EASTERN MASSASAUGA RATTLESNAKES

THE POPULATION, EVOLUTIONARY AND CONSERVATION GENETICS OF THE EASTERN MASSASAUGA RATTLESNAKE (*SISTURUS CATENATUS CATENATUS*)

By

REBECCA L. PEARCE, B.Sc.

A Thesis Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Master of Sciences

McMaster University

September 2001

MASTER OF SCIENCES 2001
(Biology)

McMaster University
Hamilton, Ontario

TITLE: The Population and Evolutionary Genetics of the Eastern Massasauga Rattlesnake (*Sistrurus catenatus catenatus*)

AUTHOR: Rebecca L. Pearce, B.Sc. (Queen's University)

SUPERVISOR: Dr. H. Lisle Gibbs

NUMBER OF PAGES: x, 126

ABSTRACT

This thesis assesses two distinct aspects of the genetics of the eastern massasauga rattlesnake (*Sistrurus catenatus catenatus*). Firstly, work relating to the population and evolutionary genetics of the eastern massasauga rattlesnake, and the conservation implications of the results are presented. In this study, variation at a portion of mitochondrial DNA (mtDNA), and at two nuclear intron DNA loci was examined for snakes from various populations, in order to examine the partitioning of genetic diversity among populations. Seven mtDNA haplotypes were identified, most found in one or two U.S. populations of snakes. Intron variation was low and phylogenetically uniformative. Neither mitochondrial nor nuclear DNA showed any phylogenetic structuring. Results also suggest that eastern massasauga rattlesnakes have undergone a post-Pleistocene population expansion from a glacial refugium. From a conservation perspective, three geographical groupings of populations that showed significant genetic differentiation were identified, suggesting that these groups should be managed independently, in order to conserve within-species variation.

Secondly, the concerted evolution of two control-region-like sequences in eastern massasauga mitochondria is discussed. In this study, two entire control-region-like sequences were sequenced in 10 eastern massasauga rattlesnakes, one found in the cytb-12S rRNA region, and the other in the IQM tRNA cluster, as well as the tRNA sequences flanking these regions. Also, the sequence of the last half of both control regions for 28 individual snakes was compared. Both control regions were found to have a divergence of less than 0.5% within individuals. Among the 28 individuals only three showed different sequences between the two control regions, and these sequences differed at only one nucleotide. The similarity of these separate regions within individuals could be attributed to concerted evolution, potentially due to gene conversion. A minimum gene conversion rate of one conversion per 1.2 million years is estimated for the sequences.

ACKNOWLEDGEMENTS

There are a great number of people who provided me with academic, financial, and social support during the completion of my Master's thesis. I would like to thank my supervisor, Lisle Gibbs, for providing me with the opportunity to work on this very interesting project, and for offering me financial and academic support. The members of the Gibbs lab, Liliana DeSouza, Leanne Blanchard, Anoma Patirana, Marylene Boulet, Natalie Bulgin, Bonnie Woolfenden, Kevin Shonk, and Kevin Ellison, have been fantastic colleagues and friends, and provided me with an enriching and stimulating environment in which to conduct my research. I would like to thank Richard Morton for agreeing to sit on my defense committee and for offering invaluable suggestions and advice, and Susan Dudley for participating in my defense. Many thanks to Chris Parent, for showing me how to bleed a rattlesnake! I would like to acknowledge my friends in the department and elsewhere, especially my roommates in Hamilton – Victoria, Marc, Monica, Jen and Tim - I have and will always treasure your friendship. Finally, I would like to thank my family, particularly my parents, Bill and Linda, my sister Jessica, and Jason Baardsnes. Your love, patience, and financial and emotional support have carried me through the last two years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1
General Introduction

Conservation Genetics

Frankham (1995) argues that there are seven major genetic issues in conservation biology. These include inbreeding depression, accumulation and loss of deleterious mutation, loss of genetic variation in small populations, genetic adaptation to captivity and its effects on reintroduction success, outbreeding depression, fragmentation of populations and reduction in migration, and taxonomic uncertainties and introgression. The use of molecular markers can provide information that is useful in addressing two types of questions related to these issues: questions related to genetic conservation, which is the use of genetic information to measure and manage genetic diversity for its own sake, and those related to molecular ecology, which uses genetic analysis as a complement to ecological studies of demography (Moritz 1994a). From a molecular ecology perspective, information on the genetic structure of populations can contribute to a better understanding of the ecology and biogeography of a species by providing information about gene flow and genetic isolation that cannot be detected using traditional demographic measures (Prior *et al.* 1996). Genetic markers can identify populations that are in need of new donor individuals due to reductions in population size, populations that have individuals that can act as donors to more vulnerable populations, and populations that may further warrant demographic or environmental consideration (Haig 1998). Molecular genetic information can also be used to appraise phylogenetic discontinuities within species or subspecies, which can provide a better foundation for the recognition and management of diversity (Avise 1989). Defining the structure of populations can lead to more accurate population management at the beginning of recovery efforts (Haig 1998).

One of the most difficult questions in assessing population differentiation is how genetic distinctness should be measured. Since a species' genetic variability represent the reservoir upon which future evolutionary potential depends, it is important to identify how genetic diversity is partitioned within a species, and to conserve as much of it as possible (Waples 1995). The National Marine Fisheries Service (NMFS) has adopted a policy of protecting evolutionarily significant units (ESUs), which are defined on the basis of such criteria as isolation, and the contribution of particular populations to the ecological-genetic diversity of the species (Waples 1995). In order to clarify the meaning of ESUs, and to avoid the issue of "how much variation is enough" Moritz (1994b) proposed a new definition for ESU, as well as a management units (MUs). ESUs, which emphasize historical population structure rather than current adaptation, are defined as being reciprocally monophyletic for mitochondrial DNA (mtDNA) haplotypes, and show significant divergence of allele frequencies at nuclear loci (Moritz 1994b). MUs, which are more suitable for short-term management, are defined as populations with significant divergence of allele frequencies at nuclear or mitochondrial loci, regardless of the phylogenetic distinctiveness of the alleles (Moritz 1994b).

The evolutionary history of a species, including founder effects, vicariance events, and range expansions, can also be reflected by its present genetic makeup. The effects of Pleistocene ice ages, which had a great effect on the genetic processes and structure of various species, can often be deduced from genetic patterns (Hewitt 1996). For example, genetic markers have the potential to provide evidence for the isolation and differentiation of populations in glacial refugia (e.g. Holder *et al.* 1999), and indeed the

current population structures of such diverse organisms as the lake whitefish *Coregonus clupeaformis* (Bernatchez and Dodson 1991), common guillemot *Uria aalge* (Friesen *et al.* 1996), greenfinch *Carduelis chloris* (Merila *et al.* 1997), California mountain kingsnake *Lampropeltis zonata* (Rodriguez-Robles *et al.* 1998), and marbled murrelets *Brachyramphus marmoratus* (Congdon *et al.* 2000), have been attributed to expansions from glacial refugia. As population history can potentially play a large role in determining levels of population differentiation (e.g. Avise *et al.* 1998, Hewitt 1996), one of the greatest challenges in population genetic studies is to separate population history from contemporary processes. For example, if populations do not exist long enough to approach equilibrium between gene flow and drift, then population structure should be interpreted in terms of history, rather than present-day processes (Bohanak 1999). Also, the time scale over which genealogical analyses are appropriate, and over which the inference of demographic information is possible, is dependent on the mutation rate of the markers available (Milligan *et al.* 1994).

Molecular Markers

The introduction of polymerase chain-reaction (PCR)-based molecular techniques has increased the availability of DNA-based genetic markers for population and evolutionary research. One widely-based marker is mtDNA sequence data (e.g. Avise 1989). For a number of reasons, mitochondrial DNA (mtDNA) is far more sensitive than nuclear DNA at distinguishing recent genetic population structure and demographic events in the recent past. MtDNA has a high copy number due to the large number of mitochondria per cell (Wilson *et al.* 1985). MtDNA is inherited maternally, and appears

to be haploid (Wilson *et al.* 1985). Wilson *et al.* (1985) reported that the mean rate of divergence over the whole mtDNA is about 2% per million years (my), meaning it evolves at least 5 to 10 times faster than single-copy nuclear DNA (Brown *et al.* 1979). Zamudio and Greene (1997) provided an estimate of 0.47-1.32% per my for overall mtDNA divergence rates for small to medium-sized ectotherms, including snakes.

The circular, double-stranded DNA of mitochondria contains 37 genes coding for 22 tRNAs, 13 mRNAs, and 2 rRNAs, as well as a non-coding control region; the overall rate of mtDNA evolution is high, and the molecule as a whole can exhibit considerable mutation rate heterogeneity (Wilson *et al.* 1985). The high rate of evolution of mtDNA is thought to be due to its inefficient repair mechanism, and a lack of mismatch repair system, which results in a higher rate of transitions over transversions (Wilson *et al.* 1985). MtDNA lacks recombination, so it can essentially be represented as a single supergene (Friesen *et al.* 1997). This supposed lack of recombination, and its haploidy, make mtDNA useful as a genealogical tool (Wilson *et al.* 1985). Since mtDNA is haploid and maternally inherited, it has 1/4 the effective population size (N_e) of nuclear DNA, which means it is more sensitive at detecting demographic events such as bottlenecks, in which populations can lose much of their mtDNA variability without losing a significant portion of their nuclear variability (Avise 1994). All of these characteristics makes mtDNA an extremely useful molecular tool, and the phylogenies of mtDNA haplotypes within a species can normally be estimated by reasonable evolutionary criteria such as cladistic analysis or parsimony, and the results compared to expectations of theoretical demographic models to make inferences about population

histories (Avise 1990). Due to its high rate of evolution, mtDNA is most useful for examining relationships among species and populations that have diverged recently – within the past 5 to 10 million years (Brown *et al.* 1979).

Within the mitochondria, the control region is a non-coding region containing three-strand displacement loop (D loop), and is thought to be the most rapidly evolving region of the mtDNA molecule (Baker and Marshall 1997). The D-loop contains the heavy strand origin of replication, and transcription promoters for both strands (Baker and Marshall 1997). Rates of substitution in avian species have been found to be as high as 21.8% per my (Quinn *et al.* 1991), which is 10 times the average rate for the rest of the mtDNA genome. The control region can be divided into three parts, parts III and I corresponding to the 3' and 5' ends, respectively, and part II, the conserved central block (Wenink *et al.* 1994).

Another group of DNA markers that are increasingly used are nuclear DNA introns. Nuclear introns are noncoding sequences in nuclear DNA, also called intervening sequences (IVS) (Nei 1987). The function of introns is not known, but they could possibly be historical remnants with no important function (Nei 1987). It has also been suggested that recombinational events in introns may facilitate the evolution of genes by permitting the construction of new genes from portions of existing ones (Gilbert 1978). From an evolutionary genetics standpoint, introns are useful for dating past evolutionary events during certain historical timeframes, as they evolve more rapidly than nondegenerate sites in protein-coding nuclear genes, at a rate of approximately 1% per my (Avise 1994). Thus, introns would be expected to provide more information about

more recent evolutionary events than many types of coding nuclear DNA. Analysis of intron sequence can provide high-resolution allelic phylogenies comparable to those from mtDNA studies, and PCR primers that anneal to exons of highly conserved nuclear genes have been found to be of broad taxonomic utility (e.g. Palumbi and Baker 1994, Friesen *et al.* 1997, Condgon *et al.* 2000).

Methods for Assaying Variation

Combined with sequencing, analysis of single-stranded conformational polymorphisms (SSCP) can be an efficient tool to detect variation, as the technique can be used to survey a large number and variety of genes to identify individuals with unique genotypes, which can then be sequenced (Friesen *et al.* 1997). SSCP analysis is based on the behaviour of DNA during electrophoresis through an acrylamide gel (Lessa and Applebaum 1993). SSCP analysis uses the principle that the molecular conformation of single stranded DNA depends on its nucleotide sequence, and even one sequence difference will alter this conformation, and consequently, the molecule's mobility through a gel (Lessa and Applebaum 1993). The technique has its limits. For strands of 300 to 450 base pairs, the probability of detecting differences in mobility is only 89% (Lessa and Applebaum 1993). Also, the optimal conditions for SSCP analysis are often difficult to determine and replicate (Lessa and Applebaum 1993).

Assessment of population structure using multiple DNA markers is important for a number of reasons. As a gene tree from on a single genetic locus represents only a small fraction of the organismal phylogeny (Avise 1994), reliance on a single locus greatly diminishes power to detect significant spatial or temporal structure. Genetic drift in

populations involves random change in gene frequencies and these changes will not occur in exactly the same way among independent loci (Palumbi and Baker 1994). Differences in the dispersal of males versus females means that the population structure of maternally inherited mtDNA may differ from that of biparentally inherited nuclear DNA (Palumbi and Baker 1994). For this reason, mtDNA markers should be used in conjunction with nuclear markers in order to obtain a more complete picture of population divergence and evolutionary history within a species.

Study Species: Eastern Massasauga Rattlesnake (Sistrurus catenatus catenatus)

There are three recognized species in the genus *Sistrurus*, including the pygmy rattlesnake (*S. miliarius*), which is found in southeastern North America, the massasauga rattlesnake (*S. catenatus*), which ranges from southcentral North America northeast to the Great Lakes Region, and *S. ravus*, known as Mexican pygmy rattlesnakes, which are found in Mexico (Cook 1992), however doubt has been raised about this latter species belonging to the genus *Sistrurus*. (Cook 1992). There are three subspecies of *Sistrurus catenatus*, including the desert grassland massasauga rattlesnake (*Sistrurus c. edwarsii*) the western massasauga rattlesnake (*Sistrurus c. tergiminus*), and the eastern massasauga rattlesnake (*Sistrurus c. catenatus*). While the two previous subspecies have a more southern range, the eastern massasauga is found only in northeastern North America. The range of the eastern massasauga is not uniform, but rather discreet and localized. In Canada, the two largest extant populations of the eastern massasauga occur in Ontario along the Bruce Peninsula, and inland portions of the eastern shore of Georgian Bay (Prior *et al.* 2001). Eastern massasaugas are found in two other smaller populations in

Ontario: Wainfleet Marsh, located in Southeastern Ontario, and Ojibway Prairie, located in the city of Windsor (Prior *et al.* 2001). In the United States, eastern massasaugas are found in small, isolated populations in ten states: New York, Pennsylvania, Ohio, Michigan, Indiana, Illinois, Iowa, Wisconsin, Minnesota and Missouri.

As members of the Family Viperidae, eastern massasaugas have the hallmark physical characteristics of pitvipers, including head-sensitive pits between the nostril and eye, a triangular head, diamond-shaped eyes, and a stubby rattle. They have gray or brownish bodies, with dark dorsal blotches, and three rows of smaller dark spots on each side of their bodies. Eastern massasaugas sexually mature at an approximate age of three to four years, and are active from mid-April until late October, although widespread variation in timing of reproduction, and hibernation and activity periods has been noted in snakes from different geographical localities (Seigel 1986). In the spring, eastern massasaugas are commonly found in upland habitats that have confierous cover in a low tree canopy density, and in the summer, they prefer wetlands with high vegetation coverage, such as marshes or fens (Hutchinson *et al.* 1992). Prey items include mammals, other snakes, and occasionally amphibians (Seigel 1986).

Glaciation has likely played a role in the present-day distribution of the eastern massasauga. The Wisconsin glaciation, which began approximately 100 000 years before present, began to retreat about 18 000 years ago, affected species in North America due to lower mean temperatures, and ice-sheet advances, which may have fragmented species' ranges. (Pielou 1991). The Hypsithermal, a period of drier, warmer temperatures, was present 5 000 years ago, or 13 000 years after the Wisconsin retreat

began. Some species that are found in the Great Lakes basin have been regarded as relicts of a prairie peninsula corridor thought to have existed during this time (Pielou 1991). Indeed, eastern massasaugas are cited as a species with a typical corridor distribution in that they existed in this dry corridor through southern Ontario (Pielou 1991). The present disjunct localities of massasauga may be reflection of climatic cooling and vegetational changes after the Hypsithermal, as well as habitat change wrought by human settlers (Cook 1992).

Historically, the range of the eastern massasauga in Ontario extended throughout the mesic prairies and wetlands that were common in southwestern and west-central Ontario. However, drastic modern-day reduction in range has created geographically isolated populations (Beltz 1992). Many populations of rattlesnakes are becoming more disjunct, as corridors of acceptable habitat disappear (Moran 1992). While accurate size estimates are not available for populations (Prior *et al.* 2001), a decrease in range suggests a decline in population number. For example, historically, eastern massasaugas were found in nineteen populations from six counties in Pennsylvania, while in 1977, they were found in six extant population in three counties (Beltz 1992). These changes are due to different factors, including habitat fragmentation and destruction and land modification (Parent and Weatherhead 2000, Seigel *et al.* 1998, Beltz 1992), vehicular traffic (Seigel 1986), and site disturbance (Parent and Weatherhead 2000). In 1990, the eastern massasauga rattlesnake was protected by Ontario Regulation 263/90 under the Game and Fish Act, and in April, 1991, it was declared a threatened species in Canada by the Committee on the Status of Endangered Wildlife in Canada (COSWEIC) (Beltz

1992). The eastern massasauga is currently a candidate for protection under the U.S. Endangered Species Act (http://endangered.fws.gov/wildlife.html).

Pitvipers in particular may be highly vulnerable to extinction due to their restricted distribution, susceptibility to habitat destruction, low reproductive rates, and high mortality due to human persecution and destruction of individuals and habitats (Seigel *et al.* 1998). Protected areas are vulnerable to succession, and may naturally develop into habitat that is unsuitable for rattlesnakes. The complexity of the massasauga's bio-requirements also makes it difficult to pinpoint specific habitat for protection (Moran 1992). There is also a lack of knowledge regarding the life history and critical habitat relationships of the massasauga, which limits the ability to manage them effectively (Hay 1992). Genetic data can provide information about population structure, as well as factors such as effective population sizes, and gene flow, the knowledge of which can contribute to management practices.

Few studies of the genetic structure of snake populations have been undertaken, most likely due to the secretive behaviour and low population densities of most snake species, which make it difficult to obtain large samples (Bushar *et al.* 1998). Genetic studies of snakes are few and far between, and include work on the black rat snake *Elaphe obsoleta* (Prior et al. 1996, Burbrink *et al.* 2000), the bushmaster *Lachesis muta* (Zamudio and Greene 1997), California mountain kingsnake *Lampropeltis zonata* (Rodriguez-Robles *et al.* 1998), the timber rattlesnake *Crotalus horridus* (Bushar et al. 1998), the pitviper clade *Agkistrodon* (Parkinson *et al.* 2000), the western rattlesnake *Crotalus viridis* (Pook *et al.* 2000), and the bockadam snake *Cerberus rynchops* (Karns *et*

al. 2000). The majority of intraspecific studies (Prior *et al.* 1996, Gibbs *et al.* 1997, Bushar *et al.* 1998), have shown that geographically isolated population exhibit high genetic divergence. Phylogenetic studies have shown the presence of well-differentiated evolutionary lineages in the bushmaster (Zamudio and Greene 1997), cantil *Agkistrodon binineatus* (Parkinson *et al.* 2000), western rattlesnake (Pook *et al.* 2000), and rat snake (Burbrink *et al.* 2000).

In 1991 a conservation genetics study of eastern massasauga rattlesnakes was initiated in order to help evaluate the conservation status of populations, and to contribute information that might assist in the long-term management of extant populations. The most recent results of this work are presented in Gibbs *et al.* (1997), who used six microsatellite loci to examine levels of differentiation amongst three Ontario and two U.S. populations of eastern massasauga rattlesnakes. Results of this study showed that geographically separated populations of massasaugas were genetically distinct, and that each population contained a large portion of the overall variation found in the subspecies (Gibbs *et al.* 1997). There was also evidence for fine-scale genetic differentiation within populations (Gibbs *et al.* 1997). While knowing how genetically distinct isolated populations are from one another can contribute to the conservation of the eastern massasauga, an important issue not addressed by this study was whether any significant phylogenetic relationships, or distinct evolutionary lineages, occur amongst populations of this subspecies (Gibbs *et al.* 1997). Likewise, the pattern of large amounts of structure needs to be confirmed with other nuclear DNA markers.

Thesis Outline

This thesis examines geographic patterns of genetic variation in the eastern massasauga rattlesnake using nuclear intron and mitochondrial control region sequence. The three main objectives of this study were to 1) determine whether any phylogenetic relationships exists between populations of eastern massasauga rattlesnakes based on mtDNA variation, 2) compare levels of divergence obtained using different molecular markers, namely nuclear introns, mitochondrial DNA, and microsatellites, and 3) assess variation from individuals from a greater number of populations than in Gibbs *et al.*, (1997). In particular, snakes from small U.S. populations not surveyed by Gibbs *et al.* 1997 were included in this study.

Chapter 2 describes patterns of genetic differentiation in the mitochondrial control region and two nuclear introns among eastern massasauga populations in Canada and the United States. I use this information to assess the distribution of genetic variation among populations, and to determine if any evolutionarily distinct lineages are present. In Chapter 2, I also make inferences about the evolutionary history of the eastern massasauga rattlesnake, possible factors influencing the subspecies' contemporary population structure, and make management recommendations based on the genetic results.

On a distinctly different note, Chapter 3 examines two control-region-like sequences in the mitochondria of eastern massasauga rattlesnakes, which appear to have undergone concerted evolution possibly due to gene conversion. Sequence similarities between regions within individuals are investigated, and a minimum rate of gene

conversion is estimated. Potential mechanisms of sequence homogenization are also discussed.

LITERATURE CITED

Avise, J.C. 1989. A role for molecular genetics in the recognition and conservation of endangered species. *Trends in Ecology and Evolution* 4: 279-281.

Avise, J.C. 1990. Molecular population structure and the biogeographic history of a region fauna: a case history with lessons for conservation biology. *Oikos* 63: 62-76.

Avise, J. 1994. Mitochondrial DNA polymorphism and a connection between genetics and demography of relevance to conservation. *Conservation Bi*ology 9 686-690.

Avise, J.C., D. Walker, and G.C. Johns. 1998. Speciation durations and Pleistocene effects on vertebrate phylogeography. *Proceedings of the Royal Society of London Series B* 265: 1707-1712.

Beltz, E. 1992. Distribution and status of the eastern massasauga rattlesnake, *sistrurus catenatus catenatus*, in the United States and Canada. In: *The International Symposium and Workshop on the Conservation of the Eastern Massasauga Rattlesnake Sistrurus catenatus catenatus* (eds Johnson, B., and V. Menzies). Metro Toronto Zoo, Toronto.

Bernatchez, L., and J.J. Dodson. 1991. Phylogeographic structure in mitochondrial DNA of the lake whitefish (*Coregonus clupeaformis*) and its relation to Pleistocene glaciations. *Evolution* 454: 1016-1035.

Bohanak, A.J. 1999. Dispersal, gene flow, and population structure. *The Quarterly Review of Biology* 74: 21-45.

Brown, W.L., M. George, Jr., and A.C. Wilson. 1979. Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America* 76: 1976-1971.

Burbrink, F.T., R. Lawson, and J.B. Slowinski. 2000. Mitochondrial DNA phylogeography of the polytypic north American rat snake (*Elaphe obsoleta*): a critique of the subspecies concept. *Evolution* 54: 2107-2118.

Bushar, L.M., H.K. Reinert, and L. Gelbert. 1998. Genetic variation and gene flow within and between location populations of the timber rattlesnake, *Crotalus horridus*. *Copeia* 54: 411-422.

Congdon, B.C., J.F. Piatt, K. Martin, and V.L. Friesen. 2000. Mechanisms of population differentiation in marbled murrelets: historical versus contemporary processes. *Evolution* 54: 974-986.

Cook, F.R. 1992. After an ice age: Zoogeography of the massasauga within a Canadian herpetofaunal perspective. In: *The International Symposium and Workshop on the Conservation of the Eastern Massasauga Rattlesnake Sistrurus catenatus catenatus* (eds Johnson, B., and V. Menzies). Metro Toronto Zoo, Toronto.

Frankham, R., 1995. Conservation genetics. *Annual Review of Genetics* 29: 305-327.

Friesen, V.L., B.C. Congdon, H.E. Walsh, and T.P. Birt. 1997. Intron variation in marbled murrelets detected using analyses of single-stranded conformational polymorphisms. *Molecular Ecology* 6: 1047-1058.

Friesen, V.L., W. A. Montevecchi, A.J. Baker, R.T. Barrett, and W.S. Davidson. 1996. Population differentiation and evolution in the common guillemot *Uria aalge*. *Molecular Ecology* 5: 793-805

Gibbs, H.L., K.A. Prior, P.J. Weatherhead, and G. Johnson. 1997. Genetics structure of populations of the threatened eastern massasauga rattlesnake, *Sistrurus c. catenatus*: evidence from microsatellite DNA markers. *Molecular Ecology* 6: 1123-1132

Gilbert, W. 1978. Why genes in pieces? *Nature.* 271: 501.

Haig, S. M. 1998. Molecular contribution to conservation. *Ecology* 79: 413-425.

Hay, R. 1992. Summary report of management concerns and guidelines for the eastern massasauga rattlesnake (*Sistrurus catenatus catenatus*). In: *The International Symposium and Workshop on the Conservation of the Eastern Massasauga Rattlesnake Sistrurus catenatus catenatus* (eds Johnson, B., and V. Menzies). Metro Toronto Zoo, Toronto.

Hewitt, G.M. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* 58: 247-276.

Holder, K., R. Montgomerie, and V.L. Friesen. 1999. A test of the glacial refugium hypothesis using patterns of mitochondrial and nuclear DNA sequence variation in rock ptarmigan (*Lagopus mutus*). *Evolution* 53: 1936-1950.

Hutchinson, B., C. Duchesne, and H. Schryver. 1992. GIS analysis of habitat use by massasauga rattlesnakes in Bruce Peninsula National Park. In: *The International Symposium and Workshop on the Conservation of the Eastern Massasauga Rattlesnake Sistrurus catenatus catenatus* (eds Johnson, B., and V. Menzies). Metro Toronto Zoo, Toronto.

Karns, D.R., A. O'Bannon, H.K. Voris, and L.A. Weigt. 2000. Biogeographical implications of mitochondrial DNA variation in the bockadam snake (*Cerberus rynchops*, Serpentes: Homalopsinae) in southeast Asia. *Journal of Biogeography* 27: 391-402.

Lessa, E.P., and G. Applebaum. 1993. Screening techniques for detecting allelic variation in DNA sequences. *Molecular Ecology* 2:119-129.

Merila, J., M. Bjorklund, and A.J. Baker. 1997. Historical demography and present day population structure of the greenfinch, *Carduelis chloris*- an analysis the mtDNA control-region sequences. *Evolution* 51: 946-956.

Milligan, B.G., J. Leebens-Mack, and A.E. Strand. 1994. Conservation genetics: beyond the maintenance of marker diversity. *Molecular Ecology* 3: 423-435.

Moran, T. 1992. Current status of eastern massasauga rattlesnakes in Michigan. In: *The International Symposium and Workshop on the Conservation of the Eastern Massasauga Rattlesnake Sistrurus catenatus catenatus* (eds Johnson, B., and V. Menzies). Metro Toronto Zoo, Toronto.

Moritz, C. 1994a. Applications of mitochondrial DNA analysis in conservation: a critical review. *Molecular Ecology* 3: 401-411.

Moritz, C. 1994b. Defining 'evolutionarily significant units' for conservation. *Trends in Ecology and Evolution* 9: 373-375.

Nei, M. 1987. *Molecular Evolutionary Genetics.* Columbia University Press, New York.

Pielou, E.C. 1991. *After the Ice Age*. University of Chicago Press, Chicago

Palumbi, S.R., and C.S. Baker. 1994. Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. *Molecular Biology and Evolution* 11: 426-435.

Parent, C., and P.J. Weatherhead. 2000. Behavioral and life history responses of eastern massasauga rattlesnakes (*Sistrurus catenatus catenatus*) to human disturbance. *Oecologia* 125: 170-178.

Pook, C.E., W. Wuster, and R.S. Thorpe. 2000. Historical biogeography of the western rattlesnake (Sepentes: Viperidae: *Crotalus viridis*), inferred from mitochondrial DNA sequence information. *Molecular Phylogenetics and Evolution* 15: 269-282.

Prior, K.A., H.L. Gibbs, and P.J. Weatherhead. 1996. Population genetic structure in the black rate snake: implications for management. *Conservation Biology* 11: 1147-1158.

Prior, K.A., P.F. Weatherhead, B. Hutchinson, R. Black, B. Crins, B. Gray, B. Johnson, J. Middleton, P.Pratt, D. Sweiger, M. Villeneuve, and W. Weller. 2001. National recovery plan for the eastern massasauga. http://www.terra-plex.com/sin/pub_docs/plansum.doc

Quinn, T.W., G.F. Shields, and A.C. Wilson. 1991. Affinities of the Hawaiian goose based on two types of mitochondrial DNA data. *Auk* 108: 585-593.

Rodriguez-Robles, J.A., D.F. Denardo, and R.E. Staub. 1998. Phylogeography of the California mountain kingsnake *Lampropeltis zonata* (Colubridae). *Molecular Ecology* 8: 1923-1934.

Seigel, R.A. 1986. Ecology and conservation of an endangered rattlesnake, *Sistrurus catenatus*, in Missouri, USA. *Biological Conservation* 35: 333-346.

Seigel, R.A., C. A. Sheil, and J.S. Doody. 1998. Changes in a population of an endangered rattlesnake *Sistrurus catenatus* following a severe flood. *Biological Conservation* 83: 127-131.

Waples, R.S. 1995. Evolutionarily significant units and the conservation of biological diversity under the endangered species act. *American Fisheries Society Symposium* 17: 8-27.

Wenink, P.W., A.J. Baker, and M.G.J. Tilanus. 1993. Hypervariable control-region sequences reveal global population structuring in a long distance migrant shorebird, the Dunlin (*Calidris alpina*). *Proceedings of the Nation Academy of Sciences of the United States of America* 90: 94-98.

Wilson, A.C., R.L. Cann, S.M. Carr, M. George, U.B. Gyllensten, K.M Helm-Bychowski, R.C. Higuchi, S.R. Palumbi, E.M. Prager, R.D. Sage, and M. Stoneking. 1985. Mitochondrial DNA and two perspectives on evolutionary genetics. *Biological Journal of the Linnean Society* 26: 375-400.

Zamudio, K.R., and H.W. Greene. 1997. Phylogeography of the bushmaster (*Lachesis muta*: Viperidae): implications for neotropical biogeography, systematics, and conservation. *Biological Journal of the Linnean Society* 62: 421-442.

Chapter 2
Population and Evolutionary Genetics of the Eastern Massasauga Rattlesnake (*Sistrurus catenatus catenatus*) Assessed by Mitochondrial DNA and Nuclear Introns: Historical Biogeography and Conservation Implications

Keywords: eastern massasauga rattlesnake, mitochondrial control region, nuclear intron, sequence, conservation, historical demography, contemporary population structure

ABSTRACT

The eastern massasauga rattlesnake (*Sistrurus catenatus catenatus*) has recently experienced extreme reductions in population size throughout its range in North America. Results from previous investigation of genetic differentiation between populations using microsatellites revealed significant, and even fine-scale, genetic structuring among populations. In order to further examine the partitioning of genetic diversity among snake populations, variation at a 528 bp portion of mitochondrial DNA (mtDNA), and at two nuclear intron DNA loci, was examined for snakes from eleven populations throughout its range. Seven mtDNA haplotypes were identified, and the majority of these were found in one or two U.S. populations of snakes. Intron variation was low and phylogenetically uniformative, as each locus gave contrasting patterns of variation. Neither mitochondrial nor nuclear DNA showed any phylogenetic structuring. The broad distribution of haplotypes, low haploypte and nucleotide diversity, and a star-like phylogeny of mtDNA haplotypes suggest that eastern massasauga rattlesnakes have undergone a post-Pleistocene population expansion. From a conservation perspective, three geographical groupings of populations that showed significant genetic differentiation were identified and I suggest that these groups should be managed independently, in order to conserve within-species variation.

INTRODUCTION

Information provided by molecular markers can contribute to the management of taxa that are under conservation threat. As genetic variation is necessary for evolution to occur (Lande and Barrowclough 1987), population genetic differentiation can indicate the potential for local adaptation or future speciation (Bohanak 1999). Molecular genetic information can be used to appraise phylogenetic discontinuities within species or subspecies, which can provide a better foundation for the recognition and management of diversity (Avise 1989). Identifying the structure of populations can contribute to management efforts by identifying unique genetic components that should be conserved in order to maintain the evolutionary potential of a species or sub-species (Waples 1995). For example, Moritz (1994a,b) has proposed two units of conservation based on genetic data, the evolutionarily significant unit (ESU), and management unit (MU), that consolidate the fields of genetics and conservation management. Studies of a number of species, including the harbor seal *Phoca vitulina* (Stanley *et al.* 1996), carp (Lu *et al.* 1996), and beluga whales *Deliphinapterus leucas* (Brown Gladden *et al.* 1999) have demonstrated the utility of these criteria in defining units of conservation in species with high population subdivision.

Eastern massasauga rattlesnakes (*Sistrurus catenatus catenatus*) are small, venomous snakes of the family Viperidae, that populate both upland and wetland habitats in northeastern North America. In Canada, eastern massasaugas are found in four populations in Ontario, including two large, extant populations on the eastern shore of Georgian Bay and along the Bruce Peninsula, as well as smaller populations in Wainfleet

Marsh (southeastern Ontario), and Ojibway Prairie (Windsor) (Prior *et al.* 2001) (Figure 1). In the United States, eastern massasaugas are found in 10 states, including New York, Pennsylvania, Ohio, Michigan, Indiana, Illinois, Iowa, Wisconsin, Minnesota and Missouri.

Historically, the range of the eastern massasauga was widespread and continuous (e.g. Beltz 1986). However, recent reductions in ranges due to the disappearance of acceptable habitat has resulted in eastern massasauga populations becoming smaller and more disjunct (Moran 1992). Threats to eastern massasauga populations include habitat fragmentation and destruction and land modification (Parent and Weatherhead 2000, Seigel *et al.* 1998, Beltz 1992), vehicular traffic (Seigel 1986), and site disturbance (Parent and Weatherhead 2000). In 1990, the eastern massasauga rattlesnake was protected by Ontario Regulation 263/90 under the Game and Fish Act, and in April, 1991, it was declared a threatened species in Canada by the Committee on the Status of Endangered Wildlife in Canada (COSWEIC) (Beltz 1992). In the United States, the eastern massasauga is a candidate for protection under the U.S. Endangered Species Act (http://endangered.fws.gov/wildlife.html).

A recent genetics study (Gibbs *et al.* 1997), in which six microsatellite loci were used to examine levels of differentiation amongst three Ontario and two U.S. populations of eastern massasauga rattlesnakes indicated that geographically separated populations of massasaugas were genetically distinct. Each population was found to contain a large portion of the overall variation found in the subspecies (Gibbs *et al.* 1997).

One important conservation issue not addressed by the previous study is whether any significant phylogenetic relationships, or distinct evolutionary lineages, occur amongst populations of this subspecies (Gibbs *et al.* 1997). This information can help in the prioritization of populations for conservation (Moritz 1994a,b). This study examines the phylogenetic relationships among Canadian and U.S. populations of snakes with the use of mtDNA and nuclear intron sequence. We then use this data to make inferences about the evolutionary history, female-mediated gene flow, and current geographical distribution of mtDNA haploytpes for the eastern massasauga rattlesnake. Genetic data is also used to delimit units of conservation for eastern massasaugas using current operational definitions.

MATERIALS AND METHODS

DNA Sampling, Extraction and Amplification

Blood, shed skin, or tissue samples were obtained from 314 eastern massasauga rattlesnakes from 13 different sites in Ontario and the United States. Samples were preserved in Queen's Lysis Buffer for transport and storage (blood), or dried (skin and tissue) and once received were kept at -20°C for long-term storage. To extract DNA from blood samples, blood was digested overnight at 65°C in a solution of 1mg/ml proteinase K (ICN Biomedicals, Aurora, Ohio), 100mM Tris Cl (pH 8.0), 10mM EDTA (pH 8.0), 100mM NaCl, and 1% SDS. DNA was isolated from the digested samples by extracting them twice with equal volumes of Tris-saturated phenol (pH 8.0) and once with 24:1 chloroform:isoamyl alcohol. Extracted DNA was then diluted 1:10 with double distilled water prior to amplification. Skin and tissue samples were minced using

a sterile razor blade, and extracted using a Dneasy Tissue Kit™ (Qiagen). Enriched mitochondrial DNA was also extracted from frozen tissue samples with an mtDNA Extractor CT Kit (Wako).

DNA was amplified using the polymerase chain reaction (PCR) (Mullis and Faloona 1987) in 30µl volumes containing 75mM Tris-HCl, 20mM $(NH_4)_2SO_4$, and 0.01% Tween 20, 0.33mM each of the four dNTPs, 15 pmol each of a forward and reverse primer, 8% DMSO, and 0.6 units *Thermus aquaticus* (Taq) polymerase (MBI Fermentas) in either a GeneAmp ™ 9700 (Applied Biosystems), or PTC ™ 900 thermal cycler (MJ Research) with Hot Bonnet, or a Perkin Elmer 480 thermal cycler, with an oil overlay. Amplification followed standard temperature profiles (denaturation at 94°C for 45s, annealing for 45s, extension at 72°C for 45s), with primer-specific annealing temperatures (see Table 1). The entire cytochrome b-control region area was amplified by using slightly different reaction conditions. These "long" PCRs were performed in 30µl volumes containing 75mM Tris-HCl, 20mM $(NH_4)_2SO_4$, and 0.01% Tween 20, 3% BSA, 5% DMSO, 0.58mM each of the four dNTPs, 25pmol of forward and reverse primers, and 1 unit *Thermus aquaticus* (Taq) polymerase (MBI Fermentas) in a Perkin Elmer 2400 thermal cycler. Long PCRs were amplified with an annealing temperature of 48°C.

Design of mitochondrial primers for eastern massasauga rattlesnakes

In order to find the most variable mitochondrial region to sequence for population-level analyses, I first surveyed levels of variation within different mitochondrial regions

for 10 snakes chosen from populations across the range of the sub-species (Figure 2). The primers 12SAL and 12SBH (Lorie Collins, unpubl.) (Table 1) were used to amplify a portion of the 12sRNA gene. The primers ND4L and LEUH (Arevalo *et al.* 1994) (Table 1) were used to amplify a portion of the ND4 gene, as well as $tRNA^{His}$ and $tRNA^{Ser}$. In *Dinodon semicaritinatus*, the entire control region, as well as $tRNA^{Leu}$ and $tRNA^{Pro}$ have been duplicated, and the sequence inserted within the isoleucine-glutamine-asparagine (IQM) tRNA gene cluster (Kumazawa *et al.* 1996, 1998). The original control region, designated control region 1, is located between cytochrome *b* and $tRNA^{Phe}$, and the duplicate, control region 2, is located between $tRNA^{Pro}$ and $tRNA^{Leu}$, within the IQM. In order to ensure that only control region 1 was amplified, PCR primers specific for amplifying cytochrome *b*, $tRNA^{Thr}$, and the mitochondrial control regions of the eastern massasauga rattlesnake, and the closely related desert massasauga rattlesnake (*Sistrurus catenatus edwarsii*) were designed as follows. Firstly, an approximately 2.5 kb fragment consisting of cytochrome *b*, $tRNA^{Thr}$, and the mitochondrial control region was amplified using a light-stranded primer situated in $tRNA^{Gln}$ near the 3' end of cytochrome *b* (L14919, R. Lawson) and a heavy-stranded primer situated in $tRNA^{Phe}$ near the 3' end of the mitochondrial control region (Phe1H, R. Lawson) (Table 1). PCR products were subjected to electrophoresis in 1% agarose gels, gel purified, and sequenced using the chain termination method (Sanger *et al.* 1977) with ThermoSequenase^{TM} radiolabelled kits (Amersham), according to the manufacturer's suggested protocol. The sequencing reaction products were electrophoretically separated in 5% polyacrylamide gels, and visualized by autoradiography. Sequences were manually aligned using the program

ESEE (Cabot and Beckenbach 1989), and used to develop a series of eastern massasauga rattlesnake-specific primers (Table 1). These nested primers were used in combination with the previously mentioned primers to amplify overlapping segments of varying lengths, comprising cytochrome *b*, $tRNA^{Thr}$, a non-coding spacer region and the control region in their entirety. These overlapping sequences were used to develop the eastern massasauga rattlesnake-specific primer CR1703L, which was used along with Phe1H to selectively amplify a 528 bp fragment of Domain II and Domain III of the mitochondrial control region (509 bp) and $tRNA^{Phe}$ (19 bp) for 64 individuals (Figure 1). The entire control region 2 was also amplified and sequenced for 10 individuals, using the primer set 4221L, located in $tRNA^{Ile}$ and 4419H, located in $tRNA^{Met}$ (Macey *et al.* 1999).

SSCP analysis and sequencing of nuclear introns

Variation at two nuclear intron loci was assessed by analysis of single-stranded conformational polymorphisms (SSCP), and direct sequencing of detected variants. The primers GapdL890 and GapdH950 (Friesen *et al.* 1997) (Table 1), which were designed to anneal to exons 11 and 12 of the glyceraldehyde-3-phosphate dehydrogenase gene, were used to amplify 234 bp of intron XI of this gene for 264 individuals. Primers CBAL and CBAH (Friesen *et al* 1999) (Table 1), were used to amplify 237 bp of intron 3 of the cytoplasmic β-actin gene for 273 individuals.

Products to be used in SSCP screening were amplified using standard PCR protocols, with 0.05 μL of 10 μ Ci/μL α-^{33}-P dATP (Amersham) incorporated into the reaction mixture. The amplified, radiolabelled DNA fragments were denatured by heating to 94°C for 3-5 minutes, snap-cooled in an ice bath, and subjected to

electrophoresis through non-denaturing 5% polyacrylamide gels at a constant temperature of 4°C. The resulting products were visualised through autoradiography, and individual genotypes at each locus were determined manually based on the obtained banding profile.

Once variants at each locus were identified as above, the exact nature of the variation was quantified by direct sequencing of haplotypes for homozygotes (as above). The sequence of each allele for heterozygous individuals was determined by "subtracting" one sequence from another. The introns of glyceraldehyde-3-phosphate dehydrogenase and cytoplasmic-β-actin were sequenced with the primers GapdL890 and GapdH950, and CBAL and CBAH, respectively.

Tests of Assumptions

The methods used to estimate population genetic structure and gene flow make a number of simplifying assumptions, including the assumption loci are unlinked, variation is neutral to selection, and population sizes are stable, and have not undergone recent bottlenecks or expansions. Arlequin (version 2.0; Schneider *et al.* 2000) was used to test for linkage disequilibrium among intron loci. Selection, nonrandom mating, population subdivision, small population size, mutation and null alleles can result in deviations of genotype frequencies from Hardy-Weinberg frequencies, and selection, mutation rate heterogeneity, and population expansions and cause Tajima's D to be significantly different from zero (Tajima 1989, Aris-Brosou and Excoffier 1996). Genotype frequencies at each intron locus were tested for deviations from Hardy-Weinberg equilibrium for individual populations, and for the total sample with Arlequin (version 2.0; Schneider *et al.* 2000). Neutrality of variation was tested in both the control region

and the CBA and GAPD introns using Tajima's *D* statistic (Tajima 1989) with Arlequin (version 2.0; Schneider *et al.* 2000).

Phylogenetic analysis

To assess phylogenetic relationships among haplotypes, a maximum-likelihood (ML) analysis was performed using TREE-PUZZLE (version 5.0; Strimmer and von Haeseler 1996) using a transition:transversion ratio empirically estimated from the data set, with the desert massasauga rattlesnake designated as the outgroup taxon. A minimum-spanning network was constructed using the number of pairwise substitutions as the phenetic measure using the statistical package Arlequin (version 2.0; Schneider *et al.* 2000).

Isolation by distance

Isolation by distance (Wright 1943) was tested using the ISOLDE subprogram of the program GENEPOP (web-based version; Raymond and Rousset 1996) which uses a rank correlation coefficient to test for positive correlations between geographic distance and genetic distance. Here, ϕ_{st} values (defined below) from pairwise comparisons of sampling sites were used, and approximate geographic distances between sites were calculated using How Far Is It? (http://www.howfarisit?.com). Populations from Northern Ontario were grouped due to their geographic proximity, and apparent panmixia (see below).

Analyses of population genetic structure

A hierarchical analysis of molecular variation (AMOVA; Excoffier *et al.* 1992) was used to determine the extent of geographic division of mitochondrial haplotypes and nuclear introns. For mtDNA, sequences of individual haplotypes were used, while for introns, allele sequences were used as input to the program. The program Arlequin (version 2.0; Schneider *et al.* 2000) was then used to calculate ϕ statistics, which are analogs of F-statistics (Wright 1951). ϕ_{st} is the correlation of random genotypes within a population relative to that from the whole species, ϕ_{ct} is the correlation of random genotypes within a group of populations relative to that drawn from the entire species, which measures proportion of genetic variations among groupings of populations, and ϕ_{sc} is the correlation of random genotypes within populations relative to that within a regional grouping of populations, which measures proportion of variation among populations within a region. The grouping that maximized the among-group variance component was considered the group that best described the subdivision among the populations analyzed (Stanley *et al.* 1996). Estimates were derived using Kimura's 2-parameter distance (Kimura 1980), and significance levels were corrected using sequential Bonferroni corrections. PAUP* (version 4.0; Swofford *et al.* 1998) was unable to estimate a gamma value for the mitochondrial data, so a rate parameter (α) of 0.05 for the gamma correction was used. Values of 0 and 0.1 were also tested, but gave results very similar to a rate parameter of 0.05. The same gamma correction was used for the intron data. Statistical significant of all estimates of ϕ-statistics were tested by randomization using 1000 permutations. Estimates were also derived using simple p-

values to check for concordance with values estimated using Kimura's 2-parameter distance, and were found to be analogous.

Descriptive statistics for mtDNA

In order to determine whether eastern massasauga rattlesnakes were genetically variable or genetically impoverished within populations and regions, nucleotide diversity was estimated from control region sequence using Nei and Tajima's (1981) index defined as:

$$\pi = (n/[n\text{-}1])\Sigma x_i x_j \pi_{ij} \quad (1)$$

where x_i is the frequency of the *i*th haplotype, and π_{ij} is the percent sequence divergence between the *i*th and *j*th haplotypes. Haplotypic diversity was also calculated (H_s; Nei 1987):

$$H_s = (n/[n\text{-}1])(1\text{-}x_i^2) \quad (2)$$

where x_i is the frequency of the *i*th haplotype in a sample of *n* individuals. Genetic divergence among populations was estimated using Nei's genetic distance, δ, for the control region:

$$\delta = \pi_{xy} - 0.5(\pi_x + \pi_y) \quad (3)$$

where π_{xy} is the mean pairwise divergence between individual snakes of groups *x* and *y*, and π_x and π_y are the mean pairwise divergences among individuals within groups *x* and *y* (Nei 1987). Long-term female effective population size was estimated from the equation

$$N_f = \pi/(rg) \quad (4)$$

(Wilson *et al.* 1985) where r is the evolutionary rate (substitutions/site/yr) and g is the mean generation time (generations/yr). Gene flow among populations was estimated from ϕ_{st} using the equation:

$$N_f m = 0.5[(1/\phi_{st}) - 1] \quad (5)$$

(Slatkin 1993). This equation assumes that populations are in equilibrium between mutation, migration, and gene flow. If populations are in genetic equilibrium, population structure is a reflection of current levels of gene flow and drift, and if populations are not in equilibrium, estimates likely reflect historic conditions. Populations should be in genetic equilibrium if time since divergence (t), calculated as

$$t = \delta/r; \quad (6)$$

exceeds $4N_f$ generations (Avise *et al.* 1984), or when $\delta > 4\pi/g$ (Kidd and Friesen 1998).

As estimates of gene flow from ϕ_{st} assume symmetrical migration rates between populations, which may not be true, gene flow was also estimated using the program Migrate (version 1.1; Beerli and Felsenstein 1998). This program estimates bidirectional gene flow using a maximum-likelihood approach (Rannala and Hartigan 1996) based on coalescent theory (Kingman 1982), which infers measures of gene flow inferred from the phylogeny of sampled individuals looking backwards in time. Bidirectional gene flow between groupings of populations using both mtDNA and intron sequence data, was derived using the coalescent and maximum-likelihood estimators from Migrate, using search strategies of 20 short chains with 500 sampled genealogies, and 2 long chains with 5000 genealogies sampled (mtDNA), and 30 short chains with 500 sampled genealogies,

and 4 long chains with 5000 genealogies samples (introns). An initial theta (θ) value, calculated as

$$\theta = N_f\mu \quad (7)$$

was estimated from F_{st}, and Migrate was re-run using the new generated θ value until convergent migration rates were obtained.

Comparison of variation from intron and microsatellite data

In order to compare patterns of variation obtained from nuclear intron and microsatellite data from the same individuals, an AMOVA using intron allele frequencies was run on a subset of 208 individuals that were also analyzed at six microsatellite loci (Gibbs *et al.*, 1997).

RESULTS

Characterization of surveyed mitochondrial variation

Levels of variation were assessed at the 12sRNA gene (278 bp), the ND4 gene and $tRNA^{His}$ and $tRNA^{Ser}$ (654 bp), cytochrome *b* (1117 bp), $tRNA^{Thr}$ (65 bp), non-coding region (51 bp) and the mitochondrial control region (1023 bp). No variation was found at the ND4 gene, $tRNA^{His}$, $tRNA^{Ser}$, or 12sRNA gene among the 10 individuals surveyed (see Appendices I and II, Figure 3). Among the 10 individuals surveyed only three variable positions were identified in the cytochrome *b* gene. The most variable mitochondrial region was identified as an approximately 500 bp section encompassing part of domain II and III of the control region, and a portion of $tRNA^{Phe}$. Sixty-four individuals representative of the populations across the range of this subspecies (Figure

1) were sequenced at this region. We identified 7 variable positions in 528 bp of control region and $tRNA^{Phe}$ which define seven distinct haplotypes among 64 eastern massasauga rattlesnakes (Table 2) (see Appendix III). The number of substitutions between haplotypes varied from 1 to 4, corresponding to Kimura 2-parameter distances of 0.19% to 0.76%, respectively.

The most common haplotype, A, was found in nine of the twelve populations sampled (Table 3). Four of the populations, Bruce Peninsula National Park (BPNP), Georgian Bay Islands National Park (GBINP), New York, and Wainfleet were monomorphic for this haplotype. The other six haplotypes were restricted to between one and three populations: B was found in Illinois, Wisconsin and Indiana, C in Ojibway and Ohio, D in Michigan, E in Illinois and Indiana, and F in Pennsylvania. One private haplotype, G, was found in only one individual from Killbear Provincial Park (Figure 5). The overall haplotype diversity was 0.588 (Table 4). Indiana showed the highest haplotype diversity, while BPNP, GBINP, New York, Wisconsin and Wainfleet had the lowest, as each of these populations was monomorphic for one haplotype. Nucleotide diversities ranged from 0.000758 to 0.003788 (Table 4).

Evidence for mitochondrial origin of control region sequences

In conducting surveys of mtDNA variation, a concern is that the presumed mtDNA sequence may have a nuclear DNA copy (Sorenson and Quinn 1998). To confirm that the 528 bp control region fragment sequenced here was of mtDNA origins, I did the following: first, sequence autoradiographs were scrutinized for evidence of multiple bands indicating more than one sequence; second, sequences of protein-coding genes

(ND4 and cytochrome *b*) contained in the fragment from which the control region sequences were obtained were translated to confirm that the reading frames were free of insertions or deletions that were not a multiple of three; third, the tRNA genes were mapped by hand ($tRNA^{Thr}$, $tRNA^{His}$, $tRNA^{Ser}$) onto secondary structure models and checked for any substitutions that would be incompatible with the predicted structures; and fourth sequence was compared to that obtained from an enriched mtDNA sample, and found to be identical. The control region had a base composition (A=25%, G=15%, C=24%, and T=36% similar to that reported for other species of snakes (Kumazawa *et al.* 1996). A number of conserved structural features were also present in the control region (C-rich sequence, hairpins 1, 2, and 3, repetitive sequences 1 and 2, conserved sequence blocks 1 and 3) that show a high degree of similarity to analogous regions in the western rattlesnake (*Crotalus viridis*) (Kumazawa *et al.* 1996) (Figure 3).

Tests of Assumptions

No significant cases of linkage disequilibrium were found within populations. There was found to be an excess of heterozygotes ($P<0.0001$) in both the CBA and GAPD introns for the total sample, suggesting that allele frequencies are not in Hardy-Weinberg equilibrium. Tajima's *D* was not significant for either the introns, or the control region sequence.

Phylogeographic reconstruction for mtDNA

The minimum-spanning network was star-like, with no apparent clustering of haplotypes by geographic location (Figure 4). The central haplotype from which the

majority of the haplotypes radiated is A, suggesting it is an extant ancestral sequence from which other haplotypes have recently arisen via mutations. The maximum-likelihood tree resulted in an unresolved polytomy, and is not shown.

Population structure and gene flow from mtDNA

Overall, genetic differentiation among eastern massasauga populations was substantial (ϕ_{st} = 0.70, P<0.0001). Following sequential Bonferroni corrections, pairwise values of ϕ_{st} between a number of populations were also found to be significant (Table 5). The highest ϕ_{ct} value (0.59, P<0.01) was obtained when populations were divided into three groups, the first consisting of GBINP, BPNP, Killbear, Wainfleet, New York, and Pennsylvania, the second consisting of Michigan, Ojibway and Ohio, and the third consisting of Wisconsin, Illinois and Indiana. This high ϕ_{ct} value suggests that this is the most probable geographic subdivision (Stanley *et al.* 1996). Several other geographic grouping of populations gave high ϕ_{ct} values (see Appendix IV), but for the purposes of analyses, I chose to use the grouping with the best value. Estimates of sequence divergence between groups of populations showed that differentiation between each population group was significant (Table 6).

To raise sample sizes for estimates of gene flow, I estimated values of N_m using Migrate for pairwise comparisons between the groups of populations discussed above. Estimates of gene flow using Migrate (Table 7) suggest that migration rates are low and asymmetrical between these population groups. However, high gene flow was noted from group one (BPNP, GBINP, KPP, Wainfleet, New York, Pennsylvania) to group 2

(Michigan, Ojibway, Ohio) and group two to group three (Wisconsin, Illinois, Indiana), whereas gene flow estimates in the other direction were all zero. This suggests that gene flow may be asymmetrical and occurs, or historically occurred, from East to West (Figure 6).

Assuming a generation time of 4 years (K. Prior, pers. comm.), and divergence rates of 1.32% (estimated for reptilian mtDNA; Zamudio and Greene, 1997) and 20% (estimated for Domain I of geese; Quinn 1992), eastern massasauga rattlesnakes from Killbear Provincial Park showed the lowest N_e while those from Illinois had the highest (Table 4). These estimates are very imprecise, as the actual N_e of these populations is most likely on an order of magnitude lower than the estimated values (e.g. Seigel *et al.* 1998), and the largest contemporary populations of eastern massasaugas are found in Northern Ontario. When populations were grouped, the group consisting of populations from Illinois and Wisconsin showed the highest N_e, while the grouping of populations from Northern Ontario, New York and Pennsylvania showed the lowest N_e. Correspondingly, the highest nucleotide diversities (π) were found in the small U.S. populations that were studied. Estimates of effective population size calculated from Migrate are comparable in size to those estimated from π, with the exception that the Wisconsin-Illinois group was found to have the smallest N_e (Table 8).

Most pairwise estimates of δ were less than $4\pi/g$, indicating that some populations may not have reach equilibrium between mutation, migration and genetic drift (Table 4). The star-like phylogeny observed for the minimum-spanning network also suggests recent population expansion.

Isolation by distance

Levels of genetic divergence in mtDNA between populations estimated by ϕ_{st} were found to be significantly correlated with distance between sampling sites ($P<0.05$), suggesting an isolation by distance effect, which could help explain the pattern of differentiation found among eastern massasaugas. In species in which individuals are grouped into discreet populations (an "island" model), the level of migration between these populations is partially a function of the distance between them (Wright 1943). In this case some of the genetic structure in a species can be predicted by demographic structure (Wright, 1943, Slatkin 1993).

Nuclear intron variation

For the CBA intron, we identified five variable sites in 237 bp of sequence, which defined three alleles among 273 individuals (Table 9) (Appendix V). The most common allele, B, was found in all populations surveyed. Allele A was found in only three populations, and allele C in six populations (Table 10). One allele, A, also had a 5 bp insert. For the GAPD intron, three variable sites were identified in 234 bp of sequence, defining two alleles among 264 individuals (Table 11) (Appendix VI). Allele A was found in all populations, and allele B all but three populations (Table 12).

Population structure and gene flow from introns

Separately each intron locus gives similar levels of among-population variation (a ϕ_{st} of 0.188 for CBA; ϕ_{st} of 0.184 for GAPD). However, when intron data is pooled, the among-population component of variation (ϕ_{st} of 0.0006), is much lower. This suggests

that each intron is detecting different patterns of differentiation (Tables 13 and 14), and pooling data essentially homogenizes the patterns of variation for each intron (Table 15). Introns do not show a significant among-group component of variation when considered separately, or when pooled (Table 16).

Values of N_m were estimated for introns using Migrate for pairwise comparisons between the same groups of populations discussed above. Estimates suggest that migration rates are high between geographically adjacent groupings of populations, with the exception of gene flow from group three to group two (Table 17). Gene flow appears to be somewhat asymmetrical, occurring from east to west, but also from west to east, in the case of groups one and two.

Comparison of variation from intron and microsatellite data

The ϕ_{st} value obtained from intron data from the same individuals that were analyzed by Gibbs *et al.* (1997) was 0.047 ($P<0.05$), as compared with an F_{st} of 0.164 found for microsatellites.

DISCUSSION

There are three significant results of this work. Eastern massasauga populations show unusually low levels of variation in mtDNA. The patterns of differentiation that are present suggest that recent history has strongly effected patterns of variation. Finally, we found substantial differences in the patterns of differentiation revealed by different markers. These results, as well as their conservation implications, are discussed below.

Low levels of variation in mtDNA

While conclusive calculations of normal levels of genetic variation in snakes have been difficult to estimate due to a lack of concordance between markers and methods used, as well as differences in population history, levels of intraspecific variation found in eastern massasaugas appear to be unusually low. Other studies of snakes invariably show higher levels of variation. For example, in the black rat snake (*Elaphe obsoleta*), Burbrink *et al.* (2000) found 26 distinct haploytpes among 68 individuals from 1034 bp of control region sequence. Zamudio and Greene (1997) found 9 unique haplotypes in 528 bp of ND4 and cytochrome *b* sequence among sixteen bushmasters (*Lachesis muta*). Pook *et al.* (2000) found 37 unique haploytpes in 1658 bp of cytochrome *b* and ND4 sequence among 68 western rattlesnake (*Crotalus viridis*). Divergence between haplotypes also seems to be small in eastern massasauga rattlesnakes. For example, uncorrected sequence divergence ranged between 0.2 and 6.1% among haplotypes (*C. viridis*) (Pook *et al.* 2000). Genetic distances estimated from mtDNA for twenty-six bockadam snakes (*Cerberus rynchops*) ranged from 0.6 to 8.6% (Karns *et al.* 2000). These values compare to Kimura 2-parameter distances of 0.19 to 0.76% obtained for eastern massasauga rattlesnake mtDNA haploytpes.

Low levels of variation found by mtDNA could be due to a reduced rate of mtDNA evolution in the eastern massasauga rattlesnake (Zamudio and Greene 1997), although recent work has suggested that snakes as group have undergone an accelerated rate of mtDNA evolution when compared to other taxa, potentially due to instability associated with mtDNA gene arrangements (Kumazawa *et al.* 1998). It has been suggested that this

increased rate of evolution could be due to the presence of two control region-like sequences in snake mitochondria. Kumazawa *et al.* (1996) demonstrated the presence of two, evolutionarily long-lived, control regions in a number of snake species that appear to have evolved in a concerted manner. Frequent gene conversion was suggested as a mechanism by which sequence in these two regions is homogenized. A similar pattern of concerted evolution has been demonstrated in eastern massasaugas (Pearce and Gibbs, unpubl.). This homogenization may actually have an effect on the evolution of control region sequence opposite to that suggested by Kumazawa et al. (1996) in that it could decrease the rate of control region evolution. Further intraspecific studies of mitochondrial variation in other snakes with two control regions would potentially help elucidate the effects of concerted control region evolution. Another intriguing explanation for the few mtDNA haplotypes found in this study could be due to sampling. Bushar *et al.* (1998) suggested that the behaviour of gravid females and neonates may contribute to reduced genetic variation that was found to be present in hibernacula, in that they do not move far, and neonates find hibernacula by tracking their mother or other adult snakes. While eastern massasaugas do not overwinter in group hibernacula, these snakes are often sampled from the same sites, including known hibernation sites, year after year. While no studies have investigated the relatedness of individual snakes within any geographic area, it could be that these snakes are maternally related, which would result in reduced mtDNA variation within population. However, Gibbs *et al.* (1997) did find typical levels of variation at microsatellite loci, which would argue against sampling as an explanation for reduced mitochondrial variation.

MtDNA population structure and geographic patterns of variation

Results from phylogenetic analysis found no evidence for structure among populations of eastern massasauga rattlesnakes from mtDNA. Levels of gene flow estimated from Migrate show relatively high levels of gene flow from groupings of populations in the east to groupings in the west (Figure 6).

The presence of high gene flow between geographically distant populations as suggested by Migrate is debatable. Eastern massasaugas have a fairly small activity range, which were found by one study to average 0.25km^2 (Weatherhead and Prior 1992), which makes it highly unlikely that contemporary levels of gene flow are high. Migrate indicates that asymmetries in gene flow exists, and estimates of N_e suggest that effective populations sizes vary greatly among populations, which violates a number of assumptions made for the inference of N_m. Although coalescent analyses are thought to be less sensitive to small sample sizes than other approaches (Congdon *et al.* 2000), one of the groupings of population used in the Migrate analysis consisted of only 10 individuals, so results should be interpreted cautiously. Also, as the populations that were sampled in this study are in areas that were glaciated during the Pleistocene, it is unlikely that populations are in mutation-drift equilibrium. Thus estimates of demographic parameters obtained from mtDNA will likely reflect historical conditions rather than contemporary levels of gene flow and effective populations size.

Estimates of sequence divergence between populations were somewhat inconsistent with geography, as some geographically close populations showed higher levels of divergence than geographically distant populations (Table 9). The high mtDNA ϕ_{st}

values found are likely due to the presence of rare haplotypes in U.S. populations of snakes (Figure 5), and in turn, the rarity of these haplotypes could be due to the small sample sizes used in this study, which can lead to sampling error. It has been demonstrated that ϕ_{st} is fairly sensitive to small sample sizes (e.g. Congdon *et al.* 2000). In the future, samples from more individuals from across the range of the eastern massasauga, in particular from smaller U.S. populations of snakes should be obtained and analyzed in order to confirm and further clarify the patterns observed in this study.

Nuclear intron variation

Levels of gene flow estimated from nuclear intron sequence suggest high levels of gene flow from eastern to western populations, and also from centrally sampled populations to populations in the east. These estimates should also be interpreted with caution, as, similar to that for mtDNA, one group consisted of a relatively low number of individuals. Also, levels of variation found for both intron loci were also extremely low, bringing into question their utility in estimating population parameters such as gene flow.

Population differentiation is far less defined for the CBA and GAPD introns than for mtDNA, as only a few alleles were found for each intron locus, and these alleles were shared among most populations. Indeed, levels of variation found for the introns used in this study were lower than those found in other published studies. For example, Holder et al. (1999) identified 25 alleles described by 20 variable sites in the GAPD intron for rock ptarmigan (*Lagopus mutus*), and Congdon *et al.* (2000) identified 14 GAPD alleles in 120 marbled murrelets (*Brachyramphus marmoratus*). Potential reasons for this are discussed below.

Differences in estimates obtained from different molecular markers

In this study, great discrepancies was found in results obtained from mitochondrial and nuclear markers. ϕ_{st} and ϕ_{ct} estimates were much higher from mtDNA than introns, and further, when a comparison was made between ϕ_{st} values from introns and F_{st} values from microsatellites, introns showed lower levels of differentiation, with a ϕ_{st} of 0.047 for introns compared to an F_{st} of 0.164 for microsatellites (Gibbs et al. 1997). Each intron locus also gave different patterns of among-population variation, and different levels of gene flow. Values of gene flow using introns should be four times as high as estimates from mtDNA (Palumbi and Baker 1994), and because the effective population size of mtDNA is ¼ as large as for nuclear genes, more population structuring should be expected for mtDNA (Merila *et al.* 1997). Results found for eastern massasauga rattlesnakes do show this general trend, however there are distinct differences between patterns of gene flow and population structure suggested by mtDNA and introns.

A number of studies have found varying levels of concordance between genetic results obtained from different markers. Holder *et al.* (1999) found that Fst values derived from pairwise comparisons of rock ptarmigan populations for control region haplotypes and introns were weakly but significantly correlated, suggesting that patterns of population structure are not entirely gene specific, but reflect the historic and demographic processes influencing the history of the species. Similarly, Palumbi and Baker (1994) found that geographic patterns between ocean populations of the humpback whale *Megaptera novaengliae* found from the actin intron paralleled results found with

mtDNA. However, over a smaller spatial scale, mtDNA and nuclear DNA results showed contrasting patterns (Palumbi and Baker 1994). Brown Gladden *et al.* (1999) reported that analysis of beluga whale (*Delphinapterus leucas*) nDNA showed less heterogeneity among groups than analysis of mtDNA, with an nDNA ϕst of 0.072 compared with an mtDNA ϕst value of 0.409.

Alternative markers sometimes give different results for a variety of reasons. One cause of variation among markers can be due to sampling errors associated with the small fraction of the genome represented in any assay, and the possibility that the markers used were influenced by different evolutionary forces (e.g. selection vs. neutral drift) (Avise 1994), however, the results of the neutrality test (Tajima's *D*) gave no reason to reject the null hypothesis of neutrality for introns or mtDNA. The lack of population structure found with the CBA and GAPD introns could simply reflect the fact that they lack the resolving power of mtDNA or microsatellites to distinguish structure, due to, for example, low substitution rates. The fact that the intron loci used in this study are not in Hardy-Weinberg equilibrium suggests that demography could be an explanation for the lack of structure found with introns. Higher levels of structure suggested by microsatellites could be due to a higher mutation rate in the microsatellites studied, which could increase allele frequency divergence among populations (Allendorf and Seeb 2000). While it is difficult to pinpoint the exact cause of differences in results obtained from different markers, this study emphasizes that comparisons of results obtained from these markers can be difficult to interpret. Studying several loci, in particular a large

number of intron loci, may be necessary before any conclusions about population structure, history and conservation are reached.

Lack of phylogenetic structure and the inference of historical processes

Organisms with low mobility and strict ecological requirements, such as the eastern massasauga rattlesnake, may be useful for studies of historical biogeography, as they are often subdivided in genetically isolated populations (Froufe *et al.* 2000). While no evidence for phylogenetic structure was found for eastern massasaugas, several striking patterns indicative of historical population processes are obvious. A minimum-spanning network of mtDNA haplotypes (Figure 4) is characterized by a star-like cluster with the most common haplotypes, A, as the hub, and few mutational steps between mtDNA haplotypes. Haplotype A was found in the majority of populations, while the other six haplotypes were found in only one or two other populations, the majority from small, U.S. populations of eastern massasaugas. Northern Ontario populations of rattlesnakes, which have by far the largest population sizes, were found to have the lowest levels of within-population variation. Indeed, populations from BPNP, GBINP, Wainfleet Bog, and also New York, were monomorphic for the most common mtDNA haplotype (Figure 5).

As the most common, widespread haplotypes are often the oldest (Fry and Zink 1998), and results from Migrate suggest historical unidirectional gene flow from East to West, this pattern implies a range expansion from a single, eastern refugium. This is consistent with the fact that eastern massasaugas are cited as a species with a typical corridor distribution, in that there is evidence that they existed in a prairie peninsula

corridor thought to have been present in the Great Lakes basin during the Hypsithermal, a period of drier, warmer temperatures present 7 000 years ago, or 13 000 years after the Wisconsin retreat (Pielou 1991). There is no evidence that eastern massasaugas expanded from multiple refugia, such as present-day distribution of multiple divergent genetic lineages coinciding with refugia (e.g Holder *et al.* 1999), or high π values in certain populations due to dispersing haplotypes from different refugia meeting in zones of contact (e.g. Fry and Zink 1998).

Similar patterns of variation, including a star-like relationship among mtDNA haplotypes with the most common haplotype as a hub, low nucleotide divergence among haplotypes, and low haplotype and nucleotide diversity with populations have been found for a number of other species, including the lake whitefish *Coregonus clupeaformis* (Bernatchez and Dodson 1991), wood lemming *Myopus schisticolor* (Federov *et al.* 1996), harbour seal *Phoca vitulina* (Stanley *et al.* 1996), European greenfinch *Carduelis chloris* (Merila *et al.* 1997), European perch *Perca fluviatilis* (Nesbo *et al.* 1999), song sparrow *Melospiza melodia* (Fry and Zink 1998), and northern clingfish *Gobbiesox maeandricus* (Hickerson and Ross 2001) and attributed to range expansion from glacial refugia. Despite the fact that many expansions are hypothesized to have occurred from more Southern refugia, Bernatchez (1997) found the evidence for expansion of rainbow smelt (*Osmerus mordax*), from two distinct glacial refugia, one of which was north of the Atlantic coastal plains. Studies investigating effects of Pleistocene glaciations have found high levels of genetic diversity in populations derived from refugia, and the loss of genetic diversity in northern recolonized areas (Froufe *et al.* 2000). This supports the

prediction that a refugial population spreading from its leading edge will experience a series of bottlenecks that will reduce genetic diversity (Hewitt 1996). While the opposite was found in this study, in that populations furthest from the hypothesized refugium had the most within-population diversity, one explanation for the present pattern of diversity is that populations experienced divergence *in situ* following postglacial colonization by a limited number of founders with the most common haplotypes (e.g. Federov *et al.* 1996). An equally plausible hypothesis is that the direction of historical migration events is actually in the other direction, from west to east. If not enough time has passed for populations to reach reciprocal monophyly, then the pattern of geographical distribution of haplotypes could be due to ancestry, in that an individual with a haplotype ancestral to those found in the more western populations may have migrated into a more eastern population (see Milot *et al.* 2000). Since Migrate uses estimates the number of migration events in the history of a set of sequences necessary for their current geographic distribution to be consistent with their phylogeny (Slatkin and Maddison 1989), the presence of a haplotype in the west which is ancestral to those in the east could indicate migration from west to east. The failure to find such a haplotype in this study could be due to sampling error, in that small sample sizes were obtained for most of the more western U.S. populations.

A species with a restricted dispersal should exhibit a pattern of isolation by distances if it has had time to approach equilibrium, or if the species was in its present range for a substantial time (Slatkin 1993). As it is unlikely that eastern massasauga populations are in equilibrium (see above), the observation that genetic distances between

snake populations are correlated with geographic distances suggests that isolation between populations of this subspecies is the norm, and that populations have been separated for a long period of time.

Conservation Implications

While no evidence was found for the classification of ESUs or MUs (Moritz 1994a, b) among populations of eastern massasauga rattlesnakes, this study identified three divergent groupings of populations, one eastern, one central, and one western that showed high levels of differentiation (Figure 6). This division of populations reflects the presence of relatively unique mtDNA haplotypes found in some U.S. populations in the states of Pennsylvania, Michigan, and Illinois. A perhaps less significant unit for conservation may be defined as those characterized by unique haplotypes (Stanley *et al.* 1996), so by this definition, the three groupings of populations could be defined as units that should be managed independently.

Attempts to determine significantly different units for conservation purposes can lead to two types of statistical errors. A type I error assumes more differentiation than actually exists, and may result in splitting the populations too finely for management (Brown Gladden *et al.* 1999). A type II error is a failure to recognize demographically independent populations due to a lack of power in genetic analyses, which may lead to inappropriate pooling of populations into a single management unit. This can lead to extinction of populations due to lack of dispersal from neighbouring regions (Brown Gladden *et al.* 1999). The identification of ESUs and MUs can be susceptible to error due to the sampling of too few nucleotides, or nuclear loci, which can lead to failure to

recognize important genetic patterns (Moritz 1994a). While Gibbs et al. (1997) suggested that Northern Ontario, New York, and Ohio populations of eastern massasaugas should be managed as demographically independent units based on microsatellite data, intron and mtDNA data suggest that most of these populations are essentially panmictic. In combining results from this study and that of Gibbs *et al.* (1997), we suggest the maintenance of the smaller U.S. populations, including New York, Ohio, Illinois, Michigan, Wisconsin, and Ontario's Ojibway Prairie as separate management units, and all other Ontario populations as one management unit. This will ensure the preservation of the population-level variation that constitutes an important part of the subspecies-wide variation for the eastern massasauga rattlesnake.

Finally, one the concerns adversely affecting massasaugas is that of gene pool "contamination", or the introduction of specimens from one population to another which could change the genetic makeup of these populations (Hay 1992). While this study has shown that a number of eastern massasauga populations are essentially genetically identical for the markers examined, recent published work on the timber rattlesnake *Crotalus horridus*, has shown that these snakes show low survival rates when translocated distances of eight kilometers or more (Reinert and Rupert 1999). Low survival rates have also been suggested for translocated eastern massasauga rattlesnakes (C. Parent pers. comm.). This is an excellent illustration of how life history traits and demography can often take precedence over genetics in management practices.

LITERATURE CITED

Allendorf, F.W., and L.W. Seeb. 2000. Concordance of genetic divergence among sockeye salmon populations at allozyme, nuclear DNA, and mitochondrial DNA markers. *Evolution* 54: 640-651.

Arevalo, E., S.K. Davis, and J.W Sites Jr. 1994. Mitochondrial DNA sequences divergences and phylogenetic relationships among eight chromosome races of the *Sceloporus grammicus* complex (Phrynosomatidae) in central Mexico. *Systematic Biology* 43: 387-418.

Aris-Brosou, S., and L. Excoffier. 1996. The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. *Molecular Biology and Evolution* 13: 494-504.

Avise, J.C., J.E. Neigel, and J. Arnold. 1984. Demographic influences on mitochondrial DNA lineage survivorship in animal populations. *Journal of Molecular Evolution* 20: 99-105.

Avise, J.C. 1989. A role for molecular genetics in the recognition and conservation of endangered species. *Trends in Ecology and Evolution* 4: 279-281.

Avise, J. 1994. Mitochondrial DNA polymorphism and a connection between genetics and demography of relevance to conservation. *Conservation Biology* 9: 686-690.

Avise, J.C., D. Walker, and G.C. Johns. 1998. Speciation durations and Pleistocene effects on vertebrate phylogeography. *Proceedings of the Royal Society of London Series B* 265: 1707-1712.

Beerli, P and J. Felsenstein. 1989. Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* 152: 763-773.

Beltz, E. 1992. Distribution and status of the eastern massasauga rattlesnake, *sistrurus catenatus catenatus*, in the United States and Canada. In: *The International Symposium and Workshop on the Conservation of the Eastern Massasauga Rattlesnake Sistrurus catenatus catenatus* (eds Johnson, B., and V. Menzies). Metro Toronto Zoo, Toronto.

Bernatchez, L., and J.J. Dodson. 1991. Phylogeographic structure in mitochondrial DNA of the lake whitefish (*Coregonus clupeaformis*) and its relation to Pleistocene glaciations. *Evolution* 45: 1016-1035.

Bernatchez, L. 1997. Mitochondrial DNA analysis confirms the existence of two glacial races of rainbow smelt *Osmerus mordax* and their reproductive isolation in the St. Lawrence River estuary (Quebec, Canada). *Molecular Ecology* 6: 73-83.

Brown Gladden, J.G., M.M. Ferguson, M.K. Friesen, and J.W. Clayton. 1999. Population structure of north American beluga whales (*Delphinapterus leucas*) based on nuclear DNA microsatellite variation and contrasted with the populations structure revealed by mitochondrial DNA variation. *Molecular Ecology* 8: 347-363.

Burbrink, F.T., R. Lawson, and J.B. Slowinski. 2000. Mitochondrial DNA phylogeography of the polytypic North American rat snake (*Elaphe obsoleta*): a critique of the subspecies concept. *Evolution* 54: 2107-2118.

Bushar, L.M., H.K. Reinert, and L. Gelbert. 1998. Genetic variation and gene flow within and between location populations of the timber rattlesnake, *Crotalus horridus*. *Copeia* 54: 411-422.

Cabot, E. and A.T. Beckenbach. 1989. Simultaneous editing of multiple nucleic acid and protein sequences with Esee. *Computing and Applied Biosystems* 5: 233-234.

Congdon, B.C., J.F. Piatt, K. Martin, and V.L. Friesen. 2000. Mechanisms of population differentiation in marbled murrelets: historical versus contemporary processes. *Evolution* 54: 974-986.

Federov, V., M. Jaarola, and K. Fedga. 1996. Low mitochondrial DNA variation and recent colonization of Scandinavia by the wood lemming *Myopus schisticolor*. *Molecular Ecology* 5: 577-581.

Froufe, A.E., J.W. Arntzen, and N. Ferrand. 2000. Genetic subdivision, glacial refugia and postglacial recolonization in the golden-striped salamander, *Chioglossa lusitanica* (Amphibia:Urodela). *Molecular Ecology* 9:771-781.

Friesen, V.L., B.C. Congdon, H.E. Walsh, and T.P. Birt. 1997. Intron variation in marbled murrelets detected using analyses of single-stranded conformational polymorphisms. *Molecular Ecology* 6: 1047-1058.

Friesen, V.L., B.C. Congdon, M.G. Kidd and T.P. Birt. 1999. General PCR primers for the analysis of nuclear introns in vertebrates. *Molecular Ecology* 8: 2147-2149.

Fry, A.J., and R. M. Zink. 1998. Geographic analysis of nucleotide diversity and song sparrow (Aves: Emberizidae) population history. *Molecular Ecology* 7: 1303-1313.

Gibbs, H.L., K.A. Prior, P.J. Weatherhead, and G. Johnson. 1997. Genetic structure of populations of the threatened eastern massasauga rattlesnake, *Sistrurus c. catenatus*: evidence from microsatellite DNA markers. *Molecular Ecology* 6: 1123-1132.

Hawks, J., K. Hunley, S.H. Lee, and M. Wolpoff. 2000. Population bottlenecks and pleistocene human evolution. *Molecular Biology and Evolution* 17: 2-22.

Hay, R. 1992. Summary report of management concerns and guidelines for the eastern massasauga rattlesnake (*Sistrurus catenatus catenatus*). In: *The International Symposium and Workshop on the Conservation of the Eastern Massasauga Rattlesnake Sistrurus catenatus catenatus* (eds Johnson B, and V. Menzies) Metro Toronto Zoo, Toronto.

Hewitt, G.M. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* 58: 247-276.

Hickerson, M.J., and J.R.P. Ross. 2001. Post-glacial population history and genetic structure of the northern clingfish (*Gobbiesox maeandricus*) revealed from mtDNA analysis. *Marine Biology* 138: 407-419.

Holder, K., R. Montgomerie, and V.L. Friesen. 1999. A test of the glacial refugium hypothesis using patterns of mitochondrial and nuclear DNA sequence variation in rock ptarmigan (*Lagopus mutus*). *Evolution* 53: 1936-1950.

Karns, D.R., A. O'Bannon, H.K. Voris, and L.A. Weigt. 2000. Biogeographical implications of mitochondrial DNA variation in the bockadam snake (*Cerberus rynchops*, Serpentes: Homalopsinae) in southeast Asia. *Journal of Biogeography* 27: 391-402.

Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16: 111-120.

Kingman, J.F.C. 1982. The coalescent. *Stochastic Processes and their Applications* 13: 235-248.

Kumazawa, Y., H. Ota, M. Nishida, and T. Ozawa. 1996. Gene rearrangements in snake mitochondrial genomes: highly concerted evolution of control-region-like sequences duplicated and inserted into a tRNA gene cluster. *Molecular Biology and Evolution* 13: 1242-1254.

Kumazawa, Y., H. Ota, M. Nishida, and T. Ozawa. 1998. The complete nucleotide sequence of a snake (*Dinodon semicarinatus*) mitochondrial genome with two identical control regions. *Genetics* 150: 313-329.

Lande, R., and G.F. Barrowclough. 1987. Effective population size, genetic variation, and their use in population management. In: *Viable Populations for Conservation* (ed Soule, M.), pp87-124. Cambridge University Press, Cambridge, New York.

Lessa, E.P., and G. Applebaum. 1993. Screening techniques for detecting allelic variation in DNA sequences. *Molecular Ecology* 2:119-129.

Macey, J.R., J.A. Schulte II, A. Larson, B.S. Tuniyev, N. Orlov, and T.J. Papenfuss. 1999. Molecular phylogenetics, tRNA evolution, and historical biogeography in anguid lizards and related taxonomic families. *Molecular Phylogenetics and Evolution* 12: 250-272.

Merila, J., M. Bjorklund, and A.J. Baker. 1997. Historical demography and present day population structure of the greenfinsh, *Carduelis chloris* - an analysis of mtDNA control-region sequences. *Evolution* 51: 946-956.

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

Moran, T. 1992. Current status of eastern massasauga rattlesnakes in Michigan. In: *The International Symposium and Workshop on the Conservation of the Eastern Massasauga Rattlesnake Sistrurus catenatus catenatus* (eds Johnson, B., and V. Menzies). Metro Toronto Zoo, Toronto.

Moritz, C. 1994a. Applications of mitochondrial DNA analysis in conservation: a critical review. *Molecular Ecology* 3: 401-411.

Moritz, C. 1994b. Defining 'evolutionarily significant units' for conservation. *Trends in Ecology and Evolution* 9: 373-375.

Mullis, K., and F.A. Faloona. 1987. Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. *Methods in Enzymology* 155: 335-350.

Nei, M. 1987. *Molecular Evolutionary Genetics.* Columbia University Press, New York.

Nei, M., and F. Tajima. 1981. Maximum-likelihood estimation of the number of nucleotide substitutions from restriction sites data. *Genetics* 105: 207-217.

Nesbo, C.L., T. Fossheim, L.A. Vollestad, and K.S. Jakobsen. 1999. Genetic divergence and phylogeographic relationships among European perch (*Perca fluviatilis*) populations reflect glacial refugia and postglacial colonization. *Molecular Ecology* 8: 1387-1404.

Palumbi, S.R., and C.S. Baker. 1994. Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. *Molecular Biology and Evolution* 11: 426-435.

Parent, C., and P.J. Weatherhead. 2000. Behavioral and life history responses of eastern massasauga rattlesnakes (*Sistrurus catenatus catenatus*) to human disturbance. *Oecologia* 125: 170-178.

Pook, C.E., W. Wuster, and R.S. Thorpe. 2000. Historical biogeography of the western rattlesnake (Sepentes: Viperidae: *Crotalus viridis*), inferred from mitochondrial DNA sequence information. *Molecular Phylogenetics and Evolution* 15: 269-282.

Porter, A.H. 1998. Refugees from lost habitat and reorganization of genetic population structure. *Conservation Biology* 13: 850-859.

Prior, K.A., P.J. Weatherhead, B. Hutchinson, R. Black, B. Crins, B. Gray, B. Johnson, J. Middleton, P.Pratt, D. Sweiger, M. Villeneuve, and W. Weller. 2001. National recovery plan for the eastern massasauga. http://www.terra-plex.com/sin/pub_docs/plansum.doc

Quinn, T.W. 1992. The genetic legacy of mother goose – phylogeographic patterns of lesser snow goose *Chen caerulescens caerulescens* maternal lineages. *Molecular Ecology* 1: 105-117.

Rannala, B., and J.A. Hartigan. 1996. Estimating gene flow in island populations. *Genetical Research* 67: 147-158.

Raymond, M., and F. Rousset. http://wbiomed.curtin.edu.au/genepop/index.html.

Reinert, H.K., and R.R. Rupert, Jr. 1999. Impacts of translocation on behavior and survival of timber rattlesnakes, *Crotalus horridus*. *Journal of Herpetology* 33: 45-61.

Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, United States of America.* 74: 5463-5467.

Schneider, S., D. Roessli, and L. Excoffier. 2000. Arlequin: a software for population genetics data analysis Vers 2.00. Genetics and Biometry Lab, Department of Anthropology, University of Geneva.

Seigel, R.A. 1986. Ecology and conservation of an endangered rattlesnake, *Sistrurus catenatus*, in Missouri, USA. *Biological Conservation* 35: 333-346.

Seigel, R.A., C. A. Sheil, and J.S. Doody. 1998. Changes in a population of an endangered rattlesnake *Sistrurus catenatus* following a severe flood. *Biological Conservation* 83: 127-131.

Slatkin, M. 1993. Gene flow and population structure. In: *Ecological Genetics* (ed Real, L.A.), pp3-17. Princeton University Press, Princeton, USA.

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

Slatkin, M., and W.P. Maddison. 1989. A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics* 123: 603-613.

Sorenson, M.D., and T.W. Quinn. 1998. Numts: a challenge for avian systematics and population biology. The Auk 115: 214-221.

Stanley, H.F., S. Casey, J.M. Carnahan, S. Goodman, J. Harwood, and R.K. Wayne. 1996. Worldwide patterns of mitochondrial DNA differentiation in the harbor seal (*Phoca vitulina*). *Molecular Biology and Evolution* 13: 368-382.

Strimmer, K, and A. von Haeseler. 1997. Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. *Molecular Biology and Evolution* 13: 964-969.

Swofford, D.L. 1998. Phylogenetic analysis using parsimony and other methods, version 4.0. Sinauer Associates, Sunderland, Massachusetts.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585-595.

Waples, R.S. 1995. Evolutionarily significant units and the conservation of biological diversity under the endangered species act. *American Fisheries Society Symposium* 17: 8-27.

Weatherhead, P.J., and K.A. Prior. 1992. Preliminary observations of habitat use and movements of the eastern massasauga rattlesnake (*Sistrurus c. catenatus*). *Journal of Herpetology* 26: 447-452.

Wilson, A.C., R.L. Cann, S.M. Carr, M. George, U.B. Gyllensten, K.M Helm-Bychowski, R.C. Higuchi, S.R. Palumbi, E.M. Prager, R.D. Sage, and M. Stoneking. 1985. Mitochondrial DNA and two perspectives on evolutionary genetics. *Biological Journal of the Linnean Society* 26: 375-400.

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

Wright, S. 1951. The genetical structure of populations. *Annals of Eugenetics* 15: 323-354.

Zamudio, K.R., and H.W. Greene. 1997. Phylogeography of the bushmaster (*Lachesis muta*: Viperidae): implications for neotropical biogeography, systematics, and conservation. *Biological Journal of the Linnean Society* 62: 421-442.

Table 1. Sequences of primers used in this study. H and L refer to heavy and light strands, respectively. Primers were created in this study unless indicated in footnotes.

Primers	Sequence (5'→3')
L14919[a]	AACCACCGTTGTA/TAA/TTCAACT
Phe1H[a]	CCATCTTAGCATC/TTTCAGTGCTA
Cytb1L[b]	AGGCCTACTAGAAAACAAAATCTC
CSB2L[b]	GGCTAATAAGCGGGAATTCTACTA
CSB1L[b]	TCTCTTAATGCTTGTTATACATA
CSB2H[b]	GTAAAAAAATGTAGTTACAGCGGG
12SAL[b]	AAACTGGGATTAGATACCCCACTAT
12SBH[b]	AGAGGGTGACGGGCGGTGTGT
ND4L[c]	TGACTACCAAAAGCTCATGTAGAAGC
LEUH[c]	CATTACTTTTACTTGGATTTGCACCA
L4221[d]	GGTATGGGCCCGATAGCTT
H4419[d]	AGGGACTACCTTGATAGAGT
CBAH	TACATGGCTGGGGTGTTGAAGG
GAPDL[f]	ACCTTTAATGCGGGTGCTGGCATTGC
GAPDH[f]	CATCAAGTCCACAACACGGTTGCTGTA
Cytb15110L	TTTATATTCACATCGCACGAGG
Cytb15810L	CCCATACTTAGGAACAACCCTCAC
Cytb16110L	CTGAAAACTTCTCAAAAGCCAACC
Cytb16410L	GGCCAACTAGCCTCTATCCTATACT
Cr1500H	GGTTTGAAGTGGATAGCCATGG
Cr1703L	CGCACCCTTTATCCTGGTACATTA
CrIIH	AAGATCATTTCCCGGATACA
CBA550L	CCCCCAATAAAATACTAACTTTGGC

[a] R. Lawson, unpubl.
[b] L. Collins, unpubl.
[c] Arevalo *et al.* (1994)
[d] Macey *et al.* (1999)
[e] Friesen et al. (1999)
[f] Friesen et al. (1997)

Table 2. Variable sites in the control region of eastern massasauga Rattlesnakes. Dots denote identity with haploytpe A. See appendix for entire sequence.

	Variable Site Location
	112455
	2699501
Haplotype	7356897
HapA	CCTACAT
HapB	GT..T..
HapC	G.....
HapD	.....T.
HapE	G.A....
HapF	...T...
HapG	......G

Table 3. Distribution of mitochondrial haplotypes among populations of eastern massasauga rattlesnakes.

Population	Haplotype						
	A	B	C	D	E	F	G
BPNP	8	-	-	-	-	-	-
GBINP	5	-	-	-	-	-	-
PA	1	-	-	-	-	3	-
NY	9	-	-	-	-	-	-
OH	6	-	4	-	-	-	-
IL	-	3	-	-	1	-	-
MI	2	-	-	3	-	-	-
WI	-	2	-	-	-	-	-
WAIN	4	-	-	-	-	-	-
KPP	4	-	-	-	-	-	1
IND	-	2	-	-	2	-	-
OJIB	1	-	3	-	-	-	-

BPNP = Bruce Peninsula National Park
GBINP = Georgian Bay Islands National Park
PA = Pennsylvania
NY = New York
OH = Ohio
IL = Illinois
MI = Michigan
WAIN = Wainfleet Bog
KPP = Killbear Provincial Park
IN = Indiana
OJIB = Ojibway Prairie

Table 4. Number of private haplotypes (n_H), haplotype (H_s), and nucleotide (π) diversities for mtDNA sequence, and effective population sizes (N_f) in number of females estimated from π using age of first reproduction of four years, and mutation rates of 1.32%-20%/Myr.

	n_H	H_s	π	N_f
BPNP	0	0	0	
GBINP	0	0	0	
PENN	1	0.5+/-0.2652	0.0009+/-0.001	1125 –17045
NY	0	0	0	
OHIO	0	0.533+/-0.0947	0.001+/-0.001	1250 - 18939
ILL	0	0.536+/-0.1232	0.003+/-0.002	3750 - 56818
MICH	1	0.6+/-0.1753	0.001+/-0.001	1250 - 18939
WIS	0	0	0	
WAIN	0	0	0	
KPP	1	0.4+/-0.2373	0.0007+/-0.0009	875 – 13257
OJIB	0	0.5+/-0.2652	0.0009+/-0.001	1125 - 17045
ALL	3	0.588+/-0.0663	0.002+/-0.001	2500 - 37878

Table 5. Estimates of sequence divergence between populations (δ - above diagonal), and among-population component of genetic variation (ϕ_{st} – below diagonal) for mtDNA sequence.

	BPNP	*GBINP*	*PA*	*NY*	*OH*	*IL*	*MI*	*WI*	*WAIN*	*KPP*	*OJIB*
BPNP	--	0	0.00097	0	0.00026	0.0018	0.00058	0.0062	0	0	0.00097
GBINP	0	--	0.00097	0	0.00026	0.0037	0.00058	0.0062	0	0	0.00097
PA	0.76	0.73	--	0.00097	0.0012	0.0049	0.0016	0.0074	0.00097	0.00099	0.002
NY	0	0	0.8	--	0.00026	0.0037	0.00058	0.0062	0	0	0.00097
OH	0.27	0.25	0.55	0.31	--	0.0024	0.00087	0.0048	0.00026	0.00027	6.50568E-05
IL	0.67*	0.65*	0.64	0.71*	0.54*	--	0.0045	0.00067	0.0037	0.0038	0.0017
MI	0.57*	0.54*	0.6	0.62	0.45	0.63	--	0.007	0.00058	0.0006	0.0016
WI	1*	1	0.91	1	0.83	0	0.88	--	0.0062	0.0063	0.0041
WAIN	0	0	0.67	0	0.18	0.6	0.45	1	--	0	0.00097
KPP	0.073*	0.04	0.54	0.13	0.21	0.6*	0.38	0.91	0	--	0.00099
OJIB	0.76	0.74	0.67	0.8	0.055	0.36	0.6	0.84	0.67	0.053	--

* =P<0.05 (corrected for multiple comparisons)

Table 6. Estimates of sequence divergence between groups of populations (δ - above diagonal), and among-population component of genetic variation (ϕ_{ct} – below diagonal) for mtDNA sequence.

	group1	group2	group3
group1	--	0.0002753	0.0037974
group2	0.27066*	--	0.0026581
group3	0.81473*	0.59117*	--

group1 = KPP, BPNP, GBINP, Wainfleet, Pennsylvania, New York
group2 = Michigan, Ojibway, Ohio
group3 = Illinois and Wisconsin
* = $P<0.05$

Table 7. Mean gene flow (1% and 99% confidence values) between groupings of populations, and estimates of effective population size in number of females estimated from mtDNA sequence using Migrate. Values represent gene flow from groups on the horizontal to groups on the vertical.

	N_fm - group1, x	N_fm - group2, x	N_fm - group3, x	N_f
group1	--	0	0	4400-60606
group2	2.4284 (0.52161-9.49062)	--	0	4800-72727
group3	0 (0-0.312799)	0.4919 (0.0300556-0.694403)	--	3650-55303

Group1 = KPP, BPNP, GBINP, Wainfleet, Pennsylvania, New York
Group2 = Michigan, Ojibway, Ohio
Group3 = Illinois, Wisconsin

Table 8. Number of private haplotypes (n_H), haplotype (H_s), and nucleotide (π) diversities for groups of populations, and effective population size (N_f) in number of females estimated from π using an age of first reproduction of four years, and mutation rates of 1.32%-20%/My.

	n_H	H_s	π	N_f
group1	2	0.2134+/-0.0879	0.0004+/-0.0005	500 - 7575
group2	1	0.6491+/-0.0614	0.0015+/-0.001	1875 - 28409
group3	0	0.4667+/0.13181	0.0027+/-0.002	3375 - 51136

Group1 = KPP, BPNP, GBINP, Wainfleet, Pennsylvania, New York
Group2 = Michigan, Ojibway, Ohio
Group3 = Illinois, Wisconsin

Table 9. Variable sites in the CBA intron sequence for massasauga rattlesnakes. Dots denote identity with allele A. See appendix for entire sequence.

		Variable Site Location
		12
		23431
Allele		99244
CBA	A	CTCTT
CBA	B	TCA.C
CBA	C	TCAAC

Table 10. Distribution of CBA intron alleles among populations of eastern massasauga rattlesnakes.

Population	Allele		
	A	B	C
BPNP	-	70	20
GBINP	-	56	6
PA	-	2	8
NY	-	62	-
OH	2	32	-
IL	-	16	-
MI	1	13	-
WI	-	4	-
GEOGR	-	13	1
KPP	58	162	8
OJIB	-	10	2

Table 11. Variable sites in the GAPD intron of eastern massasauga rattlesnakes. .

	Variable Site Location	
	11	
	214	
Allele	190	
GAPD A	GCC	
GAPD B	AAT	

Table 12. Distribution of GAPD intron alleles among populations of eastern massasauga rattlesnakes.

Population	Allele	
	A	B
BPNP	38	44
GBINP	59	5
PENN	6	4
NY	58	-
OHIO	19	15
ILL	14	-
MICH	13	1
WIS	1	3
GEORG	12	4
KPP	154	64
OJIB	12	-

Table 13. Estimates of among-population component of genetic variation (ϕ_{st}) for CBA intron sequence.

	BPNP	GBINP	PA	NY	OH	IL	MI	WI	GEORGE	KPP	OJIB
BPNP	--										
GBINP	0.041	--									
PA	0.47*	0.71*	--								
NY	0.18*	0.08	0.93*	--							
OH	0.13*	0.067	0.55*	0.06	--						
IL	0.11	0.026	0.82*	0	0	--					
MI	0.14	0.09	0.53*	0.14	0	0.0099*	--				
WI	0.015	0	0.69	0	0	0	0	--			
GEORG	0.028	0	0.69	0.14	0.0043	0.0099*	0.0063*	0	--		
KPP	0.2*	0.17*	0.32	0.16*	0.081	0.12	0.046	0.035	0.12	--	
OJIB	0	0	0.53	0.35	0.041	0.12	0.034	0	0	0.13	--

* = P<0.05 (corrected for multiple comparisons)
Georg. = Georgian Bay

Table 14. Estimates of among-population component of genetic variation (ϕ_{st}) for GAPD intron sequence.

	BPNP	GBINP	PA	NY	OH	IL	MI	WI	GEORGE	KPP	OJIB
BPNP	--										
GBINP	0.37*	--									
PA	0	0.32	--								
NY	0.49*	0.059	0.68*	--							
OH	0	0.32*	0	0.50*	--						
IL	0.39*	0.12	0.42	0	0.34	--					
MI	0.30*	0	0.21	0.13	0.22	0.0099*	--				
WI	0	0.72	0.055	0.96*	0.046	0.87*	0.67*	--			
GEORG	0.11	0.10	0	0.41*	0.030	0.2	0.044	0.31	--		
KPP	0.11*	0.11	0	0.20*	0.033	0.16	0.078	0.27	0	--	
OJIB	0.37*	0	0.36	0	0.31	0	0	0.83	0.16	0.15	--

* = P<0.05 (corrected for multiple comparisons)

Table 15. Estimates of among-population component of genetic variation (ϕ_{st}) for CBA and GAPD intron data pooled.

	BPNP	GBINP	PENN	NY	OH	IL	MI	WI	GEORGE	KPP	OJIB
BPNP	--										
GBINP	0.074*	--									
PA	0.092	0.18	--								
NY	0.11*	0.00039	0.26*	--							
OH	0.0063	0.046	0.16*	0.067	--						
IL	0.093	0	0.23	0	0.054	--					
MI	0.066	0	0.18*	0	0.027	0	--				
WI	0	0.12	0.15	0.16	0	0.15	0.1	--			
GEORG	0.019	0.0092	0.13	0.0098	0	0	0	0.026	--		
KPP	0.034*	0.031*	0.13*	0.052*	0.014	0.039	0.013	0.032	0	--	
OJIB	0.074	0	0.14	0	0.053	0	0	0.13	0	0.026	--

* = P<0.05 (corrected for multiple comparisons)

Table 16. Estimates of among-group component of genetic variation (ϕ_{CT}) for CBA, GAPD, and pooled intron data, respectively.

	group1	group2	group3
group1	--		
group2	0.017, 0, 0	--	
group3	0.042, 0.014, 0.046	0, 0.0038, 0	--

group1 = KPP, BPNP, GBINP, Georgian Bay, Pennsylvania, New York
group2 = Michigan, Ojibway, Ohio
group3 = Illinois and Wisconsin
* = P<0.05 (corrected for multiple comparisons)

Table 17. Mean gene flow (1% and 99% confidence values) between groupings of populations estimated from intron sequence using Migrate. Values represent gene flow from groups on the horizontal to groups on the vertical.

	N_fm – group1, *x*	N_fm – group2, *x*	N_fm – group3, *x*
Group1			
CBA	--	8.5	0
GAPD	--	19.15	0
Pooled	--	8.6 (3.69-10)	0
Group2			
CBA	9.6	--	0
GAPD	0	--	0
Pooled	3.02 (1.79-4.53)	--	0
group3			
CBA	0	0.34	--
GAPD	0	275	--
Pooled	0	11.4 (5.1-18.91)	--

Group1 = KPP, BPNP, GBINP, Georgian Bay, Pennsylvania, New York
Group2 = Michigan, Ojibway, Ohio
Group3 = Illinois, Wisconsin

FIGURE CAPTIONS

Figure 1. Map showing the approximate range of the eastern massasauga rattlesnake (circle), as well as the location and number of individuals used for the mitochondrial DNA analysis in this study. Population abbreviations are as follows: KPP = Killbear Provincial Park, GBINP = Georgian Bay Islands National Park, BPNP = Bruce Peninsula National Park, Wainfleet = Wainfleet Bog, and Ojibway = Ojibway Prairie.

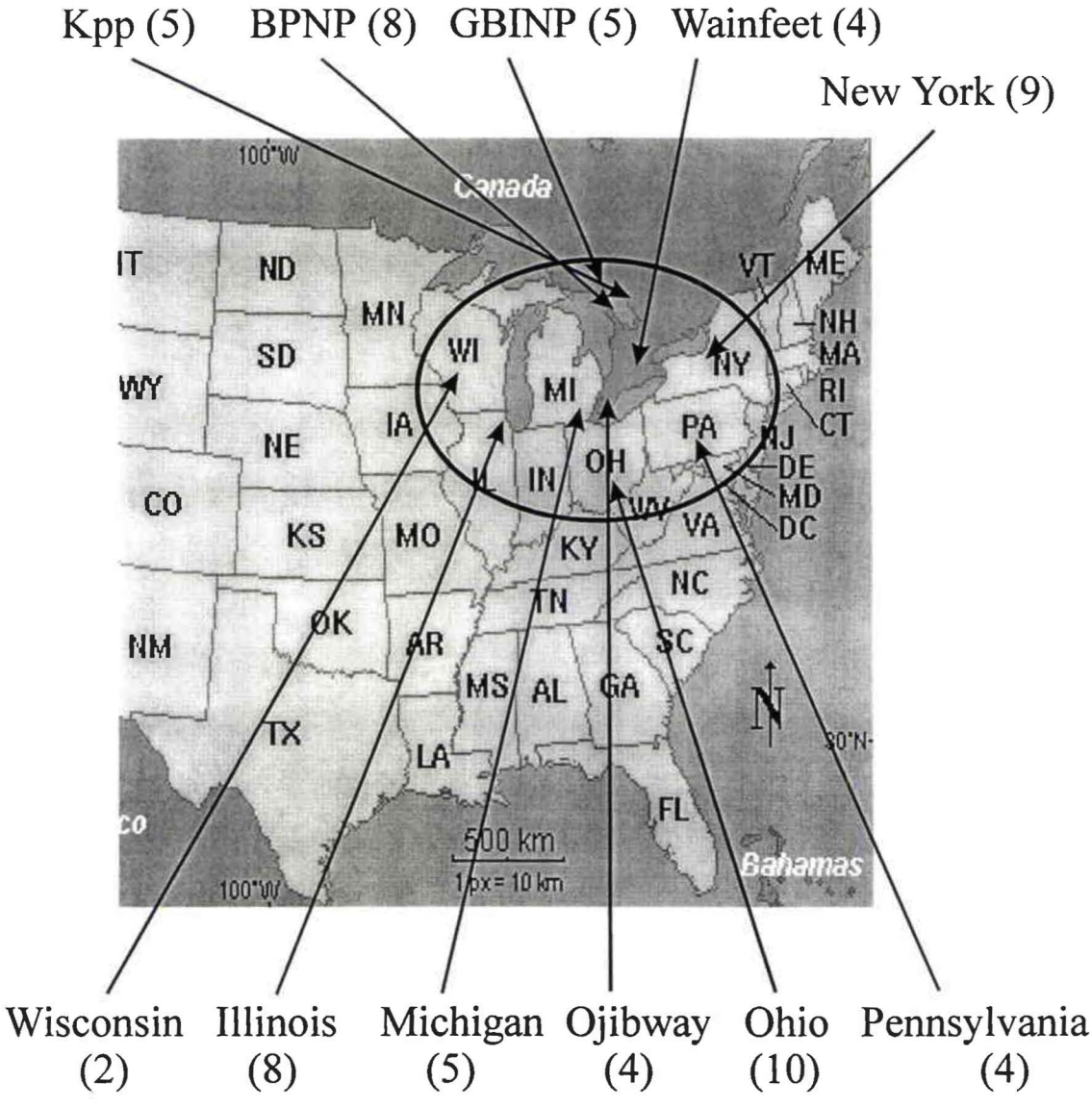
Figure 2. Diagram of the primers used to amplify the cytochrome *b*-control region complex, ND4-tRNA region, 12S, and control region II. H and L refer to heavy and light strands, respectively. Numbers indicate the approximate position of the primer in the region being amplified.

Figure 3. Cytochrome *b*, $tRNA^{Thr}$, non-coding region, and mitochondrial control region sequence for eastern massasauga rattlesnakes. Primer sequences are in bold, and conserved regions are underlined. The amino acid sequence of cytochrome *b* is shown.

Figure 4. Minimum-spanning network of eastern massasauga rattlesnake haplotypes. The number of substitutions between haplotypes is given next to branches, and the location and number of individuals with given haplotypes are denoted.

Figure 5. Map showing the distribution of control region haplotypes among eastern massasauga rattlesnake populations.

Figure 6. Schematic diagram showing levels of gene flow based on mtDNA (solid lines) and nuclear intron (dotted lines) differentiation between geographical groupings of eastern massasauga rattlesnake populations. Number represent gene flow in number of individuals per generation. N_m estimated from Migrate for mtDNA is also given within circles.


Kpp (5)
BPNP (8)
GBINP (5)
Wainfeet (4)
New York (9)
Canada
100°W
ND
MN
SD
WI
MI
NY
VT
ME
NH
MA
RI
CT
NJ
DE
MD
DC
WY
IA
NE
IL
IN
OH
PA
WV
VA
CO
KS
MO
KY
NC
TN
OK
AR
SC
NM
MS
AL
GA
TX
LA
FL
N
500 km
1 px = 10 km
Bahamas
Wisconsin (2)
Illinois (8)
Michigan (5)
Ojibway (4)
Ohio (10)
Pennsylvania (4)

Light Strand
Cytb15110L
Cytb16110L
Cytb1L
Cr1500H
CSB1L CR2H
12SAL
L14919
Cytb15810L
Cytb16410L
CSB2L
Cr1703L
CSB2H
Phe1H
12SBH
tRNA Glu
Cytochrome b
tRNA Thr
Non-coding region
Control Region I
tRNA Phe
12S rRNA
tRNA Val
L4221
H4419
Heavy Strand
tRNA Ile
tRNA Pro
Control Region II
tRNA Leu
tRNA Gln
tRNA Met
ND4L
LEUH
ND4L
tRNA His
tRNA Ser
tRNA Leu

```
    TRNA(Glu)¬ ⌐cytb→
L14919→1 aacaaaaacatgccccaccaacacctactcatacaatttaacctcctccccgtaggatcg
       1          M  P  H  Q  H  L  L  M  Q  F  N  L  L  P  V  G  S

      61 aacatctccacctgatggaacttcggatctatacttctatcctgtttaataattcaaatt
      18  N  I  S  T  W  W  N  F  G  S  M  L  L  S  C  L  M  I  Q  I

     121 actaccggcttcttcctggcaatacactacacagccaatatcaatatagccttctcctct
      38  T  T  G  F  F  L  A  M  H  Y  T  A  N  I  N  M  A  F  S  S

     181 atcgtacacatctcccgagatgtgccctacggctgaaccatacaaaacacacacgctatc
      58  I  V  H  I  S  R  D  V  P  Y  G  W  T  M  Q  N  T  H  A  I

                                          cytb15110L→
     241 ggcgcatctctattcttcatttgcatttatattcacatcgcacgaggaatctactatggc
      78  G  A  S  L  F  F  I  C  I  Y  I  H  I  A  R  G  I  Y  Y  G

     301 tcttaccttaacaaagaagtttgactgtccggcaccacccttctaattgtcctaatagct
      98  S  Y  L  N  K  E  V  W  L  S  G  T  T  L  L  I  V  L  M  A

     361 accgccttcttcggttatgttctaccatgaggccagatatcattctgggcagctacagtg
     118  T  A  F  F  G  Y  V  L  P  W  G  Q  M  S  F  W  A  A  T  V

                                          cytb15810L→
     421 atcacaaatttactaaccgcaatcccatacttaggaacaaccctcaccacatgactctga
     138  I  T  N  L  L  T  A  I  P  Y  L  G  T  T  L  T  T  W  L  W

     481 gggggctttgcaattaatgacccaacactaacccgattctttgccctacacttcatcctc
     158  G  G  F  A  I  N  D  P  T  L  T  R  F  F  A  L  H  F  I  L

     541 cccttcgccattatttctatttcatcaatccacattattcttctacacaatgaaggctcc
     178  P  F  A  I  I  S  I  S  S  I  H  I  I  L  L  H  N  E  G  S

     601 aataacccactaggcacaaactcagacatcgacaaaatcccatttcacccatatcattcc
     198  N  N  P  L  G  T  N  S  D  I  D  K  I  P  F  H  P  Y  H  S

     661 tacaaagatatcctcatatttacaacaatagtctctatttttattcatcgtcctctcactc
     218  Y  K  D  I  L  M  F  T  T  M  V  S  I  L  F  I  V  L  S  L

                                          cytb16110L→
     721 tacccggacatcttcaacgaccctgaaaacttctcaaaagccaacccttaattacacca
     238  Y  P  D  I  F  N  D  P  E  N  F  S  K  A  N  P  L  I  T  P

     781 caacacattaagcccgaatgatactttctattcgcctacggaatccttcgatcaatcccg
     258  Q  H  I  K  P  E  W  Y  F  L  F  A  Y  G  I  L  R  S  I  P

     841 aacaaactcgggggagttatagccctcatcctgtctgtcgctatcctattcacagcaccc
     278  N  K  L  G  G  V  M  A  L  I  L  S  V  A  I  L  F  T  A  P
```


901 ttcacccacacctcccacacccgatctataatgttccgacctatcatacagctaatattc
298 F T H T S H T R S M M F R P I M Q L M F
961 tgaacctttattactacattcattattattacctgaacagccaccaaaccagtagaaccc
318 W T F I T T F I I I T W T A T K P V E P
cytb16410L→
1021 ccattcacagaaatcggccaactagcctctatcctatacttcatattctttataacaaac
338 P F T E I G Q L A S I L Y F M F F M T N
cytb1L→ ←cytb¬⌐tRNA(Thr)→
1081 cccctactaggcctactagaaaacaaaatctcagacctcacctgctctaatagcttagac
358 P L L G L L E N K I S D L T
←tRNA(Thr)¬⌐non-coding
1141 tctcaaagcattgtttttgtaaaccaaagccggatattccttagagcaatcgacttccac
region→ non-coding region¬⌐control region→
1201 tttactatattttaaaccccccaaaaaccactatctggggacccaagccaaaaaaatactc
⌐ C-rich seq. ¬ Hairpin 1 Hairpin 2
1261 tcctaggaccccccccctaccccccattaatttgggtcccgaattcggccttatatgtact
Hairpin 3
1321 ctttacatatagggtcctcattgtcgctatgtataataatacattaatcgttttgcccca
CSB2L→
1381 tggctaataagcgggaattctactataattaaatatatacaaaactggctcattaacatc
1441 acttcctctcctcattttctggtcgttccatttaaaataggctgtcccttattagtaacc
←Cr1500H
1501 atggctatccacttcaaaccggtgtcccgtgatttaacccttcccgtgaaatcctctatc
⌐
1561 ctttcacctcaggcatacagtcccgcttctcacgtccatatattgtaactcctccgttt
repetitive sequence 1 ¬
1621 atgctctttccaaggccgctggttacaccttcaagggcatctcaatggtccggaaccacc
⌐ repetitive sequence 2 ¬ Cr1703L→
1681 ccgccttacttgctctttccaaggcctatggtcgcacccttttatcctggtacattaagtc
1741 tcatgttcttatcacgtatgcttgttccacccctggttggctttttatcggtacctttc
1801 acctgacacccatatatgctcgttaccgtccccctcaccggggtagacctctagtccagg
CSB1L→
⌐cons.seq.block 1
1861 tggagctatgttcttggtctggcactttcccctatagggatacatctcttcatgctcgtt
¬
1921 atacatatcgttctacattcctagaaaattccattatttttttattataaaaatcccgctg
←CSB2H

1981 **taactacatttttttta**ccctgtttttttatttttcaccaaaattaatcccactttcgtat

⌐cons.seq.block 3¬

2041 actaaaaatattaacccgaaataaaacaatctttttgctcggtgattttatttttgccgc

←**Cr2H**

2101 cccgtgaaaaaaaattcaaataaatataaaaatcgcccaccctatttt**tgtatccgggaa**

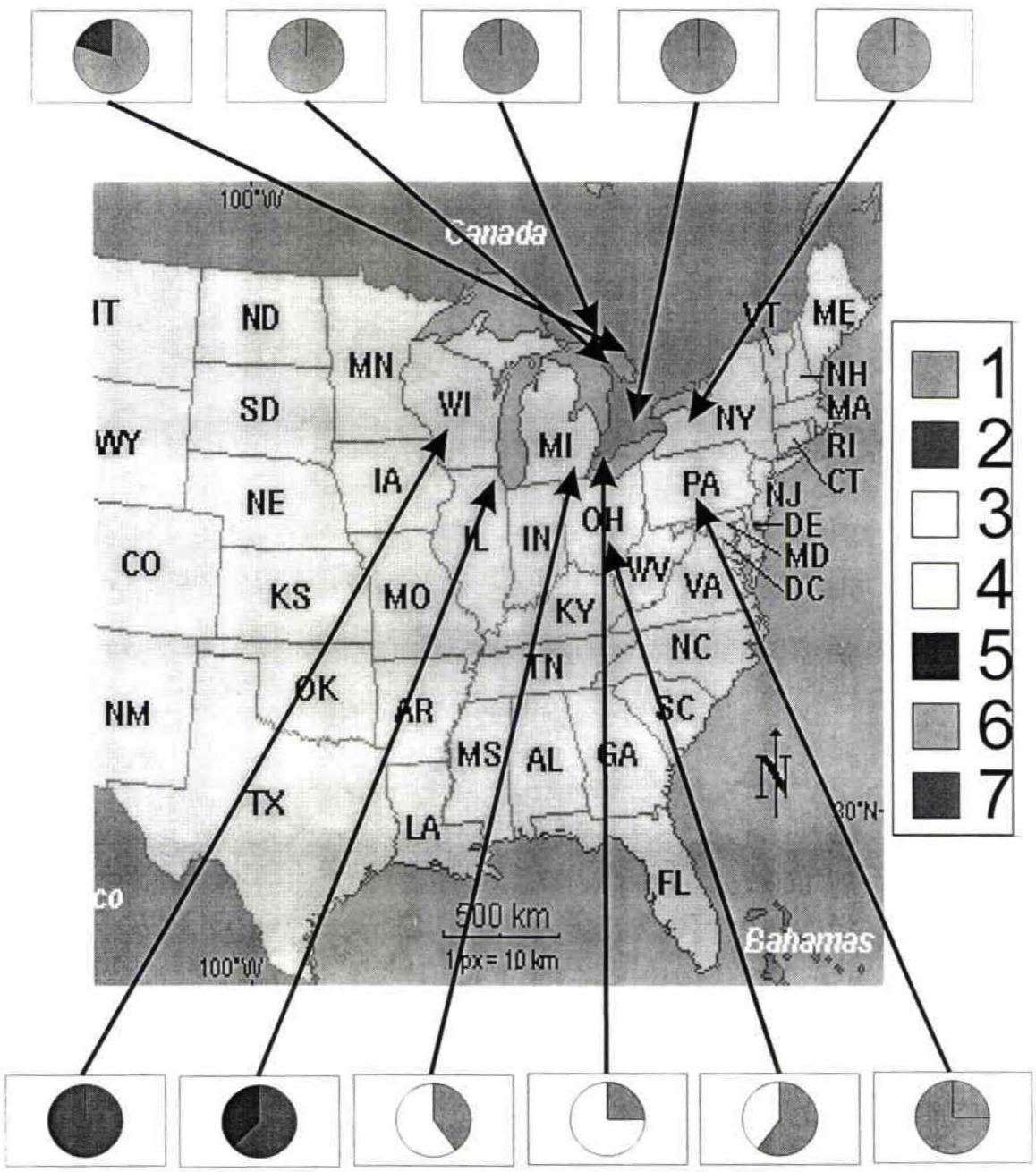
2161 **atgatctt**cttccctcggaggggacgtttcaccgaggctgagggccgaaatcagcacagc

2221 cgaatttatctttatttttttaacctctccagaatttttatacagttgtattgtagcttaa

←control region¬

2281 a

B
(7)
5-IL,
2-WI
2
D
(3)
MI
1
E
(3)
3-IL
1
C
(7)
3-Ojib,
4-OH
1
A
(40)
21-North Ontario,
9-NY, 2-MI,
1-Ojib, 1-PA,
6-OH,
1
G
(1)
KPP
1
F
(3)
PA



Distribution of haplotypes among populations

East
West
0
0
2.42
3.02
0.5
11.4
Northern
Ontario
Pennsylvania
New York
mt DNA
4400-60606
Michigan
Ojibway
Ohio
mt DNA
4800-72727
Illinois
Wisconsin
mt DNA
3650-55303
8.6
0
0
0
0
0

Appendix I. Sequence of portion of 12S rRNA examined in this study.

```
  1 AACATACAATTAAACTACTAATTGTCCGCCAAACAACTACGAGTAGTACTTAAAACTTAA 60
 61 AAGACTTGACGGTGCTTCACCACGCCCTAGAGGAGCCTGTCTAGCAACCGATAATCCACG 120
121 ATTAACCCAGCCTCCTCTGGCCTAACAGTCTATATACCGCCGTCGCCAGCTTACCTTGTA 180
181 AAAGAAATAAAGTAAGCCAAACAGTATTTTCCCTAAAACGACAGGTCGAGGTGTAACTAA 240
241 TGAGGAGGACTAAGATGGGCTACATTCTCCTAACCGAG                       278
```

Appendix II. Sequence of portion of ND4, $tRNA^{His}$, $tRNA^{Ser}$ examined in this study. Dashes denote $tRNA^{His}$ and $tRNA^{Ser}$, respectively.

```
  1  TAGTCCTAGCAGCAATTCTTCTTAAACTGGGAGGGTATGGTATTATTCGCATAATACAAA 60
 61  TTTTTCCTACAACAAAAACTGACATCTTCCTCCCATTCATTGTACTAGCCCTCTGAGGGG 120
121  CCATCCTAGCCAACTTAACATGCCTGCAACAAACAGATCTAAAATCACTAATCGCCTACT 180
181  CCTCTATTAGCCATATGGGCCTAGTAGTAGCTGCGATCATCATCCAAACCCCGTGGGGGC 240
241  TCTCCGGAGCTATAGCACTTATAATTGCACATGGCTTTACCTCCTCAGCACTTTTCTGCC 300
301  TAGCCAACACAACCTACGAACGCACACACACACGAATATTAATTACCGCCTCCTACTCCC 360
361  TACACATATTTCTATCAACACAAATAGGACCAACCCTGCTCAACAACCTAACAGAACCCA 420
421  CACACTCCCGAGAACACCTACTAATAATTCTACACATTACCCCCTTATTAATGCTCTCTT 480
481  TAAAACCAGAATTAATCATCAGA-GTGCACGTAATTTAAAAAAAATATCAAGTTGTGACC 540
541  TTGAAAATAGATTAACCTCGTTCACC-GAGAGGTACAGAAGACCTGCTAACTCTTTAATC 600
601  TGGTAAACACACCAGCCCTCTCTTCTATCAAGGGAGAATAGTTATTCCATTGGTCT     654
```

Appendix III. Sequence of control region haplotypes found for 64 eastern massasauga rattlesnakes.

```
A   1 CGTATGCTTGTTCCACCCCTGGTTGGCTTTTTTATCGGTACCTTTCACCTGACACCCATA 60
B   1 CGTATGCTTGTTCCACCCCTGGTTGGGTTTTTTATCGGTACCTTTCACCTGACACCCATA
C   1 CGTATGCTTGTTCCACCCCTGGTTGGGTTTTTTATCGGTACCTTTCACCTGACACCCATA
D   1 CGTATGCTTGTTCCACCCCTGGTTGGCTTTTTTATCGGTACCTTTCACCTGACACCCATA
E   1 CGTATGCTTGTTCCACCCCTGGTTGGGTTTTTTATCGGTACCTTTCACCTGACACCCATA
F   1 CGTATGCTTGTTCCACCCCTGGTTGGCTTTTTTATCGGTACCTTTCACCTGACACCCATA
G   1 CGTATGCTTGTTCCACCCCTGGTTGGCTTTTTTATCGGTACCTTTCACCTGACACCCATA

A  61 TATGCTCGTTACCGTCCCCCTCACCGGGGTAGACCTCTAGTCCAGGTGGAGCTATGTTCT 120
B  61 TATGCTCGTTACCGTCCCCCTCACCGGGGTAGACCTCTAGTCCAGGTGGAGCTATGTTCT
C  61 TATGCTCGTTACCGTCCCCCTCACCGGGGTAGACCTCTAGTCCAGGTGGAGCTATGTTCT
D  61 TATGCTCGTTACCGTCCCCCTCACCGGGGTAGACCTCTAGTCCAGGTGGAGCTATGTTCT
E  61 TATGCTCGTTACCGTCCCCCTCACCGGGGTAGACCTCTAGTCCAGGTGGAGCTATGTTCT
F  61 TATGCTCGTTACCGTCCCCCTCACCGGGGTAGACCTCTAGTCCAGGTGGAGCTATGTTCT
G  61 TATGCTCGTTACCGTCCCCCTCACCGGGGTAGACCTCTAGTCCAGGTGGAGCTATGTTCT

A 121 TGGTCTGGCACTTTCCCCTATAGGGATACATCTCTTCATGCTCGTTATACATATCGTTCT 180
B 121 TGGTCTGGCACTTTCCCCTATAGGGATACATCTCTTCATGCTTGTTATACATATCGTTCT
C 121 TGGTCTGGCACTTTCCCCTATAGGGATACATCTCTTCATGCTCGTTATACATATCGTTCT
D 121 TGGTCTGGCACTTTCCCCTATAGGGATACATCTCTTCATGCTCGTTATACATATCGTTCT
E 121 TGGTCTGGCACTTTCCCCTATAGGGATACATCTCTTCATGCTCGTTATACATATCGTTCT
F 121 TGGTCTGGCACTTTCCCCTATAGGGATACATCTCTTCATGCTCGTTATACATATCGTTCT
G 121 TGGTCTGGCACTTTCCCCTATAGGGATACATCTCTTCATGCTCGTTATACATATCGTTCT

A 181 ACATTCCTAGAAAATTCCATTATTTTTTATTATAAAAATCCCGCTGTAACTACATTTTTT 240
B 181 ACATTCCTAGAAAATTCCATTATTTTTTATTATAAAAATCCCGCTGTAACTACATTTTTT
C 181 ACATTCCTAGAAAATTCCATTATTTTTTATTATAAAAATCCCGCTGTAACTACATTTTTT
D 181 ACATTCCTAGAAAATTCCATTATTTTTTATTATAAAAATCCCGCTGTAACTACATTTTTT
E 181 ACATTCCTAGAAAAATCCATTATTTTTTATTATAAAAATCCCGCTGTAACTACATTTTTT
F 181 ACATTCCTAGAAAATTCCATTATTTTTTATTATAAAAATCCCGCTGTAACTACATTTTTT
G 181 ACATTCCTAGAAAATTCCATTATTTTTTATTATAAAAATCCCGCTGTAACTACATTTTTT

A 241 TACCCTGTTTTTTTATTTTTCACCAAAATTAATCCCACTTTCGTATACTAAAAATATTAA 300
B 241 TACCCTGTTTTTTTATTTTTCACCAAAATTAATCCCACTTTCGTATACTAAAAATATTAA
C 241 TACCCTGTTTTTTTATTTTTCACCAAAATTAATCCCACTTTCGTATACTAAAAATATTAA
D 241 TACCCTGTTTTTTTATTTTTCACCAAAATTAATCCCACTTTCGTATACTAAAAATATTAA
E 241 TACCCTGTTTTTTTATTTTTCACCAAAATTAATCCCACTTTCGTATACTAAAAATATTAA
F 241 TACCCTGTTTTTTTATTTTTCACCAAAATTAATCCCACTTTCGTATACTAAAAATTTTAA
G 241 TACCCTGTTTTTTTATTTTTCACCAAAATTAATCCCACTTTCGTATACTAAAAATATTAA

A 301 CCCGAAATAAAACAATCTTTTTGCTCGGTGATTTTATTTTTGCCGCCCCGTGAAAAAAAA 360
B 301 CCCGAAATAAAACAATCTTTTTGCTCGGTGATTTTATTTTTGCCGCCCCGTGAAAAAAAA
C 301 CCCGAAATAAAACAATCTTTTTGCTCGGTGATTTTATTTTTGCCGCCCCGTGAAAAAAAA
D 301 CCCGAAATAAAACAATCTTTTTGCTCGGTGATTTTATTTTTGCCGCCCCGTGAAAAAAAA
E 301 CCCGAAATAAAACAATCTTTTTGCTCGGTGATTTTATTTTTGCCGCCCCGTGAAAAAAAA
F 301 CCCGAAATAAAACAATCTTTTTGCTCGGTGATTTTATTTTTGCCGCCCCGTGAAAAAAAA
G 301 CCCGAAATAAAACAATCTTTTTGCTCGGTGATTTTATTTTTGCCGCCCCGTGAAAAAAAA
```

```
A 361 TTCAAATAAATATAAAAATCGCCCACCCTATTTTTGTATCCGGGAAATGATCTTCTTCCC 420
B 361 TTCAAATAAATATAAAAATCGCCCACCCTATTTTTGTATCCGGGAAATGATCTTCTTCCC
C 361 TTCAAATAAATATAAAAATCGCCCACCCTATTTTTGTATCCGGGAAATGATCTTCTTCCC
D 361 TTCAAATAAATATAAAAATCGCCCACCCTATTTTTGTATCCGGGAAATGATCTTCTTCCC
E 361 TTCAAATAAATATAAAAATCGCCCACCCTATTTTTGTATCCGGGAAATGATCTTCTTCCC
F 361 TTCAAATAAATATAAAAATCGCCCACCCTATTTTTGTATCCGGGAAATGATCTTCTTCCC
G 361 TTCAAATAAATATAAAAATCGCCCACCCTATTTTTGTATCCGGGAAATGATCTTCTTCCC

A 421 TCGGAGGGGACGTTTCACCGAGGCTGAGGGCCGAAATCAGCACAGCCGAATTTATCTTTA 480
B 421 TCGGAGGGGACGTTTCACCGAGGCTGAGGGCCGAAATTAGCACAGCCGAATTTATCTTTA
C 421 TCGGAGGGGACGTTTCACCGAGGCTGAGGGCCGAAATCAGCACAGCCGAATTTATCTTTA
D 421 TCGGAGGGGACGTTTCACCGAGGCTGAGGGCCGAAATCAGCACAGCCGAATTTATCTTTA
E 421 TCGGAGGGGACGTTTCACCGAGGCTGAGGGCCGAAATCAGCACAGCCGAATTTATCTTTA
F 421 TCGGAGGGGACGTTTCACCGAGGCTGAGGGCCGAAATCAGCACAGCCGAATTTATCTTTA
G 421 TCGGAGGGGACGTTTCACCGAGGCTGAGGGCCGAAATCAGCACAGCCGAATTTATCTTTA

A 481 TTTTTTAACCTCTCCAGAATTTTTATACAGTTGTATTGTAGCTTAAA             527
B 481 TTTTTTAACCTCTCCAGAATTTTTATACAGTTGTATTGTAGCTTAAA
C 481 TTTTTTAACCTCTCCAGAATTTTTATACAGTTGTATTGTAGCTTAAA
D 481 TTTTTTAACCTCTCCAGAATTTTTATACTGTTGTATTGTAGCTTAAA
E 481 TTTTTTAACCTCTCCAGAATTTTTATACAGTTGTATTGTAGCTTAAA
F 481 TTTTTTAACCTCTCCAGAATTTTTATACAGTTGTATTGTAGCTTAAA
G 481 TTTTTTAACCTCTCCAGAATTTTTATACAGTTGTATGGTAGCTTAAA
```

Appendix IV. Groupings of populations with high among-group components of variation (ϕ_{ct}). Groupings are listed from highest ϕ_{ct} to lowest.

1. $\phi_{ct} = 0.52$	1. GBINP KPP BPNP WAIN	2. NY PENN	3. MICH OJIB OHIO	4. WIS ILL
2. $\phi_{ct} = 0.47$	1. GBINP KPP BPNP WAIN NY	2. PENN MICH OJIB	3. WIS ILL OHIO	
3. $\phi_{ct} = 0.40$	1. GBINP KPP BPNP	2. NY WAIN PENN	3. MICH OJIB	4. WIS ILL OHIO

Appendix V. Sequence of CBA intron alleles.

```
CBA A    1 CCCCCAATAAAATACTAACTTTGGCTTGCTCTGAGTGATTTCTATCATGTTTTCTTAAAA 60
CBA B    1 CCCCCAATAAAATACTAACTTTGGCTTGTTCTGAGTGACTTA?????TGTTTTCTTAAAA
CBA C    1 CCCCCAATAAAATACTAACTTTGGCTTGTTCTGAGTGACTTA?????TGTTTTCTTAAAA

CBA A   61 TTGTAAAACTGAAAATTTGCCCAATACGGAAAGGGTGTGGTTTGCCTTAACACCTGCAAT 120
CBA B   56 TTGTAAAACTGAAAATTTGCCCAATACGGAAAGGGTGTGGTTTGCCTTAACACCTGCAAT
CBA C   56 TTGTAAAACTGAAAATTTGCCCAATACGGAAAGGGTGTGGTTTGCCTTAACACCTGCAAT

CBA A  121 GAAGGTTTGTATCTGTCCAGCATGTGTTAGATTTTTCCAGTGGTGTTTTCTGCGCATGGT 180
CBA B  116 GAAGGTTTGTATCTGTCCAGCATGTGTTAGATTTTTCCAGTGGTGTTTTCTGCGCATGGT
CBA C  116 GAAGGTTTGTATCAGTCCAGCATGTGTTAGATTTTTCCAGTGGTGTTTTCTGCGCATGGT

CBA A  181 GTTGATGGGGCCTTCTGTTGGCCACGATCTTCATGGCCCTCTCTCTCCTCCCATCTA    237
CBA B  176 GTTGATGGGGCCTTCTGTTGGCCACGATCTTCACGGCCCTCTCTCTCCTCCCATCTA
CBA C  176 GTTGATGGGGCCTTCTGTTGGCCACGATCTTCACGGCCCTCTCTCTCCTCCCATCTA
```

Appendix VI. Sequence of GAPD intron alleles.

```
GAPD A    1 CTCAATGATCATTTTGTCAAGCTGATCTCCTGGTAAGAGATGAAAAGGGGTAGATGATGG 60
GAPD B    1 CTCAATGATCATTTTGTCAAACTGATCTCCTGGTAAGAGATGAAAAGGGGTAGATGATGG

GAPD A   61 CTGTTAGTTGGCTTATTGGATTTGGGAAGAGATGGAGGTGCAACCTTTACTATTCTGCCG 120
GAPD B   61 CTGTTAGTTGGCTTATTGGATTTGGGAAGAGATGGAGGTGCAACCTTTACTATTCTGCAG

GAPD A  121 GATAGCAGAGGTGGGAGTACGTGGCGGCAAGAAATGGCTTCATGCCGCTCTTTTTATTGC 180
GAPD B  121 GATAGCAGAGGTGGGAGTATGTGGCGGCAAGAAATGGCTTCATGCCGCTCTTTTTATTGC

GAPD A  181 TGCTATTTTCCTCAAGACTAAAAGAAACTTTTCTTTCATAGGTATGATAATGAA       234
GAPD B  181 TGCTATTTTCCTCAAGACTAAAAGAAACTTTTCTTTCATAGGTATGATAATGAA
```

A Tale of Two Control Regions:
Evidence for Concerted Evolution of Two Control-Region-Like Sequences in the Eastern Massasauga Rattlesnake (*Sistrurus catenatus catenatus*)

Keywords: eastern massasauga rattlesnake, mitochondrial control region, concerted evolution, gene conversion, mitochondrial recombination

ABSTRACT

Mitochondrial gene rearrangements, such as the presence of two control region-like sequences, have been demonstrated for a variety of taxa, and their presence can be used to infer mechanisms and rates of recombination in mitochondria. We sequenced two entire control region-like sequences in 10 eastern massasauga rattlesnakes, *Sistrurus catenatus catenatus*, one found in the cytb-12S rRNA region, and the other in the IQM tRNA cluster, as well as the tRNA sequences flanking these regions. We also compared the sequence of the last 509 bp of both control regions for 28 individual snakes. Both control regions were found to be nearly identical within individuals, with a divergence of less than 0.5%. Among the 28 individuals only three showed different sequences between the two control regions, and these sequences differed at only one nucleotide. This suggests that the similarity of these separate regions within individuals is due to the fact that they have undergone concerted evolution, potentially due to gene conversion. A minimum gene conversion rate of one conversion per 1.2 million years is estimated for the sequences.

INTRODUCTION

Mitochondria are small, cellular organelles that contain genes for 13 proteins, two ribosomal RNAs, and 22 transfer RNAs, as well as a non-coding control region that contains the sites for the initiation of mitochondrial DNA (mtDNA) replication and RNA transcription (Brown 1983). MtDNA is maternally transmitted, so from a population-genetic perspective, it can be considered a system of sexually isolated demes, or clonal lineages (Howell 1997), and one of its great assets is that it is thought to be inherited as one, non-recombining molecule. Indeed, it was once thought that the mitochondria contain sequences that occur only once per genome, and that these genomes are less susceptible to sequence rearrangements than nuclear DNA (Brown 1983). Mitochondrial gene order and size were thought to be highly conserved features in vertebrates (Brown 1983), in that larger genomes were selected against, potentially due to the increased amount of time required to complete replication of a larger mtDNA molecule (Rand and Harrison 1986). More recent work, however, has suggested that mitochondrial have the potential to recombine, and that mitochondrial gene organization and size vary among species.

While there may be good theoretical reasons for expecting natural selection to reduce recombination, it seems to be a universal feature of DNA (Maynard Smith 1977). Thyagarajan *et al.* (1996) tested the hypothesis that recombinational DNA repair occurs within the mammalian mitochondria by extracting mitochondrial proteins from mammalian cells, and observing their effects on plasmid DNA substrates. As these proteins were found to catalyze homologous recombination of the plasmid substrates, the

authors concluded that mammalian mitochondria may have a homologous recombination-based process to repair damaged DNA. Lunt and Hyman (1997) analyzed variable tandem number repeats (VNTRs) in the mtDNA of the phytonematode *Meloidogyne javanica*, and found frequent deletions in the center of the array, indicating that genetic recombination may be occurring. Awadalla *et al.* 1999 found that linkage disequilibrium (LD) in human (*Homo sapiens*) mtDNA declines as a function of distance between sites, which they concluded could only be attributed to recombination. This work, however, has been criticized in other studies that found no significant relationship between LD and distance, and thus no evidence for significant levels of mtDNA recombination in humans (Elson *et al.* 2001). Recently, Ladoukakis and Zouros (2001) examined 681 bp of the COIII mtDNA gene from the gonads of mussels of the family Mytilidae and Unionidae, and found that of the 24 sequences obtained, six were recombinant, suggesting that recombination could be very common in mussel mitochondrion. This recombination could be due to the doubly uniparently inheritance (DUI) mechanism of mtDNA transmission found in these mussles (Ladoukakis and Zouros 2001).

Rearrangements and duplications of mtDNA sequences have been noted in a number of vertebrate and invertebrate lineages, suggesting the possibility of recombination in mitochondria. A characterization of the mitochondria of seven species of *Chnemidophorus* lizard showed that some lizards had standard-size mtDNAs, and other had larger mtDNAs (Moritz and Brown 1987). These size differences were found to be due to direct tandem duplications of sequence, a number of which spanned the control region, and one or more rRNA gene (Moritz and Brown 1987). The mtDNA

duplications were present in only some individuals of some species, so it was suggested that this duplicated state was not characteristic of *Chnemidophorus* mtDNAs, and may be transient (Moritz and Brown 1987). Analysis of mtDNA from sexual and parthenogenetic geckos of the *Heteronotia binoei* complex revealed a system of closely related genomes that varied in length due to tandem duplications, which were attributed to replication errors during the origin of the parthenogenetic lineages (Zevering *et al.* 1991). Stanton *et al.* (1994) found three duplication of gene-encoding regions in species of whiptail lizard (genus *Cnemidophorus*), and Campbell and Barker (1999) reported five repeats of the $tRNA^{Glu}$ gene and 60bp of the NADH dehydrogenase subunit 1 gene in the mitochondria of the cattle tick *Boophilus microplus*, as well as relocation of various tRNA genes (Campbell and Barker 1999). Relocated tRNAs have been found in sea cucumber species of the genus *Cucumaria* (Arndt and Smith 1998), as well as parrots of the genus *Amazona* (Eberhard et al. 2001), the chicken (*Gallu gallus*) (Desjardins and Morais 1990), and snakes, such as the himehabu (*Ovophis okinavenis*), the western rattlesnake (*Crotalus viridis*), and the akamata (*Dinodon semicarinatus*) (Kumazawa *et al.* 1996).

One of the most unique mitochondrial duplications noted involves the mtDNA control region. Kumazawa *et al.* (1996) found that mtDNAs from snakes of the Viperidae and Colubridae families, including the himehabu, the western rattlesnake, and the akamata possessed duplicated, widely separated, control-region-like sequences that are nearly identical to each other within species. One of these control regions was found in the cytb-12S rRNA region (CR1), and the other in the IQM-related tRNA cluster (Kumazawa *et al.* 1996). This duplicate state of CR-like sequences is thought to have

been maintained in snake mtDNAs for 70 million years (Kumazawa *et al.* 1996). *B. microplus* mitochondria were found to have two large (~300bp) non-coding regions, one between the 12S rRNA and $tRNA^{Ile}$ (CR#1), a position homologous to that of the control region in other arthropods, and one between $tRNA^{Leu}$ and $tRNA^{Cys}$ genes (CR#2) (Campbell and Barker 1999). The two CRs present in *B. microplus* are nearly identical, and may have persisted and evolved in concert for over 210 Myr (Campbell and Barker 1999). The sea cucumber *Cucumaria miniata* has two copies of the putative control region, which differ by only 3% if the analysis is restricted to the first 420 nucleotides (Arndt and Smith 1998). Parrots of the genus *Amazona* have duplicated control regions that differ somewhat in length, but show a high degree of sequence similarity, and conserved structural features found in avian control regions (Eberhard *et al.* 2001). Since duplication of mitochondrial genes presumably creates redundant sequences, one or the other copy should be under more relaxed selection, and show different evolution (Zevering *et al.* 1991). This is apparently not the case for the aforementioned control region duplications. One potential explanation for the apparently concerted evolution of the duplicated control regions is that gene conversion has acted to homogenize the CR-like sequences (Kumazawa *et al.* 1996).

Gene conversion, a process in which the sequence of DNA in a region is replaced by sequence from another copy of homologous DNA, which can cause sequence similarity through concerted evolution (Charlesworth *et al.* 2001), often occurs between members of nuclear gene families. Many multigene families exist in eukaryote genomes, and continued gene conversion may help genes acquire new function due to selection

(Ohta 1997). For example, genes within the trypsin gene family show a pattern of concerted evolution, in which gamma and delta trypsin genes are nearly identical in sequence within *Drosophila melanogaster* and *Drosophila erecta*, even though the duplication event that gave rise to these genes occurred at least 12 MYA (Wang *et al.* 1999). The region coding for the α gene of hemoglobin has been in the duplicate state for at least 300 million years and concerted evolution has continually kept the duplicates from diverging greatly (Zimmer *et al.* 1980). Gene conversion has been found to occur at diverse loci, including the *Est-5* gene family in *Drosophila pseudobscura* (Mertens King 1998), the major histocompatibility complex (MHC) in the mouse (*Mus musculus*) (Högstrand and Böhme 1999), and the von Willebrand factor gene in humans (Surdhar *et al.* 2001). Only recently has gene conversion been suggested to occur in mitochondria.

This study examines and compares duplicate control regions in 28 eastern massasauga rattlesnakes *Sistrurus catenatus catenatus*. Within individuals, control regions are nearly identical, suggesting concerted evolution resulting from gene conversion between the two regions. Potential mechanisms of gene conversion are reviewed and discussed.

MATERIALS AND METHODS

DNA Sampling, Extraction and Amplification

Blood, shed skin, or tissue samples were obtained from 28 eastern massasauga rattlesnakes from 13 different sites in Ontario and the United States. Samples were preserved in Queen's Lysis Buffer for transport and storage (blood), or dried (skin and tissue), and once received were kept at -20°C for long-term storage. To extract DNA

from blood samples, blood was digested overnight at 65°C in a solution of 1mg/ml proteinase K (ICN Biomedicals, Aurora, Ohio), 100mM Tris Cl (pH 8.0), 10mM EDTA (pH 8.0), 100mM NaCl, and 1% SDS. DNA was isolated from the digested samples by extracting them twice with equal volumes of Tris-saturated phenol (pH 8.0) and once with 24:1 chloroform:isoamyl alcohol. Extracted DNA was then diluted 1:10 with double distilled water prior to amplification. Skin and tissue samples were minced using a sterile razor blade, and extracted using a Dneasy Tissue Kit™ (Qiagen). Enriched mitochondrial DNA was also extracted from frozen tissue samples with an mtDNA Extractor CT Kit (Wako).

DNA was amplified using the polymerase chain reaction (PCR) (Mullis and Faloona 1987) in 30µl volumes containing 75mM Tris-HCl, 20mM $(NH_4)_2SO_4$, and 0.01% Tween 20, 0.33mM each of the four dNTPs, 15 pmol each of a forward and reverse primer, 8% DMSO, and 0.6 units *Thermus aquaticus* (Taq) polymerase (MBI Fermentas) in either a GeneAmp ™ 9700 (Applied Biosystems), or PTC ™ 900 thermal cycler (MJ Research) with Hot Bonnet, or a Perkin Elmer 480 thermal cycler, with an oil overlay. Amplification followed standard temperature profiles (denaturation at 94°C for 45s, annealing for 45s, extension at 72°C for 45s) with primer-specific annealing temperatures (see Table 1). The entire cytochrome *b*-control region area was amplified using slightly different reaction conditions. These "long" PCRs were performed in 30µl volumes containing 75mM Tris-HCl, 20mM $(NH_4)_2SO_4$, and 0.01% Tween 20, 3% BSA, 5% DMSO, 0.58mM each of the four dNTPs, 25pmol of forward and reverse primers,

and 1 unit *Thermus aquaticus* (Taq) polymerase (MBI Fermentas) in a Perkin Elmer 2400 thermal cycler. Long PCRs were amplified with an annealing temperature of 48°C.

Amplification of Control Regions

In order to ensure that only control region 1 (located between cytochrome *b* and $tRNA^{Phe}$) was amplified, PCR primers specific for amplifying cytochrome *b*, $tRNA^{Thr}$, and the mitochondrial control regions of the eastern massasauga rattlesnake, and the closely related desert massasauga rattlesnake (*Sistrurus c. edwarsii*) were designed as follows. Firstly, an approximately 2.5 kb fragment consisting of cytochrome *b*, $tRNA^{Thr}$, and the mitochondrial control region was amplified using a light-stranded primer situated in $tRNA^{Glu}$ near the 3' end of cytochrome *b* (L14919, R. Lawson) and a heavy-stranded primer situated in $tRNA^{phe}$ near the 3' end of the mitochondrial control region (Phe1H, R. Lawson) (Table 1). PCR products were subjected to electrophoresis in 1% agarose gels, gel purified, and sequenced using the chain termination method (Sanger *et al.* 1977) with ThermoSequenase™ radiolabelled kits (Amersham), according to the manufacturer's suggested protocol. The sequencing reaction products were electrophoretically separated in 5% polyacrylamide gels, and visualized by autoradiography. Sequences were manually aligned using the program ESEE (Cabot and Beckenbach 1989), and used to develop a series of eastern massasauga rattlesnake-specific primers (Table 1). These nested primers were used in combination with the previously mentioned primers to amplify overlapping segments of varying lengths, comprising cytochrome *b*, $tRNA^{Thr}$, a non-coding spacer region and the control region in their entirety for ten eastern massasauga rattlesnakes, and four desert massasauga rattlesnakes. These overlapping

sequences were used to develop the eastern massasauga rattlesnake-specific primer CR1703L, which was used along with Phe1H to selectively amplify a 509 bp fragment of Domain II and Domain III of both control region-like sequences for 28 individuals. The control region 2 (located within the isoleucine-glutamine-methionine (IQM) tRNA gene cluster), as well as $tRNA^{Pro}$, $tRNA^{Leu}$, and $tRNA^{Gln}$ were amplified and sequenced for 10 individuals, using the primer set 4221L, located in $tRNA^{Ile}$ and 4419H, located in $tRNA^{Met}$ (Macey *et al.* 1999).

Analyses

A minimum-spanning network of mtDNA haplotypes was constructed using the number of pairwise substitutions as the phenetic measure using the statistical package Arlequin (version 2.0; Schneider *et al.* 2000). Minimum-spanning trees were constructed using the Dnapars subroutine of Phylip (vers 3.5c; Felsenstein 1989).

RESULTS

Evidence for mitochondrial origin of control region sequences

Segments of the mitochondrial control region used in the population-level analyses were amplified in greater than 1kb fragments and were sequenced manually in both directions with overlap using nested primers. To further guard against the amplification of nuclear copies, sequence autoradiographs were scrutinized for evidence of multiple bands indicating more than one sequence; sequences of protein-coding genes (ND4 and cytochrome *b*) were translated to confirm that the reading frames were free of insertions or deletions that were not a multiple of three; the tRNA genes were mapped by hand

($tRNA^{Thr}$, $tRNA^{Pro}$, $tRNA^{Gln}$, $tRNA^{Leu}$) (Figure 1) onto secondary structure models and checked for any substitutions that would be incompatible with the predicted structures; and sequence was compared to that obtained from purified mtDNA. The control region had a base composition (A=23%, G=17%, C=25%, and T=35%) similar to that reported for other species of snakes (Kumazawa *et al.* 1996). A number of conserved structural features were present in the control region (C-rich sequence, hairpins 1, 2, and 3, repetitive sequences 1 and 2, conserved sequence blocks 1 and 3) that show a high degree of similarity to analogous regions in the western rattlesnake (*Crotalus viridis*) (Kumazawa *et al.* 1996) (Figure 2).

Cytb-12S rRNA Region

For eastern massasauga rattlesnakes, the control region 1 was found to be 1023bp in length, and the desert massasauga had a control region one of 1021bp. Similar to results found by Kumazawa *et al.* (1996) for the western rattlesnake (*Crotalus viridis*), both eastern massasaugas and desert massasaugas were found to have a $tRNA^{Thr}$ gene 3' immediately downstream of the cytochrome *b* gene, followed by a non-coding sequence, and the control region 1 (Figures 2, 3). In the western rattlesnake, the non-coding sequence was found to be 130bp, while in the eastern massasauga, the sequence is 51bp. On the other hand, the desert massasauga has a non-coding region of 295bp, consisting of a 71bp sequence, of which 35bp corresponds to a partially folding $tRNA^{Pro}$, followed by two 64bp repeats, of which the first 13bp corresponds to another portion of the $tRNA^{Pro}$, one 76bp stretch of sequence that closely resembles the first two repeats, and another

20bp sequence, corresponding to another, partially overlapping portion, of the $tRNA^{Pro}$ (Figure 2).

IQM Region

For the IQM region, eastern massasaugas were found to have a sequence that can be folded into a $tRNA^{Pro}$ immediately 3' downstream of the $tRNA^{Ile}$ gene. A sequence that can be folded into a $tRNA^{Leu}$ occurs immediately 5' upstream of the $tRNA^{Gln}$ gene. This is similar to the structure found for the western rattlesnake (Kumazawa *et al.* 1996). In between the $tRNA^{Pro}$ and $tRNA^{Leu}$ genes is an insert of 1023bp, corresponding to the control region 2 (Figures 2, 3).

Conserved Identity Between Control Regions Within Individuals

Within the ten individuals sequenced in entirety at both control regions, control regions were found to be nearly identical within individuals. The same structural features found in the control region 1, including a C-rich sequence, hairpins 1, 2, and 3, repetitive sequences 1 and 2, and conserved sequence blocks 1 and 3 (Kumazawa *et al.* 1996, 1998) were identical between control regions. For the 10 individuals, the only differences between control regions were noted in three of the first four nucleotides.

tRNA Characterization

Possible secondary structures of the tRNA genes found in eastern massasauga rattlesnake mtDNA sequences are given in Figure 1. For eastern massasaugas, the sequences for $tRNA^{Pro}$ and $tRNA^{Gln}$ were found to be very similar to the sequences found by Kumazawa *et al.* (1996) for the himehabu (*Ovophis okinavensis*). However,

sequences for $tRNA^{Leu}$ and $tRNA^{Thr}$ from eastern massasaugas were quite different than those found for the himehabu. The $tRNA^{Thr}$ gene found in the eastern massasauga and desert massasauga may be nonfunctional, as it contained a large number of mismatches in the T arms. Eastern massasauga and desert massasauga $tRNA^{Thr}$ gene sequences were identical. Similar to the himehabu, $tRNA^{Pro}$ exhibits the same strikingly truncated TψC (T arm), which has only two G-C pairs in the T stem region, and $tRNA^{Leu}$ has only three bases in the loop of the T arm. Kumazawa *et al.* (1996) showed that for a variety of vertebrates, the number of nucleotides in the T arm is constant at 17 for $tRNA^{Gln}$ and $tRNA^{Leu}$. Eastern massasauga $tRNA^{Gln}$ does indeed have 17 nucleotides in the T arm, however, $tRNA^{Leu}$ has 16 nucleotides, suggesting that a 17 nucleotide-T arm in $tRNA^{Leu}$ may not necessarily be the rule.

Characterization of surveyed mitochondrial variation among control regions

We identified 6 variable positions in 509 bp of sequence defining 8 distinct haplotypes for both control regions 1 and 2 among 28 eastern massasauga rattlesnakes (Table 2). A minimum-spanning network of mtDNA haplotypes is shown in Figure 4. The number of substitutions between haplotypes varied from 1 to 4. Within individuals, all but 3 snakes had identical haplotypes at both control regions (Table 3). For the 3 snakes with different control region haploytpes, the control region 1 and 2 haplotypes differed by only one nucleotide (Table 3, Figure 4). Three minimum-spanning networks were given, each differing only in the placement of haplotype 5. The most striking observation from these networks is that haplotypes 4, 7 and 8, each of which is found in only one control region in one individual, are separated from haplotype 1, which is found

in the other control regions in these individuals, by only one substitution. For this reason, the most probably network is the one in which haplotypes 5 is most closely related to haplotype 2. Five minimum-spanning trees were found, two in which haplotype 4 was more closely related to haplotypes other than 1. It seems more likely, however, that the sequences found at the two control regions within an individual should be more closely related to each other than to sequences from different individuals, so these networks are not shown in Figure 4.

DISCUSSION

Between the eastern massasauga and desert massasauga, considerable difference in the size of the non-coding region between $tRNA^{Leu}$ and the CR1 was noted. Typically, this region contains a gene for $tRNA^{Pro}$, and indeed, sequence corresponding to a partially folding $tRNA^{Pro}$ was noted in this region in the desert massasauga rattlesnake. The sequence for this tRNA is divided by three tandem repeats of sequence. The phenomenon of tandem repeats in mitochondrial DNA has been noted in a number of species, including the parrot (Eberhard *et al.* 2001), cattle tick (Campbell and Barker 1999), lizards (Moritz and Brown 1987), and *Drosophila*, which have an (A+T)-rich region that varies in size between species, due to the repetition of a 470bp sequence that is present only once in some species (Solignac *et al.* 1986). It is likely that the duplication and insertion of $tRNA^{Pro}$ into the IQM tRNA cluster released the functional constraints on the cytb-CR1-related $tRNA^{Pro}$, allowing it to more rapidly evolve and lose function.

Interestingly, the $tRNA^{Thr}$ gene found between cytochrome *b* and the non-coding region may not be functional. It is unlikely that this is represents nuclear copy, as both the eastern massasauga, and the desert massasauga have an identical sequence for this tRNA. It could be that snakes of the genus *Sistrurus* have undergone a duplication and rearrangement resulting in the insertion of a $tRNA^{Thr}$ gene somewhere else in the mitochondrial genome, allowing the sequence for the original $tRNA^{Thr}$ to diverge. Further sequencing of mitochondria from snakes of this genus could help clarify this point.

Comparison of two control regions

For the 10 snakes in which the entire CR1and CR2 were sequenced, the two control region sequences were found to be nearly identical, save for three nucleotides in the first four nucleotides found at the 5'ends of both control regions. This is similar to the cattle tick, whose two control regions differ at only four nucleotides, all of these in the first 5bp at the 3' ends of the regions (Campbell and Barker 1999). In eastern massasaugas, the sequence of the two control regions diverges by less than 0.5%. This is smaller than the values found for the *Amazona* parrots (1.4%; Eberhard *et al.* 2001) and *Cucumaria* sea cucumbers (3%; Arndt and Smith 1998).

Both eastern massasauga control regions contained structural elements commonly found in snake control regions (Kumazawa *et al.* 1996), and there is nothing to suggest that either is nonfunctional. If only one copy was functional and gene conversion was directional, such that the functional copy always converted the nonfunctional copy, the nonfunctional copy still would be expected to accumulate changes more quickly between

conversion events (Eberhard *et al.* 2001). However, when the second half of both control regions were examined for 28 individuals, this is not observed, and several interesting patterns are obvious. CR1 and CR2 haplotypes were identical within all individuals save for three. Within these three individuals, the sequence of one control region haplotype is more closely related to the haplotype at the other control region within in the individual than to any other haplotype, and these haplotypes differ by only one nucleotide.

Using the equation:

(#gene conversion events/509bp/my) = (#mutations/509bp/my)/
(#individuals with different control regions/
#individuals with identical control regions)

the frequency of gene conversion events can be estimated. This equation assumes that gene conversion events are approximately equal to the number of mutational events, and these conversion events are not the cause of the mutations observed in the control region fragment (mutation and conversion are independent). This equation suggests that the sequence between these two control regions is being homogenized at a rate that is approximately 1/10 that of the mutation rate in the region being studied. Using a mutation rate of 1.32% which was estimated for reptilian mtDNA by Zamudio and Greene (1997), gene conversion between sequences is estimated to occur approximately once every 1.2 My. However, is has been suggested that snake mitochondrial DNAs have undergone high rates of molecular evolution (Kumazawa *et al.* 1998), so gene conversion could occur much more frequently in snakes if the mutation rate is higher. For example, Eberhard *et al.* (2001) estimated that conversion events occur only between functional portions of the parrot control region every 34 670+/- 18 400 years.

Mechanisms of Sequence Homogenization

The mechanism of concerted evolution between the two control regions in eastern massasauga rattlesnakes is likely associated with tRNA genes. All vertebrate mitochondrial gene rearrangements involve genes for tRNAs, which are stem-and-loop structures. All rearranged segments begin with a tRNA gene with respect to the direction of light-strand elongation and typical vertebrate gene order, except in birds (Macey *et al.* 1997). With the exception of the lamprey, tuatara, and ranid frogs, all rearranged segments end with a tRNA gene (Macey *et al.* 1997). Many of the duplications in lizards have boundaries that are at or near tRNA genes, or potential stem-and-loop structures (Stanton *et al.* 1994, Moritz and Brown 1987). This evidence suggests that duplications in mtDNA may be mediated by stem-and-loop structures.

All essential elements necessary for transcriptional initiation are located within the confines of the control region, including the origin of H-strand synthesis (O_H) (Clayton 1992). The origin of L-strand synthesis (O_L), however, can be found in different locations, for example in sites close to the major H-strand start site for both the human and *Xenopus* mtDNA genomes (Bogenhagen *et al.* 1986), and in the WANCY tRNA cluster (Clayton 1992). In the chicken, the O_L sequence has been lost, or was never present, but the organization of the WANCY cluster has been preserved, so some feature of the tRNA gene sequences may substitute for the O_L (Desjardins and Morais 1990). Indeed, most lineages found to lack the O_L in the typical vertebrate position are known to contain gene rearrangements (Macey *et al.* 1997). Wong and Clayton (1985, 1986) recognized that the enzyme that initiates light-strand replication appears to be signaled by

a stem-and-loop structure, so these structures can signal the start of replication of mitochondrial genomes. When bacteriophages, and other autonomous genetic elements, such as the *Haemophilus influenzae* phage HP1c1, integrate into the bacterial chromosome, target sites for site-specific integration are often putative tRNA genes, suggesting that these tRNAs may have served as recognition sequences for a primordial recombination enzyme (Rieter *et al.* 1989).

Models have been proposed to account for the concerted evolution of two control regions in the mitochondria that involve the duplication of the control region, followed by gene conversion. The duplication model was suggested by Macey *et al.* 1997, and involves displacement of the O_L. A mutational event disrupts the O_L, and an alternative site is then used for initiation of replication of the light strand. A period of instability follows in which light-strand replication occurs at positions having stem-and-loop structures (e.g. tRNAs). Then, light-strand synthesis copies the heavy strand, and after heavy-strand replication has finished, the 5' end of the nascent light strand becomes detached, and slips ahead to an alternative stem-and-loop structure. Light-strand synthesis is completed with termination occurring past the point of initiation, which often appears to be a different tRNA gene. As initiation and termination are at different sites, a tandem duplication is created. A new site for initiation and termination of the light strand becomes stabilized, preventing a recurrence of the duplication. Multiple deletions then occur, creating a gene rearrangement (Macey *et al.* 1997).

Two models of gene conversion have been proposed, the first involving a tandem duplication of the region surrounding the two control regions due to slipped-strand mispairing, giving rise to either a larger mtDNA with three control regions, or a normal-sized mtDNA with two homogeneous control regions (Moritz and Brown 1987). Most duplicated sequences are within the region that is single-stranded for a long period during mtDNA replication. It may be that tRNA genes fold while single-stranded to provide secondary structures that act as signals for duplications (Moritz and Brown 1987). Kumazawa *et al.* (1998) detected a proportion of larger mtDNA molecules, potentially generated by tandem duplications, which is in agreement with this model.

Another model of frequent gene conversion has been suggested in which the crossing over of nicked strands between two control regions within a mtDNA molecule leads to formation of a Holliday structure, and sequence of one control region, and possibly flanking tRNA genes, may be replaced by that of the other via repair of heteroduplex DNA intermediates (Kumazawa *et al.* 1998). Thyagarajan *et al.* (1996) demonstrated that mitochondrial protein extracted from mammalian cells could catalyze homologous recombination of plasmid DNA in order to repair DNA damages. This homologous repair pathway may generate only gene conversion events, and may be part of a topoisomerase/resolvase complex that separates daughter monomers at the termination of replication, and introduces superhelical turns into these monomers (Thyagarajan *et al.* 1996). Johnson and Jasin (2001) have suggested that homologous recombination may be a prominent DNA double-strand break (DSB) repair pathway in mammalian cells. In this example of gene conversion, the information from the donor

sequence is copied into the broken locus, making the repaired locus an exact copy of the donor sequence (Johnson and Jasin 2001). It has also been shown that DNA damage induces gene conversion of at MHC class II genes, suggesting that some kind of DNA-nicking activity may be needed for gene conversion to take place (Högstrand and Böhme 1999).

Thus the most likely mode for homogenization of the two control region sequences in snakes is duplication, followed by frequent gene conversion. Displacement of the O_L or slipped-strand mispairing could have lead to the duplication and rearrangement of the control region and $tRNA^{Leu}$ and $tRNA^{Pro}$. The presence of two control regions and the duplicated tRNAs could allow for gene conversion to occur, as one of the prerequisites for gene conversion to occur is the existence of two sequences with a very high degree of similarity (Högstrand and Böhme 1999). Differences between the sequences found at the two control regions indicates that this gene conversion does not occur during every cycle of replication, but only sporadically, perhaps as a repair response to DNA damage at one of the sequences.

Reasons for Maintenance of Two Functional Control Regions

Eastern massasauga rattlesnakes have apparently maintained two functional control region copies for tens of millions of years, and this is also the case in species, such as the cattle tick, where selection pressure for compact genome seems to be very strong (Campbell and Barker 1999). There have been several hypotheses put forward to explain the maintenance of two functional control region in the mitochondrial genome. Duplicate control regions may provide a replicative advantage compared to molecules

with only one control region in that these dual regions could enhance the efficiency of mitochondrial transcription and/or replication. Mitochondrial replication is known to be slow; the entire replication cycle is completed in approximately two hours, and the overall rate of polymerization is about 270 nucleotides per minute per strand, one of the slowest in vivo DNA replication rates reported to date (Clayton 1982). Perhaps two control regions could accelerate the mitochondrial replication rate. Another possible function of the second control region could involved transcriptional control of the adjacent tRNA genes. Initiation or control of transcription may be required from both sites in order to transcribe all essential tRNAs (Arndt and Smith 1998).

Implications

As mitochondrial DNA is often used for phylogenetic analysis, in part due to the fact that it is considered to be inherited as a single, non-recombining molecule, gene conversion in the mitochondrial control region could have several implications, particularly in the areas of population and evolutionary biology. If intragenic recombination occurs, different parts of sequence can have different phylogenetic histories (Scierup and Hein 2000a). Ignoring recombination in analysis can lead to an overestimation of the number of mutations, and loss of a molecular clock, as has been suggested for many data sets of viruses (Scierup and Hein 2000b). In the future, an examination of the effects of gene conversion on phylogeographic studies of species structure would be beneficial.

LITERATURE CITED

Arndt, A., and M.J. Smith. 1998. Mitochondrial gene rearrangements in the sea cucumber genus *Cucumaria*. Molecular Biology and Evolution 15: 1009-1016.

Awadalla, P., A. Eyre-Walker, and J.M. Smith. 1999. Linkage disequilibrium and recombination in hominid mitochondrial DNA. *Science* 286: 2524-2525.

Bogenhagen, D.F., B.K. Yoza, and S.S. Cairns. 1986. Identification of initiation sites for transcription of *Xenopus laevis* mitochondrial DNA. *The Journal of Biological Chemistry* 261: 8488-8494.

Brown, W.M. 1983. Evolution of animal mitochondrial DNA. In: *Evolution of Genes and Proteins* (eds. Nei, M. and R.K. Koehn), Pp 62-88. Sinauer Associates Inc., Sunderland, Massachusetts.

Cabot, E. and A.T. Beckenbach. 1989. Simultaneous editing of multiple nucleic acid and protein sequences with Esee. *Computing and Applied Biosystems* 5: 233-234

Campbell, H.J.H., and S.C. Barker. 1999. The Novel mitochondrial gene arrangement of the cattle tick, *Boophilus microplus*: fivefold tandem repetition of a coding region. *Molecular Biology and Evolution* 16: 732-740.

Charlesworth, D., B. Charlesworth, and G.A.T. McVean. 2001. Genome sequences and evolutionary biology, a two-way interaction. Trend in Ecology and Evolution 16: 235-242.

Clayton, D.A. 1982. Replication of animal mitochondrial DNA. *Cell* 28: 693-705.

Clayton, D.A. 1992. Transcription and replication of animal mitochondrial DNAs. *International Review of Cytology* 141: 217-232.

Desjardins, P. and R. Morais. 1990. Sequence and gene organization of the chicken mitochondrial genome: a novel gene order in higher vertebrates. *Journal of Molecular Biology* 212: 599-634.

Eberhard, J.R., T.F. Wright, and E. Bermingham. 2001. Duplication and concerted evolution of the mitochondrial control region in the parrot genus *Amazona. Molecular Biology and Evolution* 18: 1330-1342.

Elson, J.L., R.M. Andrews, P.F. Chinnery, R.N. Lightowlers, D.M. Turnbull, and N. Howard. 2001. Analysis of European mtDNAs for recombination. *American Journal of Human Genetics* 68: 145-153.

Felsenstein, J. 1989. PHYLIP -- Phylogeny Inference Package (Version 3.2). *Cladistics* 5: 164-166.

Högstrand, K, and J. Böhme. 1999. Gene conversion can create new MHC alleles. *Immunological Reviews* 167: 305-317.

Howell, N. 1997. MtDNA recombination: what do in vitro data mean? *American Journal of Human Genetics* 61: 18-22.

Johnson, R.D., and M. Jasin. 2001. Double-strand-break-induced homologous recombination in mammalian cells. *Biochemical Society Transactions* 29: 196-201.

Kumazawa, Y., H. Ota, M. Nishida, and T. Ozawa. 1996. Gene rearrangements in snake mitochondrial genomes: highly concerted evolution of control-region-like sequences duplicated and inserted into a tRNA gene cluster. *Molecular Biology and Evolution* 13: 1242-1254.

Kumazawa, Y., H. Ota, M. Nishida, and T. Ozawa. 1998. Complete nucleotide sequence of a snake (*Dinodon semicarinatus*) mitochondrial genome with two identical control regions. *Genetics* 150: 313-329.

Ladoukakis, E.D., and E. Zouros. 2001. Direct evidence for homologous recombination in mussel (*Mytilus galloprovincialis*) mitochondrial DNA. *Molecular Biology and Evolution* 18: 1168-1175.

Lunt, D.H., and B.C. Hyman. 1997. Animal mitochondrial DNA recombination. *Nature* 387: 247.

Macey, J.R., A. Larson, N.B. Ananjeva, Z. Fand, and T.J. Papenfuss. 1997. Two novel gene orders and the role of light-strand replication in rearrangement of the vertebrate mitochondrial genome. *Molecular Biology and Evolution* 14: 91-104.

Macey, J.R., J.A. Schulte II, A. Larson, B.S. Tuniyev, N. Orlov, and T.J. Papenfuss. 1999. Molecular phylogenetics, tRNA evolution, and historical biogeography in anguid lizards and related taxonomic families. *Molecular Phylogenetics and Evolution* 12: 250-272.

Maynard Smith, J. 1977. Why the genome does not congeal. *Nature* 268: 693-696.

Mertens King, L. 1998. The role of gene conversion in determining sequence variation and divergence in the *Est-5* gene family in Drosophila pseudoobscura. *Genetics* 148: 305-315.

Moritz, C., and W.M. Brown. 1987. Tandem duplications in animal mitochondrial DNAs: variation in incidence and gene content among lizards. *Proceedings of the National Academy of Sciences of the United States of America* 84: 7183-7187.

Ohta, T. 1997. Role of gene conversion in generating polymorphisms at major histocompatibility complex loci. *Hereditas* 127: 97-103.

Rand, D.M., and R.G. Harrison. 1986. Mitochondrial DNA transmission genetics in crickets. *Genetics* 114: 955-970.

Rieter, W.D., P. Palm, and S. Yeats. 1989. Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. *Nucleic Acids Research* 17: 1907-1914.

Schneider, S., D. Roessli, and L. Excoffier. 2000. Arlequin: a software for population genetics data analysis Vers 2.00. Genetics and Biometry Lab, Department of Anthropology, University of Geneva.

Schierup, M.H., and J. Hein. 2000a. Consequences of recombination on traditional phylogenetic analysis. *Genetics* 156: 879-891.

Scierup, M.H., and J. Hein. 2000b. Recombination and the molecular clock. *Molecular Biology and Evolution* 17: 1578-1579.

Solignac, M., M. Monnerot, and J.C. Mounolou. 1986. Concerted evolution of sequence repeats in *Drosophila* mitochondrial DNA. *Journal of Molecular Evolution* 24: 53-60.

Stanton, D.J., L.L. Daehler, C.C. Moritz, and W.M. Brown. 1994. Sequences with the potential to form stem-and loop structures are associated with coding-region duplications in animal mitochondrial DNA. *Genetics* 137: 233-241.

Surdhar, G.K., M.S. Enayat, S. Lawson, M.D. Williams, and F.G.H. Hill. 2001. Homozygous gene conversion in von Willebrand factor gene as a cause of type 3 von Willebrand disease and predisposition to inhibitor development. *Blood* 98: 248-249.

Thyagarajan, B., R.A. Padua, and C. Campbell. 1996. Mammalian mitochondria possess homologous DNA recombination activity. *The Journal of Biological Chemistry* 271: 27536-27543.

Wang, S., C. Magoulas, and D. Hickey. 1999. Concerted evolution within a trypsin gene cluster in *Drosophila*. *Molecular Biology and Evolution* 16: 1117-1124.

Wong, T.W., and D.A. Clayton. 1985. *In vitro* replication of human mitochondrial DNA: accurate initiation at the origin of light-strand synthesis. *Cell* 42: 951-958.

Wong, T.W., and D.A. Clayton. 1986. DNA primase of human mitochondria is associated with structural RNA that is essential for enzymatic activity. Cell 45: 817-825.

Zamudio, K.R., and H.W. Greene. 1997. Phylogeography of the bushmaster (*Lachesis muta*: Viperidae): implications for neotropical biogeography, systematics, and conservation. *Biological Journal of the Linnean Society* 62: 421-442.

Zevering, C.E., C. Moritz, A. Heideman, and R.A. Sturm. 1991. Parallel origins of duplications and the formation of pseudogenes in mitochondrial DNA from parthenogenetic lizards (*Heteronotia binoei*: Gekkonidae). *Journal of Molecular Evolution* 33: 431-441.

Zimmer, E.A., S.L. Martin, S.M. Beverley, Y.W. Kan, and A.C. Wilson. 1980. Rapid duplication and loss of genes coding for the α chains of hemoglobin. *Proceedings of the National Academy of Sciences of the United States of America* 77: 2158-2162.

Table 1. Sequences of primers used in this study. H and L refer to heavy and light strands, respectively. Primers were created in this study unless indicated in footnotes.

Primers	Sequence (5'→3')
L14919[a]	AACCACCGTTGTA/TAA/TTCAACT
Phe1H[a]	CCATCTTAGCATC/TTTCAGTGCTA
CSB2L[b]	GGCTAATAAGCGGGAATTCTACTA
CSB1L[b]	TCTCTTAATGCTTGTTATACATA
CSB2H[b]	GTAAAAAAATGTAGTTACAGCGGG
12SAL[b]	AAACTGGGATTAGATACCCCACTAT
12SBH[b]	AGAGGGTGACGGGCGGTGTGT
L4221[c]	GGTATGGGCCCGATAGCTT
H4419[c]	AGGGACTACCTTGATAGAGT
Cr1500H	GGTTTGAAGTGGATAGCCATGG
Cr1703L	CGCACCCTTTATCCTGGTACATTA
CrIIH	AAGATCATTTCCCGGATACA

[a] R. Lawson, unpubl.
[b] L. Collins, unpubl.
[c] Macey *et al.* (1999)

Table 2. Variable sites in the haplotypes of control regions 1 and 2 of eastern massasauga rattlesnakes. Dots denote identity with haploytpe 1.

	Variable Site Location
	11245
	269950
Haplotype	735689
HAP1	CCTACA
HAP2	G.....
HAP3	...T..
HAP4	.T....
HAP5	GT..T.
HAP6	G.A...
HAP7	.....T
HAP8	....T.

Table 3. 509 bp haplotypes found at control regions one and two within individuals.

Individual	Control Region 1 Haplotype	Control Region 2 Haplotype
6	1	1
33	2	2
37	1	1
39	3	3
40	3	3
41	3	3
79	1	1
86*	1	4
91	2	2
107	1	1
111	1	1
130*	1	8
145	1	1
147	5	5
150	5	5
151	5	5
155	5	5
156	6	6
158	6	6
159	6	6
162	2	2
165*	7	1
166	1	1
253	1	1
305	1	1
310	1	1
311	1	1
312	1	1

*individuals with different CR1 and CR2 sequence

FIGURE CAPTIONS

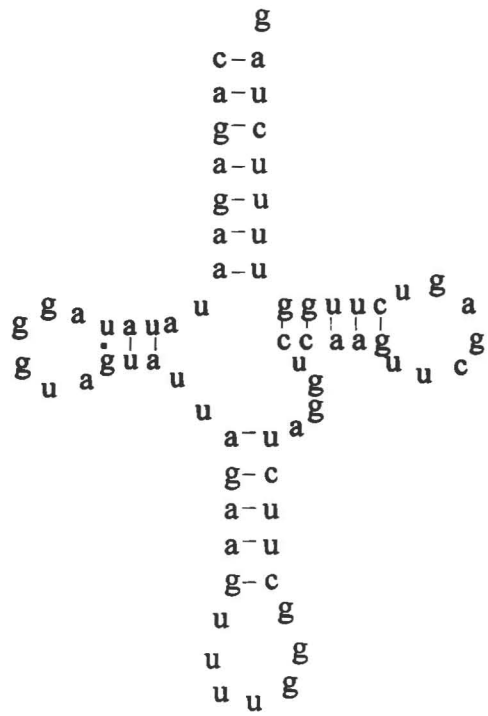
Figure 1. Putative secondary structure of four of the mitochondrial tRNA genes of *S. catenatus*. The tRNA genes are labeled with the three-letter abbreviation of the amino acid their transcripts transfer, and their full name. Arms in the tRNAs (clockwise from top): amino acid acceptor arm (AA), TC arm (T arm), anticodon arm (AC arm), and dihydrourine arm (DHU or D arm). $tRNA^{Thr}$ may be nonfunctional.

Figure 2. Nucleotide sequences of snakes for cytb-12S rRNA (CR1) and IQM-related (CR2) regions. All the nucleotides shown correspond to the light strand. Structural features and gene regions are denoted and/or underlined. Dashes denote missing sequence. Species abbreviations: SCC, eastern massasauga rattlesnake (*Sistrurus catenatus catenatus*), SCE, desert massasauga rattlesnake (*Sistrurus catenatus edwarsii*).

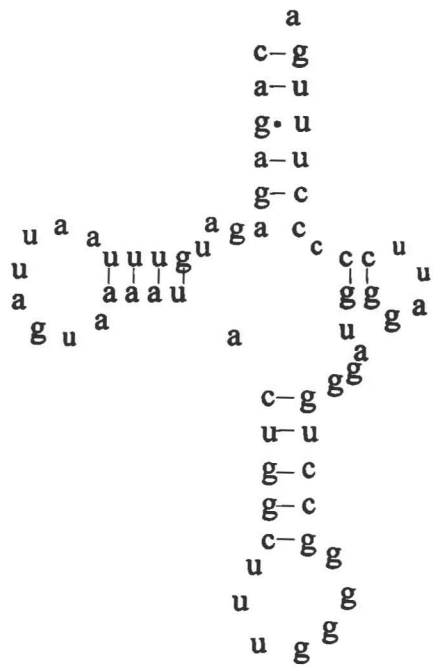
Figure 3. Aligned, pictorial representation of the sequences given in Figure 2. The location of the 509 bp fragment sequenced for 28 individuals is shown.

Figure 4. Minimum-spanning network for CR1 and CR2 mtDNA haplotypes. The number of substitutions between haplotypes is given next to the branches. Dashed lines and arrows denote the fact that three minimum-spanning networks were found, each differing in the placement of haplotype 5.

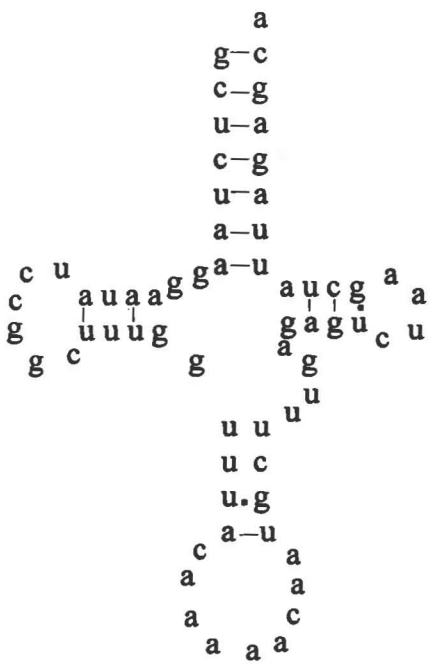
tRNA Gln (Glutamine)



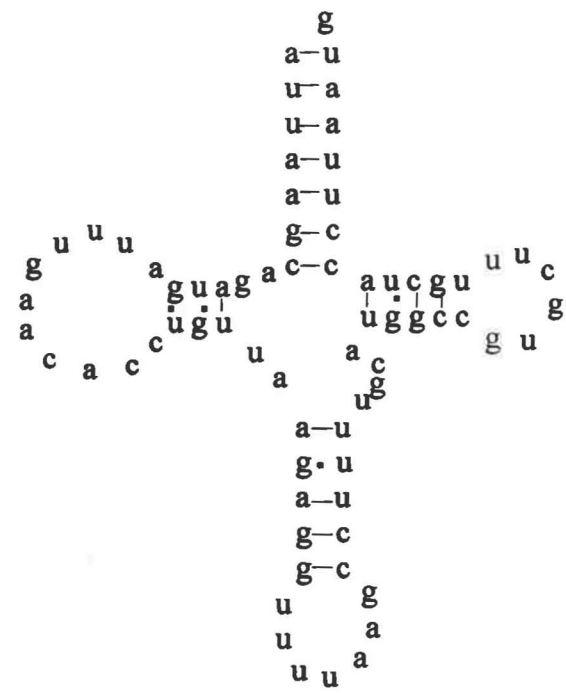
tRNA Pro (Proline)



tRNA (Threonine)



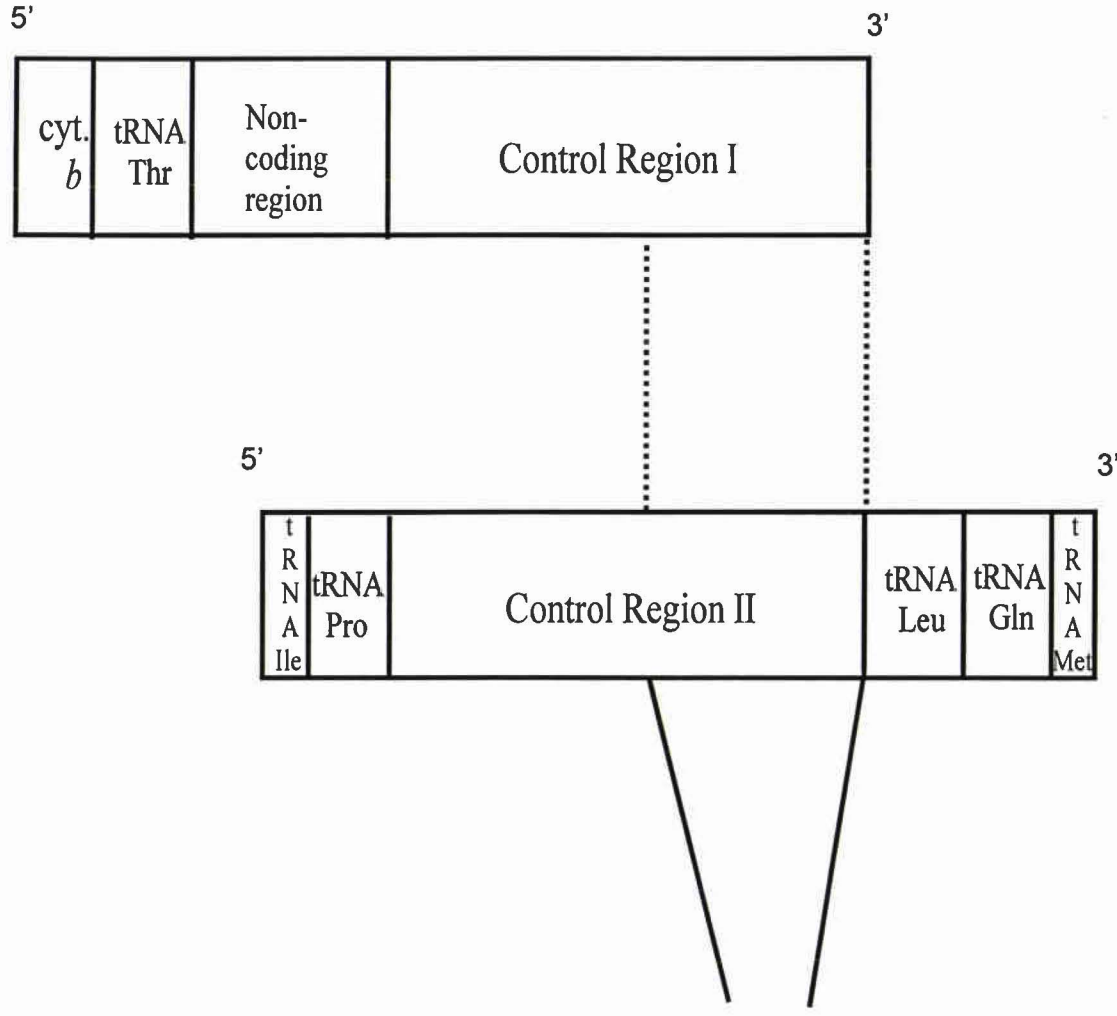
tRNA Leu (Leucine)

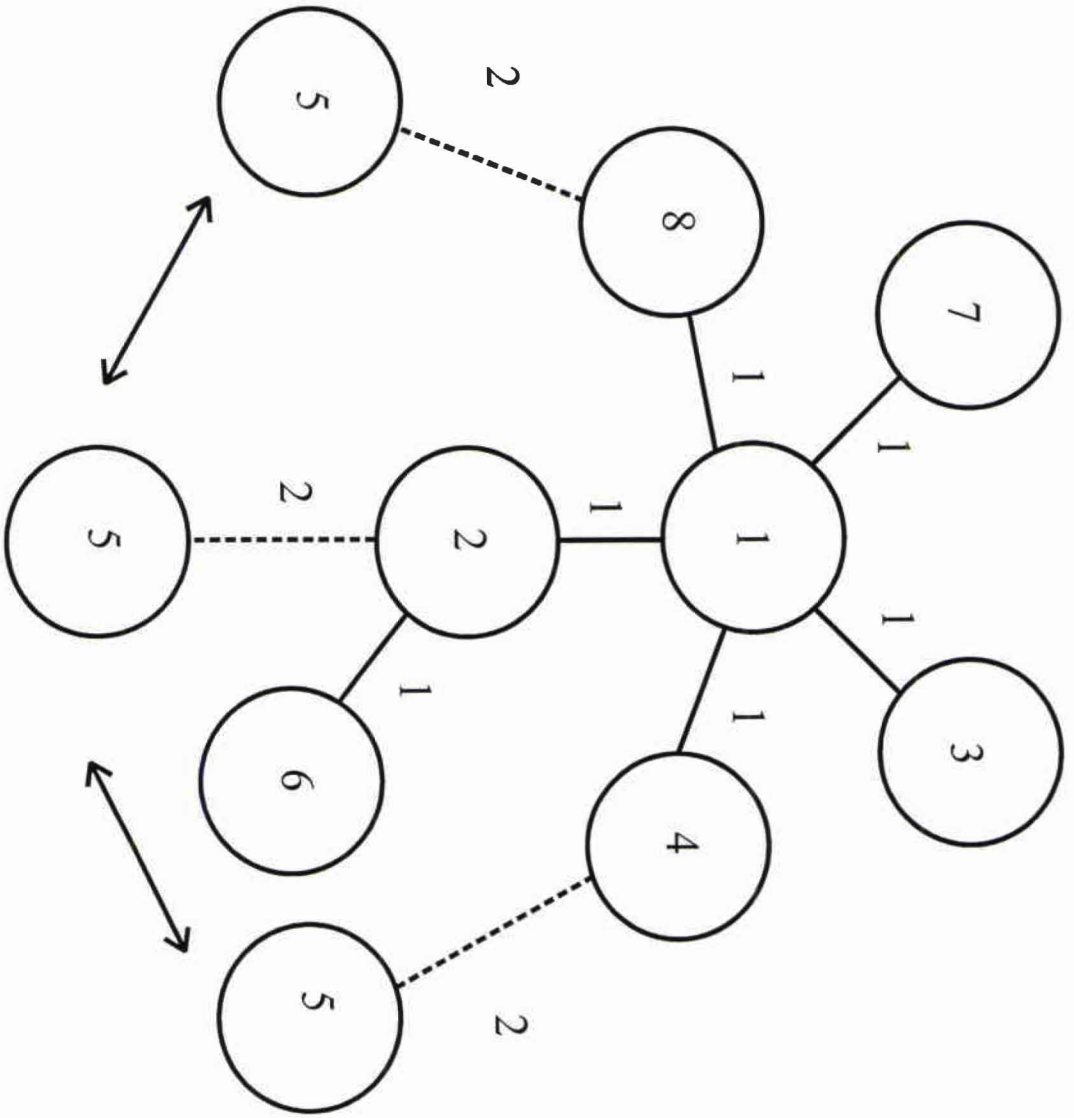


```
          ←cytb
SCC CR1   CAGCCACCAAACCAGTAGAACCCCCATTCACAGAAATCGGCCAACTAGCCTCTATCCTAT
SCC CR2
SCE CR1   CAGCCACTAAACCAGTAGAACCCCCATTCACAGAAATTGGTCAACTAGCCTCTATCCTAT


SCC CR1   ACTTCATATTCTTTATAACAAACCCCCTACTAGGCCTACTAGAAAACAAAATCTCAGACC
SCC CR2
SCE CR1   ACTTCATATTCTTCATAACAAACCCACTACTAGGCCTAGTAGAAAACAAAATCTCAAGTC


        ←cytb¬ ⌐tRNA (Thr)→
SCC CR1   TCACCTGCTCTAATAGCTTAGACTCTCAAAGCATTGTTTTTGTAAACCAAAGCCGGATAT
SCC CR2
SCE CR1   TCAAATGCTCTAATAGCTTAGACTCTCAAAGCATTGTTTTTGTAAACCAAAGCCGGATAT


        ←tRNA (Thr) ¬ ⌐ non-coding region→
SCC CR1   TCCTTAGAGCAATCGACTTCCACTTTACTATATTTTAAACCCCCAAAAACCACTATCTGG
SCC CR2
SCE CR1   TCCTTAGAGCAATCAACTTCTACTTTACTACCCTCAACCCCTTCACAATAACCCATCCCT


←non-coding region¬
SCC CR1   G
SCC CR2
SCE CR1   GGCCCCCAAAGCCAGTAAAAAACTAATTAAACTACTCTCTGAAATAAACCAATAATACTC
           partial tRNA (Pro)     └repeat 1→


SCC CR1
SCC CR2
SCE CR1   TCCTGGACACCCCAACCAGTATTTTACTAATTAAACTACTCTCTGAAATAAACCAATAAT
                          ←repeat 1┘└ repeat 2→

SCC CR1                                                  ←tRNA (Ile)┐┌
SCC CR2                                                   GCGACCCACTTCCTTC
SCE CR1   ACTCTCCTGGACACCCCAACCAGTATTTTACTAATTAAACTACTCTCTGAAATAAACCAA
                                 ←repeat 2┘└ repeat 3→

SCC CR1
SCC CR2   TATCAAAGAGGGAATCCCATCCCAGGCCCCCAAAGCCAGTATTTTACTAATTAAACTACT
SCE CR1   TAATACTCTCCTGGACACCCCCCTACCCCCCCCAACCAGTATTTTACTAATTTAACTACC
                                                  ←repeat 3┘

  ←tRNA (Pro)¬ ⌐CR-like sequence→                  ⌐C-rich  sequence¬
SCC CR1         GACCCAAGCCAAAAAAATACTCTCCTAGGACCCCCCCCTACCCCCCATTAATTTG
SCC CR2   CTCTGAAATTAAGCCAAAAAAATACTCTCCTAGGACCCCCCCCTACCCCCCATTAATTTG
SCE CR1   CTCTGAA GTTAAAGCAAAAAATACTCTCCTAGGACCCCCCCCTACCCCCCATTAATTTG
               partial  tRNA (Pro)

           hairpin 1                  hairpin 2
```

```
SCC CR1  GGTCCCGAATTCGGCCTTATATGTACTCTTTACATATAGGGTCCTCATTGTCGCTATGTA
SCC CR2  GGTCCCGAATTCGGCCTTATATGTACTCTTTACATATAGGGTCCTCATTGTCGCTATGTA
SCE CR1  GGTCCCGAATTCGGCCT----------------------------------------

           hairpin 3
SCC CR1  TAATAATACATTAATCGTTTTGCCCCATGGCTAATAAGCGGGAATTCTACTATAATTAAA
SCC CR2  TAATAATACATTAATCGTTTTGCCCCATGGCTAATAAGCGGGAATTCTACTATAATTAAA
SCE CR1  ----------------------------------------------------------A

SCC CR1  TATATACAAAACTGGCTCATTAACATCACTTCCTCTCCTCATTTTCTGGTCGTTCCATTT
SCC CR2  TATATACAAAACTGGCTCATTAACATCACTTCCTCTCCTCATTTTCTGGTCGTTCCATTT
SCE CR1  TATATACAAAACTGGGTCATTAACATTACTTCCTTTCCTCATTTTCTGGTCGTTCCATTT

SCC CR1  AAAATAGGCTGTCCCTTATTAGTAACCATGGCTATCCACTTCAAACCGGTGTCCCGTGAT
SCC CR2  AAAATAGGCTGTCCCTTATTAGTAACCATGGCTATCCACTTCAAACCGGTGTCCCGTGAT
SCE CR1  AACAAAGGCTGTTCCTTCTTAGTAACCATGGCTATCCACTTCAAACCGGTGTCCCGTGAT

SCC CR1  TTAACCCTTCCCGTGAAATCCTCTATCCTTTCACCTCAGGCATACAGTCCCGCTTCTCAC
SCC CR2  TTAACCCTTCCCGTGAAATCCTCTATCCTTTCACCTCAGGCATACAGTCCCGCTTCTCAC
SCE CR1  TTAACCCTTCCCGTGAAATCCTCTATCCTTTCACCTCAGGCATCAAGTCCCGCTTCTCAC

                                          repetitive sequence 1
SCC CR1  GTCCATATATTGTAACTCCTCCCGTTTATGCTCTTTCCAAGGCCGCTGGTTACACCTTCA
SCC CR2  GTCCATATATTGTAACTCCTCCCGTTTATGCTCTTTCCAAGGCCGCTGGTTACACCTTCA
SCE CR1  GTCCATATATTGTAACTCCTCCCGTTTATGTCCTTTCCAAGGCCGCTGGTTACACCTTCA

                                                repetitive sequence 2
SCC CR1  AGGGCATCTCAATGGTCCGGAACCACCCCGCCTTACTTGCTCTTTCCAAGGCCTATGGTC
SCC CR2  AGGGCATCTCAATGGTCCGGAACCCCCCCGCCTTACTTGCTCTTTCCAAGGCCTATGGTC
SCE CR1  AGGGCATCTCAATGGTCCGGAACCACCCCGCCTTACTTGCTCTTTCCAAGGCCTATGGTC

SCC CR1  GCACCCTTTATCCTGGTACATTAAGTCTCATGTTCTTATCACGTATGCTTGTTCCACCCC
SCC CR2  GCACCCTTTATCCTGGTACATTAAGTCTCATGTTCTTATCACGTATGCTTGTTCCACCCC
SCE CR1  GCACCCTTTATCCTGGTACATTAAGTCTCATGTTCTTATCACGTATGCTTGTTCCACCCC

SCC CR1  TGGTTGGCTTTTTTATCGGTACCTTTCACCTGACACCCATATATGCTCGTTACCGTCCCC
SCC CR2  TGGTTGGCTTTTTTATCGGTACCTTTCACCTGACACCCATATATGCTCGTTACCGTCCCC
SCE CR1  TGGTTGGCTTTTTTATCGGTACCTTTCACCTGACACCCATATATGCTCGTTACCGTCCCC

SCC CR1  CTCACCGGGGTAGACCTCTAGTCCAGGTGGAGCTATGTTCTTGGTCTGGCACTTTCCCCT
SCC CR2  CTCACCGGGGTAGACCTCTAGTCCAGGTGGAGCTATGTTCTTGGTCTGGCACTTTCCCCT
SCE CR1  CTCGCCGGGGTAGACCTCTAGTCCAGGTGGAGCTATGTTCTTGGTCTGGCACTTTCCCCT

                         conserved sequence block 1
SCC CR1  ATAGGGATACATCTCTTCATGCTCGTTATACATATCGTTCTACATTCCTAGAAAATTCCA
SCC CR2  ATAGGGATACATCTCTTCATGCTCGTTATACATATCGTTCTACATTCCTAGAAAATTCCA
SCE CR1  ATAGGGATACATCTCTTAATGCTCGTTATACATACTTTACTACGTTCCTAAAAAATTTCA
```

```
SCC CR1  TTATTTTTTATTATAAAAATCCCGCTGTAACTACATTTTTTTACCCTGTTTTTTTATTTT
SCC CR2  TTATTTTTTATTATAAAAATCCCGCTGTAACTACATTTTTTTACCCTGTTTTTTTATTTT
SCE CR1  TTATTTTTTATTATAAAAATTCCGCTGTTACTACATTTTTTTTACCCATTTTTTTATTTT


                                                conserved sequence block 3
SCC CR1  TCACCAAAATTAATCCCACTTTCGTATACTAAAAATATTAACCCGAAATAAAACAATCTT
SCC CR2  TCACCAAAATTAATCCCACTTTCGTATACTAAAAATATTAACCCGAAATAAAACAATCTT
SCE CR1  TTACCAAAATTAATCCCACTTTCGTATACTAAAAAAATTAACCCGAAATAAAACAATATT


SCC CR1  TTTGCTCGGTGATTTTATTTTTGCCGCCCCGTGAAAAAAAATTCAAATAAATATAAAAAT
SCC CR2  TTTGCTCGGTGATTTTATTTTTGCCGCCCCGTGAAAAAAAATTCAAATAAATATAAAAAT
SCE CR1  ATTTTGCTCGGAGTTTTTATTTCTGCCACCGTGAAAAAAAAATTTGATAAATATAAAAAT


SCC CR1  CGCCCACCCTATTTTTGTATCCGGGAAATGATCTTCTTCCCTCGGAGGGGACGTTTCACC
SCC CR2  CGCCCACCCTATTTTTGTATCCGGGAAATGATCTTCTTCCCTCGGAGGGGACGTTTCACC
SCE CR1  CGCCCACCCTATTTTTGTATCCGGGAAATAGTCTTCTTCCCTCGGAGGGCACGTTTCACC


SCC CR1  GAGGCTGAGGGCCGAAATCAGCACAGCCGAATTTATCTTTATTTTTTAACCTCTCCAGAA
SCC CR2  GAGGCTGAGGGCCGAAATCAGCACAGCCGAATTTATCTTTATTTTTTAACCTCTCCAGAA
SCE CR1  GGGGCTGAGGGCCGAAATCAGCACAGCCGG TTTGTCTTTATTTTTTATCATGTCCGAGA


 ←CR-like sequence┐ ┌tRNA (Leu)→
SCC CR1  TTTTTATA
SCC CR2  TTTTTATACATTAAGGTAGCAAAGCACGGCCATGCAAAAGGCTTAAAACCTCTTAACAGG
SCE CR1  TTTTTTAT


                    ←tRNA (Leu)┐ ┌tRNA (Gln)→
SCC CR1
SCC CR2  TGTTCAAATCATCTCCTTAATACTAGAAAACCAAGACTCGAACTTGGACCTAGAAGCCCA
SCE CR1


                                  ←tRNA (Gln)┐ ┌tRNA (Met)→
SCC CR1
SCC CR2  AAACTTCTAATACTACCCATAATATTTTCTAAGTAAAGTCAGCTAATA
SCE CR1
```


5'
3'
cyt. b
tRNA Thr
Non-coding region
Control Region I
5'
3'
t R N A Ile
tRNA Pro
Control Region II
tRNA Leu
tRNA Gln
t R N A Met
Approximate location of 509 bp fragment sequenced for 28 individuals


5
2
8
7
1
1
5
2
2
1
1
1
1
6
1
4
3
5
2

Appendix 1. Sequence for the eight 509bp control region haplotypes found for both CR1 and CR2 of 28 eastern massasauga rattlesnakes.

```
HAP1          1 CGTATGCTTGTTCCACCCCTGGTTGGCTTTTTTATCGGTACCTTTCACCTGACACCCATA 60
HAP2          1 CGTATGCTTGTTCCACCCCTGGTTGGGTTTTTTATCGGTACCTTTCACCTGACACCCATA
HAP3          1 CGTATGCTTGTTCCACCCCTGGTTGGCTTTTTTATCGGTACCTTTCACCTGACACCCATA
HAP4          1 CGTATGCTTGTTCCACCCCTGGTTGGCTTTTTTATCGGTACCTTTCACCTGACACCCATA
HAP5          1 CGTATGCTTGTTCCACCCCTGGTTGGGTTTTTTATCGGTACCTTTCACCTGACACCCATA
HAP6          1 CGTATGCTTGTTCCACCCCTGGTTGGGTTTTTTATCGGTACCTTTCACCTGACACCCATA
HAP7          1 CGTATGCTTGTTCCACCCCTGGTTGGCTTTTTTATCGGTACCTTTCACCTGACACCCATA
HAP8          1 CGTATGCTTGTTCCACCCCTGGTTGGCTTTTTTATCGGTACCTTTCACCTGACACCCATA

HAP1         61 TATGCTCGTTACCGTCCCCCTCACCGGGGTAGACCTCTAGTCCAGGTGGAGCTATGTTCT 120
HAP2         61 TATGCTCGTTACCGTCCCCCTCACCGGGGTAGACCTCTAGTCCAGGTGGAGCTATGTTCT
HAP3         61 TATGCTCGTTACCGTCCCCCTCACCGGGGTAGACCTCTAGTCCAGGTGGAGCTATGTTCT
HAP4         61 TATGCTCGTTACCGTCCCCCTCACCGGGGTAGACCTCTAGTCCAGGTGGAGCTATGTTCT
HAP5         61 TATGCTCGTTACCGTCCCCCTCACCGGGGTAGACCTCTAGTCCAGGTGGAGCTATGTTCT
HAP6         61 TATGCTCGTTACCGTCCCCCTCACCGGGGTAGACCTCTAGTCCAGGTGGAGCTATGTTCT
HAP7         61 TATGCTCGTTACCGTCCCCCTCACCGGGGTAGACCTCTAGTCCAGGTGGAGCTATGTTCT
HAP8         61 TATGCTCGTTACCGTCCCCCTCACCGGGGTAGACCTCTAGTCCAGGTGGAGCTATGTTCT

HAP1        121 TGGTCTGGCACTTTCCCCTATAGGGATACATCTCTTCATGCTCGTTATACATATCGTTCT 180
HAP2        121 TGGTCTGGCACTTTCCCCTATAGGGATACATCTCTTCATGCTCGTTATACATATCGTTCT
HAP3        121 TGGTCTGGCACTTTCCCCTATAGGGATACATCTCTTCATGCTCGTTATACATATCGTTCT
HAP4        121 TGGTCTGGCACTTTCCCCTATAGGGATACATCTCTTCATGCTTGTTATACATATCGTTCT
HAP5        121 TGGTCTGGCACTTTCCCCTATAGGGATACATCTCTTCATGCTTGTTATACATATCGTTCT
HAP6        121 TGGTCTGGCACTTTCCCCTATAGGGATACATCTCTTCATGCTCGTTATACATATCGTTCT
HAP7        121 TGGTCTGGCACTTTCCCCTATAGGGATACATCTCTTCATGCTCGTTATACATATCGTTCT
HAP8        121 TGGTCTGGCACTTTCCCCTATAGGGATACATCTCTTCATGCTCGTTATACATATCGTTCT

HAP1        181 ACATTCCTAGAAAATTCCATTATTTTTTATTATAAAAATCCCGCTGTAACTACATTTTTT 240
HAP2        181 ACATTCCTAGAAAATTCCATTATTTTTTATTATAAAAATCCCGCTGTAACTACATTTTTT
HAP3        181 ACATTCCTAGAAAATTCCATTATTTTTTATTATAAAAATCCCGCTGTAACTACATTTTTT
HAP4        181 ACATTCCTAGAAAATTCCATTATTTTTTATTATAAAAATCCCGCTGTAACTACATTTTTT
HAP5        181 ACATTCCTAGAAAATTCCATTATTTTTTATTATAAAAATCCCGCTGTAACTACATTTTTT
HAP6        181 ACATTCCTAGAAAAATCCATTATTTTTTATTATAAAAATCCCGCTGTAACTACATTTTTT
HAP7        181 ACATTCCTAGAAAATTCCATTATTTTTTATTATAAAAATCCCGCTGTAACTACATTTTTT
HAP8        181 ACATTCCTAGAAAATTCCATTATTTTTTATTATAAAAATCCCGCTGTAACTACATTTTTT

HAP1        241 TACCCTGTTTTTTTATTTTTCACCAAAATTAATCCCACTTTCGTATACTAAAAATATTAA 300
HAP2        241 TACCCTGTTTTTTTATTTTTCACCAAAATTAATCCCACTTTCGTATACTAAAAATATTAA
HAP3        241 TACCCTGTTTTTTTATTTTTCACCAAAATTAATCCCACTTTCGTATACTAAAAATTTTAA
HAP4        241 TACCCTGTTTTTTTATTTTTCACCAAAATTAATCCCACTTTCGTATACTAAAAATATTAA
HAP5        241 TACCCTGTTTTTTTATTTTTCACCAAAATTAATCCCACTTTCGTATACTAAAAATATTAA
HAP6        241 TACCCTGTTTTTTTATTTTTCACCAAAATTAATCCCACTTTCGTATACTAAAAATATTAA
HAP7        241 TACCCTGTTTTTTTATTTTTCACCAAAATTAATCCCACTTTCGTATACTAAAAATATTAA
HAP8        241 TACCCTGTTTTTTTATTTTTCACCAAAATTAATCCCACTTTCGTATACTAAAAATATTAA
```

```
HAP1        301 CCCGAAATAAAACAATCTTTTTGCTCGGTGATTTTATTTTTGCCGCCCCGTGAAAAAAAA 360
HAP2        301 CCCGAAATAAAACAATCTTTTTGCTCGGTGATTTTATTTTTGCCGCCCCGTGAAAAAAAA
HAP3        301 CCCGAAATAAAACAATCTTTTTGCTCGGTGATTTTATTTTTGCCGCCCCGTGAAAAAAAA
HAP4        301 CCCGAAATAAAACAATCTTTTTGCTCGGTGATTTTATTTTTGCCGCCCCGTGAAAAAAAA
HAP5        301 CCCGAAATAAAACAATCTTTTTGCTCGGTGATTTTATTTTTGCCGCCCCGTGAAAAAAAA
HAP6        301 CCCGAAATAAAACAATCTTTTTGCTCGGTGATTTTATTTTTGCCGCCCCGTGAAAAAAAA
HAP7        301 CCCGAAATAAAACAATCTTTTTGCTCGGTGATTTTATTTTTGCCGCCCCGTGAAAAAAAA
HAP8        301 CCCGAAATAAAACAATCTTTTTGCTCGGTGATTTTATTTTTGCCGCCCCGTGAAAAAAAA

HAP1        361 TTCAAATAAATATAAAAATCGCCCACCCTATTTTTGTATCCGGGAAATGATCTTCTTCCC 420
HAP2        361 TTCAAATAAATATAAAAATCGCCCACCCTATTTTTGTATCCGGGAAATGATCTTCTTCCC
HAP3        361 TTCAAATAAATATAAAAATCGCCCACCCTATTTTTGTATCCGGGAAATGATCTTCTTCCC
HAP4        361 TTCAAATAAATATAAAAATCGCCCACCCTATTTTTGTATCCGGGAAATGATCTTCTTCCC
HAP5        361 TTCAAATAAATATAAAAATCGCCCACCCTATTTTTGTATCCGGGAAATGATCTTCTTCCC
HAP6        361 TTCAAATAAATATAAAAATCGCCCACCCTATTTTTGTATCCGGGAAATGATCTTCTTCCC
HAP7        361 TTCAAATAAATATAAAAATCGCCCACCCTATTTTTGTATCCGGGAAATGATCTTCTTCCC
HAP8        361 TTCAAATAAATATAAAAATCGCCCACCCTATTTTTGTATCCGGGAAATGATCTTCTTCCC

HAP1        421 TCGGAGGGGACGTTTCACCGAGGCTGAGGGCCGAAATCAGCACAGCCGAATTTATCTTTA 480
HAP2        421 TCGGAGGGGACGTTTCACCGAGGCTGAGGGCCGAAATCAGCACAGCCGAATTTATCTTTA
HAP3        421 TCGGAGGGGACGTTTCACCGAGGCTGAGGGCCGAAATCAGCACAGCCGAATTTATCTTTA
HAP4        421 TCGGAGGGGACGTTTCACCGAGGCTGAGGGCCGAAATCAGCACAGCCGAATTTATCTTTA
HAP5        421 TCGGAGGGGACGTTTCACCGAGGCTGAGGGCCGAAATTAGCACAGCCGAATTTATCTTTA
HAP6        421 TCGGAGGGGACGTTTCACCGAGGCTGAGGGCCGAAATCAGCACAGCCGAATTTATCTTTA
HAP7        421 TCGGAGGGGACGTTTCACCGAGGCTGAGGGCCGAAATCAGCACAGCCGAATTTATCTTTA
HAP8        421 TCGGAGGGGACGTTTCACCGAGGCTGAGGGCCGAAATTAGCACAGCCGAATTTATCTTTA

HAP1        481 TTTTTTAACCTCTCCAGAATTTTTATACA                               509
HAP2        481 TTTTTTAACCTCTCCAGAATTTTTATACA
HAP3        481 TTTTTTAACCTCTCCAGAATTTTTATACA
HAP4        481 TTTTTTAACCTCTCCAGAATTTTTATACA
HAP5        481 TTTTTTAACCTCTCCAGAATTTTTATACA
HAP6        481 TTTTTTAACCTCTCCAGAATTTTTATACA
HAP7        481 TTTTTTAACCTCTCCAGAATTTTTATACT
HAP8        481 TTTTTTAACCTCTCCAGAATTTTTATACA
```