Introduction

The "prairie potholes" of the northern Great Plains are the remains of small areas of ice left after the leading edge of the main ice sheet retreated. In other parts of North America large proglacial freshwater lakes formed as the ice melted and the water was trapped (Pielou 1991). One of these, Lake Agassiz, covered present day western Minnesota, eastern North Dakota and southern Manitoba (Ashworth and Cvancara 1983, Fenton et al. 1983, Pielou 1991).

How the range expansion following the emptying of Lake Agassiz occurred and how quickly it occurred would lead to different genetic patterns. Rapid northward expansion would lead to a reduction in allelic diversity in northern populations (Nichols and Hewitt 1994; Ibrahim et al. 1996; Hewitt 2001, 2004). Rapid leptokurtic dispersal

would produce patches of homogeneity, as predicted by Ibrahim et al. (1996). If populations expanded as in Hewitt's "phalanx" model, they would retain more of their genetic diversity and produce less population structure while refugial areas would have time to both diverge genetically and become geographically subdivided (Nichols and Hewitt 1994, Hewitt 2001, 2004).

Amphibians require breeding ponds and have relatively low vagility; thus they are more likely to show evidence of phylogeographic structure on a small scale. In two more terrestrial amphibians, *Bufo calamita* (Rowe et al 2000) and *Bufo woodhousei* (Masta et al. 2003), this structure is evident at distances of a few kilometers using microsatellites to two hundred kilometers using mitochondrial DNA. A study by Brede and Beebee (2004) found *Bufo bufo* populations had significant interpopulation structure at distances of hundreds of kilometers while *Rana temporaria* populations in the study at the same distance did not. Johansson et al. (2006) found interpopulation structure in *Rana temporaria* populations in Sweden though the structure was not strongly statistically significant. *Rana iberica* populations are structured on a scale of tens of kilometers (Martinez-Solano 2005).

The objective of this study was to observe the genetic patterns in *R. sylvatica* at a regional scale. While the authors have studied this question at a fine scale (tens of kilometers) in this species (Newman and Squire 2001, Squire and Newman 2002) and more recently at a geographic range-wide scale (Chapter 2) where significant population structure and an isolation by distance pattern was found in each study, the authors have not yet investigated the intermediate, regional scale for population structure.

Methods

Sample Handling

A total of 250 samples were collected from 9 populations in Minnesota, North Dakota and Manitoba (see Figure 10 for locations). These sar ples are a subset of the samples included in Chapter 2. Tissue samples (toe tips or ha nlings (one per egg mass)) were stored and transported in 95% ethanol. DNA was extracted by a standard phenol-chloroform extraction and an ethanol precipitation (Hitchings & Beebee1998). Six microsatellite loci were used (Table 1). PCR reactions were conducted in 10ul volumes and contained 1X buffer, 0.5 uM dNTPs, 2.5 mM $MgCl_2$, 1uM unlabeled primer, 1uM fluorescently labeled primer, and 0.5 units taq DNA polymerase (Promega). The PCR reactions were carried out in a BIORAD thermocycler. For Rs23 (Newman and Squire 2001) the cycles consisted of 3 minutes at 94°C, and 3 minutes at 55°C for one cycle, 1 minute at 72°C, 1 minute at 94°C, 1 minute at 55°C for 30 cycles and a 3 minute extension at 72°C. For all other primers (Julian and King 2003) the PCR reaction cycles were 3 minutes at 94°C for one cycle, 1 minute at 94°C, 30 seconds at 54°C, and a final 1 minute at 72°C for 30 cycles, and a 2 minute extension step at 72°C. PCR products were run on a 2% agarose gel and stained with ethidium bromide (2ug/ml). Products in the correct size range were then run on an ABI (Applied Biosystems) model 3100 using 0.5ul 500ROX as a size standard in each sample. ABI output was using ABI GeneScan software (version 3.7, 2000) and allele sizes determined by ABI Genotyper.

Data Analysis

Characterization of genetic diversity (allele frequencies, genic diversity, tests for Hardy-Weinberg equilibrium) was conducted using GENEPOP version 4 (Rousset 2007).

Levels of inbreeding (F_{IS}), intrapopulation and interpopulation genetic variation were calculated and linkage disequilibrium between pairs of loci was tested for in GENEPOP using a log likelihood ratio statistic (Rousset 2007). GENEPOP was also used to test for differentiation in allele frequencies among pairs of populations using a Markov Chain Monte Carlo approximation of Fisher's exact test (Raymond and Rousset 1995). Because of the large number of pairwise tests involved, the significance of pairwise comparisons was determined using a sequential Bonferroni adjustment (Rice 1989).

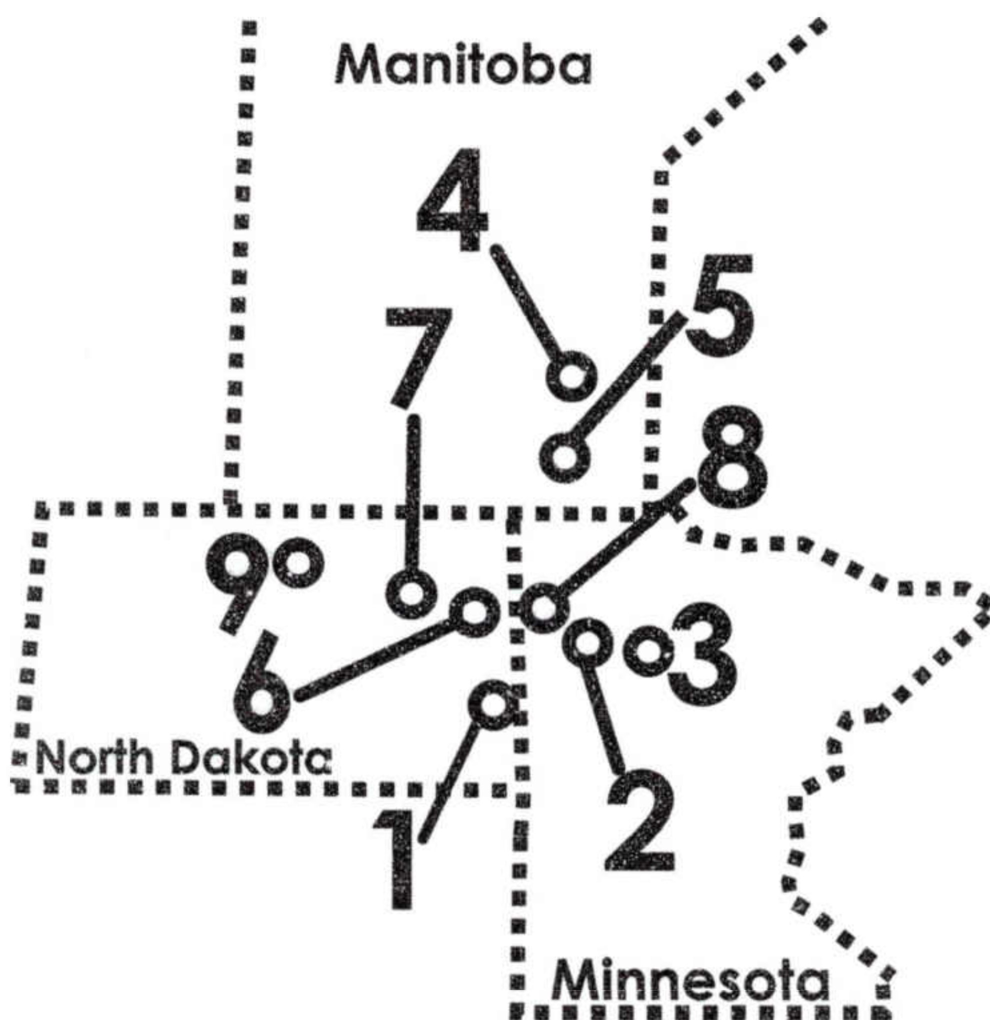


Figure 10. Sample locations and number of samples from each. 1. Sheyenne National Grasslands - 16; 2. Crookston - 20; 3. Rydell - 24; 4. Manitoba1 - 29; 5. Manitoba2 - 25; 6. Oakville - 41; 7. Hill Pond East - 47; 8. East Grand Forks - 23; 9. Turtle Mountains – 25

Pairwise F_{ST} (averaged over loci) were estimated in GENEPOP and used in analyses of isolation by distance, with a Mantel's test of significance in IBDWS 3.14 (Jensen et al. 2005). Presence and estimation of null allele frequencies was examined in GENEPOP (Rousset 2007) and MICRO-CHECKER (Van Oosterhout 2004) was used to investigate the possibility of large allele dropout, evidence of scoring errors and presence of null alleles and estimation of null allele frequencies.

Clusters or groups were delineated via Bayesian methods as in Pritchard *et al.* (2000), via STRUCTURE 2.2. STRUCTURE does not assume a mutational model or process but uses individuals' genotypes to determine the population(s) to which they should be assigned. The model accounts for Hardy–Weinberg or linkage disequilibrium by introducing population structure and attempts to find population groupings that are not in Hardy-Weinberg disequilibrium (Pritchard *et al.* 2000). The number of clusters or groups (K) was set from K=2 to K=10 to determine the likelihood of each.

Results

Genotypes were obtained for at least two loci from 204 of 250 individuals. The remaining 46 individuals with fewer than three loci genotyped were excluded from analysis (see Table 8 for the number of individuals genotyped at each locus and the number of different alleles at each locus). The number of different alleles for each population compared to the number of alleles seen among these populations is seen in Figure 11.

Inbreeding values (F_{IS}) ranged from a low in Manitoba 2 of -0.009 to a high of 0.1654 in Rydell. Please see Table 9 for all results.

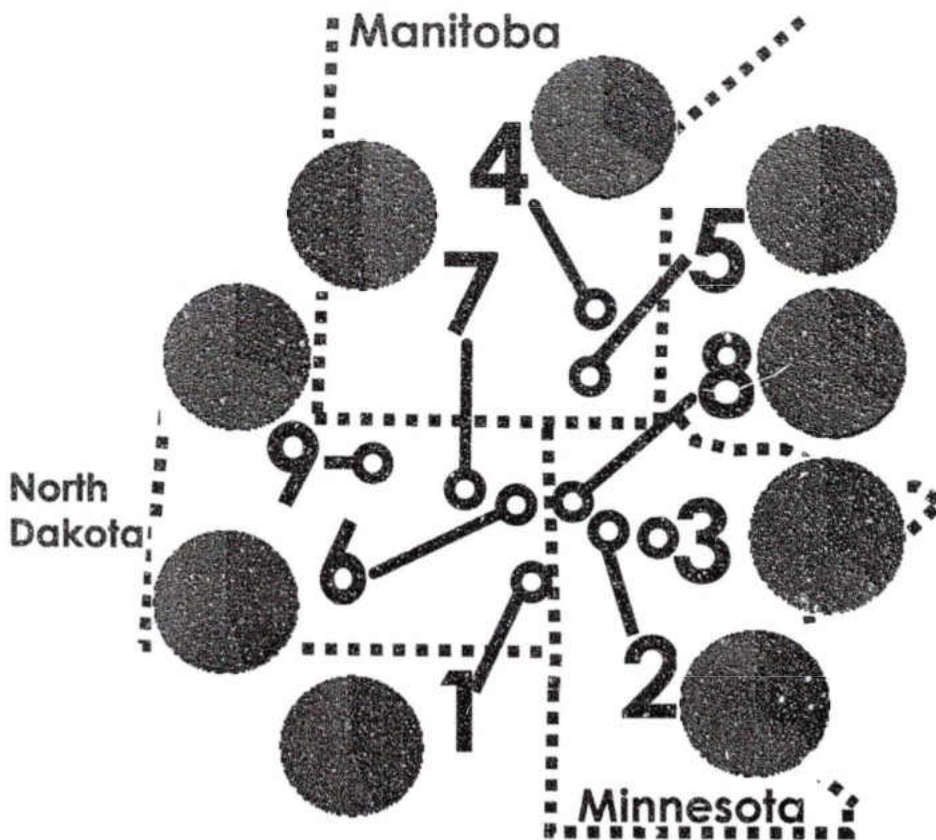


Figure 11. Alleles found in each population for all loci (red) compared to the total number of alleles found in for all populations for all loci (blue).

Table 8. Populations, with the number of different alleles (A) at each locus and the number of individuals genotyped at each locus (N).

	Rs11		**Rs23**		**Rs32**		**Rs55**		**Rs70**		**Rs88**	
	A	N	A	N	A	N	A	N	A	N	A	N
Crookston	13	15	6	19	12	18	13	20	0	0	11	19
E Grand Forks	13	12	6	14	12	10	6	6	10	7	8	13
Hill Pond E	15	18	7	37	20	30	12	34	12	19	13	26
Manitoba 1	13	10	4	19	16	15	13	19	12	15	10	11
Manitoba 2	15	16	3	2	12	15	16	14	15	12	11	13
Oakville	15	33	5	14	16	38	13	35	11	22	10	32
Rydell	15	19	5	17	9	18	11	16	8	5	11	19
Sheyenne	11	16	4	16	9	16	10	14	8	15	7	16
Turtle Mtns	12	11	5	21	0	0	12	16	9	14	9	18

Table 9. The inbreeding value (F_{IS}), within population variation (1-Q intra) and between population variation (1-Q inter) for each population.

Population	F_{IS}	1-Q intra	1-Q inter
Crookston	0.1363	0.7409	0.8525
E Grand Forks	0.0136	0.8406	0.8502
Hill Pond E	0.1257	0.7408	0.809
Manitoba 1	0.011	0.8646	0.8521
Manitoba 2	-0.009	0.8702	0.9223
Oakville	0.1414	0.6556	0.7898
Rydell	0.1654	0.711	0.8031
Sheyenne	0.1378	0.6996	0.8106
Turtle Mtns	0.0946	0.7662	0.8284

All populations were found to be significantly differentiated from all other populations except for the population pairs in Table 10 (see Appendix E for all population pairs). Manitoba 2 and East Grand Forks are not significantly differentiated from each other but are close to the 0.05 significance value. Sheyenne, both Manitoba populations, and East Grand Forks were found to be in Hardy-Weinberg equilibrium at all loci (Table 11). Crookston was in Hardy-Weinberg disequilibrium at one locus, Turtle Mountains at two loci, Rydell and Oakville at three loci and Hill Ponds at four loci.

Table 10. Population pairs with non-significant genotypic differentiation.

Populations	p-value
Manitoba 2 and East Grand Forks, MN	0.062076
Manitoba 1 and Manitoba 2	0.127297
Crookston and Rydell, MN	0.13207
Manitoba 2 and Crookston, MN	0.161127

Table 11. Global Hardy-Weinberg estimated p-values by Monte Carlo estimation by population. Non-significant values are bolded.

Population	p-value
Sheyenne	0.0013
Crookston	0.0006
Rydell	0
Manitoba 1	**0.3864**
Manitoba 2	**0.7371**
Oakville	0
Hill Pond E	0
E Grand Forks	**0.3546**
Turtle Mountains	0.0075

Two pairs of loci Rs11 and Rs88 and the pair Rs55 and Rs70 were found to be in linkage disequilibrium (Table 12). In each case only one population had a significant comparison (Oakville and Sheyenne, respectively) while each other population had a non-significant result (see Appendix F). All loci were used for further analysis.

Table 12. Linkage disequilibrium across all population pairs for each pair of loci. The significant values are bolded.

Locus pair	Chi^2	df	p-value
Rs11 & Rs23	2.5845	10	0.9896
Rs11 & Rs32	6.9218	8	0.5451
Rs11 & Rs55	0.0000	8	1.0000
Rs11 & Rs70	5.3877	4	0.2498
Rs11 & Rs88	**Infinity**	**8**	**Highly sign.**
Rs23 & Rs32	8.1788	12	0.7710
Rs23 & Rs55	11.4490	16	0.7810
Rs23 & Rs70	8.0469	8	0.4289
Rs23 & Rs88	5.8379	16	0.9898
Rs32 & Rs55	4.0752	10	0.9439
Rs32 & Rs70	1.6063	8	0.9908
Rs32 & Rs88	0.4942	12	1.0000
Rs55 & Rs70	**18.9656**	**8**	**0.0150**
Rs55 & Rs88	1.1489	14	1.0000
Rs70 & Rs88	5.0939	10	0.8848

Pairwise F_{ST} values ranged from 0.0005 between the Crookston and East Grand Forks, MN populations to 0.1469 (Table 13). Of the pairwise F_{ST} values 15 of 16 highest values involve either Oakville or Sheyenne and the highest is between these two populations. The Mantel test showed no significant correlation of geographic and genetic distance (r=-0.1616, p= 0.72; Figure 12).

Both MICRO-CHECKER and GENEPOP were used to estimate the frequency of null alleles at each locus for each population. The Brookfield 1 (Brookfield 1996) estimates were very similar to those of GENEPOP, therefore only the GENEPOP estimations of null allele frequencies are reported here. The population/loci combinations with evidence of null alleles and the estimate of the frequency of those null alleles are given in Table 14.

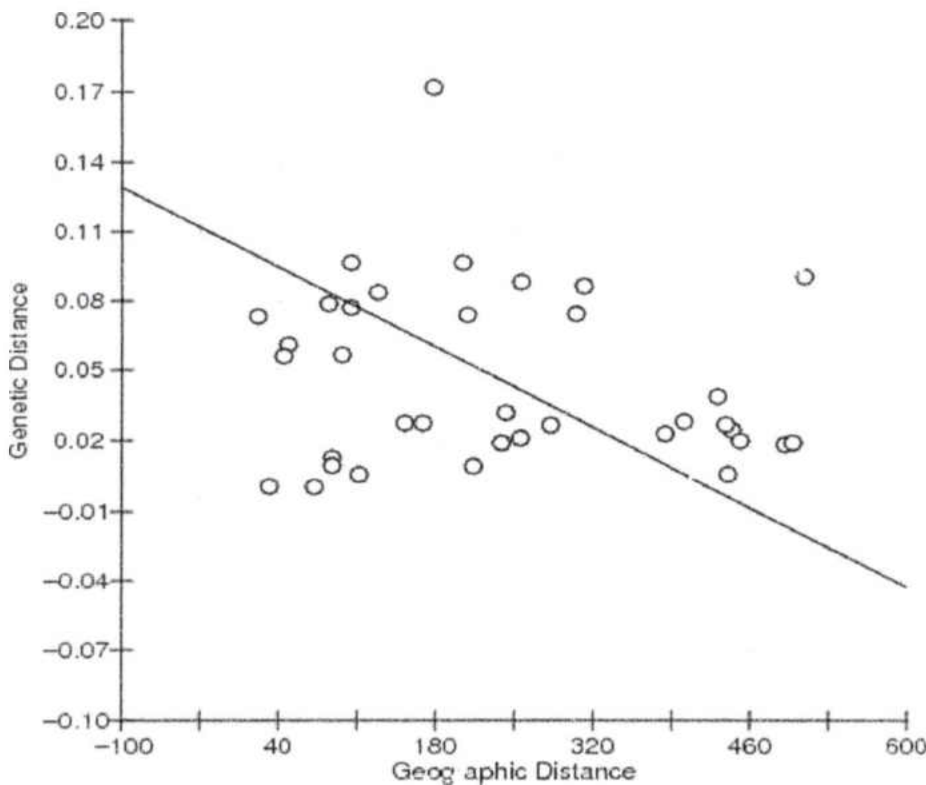


Figure 12. Pairwise F_{ST} values plotted against geographic distance in kilometers. No significant pattern of isolation be distance was found (r=-0.1616, p= 0.72). Line represents the reduced major axis regression line used to calculate slope and intercept.

Table 13. Pairwise F_{ST} values for each population pair.

	Sheyenne	Crookston	Rydell	Manitoba 1	Manitoba 2	Oakville	Hill Pond E	E Grand Forks
Crookston	0.0536							
Rydell	0.0881	0.0007						
Manitoba 1	0.0831	0.0181	0.0187					
Manitoba 2	0.0374	0.0059	0.0239	0.0057				
Oakville	0.1469	0.073	0.0716	0.0694	0.0812			
Hill Pond E	0.0881	0.0268	0.0268	0.026	0.0309	0.0574		
E Grand Forks	0.0774	0.0005	0.0125	0.0208	0.0092	0.0683	0.0095	
Turtle Mtns	0.0798	0.0225	0.0274	0.0196	0.0262	0.069	0.053	0.0188

Table 14. Frequency estimations for null alleles by population (for populations estimated to have null alleles) for each locus. GP is the GENEPOP estimate and MC is the MICROCHECKER Brookfield et al. (1996) estimate. Boxed values have a greater than 10% difference between the estimates.

	Rs11		Rs23		Rs32		Rs55		Rs70		Rs88	
Population	GP	MC	GP	MC	GP	MC	GP	MC	GP	MC	GP	MC
Sheyenne							0.210					
Crookston					0.120							
Rydell	0.160	0.174			0.112	0.121						
Oakville					0.162	0.146			0.108	0.103		
Hill Pond E			0.123	0.118	0.098	0.097			0.084	0.133		
Turtle Mtns			0.129				0.101					

The Bayesian clustering program, STRUCTURE, returned the highest probability for K=9 (K=8, -4789; K=9, -4933; K=10, -4872.1; Figure 13.) though many of the individuals could not be strongly assigned to any one population. Because the most likely number for K was also the number of population in the study no subdivision of groups was deemed necessary or seemed biologically relevant (Figure 14).

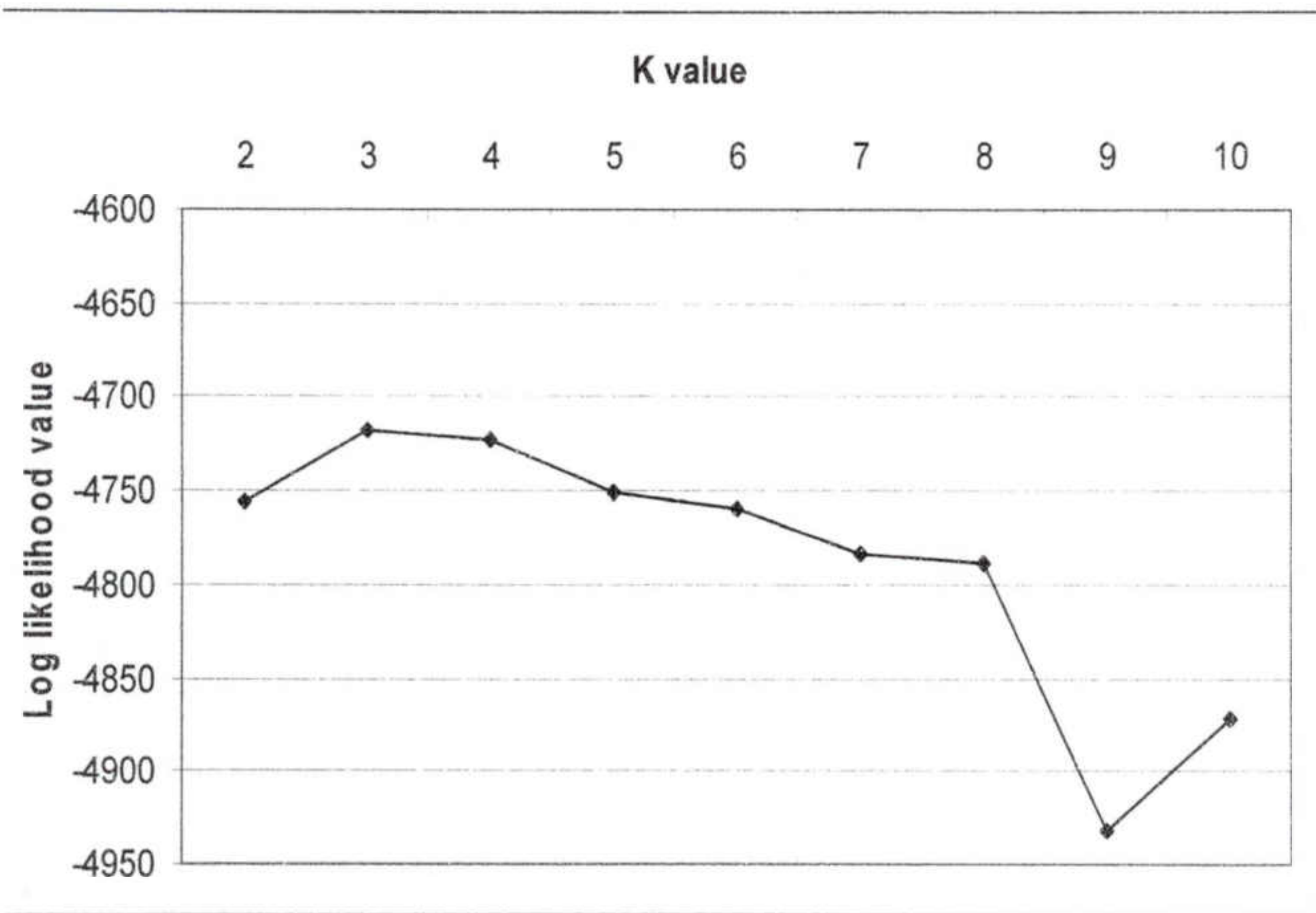


Figure 13. The log likelihood values for each value of K for all populations. The highest likelihood (and therefore lowest log likelihood) is K=9.

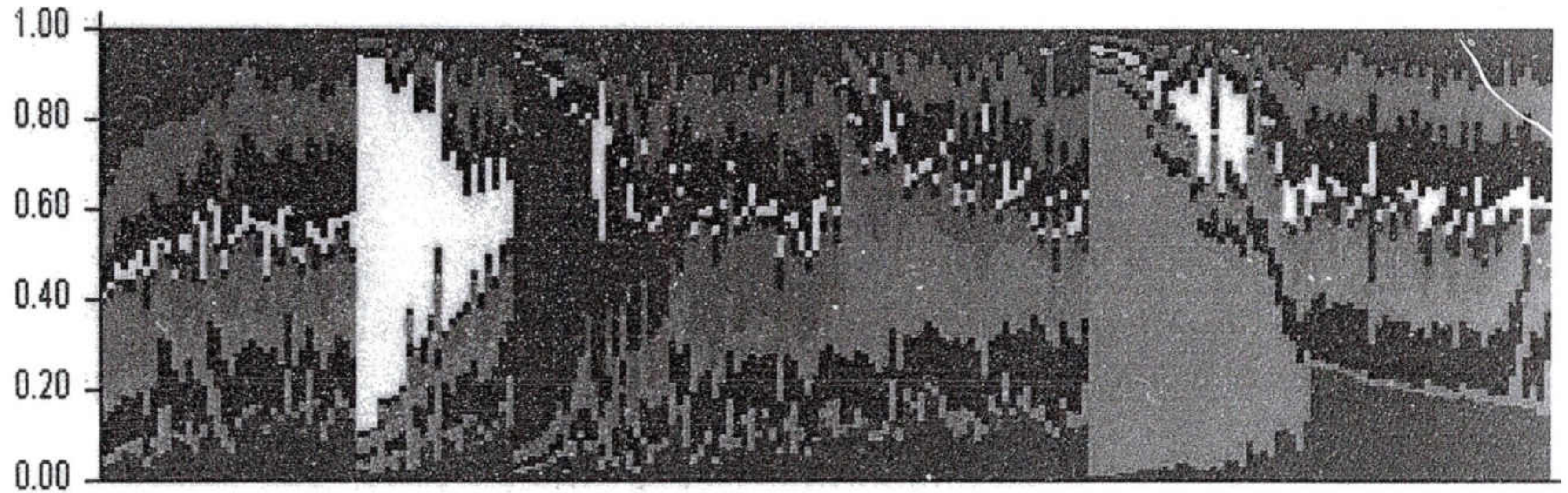


Figure 14. Visual representation of STRUCTURE results. Each individual represented by a vertical bar. The proportion of each individual's genotype that is most similar to each group is colored (see the y-axis for proportion).

Discussion

Rydell, Oakville and Hill Pond East populations were found to be in Hardy-Weinberg disequilibrium for at least half of the loci due to heterozygote deficiency (Table 11). Hardy-Weinberg disequilibrium is unlikely to be due to small sample size: Hill Pond and Oakville had the highest numbers of individuals genotyped. This deficiency may be due, in part, to the presence of null alleles; however, the estimation of null allele frequencies for those loci in those populations were not high (0.1615 and under). Another possibility is a recent decrease in population size increasing the effect of genetic drift. All three populations are in or adjacent to agricultural lands which might also limit gene flow. It is likely that inbreeding has a role in the heterozygote deficiency as all three populations have a high level of inbreeding (F_{IS}= 0.1654 Rydell, 0.1414 Oakville, 0.1257 Hill Pond East; Table 9).

The two Manitoba populations were not significantly differentiated from each other. This similarity may be due to gene flow keeping these two populations undifferentiated in spite of possible barriers to dispersal such as roadways. Alternatively, it is also possible that the two populations were founded from the same or (very similar) source population(s) and have not yet had time to differentiate. Currently there is not enough data (population size, dispersal rates, time since founding, number of founders, etc.) to determine whether gene flow or founder effects play a larger role with these populations.

Results from STRUCTURE analysis show evidence of population differentiation not found by use of the F_{ST} statistic. STRUCTURE attempts to place individuals within K populations to minimize both Hardy-Weinberg disequilibrium and admixture. It is

evident that each population is an independent unit in spite of very low F_{ST} values in a few comparisons. As a group individuals are assigned by STRUCTURE to the population from which they originated, verifying population independence.

The F_{ST} values found in this study are close to those reported in studies such as Johansson et al. (2006). In this study the authors found an overall F_{ST} value of 0.1 in *Rana temporaria* in Sweden over similar geographic distances. A global F_{ST} value of 0.1078 was found in *Rana catesbeiana* by Austin et al. (2004). A study of *Bufo bufo* and *Rana temporaria* in England found higher F_{ST} values than reported here in *Bufo bufo* (mean 0.265) but higher in *Rana temporaria* (mean 0.051) (Brede and Beebee 2004).

While *R. sylvatica* has shown significant isolation by distance at smaller geographic scales (tens of kilometers) (Newman and Squire 2001) and at the continental scale (Chapter 2) none was found at the intermediate, regional scale. The genetic pattern of this region may be reflecting the effects of landscape and geography.

Until approximately 7500 years ago proglacial Lake Agassiz covered the central portion of this region, bounded on the northeast side by the Laurentide ice sheet (Pielou 1991), that may have acted as a barrier to gene flow to wood frog colonizers. After the disappearance of Lake Agassiz the Red River of the North and other smaller lakes and rivers in the area remained as barriers. The fact that the Minnesota populations appear to be more closely related to the Manitoba populations (8 of the 9 lowest F_{ST} values are between Minnesota and Manitoba populations) than to the geographically closer North Dakota populations may be due to this barrier.

Many landscape features form barriers for amphibians. Columbia spotted frogs, *Rana luteiventris*, in the western USA have movement limited by mountain ridges

between valleys (Funk et al, 2005). The stream-breeding salamander, *Dicamptodon tenebrosus*, has lowered genetic variation and reduced population density due to forest clearing in British Columbia (Curtis and Taylor, 2004).

Sheyenne and Oakville populations are the two populations most different from each other ($F_{ST} = 0.1469$) and from all of the other populations. The Sheyenne population is the southernmost of those in this study and is at the periphery of the species range. Climatic oscillations causing successive bottlenecks and possible loss of genetic diversity in the founding of this population are possible reasons for the Sheyenne population's differentiation. Peripheral populations are subject to lower gene flow and often different natural selection pressures compared to more central populations. Both of these will have the effect of causing peripheral populations to diverge from the central populations as well (Lessica and Allendorf 1995). Increased sampling and further study would be necessary to provide more information in this particular system.

The Oakville population is found in two small ponds surrounded by a small patch of native prairie and agricultural lands. Stevens et al. (2006) found that agricultural land is a barrier for the juvenile natterjack toad (*Bufo calamita*) while Scribner et al. (2001) found agricultural land reduced genetic variation in the common toad (*Bufo bufo)*. Open landscape appeared to prevent dispersal of *Rana sylvatica* into apparently suitable breeding habitat (Gibbs 1998). Genetic drift acting in a small population coupled with a lack of gene flow may have allowed the Oakville population to differentiate from neighboring populations.

CHAPTER IV

CONCLUSIONS

The pattern of colonization of the range of *Rana sylvatica* following glaciation is complex. This pattern is not unlike that of other species including the spring peeper, *Pseudacris crucifer*; mitochondrial DNA showed four main clades and divergent haplotypes in some populations (Austin et al. 2002). *Bufo fowleri* shows a pattern at the northern limit of its range consistent with two separate colonizations, and possibly hybridization with *Bufo americanus* as well (Smith and Green 2004). Steller's jays (*Cyanocitta stelleri*) show a genetic pattern consistent with recolonization from multiple refugia (Burg et al. 2005).

Some findings are more clear. Alaska was likely a glacial refugium for *R. sylvatica*. In spite of small sample size it appears the sampled population is significantly genetically differentiated from all other populations, even those as close as Alberta, Canada. The Rocky Mountains are likely a barrier to recolonization from the Alaskan refugium. Bayesian clustering reflects this history, placing Alaskan frogs consistently in their own population though further sampling is needed to lend more support to these findings.

Other source areas were revealed during this study as well. Both New York and southern Appalachia are likely glacial refugia. The New York populations are genetically differentiated from one another and have a high number of private alleles. One of the Quebec populations is more closely related to both New York populations

than the New York populations are to each other based on F_{ST} values. Due to the geographic distance between these populations current gene flow is unlikely. In this case it is possible this Quebec population was colonized by populations from New York. The southern Appalachian populations are also differentiated from one another, indicating that these were populations isolated from one another in another glacial refugium.

Rana sylvatica shows a clear pattern of isolation by distance at a continental scale (Chapter 2), as well as a local scale (Newman and Squire 2001). At the intermediate, regional level in the northern Great Plains there is no evident isolation by distance pattern. This is not surprising given the geologic history of the region. Lake Agassiz, the enormous proglacial lake covering much of the area and bounded on the northeast side by the Laurentide ice sheet, would have acted as a barrier to gene flow. Following the disappearance of the ice sheets were periods of significant droughts (Laird et al. 1996) that would likely prevent the recolonization and maintenance of amphibian populations. After warmer, moister weather prevailed, the Red River of the North remained as a barrier, with smaller rivers and streams forming a less insurmountable barrier. That Minnesota populations appear to be more closely related to the Manitoba populations than to the geographically closer North Dakota populations may be due to these barriers.

A secondary goal of this project was to determine the usefulness of microsatellites at this scale for this species. Mitochondrial DNA was able to discern an overall pattern of eastern and western groups (Irwin unpublished data). This project has shown microsatellites' utility in this type of study, has located putative glacial refugia, and reconstructed historical patterns on a regional level.

These studies also indicate, however, microsatellites, like all genetic analysis tools, are not flawless. Hardy-Weinberg disequilibrium can have multiple causes (Wahlund effect or inbreeding for example), including one or more alleles that fails to amplify during polymerase chain reaction (PCR) due to mutations in the primer binding sites, null alleles. The likely presence of null alleles produces artifacts in the form of low numbers of heterozygotes (Selkoe and Toonen 2006). This artificial result increases the amount of Hardy-Weinberg disequilibrium seen in many of the populations, particularly at locus Rs70.

While most studies report the presence of possible null alleles and estimate their frequencies, few actually attempt to directly verify the presence of null alleles or determine the reason for their failure to amplify in PCR (Dakin and Avise 2004). Callen et al. (1993) sequenced alleles at microsatellite loci on human chromosome 16 and found 30% of the microsatellite loci sequenced had null alleles. One of these null alleles was due to an 8 base pair deletion in the primer binding region (Callen et al. 1993). Sequencing of a null allele in the grey-sided vole, *Clethrionomys rufocanus*, found two base pair substitutions in and another adjacent to the primer binding region (Ishibashi et al. 1996). Offspring that are heterozygous for a null allele may not appear to have any alleles in common with a parent who is also heterozygous for a null allele (Dakin and Avise 2004). Studies that determine the effect of those null alleles on levels of heterozygosity and populations differentiation will be important for further work in this discipline.

APPENDIX A

Sample locations and collectors:

1. Buladean, North Carolina - Gene Spears
2. 3 miles northeast of Two Rivers, Wisconsin – Joanne Kline
3. Adams County, Ohio – Jason Irwin
4. Adams County, Ohio – Jason Irwin
5. 6 km east of Dillingham, Buncombe County, North Carolina – Jim Petranka
6. 4 km north of Topton, Graham County, North Carolina – Jim Petranka
7. Lake Hill, Mont Saint-Hilaire, Quebec – Martin Ouellet
8. East Hill, Mont Saint-Hilaire ,Quebec – Martin Ouellet
9. Sheyenne National Grasslands, North Dakota – Robert Newman
10. Saskatchewan – Allison Puchniak
11. 4494035N 428695E Zone 13, Larimer County, Colorado – Ken Kehmeier
12. 46.23N 89.42W, Gogebic County, Michigan – Mitch Bergeson
13. 55.07 113 deg-min, Near South Calling Lake, Alberta, Canada – Brian Eaton
14. 40.45.515 73.44.763 deg-dec min, Long Island, New York – Russell Burke
15. 40.50.302 73.32.030 deg-dec min, Long Island, SSP New York – Russell Burke
16. Alaska – Jason Irwin
17. Crookston, Minnesota – Tina Squire
18. Near Rydell National Wildlife Refuge, Minnesota – Tina Squire
19. 50.098N 96.016W Seven Sisters Falls, Manitoba – Tina Squire
20. Highway 308 south of Highway 1, Manitoba – Tina Squire
21. 5305201N 627351E Zone 14, Grand Forks, North Dakota – Tina Squire
22. 5319839N 578583E Zone 14, Petersburg, North Dakota – Robert Newman
23. 48.05409N96.93239W, East Grand Forks, Minnesota– Tina Squire
24. 5412961N 414370E Zone 14, Turtle Mountains – Robert Newman
25. 8 km east of Townsend, Blount County, Tennessee - Jim Petranka

APPENDIX B

The inbreeding value (F_{IS}), within individual variation (1-Q intra) and within population variation (1-Q inter) for each population.

	F_{IS}	1-Q intra	1-Q inter
Alaska	-0.1604	0.7958	0.659
Alberta	0.3885	0.6414	0.8029
Colorado	0.0553	0.5068	0.5453
Crookston	0.1363	0.7409	0.8525
E Grand Forks	0.0136	0.8406	0.8502
Hill Pond E	0.1257	0.7408	0.809
Manitoba 1	0.011	0.8646	0.8521
Manitoba 2	-0.009	0.8702	0.9223
Michigan	0.0093	0.8475	0.8532
North Carolina	0.1074	0.778	0.8697
New York	0.0776	0.7361	0.8085
Oakville	0.1414	0.6556	0.7898
Ohio 1	0.2106	0.7069	0.8969
Ohio 2	0.1845	0.6551	0.8508
Quebec 1	0.1695	0.6668	0.8244
Quebec 2	0.3031	0.5967	0.8902
Rydell	0.1654	0.711	0.8031
Saskatchewan	0.3368	0.575	0.8935
Sheyenne	0.1378	0.6996	0.8106
S Appalachia 1	0.0775	0.8003	0.8759
S Appalachia 2	0.2482	0.6319	0.9159
SSP New York	0.1559	0.6029	0.7634
Turtle Mtns	0.0946	0.7662	0.8284
Wisconsin	0.1675	0.7738	0.9186

APPENDIX C

Linkage disequilibrium for all populations for loci pairs with significant linkage disequilibrium findings. Significant p-values are bolded.

Pop	Locus Pair	p-value
North Carolina	Rs11 & Rs70	0.0804
Ohio 1	Rs11 & Rs70	1
S Appalachia 1	**Rs11 & Rs70**	**0**
S Appalachia 2	Rs11 & Rs70	1
Quebec 2	Rs11 & Rs70	1
Sheyenne	Rs11 & Rs70	1
Colorado	Rs11 & Rs70	1
Oakville	**Rs11 & Rs70**	**0.04631**
North Carolina	Rs11 & Rs88	1
Ohio 1	Rs11 & Rs88	1
Ohio 2	Rs11 & Rs88	1
Quebec 2	Rs11 & Rs88	1
Sheyenne	Rs11 & Rs88	1
Colorado	Rs11 & Rs88	0.88849
Alaska	Rs11 & Rs88	1
Rydell	Rs11 & Rs88	1
Oakville	**Rs11 & Rs88**	**0**
E Grand Forks	Rs11 & Rs88	1

APPENDIX D

Pairwise F_{ST} values.

	North Carolina	Wisconsin	Ohio1	Ohio2	S Appa-lachia1	S Appa-lachia2	Quebec1	Quebec2	Sheyenne	Saskatche-wan	Colorado	Michigan
Wisconsin	0.0301											
Ohio1	0.0254	0.0222										
Ohio2	0.0258	0.0386	0									
S Appalachia1	0.0419	0.0461	0.0306	0.0406								
S Appalachia2	0.0264	0.0282	0.0129	0.0291	0.0293							
Quebec1	0.0482	0.0417	0.0454	0.0422	0.0734	0.0437						
Quebec2	0.0452	0.025	0.0371	0.0454	0.0531	0.0163	0.0208					
Sheyenne	0.1272	0.0611	0.0948	0.1255	0.1148	0.1238	0.1571	0.1068				
Saskatchewan	0.0612	0.041	0.0357	0.0519	0.0648	0.0445	0.0813	0.0565	0.1148			
Colorado	0.2481	0.2084	0.2043	0.2369	0.2362	0.2176	0.2634	0.2102	0.2557	0.3015		
Michigan	0.0854	0.0297	0.0651	0.0952	0.1034	0.0801	0.0878	0.0761	0.1024	0.0834	0.248	
Alberta	0.0886	0.0562	0.0839	0.0792	0.0924	0.0814	0.1194	0.0711	0.0907	0.043	0.2794	0.1327
New York	0.1097	0.0929	0.0932	0.1134	0.1088	0.1073	0.1429	0.0764	0.1469	0.1371	0.2721	0.1326
SSP New York	0.1417	0.1091	0.1194	0.1405	0.1366	0.1149	0.123	0.0479	0.1727	0.1504	0.2834	0.1598
Alaska	0.22	0.1692	0.1647	0.1975	0.1895	0.1902	0.2319	0.1619	0.2054	0.1728	0.3437	0.2236
Crookston	0.0964	0.0631	0.0707	0.1076	0.0921	0.0811	0.1273	0.086	0.0536	0.0601	0.2434	0.0604
Rydell	0.1201	0.0871	0.0962	0.1343	0.1213	0.1187	0.1472	0.1139	0.0881	0.0976	0.2903	0.0816
Manitoba1	0.1057	0.0806	0.0761	0.1154	0.0892	0.0967	0.1316	0.0892	0.0831	0.0723	0.2486	0.0754
Manitoba2	0.058	0.0315	0.032	0.0588	0.0583	0.0479	0.063	0.0511	0.0374	0.0561	0.2405	0.0214
Oakville	0.1432	0.1178	0.1093	0.1485	0.1141	0.1424	0.1556	0.1341	0.1469	0.1461	0.2925	0.1197
Hill Pond East	0.1188	0.0925	0.0973	0.137	0.1058	0.1164	0.1466	0.1073	0.0881	0.1162	0.2635	0.079
East Grand Forks	0.1002	0.0756	0.0746	0.112	0.0873	0.0843	0.1331	0.0921	0.0774	0.0961	0.2669	0.0712
Turtle Mountains	0.1081	0.0811	0.085	0.1249	0.0904	0.1037	0.1485	0.0983	0.0798	0.0866	0.2182	0.1003

Pairwise F_{ST} values, continued.

	Alberta	New York	SSP New York	Alaska	Crookston	Rydell	Manitoba1	Manitoba2	Oakville	Hill Pond East	East Grand Forks
Wisconsin											
Ohio1											
Ohio2											
S Appalachia1											
S Appalachia2											
Quebec1											
Quebec2											
Sheyenne											
Saskatchewan											
Colorado											
Michigan											
Alberta											
New York	0.1704										
SSP New York	0.1862	0.1178									
Alaska	0.2318	0.1381	0.1624								
Crookston	0.1315	0.1127	0.1414	0.1717							
Rydell	0.1701	0.1551	0.1767	0.2209	0.0007						
Manitoba1	0.1379	0.141	0.165	0.2083	0.0181	0.0187					
Manitoba2	0.0635	0.0851	0.1164	0.1517	0.0059	0.0239	0.0057				
Oakville	0.1969	0.1558	0.2015	0.2285	0.073	0.0716	0.0694	0.0812			
Hill Pond East	0.1744	0.1468	0.1684	0.2186	0.0268	0.0268	0.026	0.0309	0.0574		
East Grand Forks	0.1468	0.1332	0.167	0.2167	0.0005	0.0125	0.0208	0.0092	0.0683	0.0095	
Turtle Mountains	0.1458	0.1411	0.164	0.2091	0.0225	0.0274	0.0196	0.0262	0.069	0.053	0.0188

APPENDIX E

Population differentiation for each population pair arranged by Fisher's method estimation of the p-value. Non-significant p-value estimations are bolded.

Population 1	Population 2	Chi2	df	p-value
Crookston	Manitoba1	Infinity	10	Highly sign.
Crookston	Oakville	Infinity	10	Highly sign.
E Grand Forks	Turtle Mtns	Infinity	10	Highly sign.
Hill Pond E	Turtle Mtns	Infinity	10	Highly sign.
Manitoba1	E Grand Forks	Infinity	12	Highly sign.
Manitoba1	Hill Pond E	Infinity	12	Highly sign.
Manitoba1	Oakville	Infinity	12	Highly sign.
Manitoba2	Oakville	Infinity	12	Highly sign.
Oakville	E Grand Forks	Infinity	12	Highly sign.
Oakville	Hill Pond E	Infinity	12	Highly sign.
Oakville	Turtle Mtns	Infinity	10	Highly sign.
Rydell	Hill Pond E	Infinity	12	Highly sign.
Rydell	Oakville	Infinity	12	Highly sign.
Sheyenne	Crookston	Infinity	10	Highly sign.
Sheyenne	E Grand Forks	Infinity	12	Highly sign.
Sheyenne	Hill Pond E	Infinity	12	Highly sign.
Sheyenne	Manitoba1	Infinity	12	Highly sign.
Sheyenne	Manitoba2	Infinity	12	Highly sign.
Sheyenne	Oakville	Infinity	12	Highly sign.

Population differentiation for each population pair arranged by Fisher's method estimation of the p-value. Non-significant p-value estimations are bolded, continued.

Population 1	Population 2	Chi2	df	p-value
Sheyenne	Rydell	Infinity	12	Highly sign.
Sheyenne	Turtle Mtns	Infinity	10	Highly sign.
Crookston	Hill Pond E	63.742	10	0.0000
Manitoba2	Hill Pond E	51.511	12	0.0000
Rydell	E Grand Forks	41.719	12	0.0000
Manitoba1	Turtle Mtns	37.289	10	0.0001
Crookston	Turtle Mtns	32.376	8	0.0001
Manitoba2	Turtle Mtns	31.676	10	0.0005
Rydell	Manitoba1	33.920	12	0.0007
Hill Pond E	E Grand Forks	30.772	12	0.0021
Rydell	Turtle Mtns	27.396	10	0.0023
Rydell	Manitoba2	26.959	12	0.0078
Crookston	E Grand Forks	23.855	10	0.0080
Manitoba2	**E Grand Forks**	**20.274**	**12**	**0.0621**
Manitoba1	**Manitoba2**	**17.633**	**12**	**0.1273**
Crookston	**Rydell**	**15.000**	**10**	**0.1321**
Crookston	**Manitoba2**	**14.268**	**10**	**0.1611**

APPENDIX F

Linkage disequilibrium for all populations for affected loci pairs. Significant values are bolded.

Population	Locus 1	Locus 2	p-value
E Grand Forks	Rs11	Rs88	1
Oakville	**Rs11**	**Rs88**	**0**
Rydell	Rs11	Rs88	1
Sheyenne	Rs11	Rs88	1
Hill Pond E	Rs55	Rs70	0.24958
Manitoba 1	Rs55	Rs70	1
Oakville	Rs55	Rs70	0.50852
Sheyenne	**Rs55**	**Rs70**	**0.0006**

REFERENCES

Anderson LL, Hu FS, Nelson DM, Petit RJ, Paige KN (2006) Ice-age endurance: DNA evidence of a white spruce refugium in Alaska. *PNAS* 103, 12447-12450.

Alsos IG, Engelskjon T, Gielly L, Taberlet P, C. Brochmann C (2005) Impact of ice ages on circumpolar molecular diversity: insights from an ecological key species. *Molecular Ecology*, 14, 2739–2753.

Ashworth AC, Cvancara AM (1983) Paleoecology of the southern part of the Lake Agassiz basin. In JT Teller and L Clayton eds. Glacial Lake Agassiz. St. John's, Newfoundland: Geological Association of Canada.

Austin JD, Lougheed SC, Neidrauer L, Chek AA, Boag PT (2002) Cryptic lineages in a small frog: the post-glacial history of the spring peeper, *Pseudacris crucifer* (Anura: Hylidae). *Molecular Phylogenetics and Evolution*, 25, 316-329.

Avise JC (2000) Phylogeography: the history and formation of species. Cambridge, Mass: Harvard University Press.

Avise JC (2004) Molecular markers, natural history and evolution. Sunderland, Mass. Sinauer Associates, Inc.

Beebee TJC, Rowe G (2000) Microsatellite analysis of natterjack toad *Bufo calamita* Laurenti populations: consequences of dispersal from a Pleistocene refugium. *Biological Journal of the Linnean Society*, 69, 367–381.

Bernatchez L, Wilson CC (1998) Comparative phylogeography of Neartic and Paleartic fishes. *Molecular Ecology*, 7, 431-452.

Berven, KA (1990) Factors affecting population fluctuations in larval and adult stages of the wood frog (*Rana sylvatica*). *Ecology*, 71, 1599-1608.

Bonin A, Bellemain E, Eidesen PB, Pompanon F, Brochmann C, Taberlet P (2004) How to track and assess genotyping errors in population genetics studies. *Molecular Ecology*, 13, 3261-3273.

Brant SV, Orti G (2003) Phylogeography of the Northern short-tailed shrew, *Blarina brevicauda* (Insectivora: Soricidae): past fragmentation and postglacial recolonization. *Molecular Ecology*, 12, 1435-1449.

Brede EG and Beebee TJC (2004) Contrasting population structures in two sympatric anurans: implications for species conservation. *Heredity*, 92, 110–117.

Brookfield J F Y, (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology*, 5, 453-455.

Broquet T, Petit E (2004) Quantifying genotyping errors in noninvasive population genetics. *Molecular Ecology*, 13, 3601-3608.

Browder LW, Underhill JC, Merrell DJ (1966) Mid-dorsal strip in the wood frog. Journal of *Heredity*, 57, 65-67.

Burg TM, Gaston AJ, Winker K, Friesen VL (2005) Rapid divergence and postglacial colonization in western North American Steller's jays (*Cyanocitta stelleri*). *Molecular Ecology*, 14, 3745-3755.

Burg TM, Gaston AJ, Winker K, Friesen VL (2006) Effects of Pleistocene glaciations on population structure of North American chestnut-backed chickadees. *Molecular Ecology*, 8, 1513-1520.

Callen DF, Thompson AD, Shen Y, Phillips HA, Richards RI, Mulley JC, Sutherland GR (1993) Incidence and origin of "null" alleles in the (AC)n microsatellite markers. *American Journal of Human Genetics*, 52, 922-927.

Cavalli-SforzaLL, EdwardsAWF (1967) Phylogenetic analysis: models and estimation procedures. *Evolution*, 32, 550–570.

Crespi EJ, Rissler LJ, Browne RA (2003) Testing Pleistocene refugia theory: phylogeographical analysis of *Desmognathus wrighti*, a high-elevation salamander in the southern Appalachians. *Molecular Ecology*, 12, 969-984.

Conroy CJ, Cook JA (2000) Phylogeography of a post-glacial colonizer: *Microtus longicaudus* (Rodentia: Muridae). *Molecular Ecology*, 9, 165-175.

Curtis JMR, Taylor EB (2004) The genetic structure of coastal giant salamanders (*Dicamptodon tenebrosus*) in a managed forest. *Biological Conservation*, 115, 45–54.

Dakin EE, Avice JC (2004) Microsatellite null alleles in parentage analysis. *Heredity*, 93, 504–509.

Davis MB, Shaw RG (2001) Range shifts and adaptive responses to Quaternary climate change. *Science*, 292, 673-679.

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Stat. Soc.*, 39, 1-38.

DeWoody J, Nason JD, Hipkins VD (2006) Mitigating scoring errors in microsatellite data from wild populations. *Molecular Ecology Notes*, 6, 951–957.

Eisen JA (1999) Mechanistic explanations for variation in microsatellite stability within and between species. Pp. 34–48 *in* D. GOLDSTEIN and C. SCHLÖTTERER, eds. Microsatellites: evolution and applications. Oxford University Press, Oxford, England.

Fedorov VB, Goropashnaya AV, Jaarola M, Cook JA (2003) Phylogeography of lemmings (*Lemmus*): no evidence for postglacial colonization of Arctic from the Beringian refugium. *Molecular Ecology*, 12, 725-731

Fenton MM, Moran SR, Teller JT, Clayton L (1983) Quaternary stratigraphy and history in the southern part of the Lake Agassiz basin. In JT Teller and L Clayton eds. Glacial Lake Agassiz. St. John's, Newfoundland: Geological Association of Canada.

Fishbeck DW, Underhill JC (1971) Distribution of stripe polymorphism in wood frogs, *Rana sylvatica* LeConte, from Minnesota. *Copeia* 2, 253-259.

Funk WC, Blouin MS, Corn PS, Maxell BA, Pilliod DS, Amish S et al. (2005). Population structure of Columbia spotted frogs (*Rana luteiventris*) is strongly affected by the landscape. *Molecular Ecology*, 14, 483–496.

Gibbs JP (1998) Distribution of woodland amphibians along a forest fragmentation gradient. *Landscape Ecology*, 13, 263-268.

Good JM, Sullivan J (2001) Phylogeography of the red-tailed chipmunk (*Tamias ruficaudus*), a northern Rocky Mountain endemic. *Molecular Ecology*, 10, 2683–2695.

Hancock, J. 1999. Microsatellites and other simple sequences:genomic context and mutational mechanisms. Pp. 1–9 *in* D Goldstein and C Schlotterer, eds. Microsatellites: evolution and applications. Oxford University Press, Oxford, England.

Hauser L , Adcock GJ , Smith PJ, Bernal Ramírez JH, Carvalho GR (2002) Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (*Pagrus auratus*). *PNAS*, 99, 11742-11747.

Hewitt, GM (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, 68, 87-112.

Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature*, 405, 907-913.

Hewitt GM (2001) Speciation, hybrid zones and phylogeography — or seeing genes in space and time. *Molecular Ecology*, 10, 537-549.

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Royal Soc B Biological Sciences*, 359, 183-195.

Hoffman EA, Blouin MS (2004) Evolutionary history of the northern leopard frog: reconstruction of phylogeny, phylogeography, and historical changes in population demography from mitochondrial DNA. *Evolution*, 58, 145–159.

Holman JA (1995) Pleistocene Amphibians and Reptiles in North America. New York: Oxford University Press.

Ibrahim KM, Nichols RA, Hewitt GM (1996) Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity*, 77, 282-291.

Ishibashi Y, Saitoh T, Abe S, Yoshida MC (1996) Null microsatellite alleles due to nucleotide sequence variation in the grey-sided vole *Clethrionomys rufocanus*. *Molecular Ecology*, 5, 589-590.

Jensen JL, Bohonak AJ, Kelley ST (2005) Isolation by distance, web service. BMC Genetics, 6, 13.

Johansson M, Primmer CR, Merilla J (2006) History vs. current demography: explaining the genetic population structure of the common frog (*Rana temporaria*). *Molecular Ecology*, 15, 975-983.

Julian SE, King TL (2003) Novel tetranucleotide microsatellite DNA markers for the wood frog, *Rana sylvatica*. *Molecular Ecology Notes*, 3, 256–258.

Kalinowski (2002) How many alleles per locus should be used to estimate genetic distances? *Heredity*, 88, 62-65.

Klicka J, Zink RM (1997) The importance of recent ice ages in speciation: a failed paradigm. *Science*, 277, 1666-1669.

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

Laird KR, Fritz SC, Maasch KA, Cumming BF (1996) Greater drought intensity and frequency before AD 1200 in the northern Great Plains, USA. *Nature*, 384, 552-554.

Lesica P, Allendorf FW (1995) When are peripheral populations valuable for conservation? *Conservation Biology*, 9, 753-760.

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

Martınez-Solano I, Rey I, Garcia-Paris M (2005) The impact of historical and recent factors on genetic variability in a mountain frog: the case of *Rana iberica* (Anura: Ranidae). *Animal Conservation*, 8, 431-441.

Martof BS, Humphries RL (1959) Geographic variation in the wood frog *Rana sylvatica*. *The American Midland Naturalist*, 61, 350-389.

Masta SE, Laurent NM, Routman EJ (2003) Population genetic structure of the toad *Bufo woodhousii*: an empirical assessment of the effects of haplotype extinction on nested cladistic analysis. *Molecular Ecology*, 12, 1541–1554.

Milot E, Gibbs HL, Hobson KA (2000) Phylogeography and genetic structure of northern populations of the yellow warbler (*Dendroica petechia*). *Molecular Ecology*, 9, 667-681.

Nei M, Chesser RK (1983) Estimation of fixation indices and gene diversities. *Annals of Human Genetics*, 47, 253-9.

Nesbø CL, Fossheim T, Vøllestad LA, Jakobsen KS (1999) Genetic divergence and phylogeographic relationships among European perch (*Perca fluviatilis*) populations reflect glacial refugia and postglacial colonization. *Molecular Ecology*, 8, 1387–1404.

Newman RA, Squire T (2001) Microsatellite variation and fine-scale population structure of wood frogs (*Rana sylvatica*). *Molecular Ecology*, 10, 1087-1100.

Nichols RA, Hewitt GM (1994) The genetic consequences of long distance dispersal during colonization. *Heredity*, 72, 312-317.

Pielou EC (1991) After the ice age: the return of life to glaciated North America. Chicago: University of Chicago Press.

Porter KR (1969) Evolutionary status of the rocky mountain population of wood frogs. *Evolution*, 23, 163-170.

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

Pruette CL, Winkler K (2005) Northwestern song sparrow populations show genetic effects of sequential colonization. *Molecular Ecology*, 14, 1421–1434.

Raymond M, Rousset F (1995) An exact test for population differentiation. *Evolution*, 49, 1280–1283.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, 43, 223–225.

Rousset F (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*, 145, 1219–1228.

Rousset F, (2007) Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Notes*, (OnlineEarly Articles).

Rowe G, Beebee TJC, Burke T (2000) A microsatellite analysis of natterjack toad, *Bufo calamita*, metapopulations. *Oikos*, 88, 641–651.

Scribner KT, Arntzen JW, Cruddace N, Oldham RS, Burke T (2001) Environmental correlates of toad abundance and population genetic diversity. *Biological Conservation*, 98, 201–210.

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

Smith MA, Green DM (2004) Phylogeography of *Bufo fowleri* at its northern range limit. *Molecular Ecology*, 13, 3723–3733.

Squire T, Newman RA (2002) Fine-scale population structure in the wood frog (*Rana sylvatica*) in a northern woodland. *Herpetologica*, 58, 119–130.

Stebbins RC (1985) A field guide to western reptiles and amphibians. Boston: Houghton Mifflin Company.

Stevens VM, Verkenne C, Vandewoestijne S, Wesselingh RA, Baguette M (2006) Gene flow and functional connectivity in the natterjack toad. *Molecular Ecology*, 15, 2333-2344.

Stewart KW, Lindsay CC (1983) Postglacial dispersal of lower vertebrates in the Lake Agassiz region. In JT Teller and L Clayton eds. Glacial Lake Agassiz. St. John's, Newfoundland: Geological Association of Canada.

Stone KD, Flynn RW, Cook JA (2002) Post-glacial colonization of northwestern North America by the forest-associated American marten (*Martes americana*, Mammalia: Carnivora: Musteiidae). *Molecular Ecology*, 11, 2049–2063.

Taberlet P, Cheddadi R (2002) Quaternary refugia and persistence of biodiversity. *Science*, 297, 2009-2010.

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.

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

Zhang D-X, Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology*, 12, 563-584.

Zeisset I, Beebee TJC (2001). Determination of biogeographical range: an application of molecular phylogeography to the European pool frog *Rana lessonae*. *Proc Roy Soc London B*, 268: 933–938.