MOLECULAR EVOLUTION AND POPULATION GENETICS OF STURGEON (GENUS *ACIPENSER*) BASED ON MITOCHONDRIAL DNA ANALYSIS

by

James Raymond Brown

B.Sc., McGill University, 1978

M.Sc., Simon Fraser University, 1986

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in the Institute of Molecular Biology and Biochemistry

of

the Department of Biological Sciences

SIMON FRASER UNIVERSITY

June 1991

APPROVAL

Name: **JAMES RAYMOND BROWN**

Degree: Doctor of Philosophy

Title of Thesis:

MOLECULAR EVOLUTION AND POPULATION GENETICS OF STURGEON (GENUS ACIPENSER) BASED ON MITOCHONDRIAL DNA

Examining Committee:

Chairman: Dr. J.M. Webster, Professor

Dr. M.J. Smith, Professor, Senior Supervisor, Dept. Biological Sciences, SFU

Dr. A.T. Beckenbach, Associate Professor, Dept. Biological Sciences, SFU

Dr. D.L. Baillie, Professor, Dept. Biological Sciences, SFU

Dr. J. Don. McPhail, Professor, Dept. Zoology, UBC

Dr. M. Moore, Assistant Professor, Dept. Biological Sciences, SFU, Public Examiner

Dr. A.S. Harestad, Associate Professor, Dept. Biological Sciences, SFU, Public Examiner

Dr. F.W. Allendorf, Professor, Div. Biological Sciences, U. of Montana, Missoula MT, U.S.A. External Examiner

Date Approved June 28- 1991

PARTIAL COPYRIGHT LICENSE

I hereby grant to Simon Fraser University the right to lend my thesis, project or extended essay (the title of which is shown below) to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users. I further agree that permission for multiple copying of this work for scholarly purposes may be granted by me or the Dean of Graduate Studies. It is understood that copying or publication of this work for financial gain shall not be allowed without my written permission.

Title of Thesis/Project/Extended Essay

Molecular Evolution and Population Genetics of Sturgeon (Genus Acipenser) Based On Mitochondrial DNA Analysis

Author: _____
(signature)

James R. Brown
(name)

June 28, 1991
(date)

Abstract

The intra- and interspecific molecular evolution of four sturgeon species (Genus *Acipenser*) was studied using restriction enzyme analysis of the entire mitochondrial DNA (mtDNA) genome and direct sequencing of the origin of heavy strand replication or control region. Restriction enzyme analysis of 178 white sturgeon (*A. transmontanus*) revealed a total of ten distinct genotypes among Fraser and Columbia River populations. Phylogenetic analysis showed little evidence of geographical structuring, possibly as a result of high contemporary gene flow among the populations. Genetic diversity is lower in the Columbia River population even though the population was the probable source of founders for the Fraser River. Recent human intervention is suggested as a severe genetic bottleneck which has greatly reduced mtDNA diversity in Columbia River sturgeon.

The polymerase chain reaction (PCR) method was used to amplify a 462 nucleotide region of the mtDNA control region for direct DNA sequence analysis. The 34 variable nucleotide positions showed a strong bias to transitional changes. Among the 27 individuals surveyed, there were 19 genotypes. Three different phylogenetic analyses (neighbor joining, maximum likelihood and parsimony) showed that the trees derived from restriction enzyme and DNA sequence data were generally congruent. However, interior nodes of trees based on DNA sequence data were better supported statistically. The ratio of the mean percent differences estimated for the control region (2.30 %) to restriction enzymes (0.54 %) is 4.3 which is very close to that estimated for humans. Intra-molecular mutation rates for mtDNA are suggested to be similar among vertebrate species.

MtDNA restriction enzyme and control region sequence data were used to determine the phylogeny for four North American sturgeon species. Western species, *A. transmontanus* and *A. medirostris*, are more closely related to each other than to

A. fulvescens and *A. oxyrhynchus* which are found only to the east of the Continental Divide. The two western species show extensive length variation and heteroplasmy (multiple forms within an individual) of the control region. The differences in length are due to variable numbers of a perfect 82 nucleotide direct repeat. The mtDNA control region in *A. fulvescens* varies in length but heteroplasmy does not occur. No length variation occurs in *A. oxyrhynchus* mtDNA. Comparisons of the control regions among all four species reveal a mosaic of highly conserved and extremely variable sequence blocks. Inter- and intraspecific frequencies of length variants in 224 individuals were determined from hybridization of a cloned white sturgeon D-loop probe to Southern blots of whole mtDNA and densitometry of autoradiographs. Diversity indices and hierarchial statistics were used to examine the relationships of genetic drift and mutation in the maintenance of length variation.

Dedication

With Love

To Rose, Reema and Layal

Acknowledgements

I wish to thank my senior supervisor, Dr. Michael J. Smith for his support, encouragement and friendship. His enthusiastic but disciplined approach to scientific investigation imparted a valuable lesson. I also wish to thank my supervisory committee, Drs. Andy Beckenbach, David Baillie and J.D. McPhail, for their critical input. In addition, the reviews of this thesis by the external examiner, Dr. F.W. Allendorf and the two public examiners, Drs. A. Harestad and M. Moore, were much appreciated.

The following are acknowledged for providing sturgeon livers: L. Beckman, U.S. Fish and Wildlife Service; A.J. Dereksen, Manitoba Dept. of Natural Resources; S. Dionne, Quebec Ministry of Wildlife and Fisheries; S. Gvenette, University of Quebec in Montreal; G. Edmundson and D. Lane, Malaspina College; V. Swiatkiewicz, British Columbia Ministry of Environment; C. Tracy, State of Washington Dept. of Fisheries and several sturgeon sportsfishermen.

L. Fajber assisted in some of the laboratory analysis. Dr. D. O'Foighil photographed the autoradiographs. Additional computer facilities were provided by Drs. Baillie and Beckenbach. Dr. E.B. Hartwick kindly permitted me to write the thesis in his lab.

To a large extent, a thesis represents the cumulative knowledge of fellow colleagues. I am grateful for the input and support of the following co-workers: D. Bamfield, K. Beckenbach, J. Boom, E. Cabot, K. Doteval, Dr. L. Druehl, S. Gorski, D. Kowbel, G. Saunders, M. Stanhope, Dr. K. Thomas, H. Vahidi and L. Willis. In particular, I appreciate the support of current lab members, A. Arndt, I. Barriga and Dr. D. O'Foighil. I wish them every success in the future.

To a very large degree, this thesis also belongs to my wife, Rose Obeid, my two daughters, Reema and Layal, and my sister-in-law, Afifi Obeid. This thesis

would not have been possible without their care and understanding. Graduate fellowship support was provided by British Columbia Science Council Graduate Research and Engineering and Technology Award, SFU Graduate Fellowships, J. Abbott\ M. Fretwell Graduate Fellowship in Fisheries Biology, B.C. Packers Limited Fellowship and a SFU President Research Stipend. Additional funding was provided by SFU Special Research Projects, SFU President Research and NSERC grants to Drs. Beckenbach and Smith.

Table of Contents

List of Tables

List of Figures

List of Figures cont.

Introduction

Mitochondrial DNA as a Genetic Marker

The elucidation of historical relationships between organisms is fundamental to the study of evolution. To accomplish this goal, evolutionary descent must be determined by measuring characters with inheritable variation. Morphological and behavioural characteristics have been principally used to define taxonomic hierarchies. The development of molecular level markers has added a new level of sophistication in comparing the relationships between different groups of organisms (reviewed in Hillis and Moritz 1990).

Comparative immunology was the first molecular level approach to systematics. Histochemical staining and isozyme electrophoresis developed in the 1960's have become the most widely used methodology for detecting genetic variation. Amino acid sequencing increased the level of resolution of genetic variation and raised the possibility that molecular changes may occur at a constant or clock-like rate (Zuckerkandl and Pauling 1962).

The greatest resolving power of genetic variation rests with methods which deal directly with the DNA molecule. These techniques include DNA-DNA hybridizations, restriction enzyme mapping and direct DNA sequencing. Major advances in the analysis of DNA and RNA sequences have lead to their growing application in studies of both nuclear and organelle DNA.

The DNA of chloroplasts and mitochondria have received attention because of their compact structure and cytoplasmic inheritance. Both organelles are found only in the cell cytoplasm and are believed to have an endosymbiotic origin (Li and Graur 1991). For animal mitochondrial DNA (mtDNA), this restriction to the cytoplasm

results in the progeny inheriting nearly all of their mtDNA copies maternally through the egg.

The mitochondrial genome of animals consists of genes for small (12S) and large (16S) rRNAs, 13 proteins and 22 tRNAs (Brown 1983). The order of mtDNA genes differs among vertebrates, *Drosophila*, nematodes, sea urchins and sea stars (reviewed in Moritz et al. 1987; Smith et al. 1989). However, no gene rearrangements are known between vertebrate groups, with the exception of some minor changes in gallinaceous birds (Desjardins and Morais 1991). The gene organization of mtDNA is extremely compact with genes either adjacent to each other or separated by only a few nucleotides.

The only significant noncoding area is the origin of heavy strand replication also known as the control region or the D-loop region of vertebrates (Brown 1983). During DNA replication, synthesis of the heavy strand is initiated in the control region. As the DNA polymerase completes copying about two thirds of the molecule, it reaches the origin of light strand replication. At this point, synthesis of the light strand is initiated and proceeds in the opposite direction to heavy strand replication. The control region is also important as the initiation site of transcription (reviewed in Clayton 1982). In mammalian mtDNA, the entire coding sequence is transcribed as a single mRNA. The polycistronic transcript is then secondarily processed.

The maternal inheritance and rapid evolution of mtDNA makes it a compelling subject for molecular evolutionary analysis. The average rate of nucleotide substitutions in vertebrate mtDNA is estimated to be about 5 - 10 times higher than single copy nuclear DNA (Brown et al. 1979) with a bias to base transitions (changes from purine to purine or pyrimidine to pyrimidine) over transversions (changes between purine and pyrimidine; Brown 1985). The reasons for a higher rate of nucleotide change in mtDNA is unclear but could be related to deficiencies in DNA repair mechanisms (Tomkinson and Linn 1986).

According to the neutral theory of evolution, two DNA sequences will begin to accumulate nucleotide changes at a constant rate from their time of divergence (Kimura 1968). Given a high mutation rate, mtDNA comparisons should be able to detect more recent divergence events than comparisons of nuclear genes. This sensitivity of mtDNA to intraspecific levels of divergence and its facilitation of genealogical reconstruction, due to maternal inheritance, have lead to an increased use of mtDNA sequence analysis in population biology (Wilson et al. 1985).

Most population studies have employed restriction enzymes to secondarily estimate mtDNA sequence variation. Restriction enzymes cleave all DNA at specific palindromic sequences of 4, 5 or 6 nucleotides. MtDNA genotypes often show concordance with the geographical distribution of a species. This relationship between mtDNA phylogeny and biogeography provides for a possible link between evolutionary processes at the population (microevolution) and species level (macroevolution; Avise et al. 1987).

However, there is growing use of mtDNA comparisons in investigations of more ancient divergence events (Kocher et al. 1989; Meyer and Wilson 1990; Irwin et al. 1991; Ruvolo et al. 1991). Rather than using restriction enzymes which sample nucleotide sequences at random around the mtDNA genome, these studies have focused on specific mtDNA genes. Nucleotide substitution rates may vary considerably among mtDNA genes (Brown 1983) and this intergenic variation in accumulated base changes can be used to examine specific speciation events. Ribosomal RNA sequences which change slowly can be used to evaluate ancient branching points such as that between tetrapods and fishes (Meyer and Wilson 1990). At the other end of the evolutionary spectrum are population level polymorphisms observed in the human control or D-loop region (Aquadro and Greenberg 1983; Viligant et al. 1989; Horai and Hayasaka 1990).

The rapid determination of DNA sequences from a large number of individuals and very small tissue preparations has been revolutionized by the development of the polymerase chain reaction or PCR (Saiki et al. 1988). In PCR, specifically designed oligonucleotides are used to prime DNA synthesis *in vitro* in the presence of Taq DNA polymerase and deoxynucleotides. Two primers are designed to anneal on either side of the DNA sequence of interest. After several successive rounds (30 - 40) of strand denaturation, primer annealing and DNA synthesis, the targeted sequence is amplified in sufficient quantity for DNA sequence determination using the Sanger method (Sanger et al. 1977). PCR amplifications of mtDNA sequences have been used in the analysis of human populations (Viligant et al. 1989) and in comparisons of genetic variation in contemporary kangaroo rat (*Dipodomys panamintinus*) populations with that of museum specimens (Thomas et al. 1990). In the future, PCR techniques will lead to further applications of comparative DNA analysis in studies of systematics and population biology.

Phylogenetic Analysis

Concurrent with advances in the acquisition of DNA data are increasingly sophisticated approaches to the analysis of nucleotide sequences. While considerable conceptual and operational advances have been made in the analysis of DNA databases, it is still an emerging field of study and a consensus opinion has yet to be reached regarding the most suitable analytical approach (reviewed in Felsenstein 1988; Swofford and Olsen 1990). Therefore, phylogenetic inferences from molecular sequences are usually based on the results of several different statistical techniques.

Methodologies of phylogenetic analysis fall into three broad categories. The first are the distance methods which consider the proportion of similar nucleotide positions between two sequences. Estimates of the genetic distance (or divergence) separating two sequences are usually based on one of two probability models of

nucleotide change. The one parameter model considers the probability whether a nucleotide position will remain the same or change into one of the three other bases, with each of the alternative nucleotides being equally likely (Jukes and Cantor 1969). The two parameter model of Kimura (1980) also considers the probability whether a nucleotide position will remain the same or change into one of the three other bases. However, the probabilities of that change being a transition or a transversion are calculated differently. For restriction enzyme data, where the exact base composition is unknown, pairwise distances can be estimated using a maximum likelihood method based on the number of shared restriction enzymes sites and the length of recognition sequences (Nei and Tajima 1983). For multiple sequences, the distances between all pairwise comparisons are determined. These distance estimates are then used to construct a phylogenetic tree using cluster analysis (for example UPGMA) or conceptually similar types of analyses, such as the neighbor joining method (Saitou and Nei 1987). Nei et al. (1985) provided a method for calculating the standard errors about internal branch points or nodes in trees made using the UPGMA method.

The second category of phylogenetic reconstruction is maximum parsimony analysis. Parsimony analysis considers changes of nucleotide positions in all sequences in the search for the optimal tree. The optimal tree is assumed to require the fewest steps to account for the differences between sequences. This process can be computationally intensive for a large number of sequences and there are various mathematical approaches to optimize the search process (reviewed in Felsentein 1988; Swofford and Olsen 1990). There are often several equally parsimonious trees. One method to evaluate the statistical robustness of a parsimony tree is to use the bootstrap method to derive many trees from random samples (at least 100) of the data set (Felsenstein 1985). The number of times the same branching points occur in multiple bootstrap runs should indicate the relative statistical reliability of the tree.

The third method, maximum likelihood analysis (Felsenstein 1988), is a rigorous probability approach which has just recently seen more general application (Ruvolo et al. 1991). The method selects the tree which best explains the observed data or maximizes the likelihood that the data could have resulted. The advantage of the maximum likelihood method is that the statistical significance of internal nodes can be tested using the likelihood ratio test or LRT (Felsenstein 1988). In addition, the maximum likelihood values for different trees can be compared and evaluated using statistical criteria (Kishino and Hasegawa 1989). The general trend in the development of new phylogenetic methods, such as analysis of invariants (Lake 1987), is towards providing a better statistical bases for evaluating phylogenies.

If sufficient DNA sequence data are available and the rates of evolution among the lineages are approximately equal, then distance, parsimony and likelihood methods should result in the same tree topology. However, equal rates of nucleotide change among lineages cannot be assumed beforehand. Therefore, most phylogenetic studies usually cite the results of both distance matrix and maximum parsimony analyses.

The Molecular Phylogeny of Sturgeon

The white sturgeon (*Acipenser transmontanus*) was chosen as the focus of this study for two reasons. First, the migration patterns, population genetics and historical biogeography of this species is poorly understood. Analysis of mtDNA variation in white sturgeon from the Fraser and Columbia Rivers could provide new information on the population biology of this species and add to our general knowledge of gene flow in migratory fish species.

The second objective of this study was to characterize certain molecular features of mtDNA in a primitive vertebrate. Sturgeon are a relict vertebrate group believed to have evolved nearly 200 million years ago (Gardiner 1984). Comparisons

of the nucleotide sequence and organization of sturgeon mtDNA with that of other vertebrates could provide some useful insights into the evolution of vertebrate mtDNA. In an earlier collaborative effort, sequences of several white sturgeon mtDNA genes were determined and have been published elsewhere. This region included the sequences of the genes for cytochrome *b*, $tRNA^{Thr}$ and $tRNA^{Pro}$ as well as the D-loop region (Gilbert et al. 1988; Brown et al. 1989; Buroker et al. 1990). The gene organization of this region is similar to that of other vertebrates.

In this study, the distribution of mtDNA variation was first assessed in 178 white sturgeon from the Fraser and Columbia Rivers using restriction enzymes (Chapter I). The observed phylogeography and diversity of mtDNA genotypes are discussed in relation to gene flow and possible genetic bottlenecks resulting from Pleistocene glaciations and human intervention. Intraspecific comparisons of a 462 nucleotide (PCR amplified) region of the D-loop in 27 individuals provided further information on the population genetics of sturgeon and the evolution of the control region in lower vertebrates (Chapter II). The relative efficiencies of phylogenetic analysis from restriction enzyme and DNA sequence data were also assessed.

White sturgeon mtDNA shows extensive size variation both between and within individuals. A previous study had shown that this variation was due to a tandem repeated 82 nucleotide sequence in the control region (Buroker et al. 1990). In Chapter III, the molecular genetics of mtDNA size variation was examined in sturgeon populations using Southern blot analysis. The maintenance and distribution of this variation is discussed in terms of the balancing forces of mutation, genetic drift and selection.

In Chapter IV, the evolutionary relationships of four species of the genus *Acipenser* were examined using restriction enzyme analysis of entire mtDNA and comparisons of a PCR amplified DNA sequence of the control region. The extent of mtDNA size variation in the different species was also assessed using Southern blot

analysis. The molecular basis of size variation in all four species was determined from DNA sequence analysis of the repeated region of the D-loop.

Chapter I
The Effects of Pleistocene Glaciations and Human Intervention Upon Mitochondrial DNA Diversity in White Sturgeon (*Acipenser transmontanus*) Populations

Introduction

Mitochondrial DNA (mtDNA) has been widely used as a genetic marker in population studies because of its higher mutation rate relative to nuclear DNA, generally maternal inheritance and lack of sexual recombination (reviewed in Brown 1985). Several North American fish species have been subjected to mtDNA analysis including walleye (*Stizostedian vitreum*; Ward et al. 1989), American shad (*Alosa sapidissima*; Bentzen et al. 1989), North America whitefish (*Coregonus clupeaformis*; Bernatchez et al. 1989) and chinook salmon (*Oncorhynchus tshawytscha*; Wilson et al. 1987).

In the above studies, northern distributions of intraspecific mtDNA genotypes suggested possible dispersal patterns of founder groups from one or several southern refuges. Southern river drainages, located beyond the maximum excursions of Pleistocene ice sheets, have had prolonged periods of ecological stability. In contrast, northern freshwater habitats have been completely deglaciated for only the last 10-12,000 years (Smith 1981). Post-glacial dispersals from southern refuges were probably severe genetic bottlenecks, which have typically lead to a reduction in mtDNA diversity among northern populations of widely distributed species. In this chapter, mtDNA variation is described between two populations of *Acipenser transmontanus* (white sturgeon) from southern refuge and formerly glaciated river systems.

Sturgeon and their closest living relatives, paddlefish, are members of the Infraclass Chondrostei, a pre-Jurassic diversion from the teleost lineage (Lauder and Liem 1983; Gardiner 1984). Knowledge of the life history and population genetics of sturgeon species, particularly *A. transmontanus*, is sparse. The white sturgeon is the largest freshwater fish in North America, and individuals may exceed 600 kg in weight and 6 m in length (Scott and Crossman 1973). Reproductive maturity is not attained until between ages 11 to 34 years and specimens older than 70 years have been recorded (Semakula and Larkin 1968).

A. transmontanus is anadromous and readily moves between freshwater and marine coastal habitats. Conventional and radio tagging studies have yielded little information regarding the length of time white sturgeon spend at sea, their fidelity to river of birth or migration patterns (Miller 1972a, 1972b; Haynes and Gray 1981; Haynes et al. 1978; Dixon 1986). Individuals may travel distances of several hundred kilometers in marine coastal waters (Galbreath 1985). The potential exists for significant gene flow among white sturgeon populations with access to the Pacific Ocean because individuals are capable of spawning in different rivers over their life cycle.

The Fraser and Columbia Rivers have the largest *A. transmontanus* populations in the Pacific Northwest (McPhail and Lindsay 1986). Both rivers empty into the Pacific Ocean, with the mouth of the Fraser River located about 500 km north of that of the Columbia River. The southern two-thirds of the Columbia River remained ice-free during the Pleistocene glaciations (McPhail and Lindsay 1986). In comparison, the entire Fraser River was completely covered by the Cordilleran Ice Sheet as recently as 10,000-12,000 years ago (Clague and Luternauer 1983). Based on the paleogeography of the region, it is probable that Columbia River sturgeon migrating northward, through marine coastal waters, colonized the Fraser River after the last glacial retreat. Therefore, this founder effect may have

acted as a significant genetic bottleneck and cause notably lower mtDNA diversity in the Fraser River sturgeon population relative to the Columbia River population.

Human activity may have imposed a second genetic bottleneck of unknown magnitude. Nearly all North American sturgeon populations were extensively overfished from the late 1800's to the early 1900's (reviewed in Binkowski and Doroshov 1985). Later construction of large hydroelectric dams on the Columbia River in the 1930's and 1950's may have accentuated this bottleneck effect through alteration of migration routes and destruction of spawning habitat (Galbreath 1985).

In this study, mtDNA variation in *A. transmontanus* populations of the Fraser and Columbia Rivers were characterized over a three year period i) to investigate within river structuring of populations, ii) to provide estimates of existing genetic diversity in light of recent bottleneck events, and iii) to determine the extent of past and contemporary gene flow among populations.

Materials and Methods

Sample Collection and mtDNA Isolation

Mitochondrial DNA (mtDNA) was extracted from fresh sturgeon liver tissue removed in the field and transported on ice. MtDNA's were prepared from 105 sturgeon caught in the Fraser River, British Columbia, and 73 individuals from the Columbia River, Washington State. Most sturgeon were between 1 to 2 m fork length with estimated ages ranging from 9 to 30 years old (D. Lane, Malaspina College, Nanaimo, British Columbia, unpubl. data).

Columbia River sturgeon originated from the seasonal commercial fishery. State of Washington Department of Fisheries personnel collected liver samples from sturgeon in the late summers of 1987 and 1988, and in the spring of 1990. Sturgeon were caught near the mouth of the lower Columbia River ($n = 19$) and in the upper

Columbia River ($n = 54$) above Bonneville Dam, located about 230 km from the river mouth. This dam is the last impoundment of the Columbia River before it flows into the Pacific Ocean and has effectively isolated upper river sturgeon populations since the mid-1930's. Upper Columbia River sturgeon were caught from three separate reservoirs which were (in order upstream from Bonneville Dam): Bonneville Pool ($n = 25$), The Dalles Pool ($n = 24$) and John Day Pool ($n = 5$). The latter two pools were formed by dams built in the 1950's.

Commercial sturgeon fishing is prohibited on the Fraser River. Most of the liver samples came from sturgeon caught by the seasonal (May to November) sports fishery. A few sturgeon were caught either incidentally in the salmon gillnet fishery or in a separate tagging study (D. Lane pers. comm.). Sample sites ranged from the mouth of the Fraser River to the town of Lillooet, nearly 320 km upstream. The largest sample size was in 1987 ($n = 66$) with smaller sample sizes in 1988 ($n = 19$) and 1989 ($n = 20$). It is unknown whether the lower sample sizes in latter years reflect changes in the sturgeon population or a reduction in angler effort. A single mtDNA sample was also prepared from liver tissue pooled from twenty sibling juvenile sturgeon (2.5 months old) from the Sacramento River, California.

MtDNA was extracted from liver tissue within 6 - 72 hours of collection. MtDNA was isolated using a modified protocol of Lansman et al. (1981) which omitted the sucrose gradient step but included an increase in CsCl-ethidium bromide centrifugation time to 60 - 72 hours. Typical mtDNA yields were 300 - 500 *ng* per gram of homogenized tissue.

Restriction Enzyme Analysis

MtDNA comparisons were based on digests with 12 restriction enzymes with specificities for six nucleotide (*Apa*I, *Bcl*I, *Bgl*II, *Eco*RI, *Hind*III, *Pvu*II and *Xba*I), multiple six nucleotide (*Acc*I, *Ava*I, *Hae*II and *Hinc*II) and multiple five nucleotide

(*Ava*II) recognition sequences. The mtDNA genome of one individual was digested with paired combinations of several enzymes (*Ava*I, *Bcl*I, *Eco*RI, *Hind*III, *Kpn*I, *Pvu*II, *Sacc*I and *Xba*I) to map the relative locations of specific restriction sites. Enzyme digest conditions were as specified by the vendors (BRL and Pharmacia). DNA fragments were visualized under ultra-violet light after electrophoresis in ethidium-bromide stained agarose gels.

The observed intraspecific differences in DNA fragment patterns for each enzyme were due to relatively few restriction site changes. Therefore, it was possible to base divergence estimates on the presence or absence of restriction sites located on specific DNA fragments. Fragment length differences caused by the loss or gain of restriction sites were visually distinguishable from length polymorphisms found in the control or D-loop region of *A. transmontanus* mtDNA (Buroker et al. 1990). Southern blot analysis and direct sequencing have also confirmed which DNA fragment contained the control region in specific restriction enzyme digests (see Chapters II and III).

Nucleotide substitutions per site (δ) between pairs of mtDNA genotypes were estimated using the maximum likelihood method of Nei and Tajima (1983). An UPGMA tree of percent sequence divergence (d) was constructed with internal nodes plotted as $1/2d$ (Nei 1987; p. 293-296). Standard errors of internal nodes were calculated using the method of Nei et al. (1985).

The presence or absence of restriction sites genotypes were also analyzed using the Wagner parsimony method. The bootstrapped mixed parsimony algorithm (BOOT) of the PHYLIP 3.3 program package was used with 100 replicates (Felsenstein 1990).

The estimate of haplotypic or nucleon diversity (h) proposed by Nei and Tajima (1981) was used here to evaluate the diversity of mtDNA lineages within a population:

$$h = n(1 - \Sigma_{i=1}^{r} x^2_i)/(n - 1) \qquad 1.1$$

where x_i is the frequency of the ith type of mtDNA in a population of n specimens and r is the number of mtDNA types. Although this index is usually applied to the heterozygosity at nuclear loci, it is an appropriate measure for the diversity of maternal lineages (Nei 1987). Its application here permits comparisons with published mtDNA diversity estimates for other species.

A second estimate of nucleotide diversity (π) was calculated:

$$\pi = \Sigma_{ij} x_i x_j \pi_{ij} \qquad 1.2$$

where x is the frequency of the ith and jth type of mtDNA in a population and π_{ij} is the percent sequence difference between the two types (Nei and Li 1979). Therefore, this index of nucleotide diversity considers the frequency of occurrence of mtDNA types in a population as well as the divergence among mtDNA lineages.

Sensitivity of the sampling regime to detect all present mtDNA genotypes was evaluated using the combinatorial analysis of Hebert et al. (1988) as applied by Bernatchez et al. (1989). The relationship between sample size and the number of mtDNA genotypes observed in the Fraser and Columbia Rivers was determined from the cumulative number of mtDNA genotypes detected in a random choice of 60 individuals. Increments of three individuals (n = 3,6,9,12...60) were selected at random, without replacement, using the random case selection algorithm in the data management program (P1D) of the BMDP statistical software package (Dixon 1983). The mean number of genotypes detected in 10 randomizations were plotted against sample size.

Results

Restriction map of Sturgeon mtDNA

The size of *A. transmontanus* mtDNA varies between 16.1 and 16.5 kb depending upon the number of tandemly repeated 82 nucleotide sequences in the D-loop region (Buroker et al. 1990). Figure 1 shows several restriction enzyme sites, the D-loop region and genes for apocytochrome *b*, $tRNA^{Thr}$ and $tRNA^{Pro}$ (Gilbert et al. 1988; Brown et al. 1989; Buroker et al. 1990). The order and transcriptional polarity of these genes are identical with that of other vertebrates.

Intraspecific Phylogeny

The 12 restriction enzymes detected between 52 to 56 sites per *A. transmontanus* mtDNA genome. Since most of the enzymes cleaved at six nucleotide recognition sites, between 312 - 336 nucleotides or 1.9 - 2.0% of the genome was sampled. DNA fragment lengths for each enzyme are presented in Appendix A. About 42 % of individuals were heteroplasmic (i.e. had multiple copies of different mtDNA types within an individual) for length variation (Buroker et al. 1990). This length variation was localized to the control region using a cloned 0.8 kb fragment of the entire control region from *A. transmontanus* mtDNA to probe Southern blots (Chapter III).

Five restriction enzymes (*Acc*I, *Ava*I, *Ava*II, *Hinc*II and *Bcl*I) revealed intraspecific restriction site polymorphisms in *A. transmontanus*. The loss or gain of single restriction sites was observed for the enzymes *Acc*I, *Ava*II and *Hinc*II (Appendix A). There were two predominant *Ava*I fragment patterns, with a third pattern occurring in a single individual. Three individuals had a parallel loss and gain of two different *Bcl*I sites relative to the most frequent *Bcl*I fragment pattern.

Fig. 1. Restriction enzyme map of the *A. transmontanus* mtDNA genome showing the locations of known coding sequence.

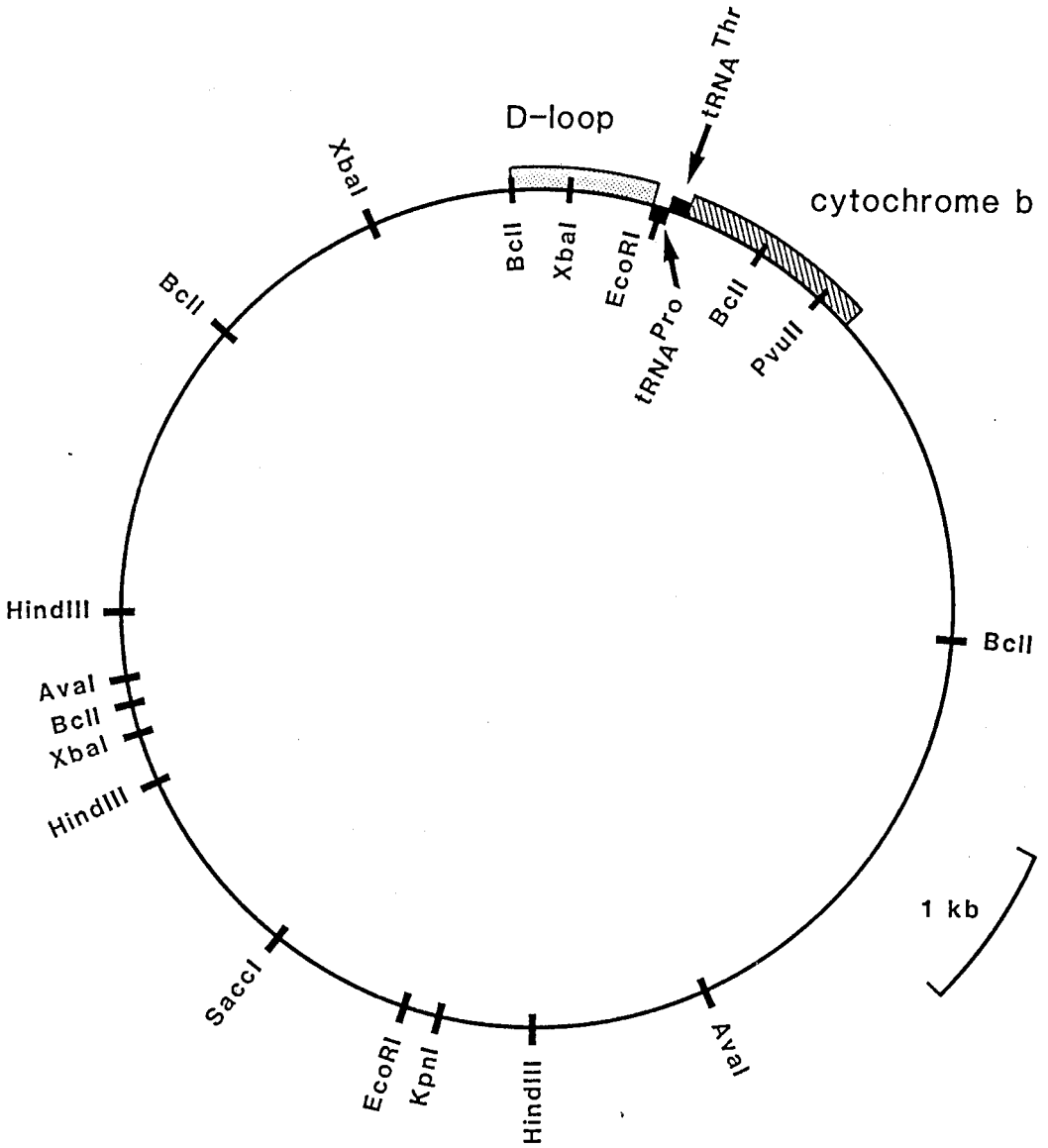
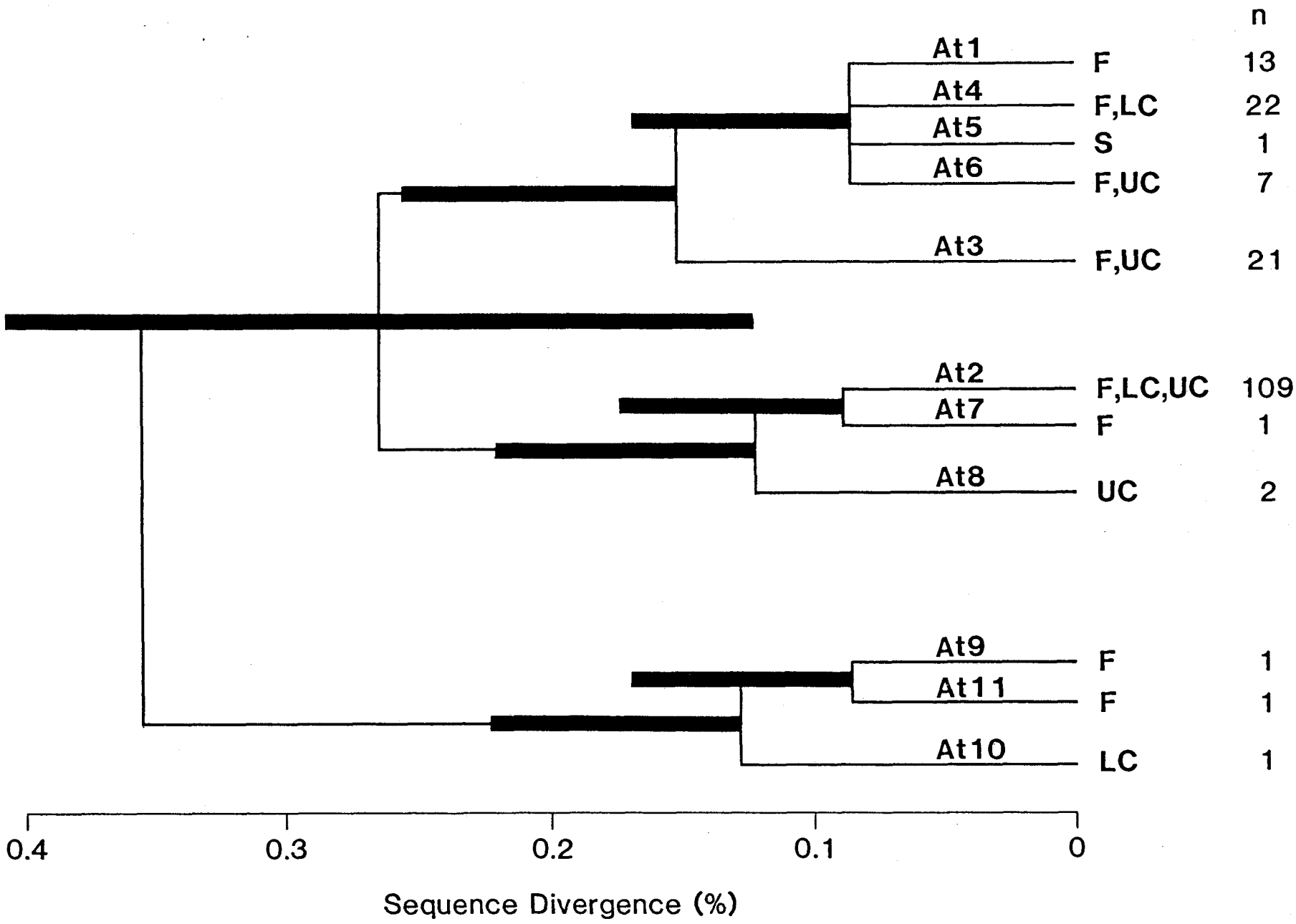

D-loop
tRNA Thr
cytochrome b
XbaI
BclI
XbaI
EcoRI
tRNA Pro
BclI
PvuII
BclI
HindIII
AvaI
BclI
XbaI
HindIII
SacI
EcoRI
KpnI
HindIII
AvaI
1 kb

Table 1. Percent pairwise sequence divergence estimates (lower diagonal) and standard errors (upper diagonal) for mtDNA lineages of A. transmontanus (At1-11).

	At1	At2	At3	At4	At5	At6	At7	At8	At9	At10	At11
At1	----	0.355	0.171	0.242	0.242	0.168	0.397	0.308	0.297	0.302	0.349
At2	0.708	----	0.308	0.246	0.246	0.302	0.174	0.177	0.397	0.404	0.355
At3	0.173	0.533	----	0.171	0.302	0.242	0.355	0.251	0.349	0.247	0.302
At4	0.344	0.350	0.173	----	0.242	0.168	0.302	0.308	0.297	0.302	0.242
At5	0.344	0.350	0.524	0.344	----	0.168	0.302	0.308	0.297	0.397	0.297
At6	0.170	0.523	0.344	0.170	0.170	----	0.349	0.355	0.238	0.349	0.297
At7	0.881	0.176	0.708	0.523	0.523	0.695	----	0.251	0.435	0.443	0.397
At8	0.533	0.179	0.357	0.533	0.533	0.708	0.357	----	0.443	0.362	0.404
At9	0.514	0.880	0.695	0.514	0.514	0.338	1.051	1.072	----	0.242	0.168
At10	0.523	0.897	0.350	0.523	0.886	0.695	1.072	0.721	0.344	----	0.171
At11	0.692	0.707	0.523	0.344	0.695	0.514	0.880	0.897	0.170	0.173	----

Fig. 2. UPGMA phylogeny of *A. transmontanus* genotypes (At1 - 11). Branching points are plotted as 1/2 *d*. Bars represent one standard error (SE) on either side of the branching point. Number of individuals (n) and occurrence in the Fraser (F), Lower Columbia (LC), Upper Columbia (UC) and Sacramento (S) Rivers are shown.


n
At1
F
13
At4
F,LC
22
At5
S
1
At6
F,UC
7
At3
F,UC
21
At2
F,LC,UC
109
At7
F
1
At8
UC
2
At9
F
1
At11
F
1
At10
LC
1
0.4
0.3
0.2
0.1
0
Sequence Divergence (%)

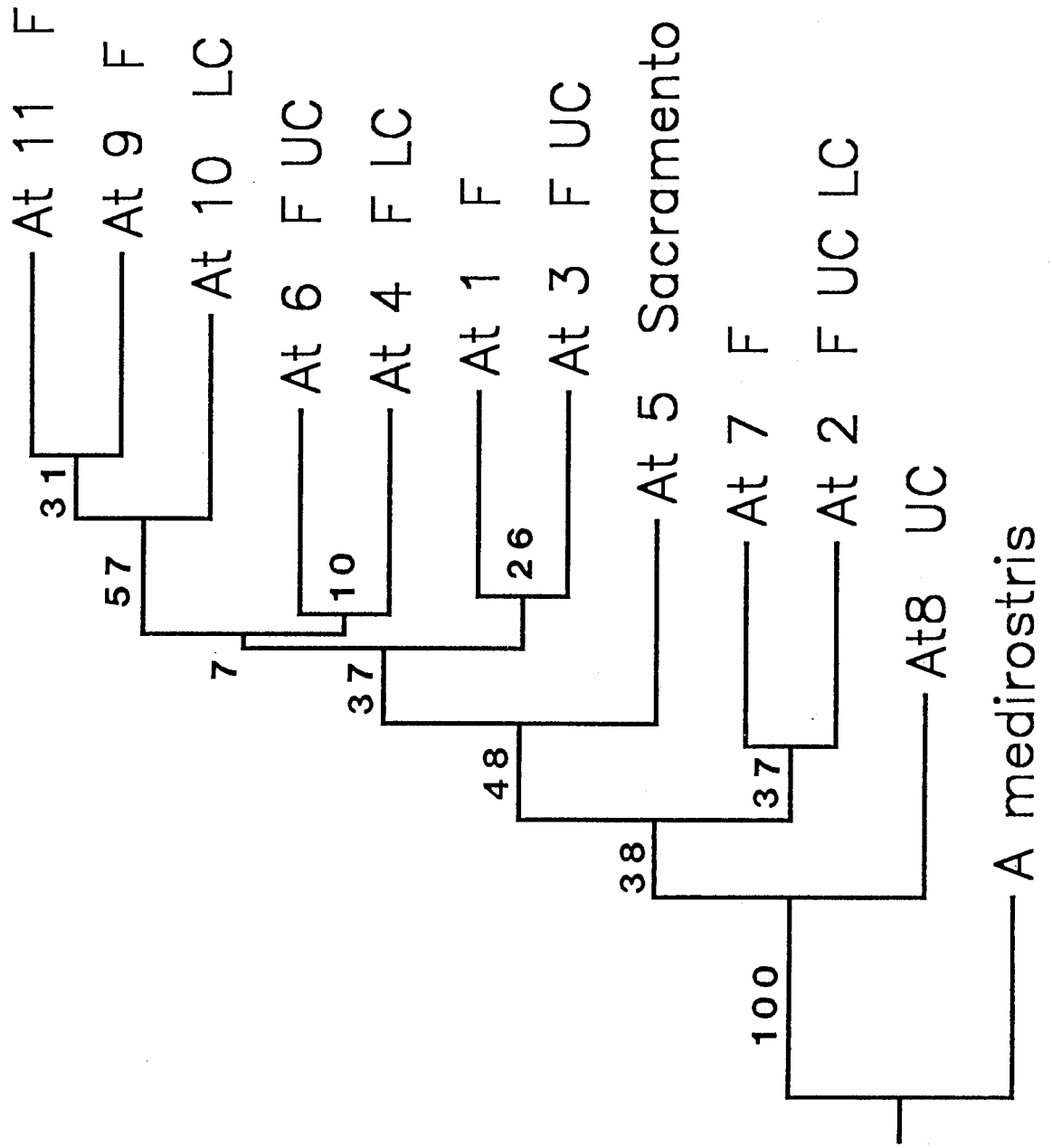
In total, 11 mtDNA genotypes or clonal lines were detected in *A. transmontanus*. Of the 179 individuals examined, 109 (61 %) shared a single mtDNA genotype, At2 (Fig. 2). The next three most common mtDNA genotypes (At1, At3 and At4) occurred in 13, 21 and 22 individuals, respectively. The seven remaining genotypes occurred in one to seven individuals.

Sequence divergence between genotypes ranged between 0.17 to 1.07 % (Table 1). UPGMA analysis revealed three major genotypic clusters (Fig. 2). Genotypes At1, At3, At4, At5 and At6 formed one cluster. The branch ordering of four of these genotypes is arbitrary since At1, At4 and At5 are equally diverged from At6. The second major cluster consists of At2, At7 and At8. The third cluster, as defined by two *BclI* restriction site differences, included three rare genotypes (At9, At10 and At11). Low estimates of intraspecific sequence divergence resulted in overlapping standard errors about the internal nodes. Therefore, the topology of the UPGMA tree is not significant.

A closely related species, *A. medirostris*, was used as an outgroup taxon in the Wagner parsimony analysis. Presence or absence data on a total of 64 restriction sites were entered into the analysis (Appendix A). In 100 bootstrap replicates, only one intra-specific branching point occurred more than 50 times (Fig. 3). However, the topology of the parsimony tree was generally congruent with that of the UPGMA tree, with similar clusterings of genotypes At2, At7 and At8 and genotypes At9, At10 and At11.

Distinctive geographical groupings of Fraser and Columbia River genotypes were not evident in either phylogenetic analysis. The pooled sibling sample from the Sacramento River was a separate genotype (At5) which clustered closely with genotypes found in the Fraser and Columbia Rivers. In the nomenclature used in Table 2, the composite genotype of this California sample was ABABA.

Fig. 3. Wagner parsimony phylogeny of *A. transmontanus* genotypes. Number of times each branching point occurred in 100 bootstrap replications are shown. Letter codes for geographical location of genotypes follows Fig. 3.


At 11 F
At 9 F
At 10 LC
At 6 F UC
At 4 F LC
At 1 F
At 3 F UC
At 5 Sacramento
At 7 F
At 2 F UC LC
At8 UC
A medirostris
31
57
7
10
26
37
48
38
37
100

Geographical Occurrence of Genotypes

The preceding phylogenetic analysis showed that topologies of the two genotypic trees were not statistically significant. Genotypes were characterized on the basis of the presence or absence of restriction sites which indicates actual nucleotide sequence changes between the mtDNA genomes of individuals. Therefore, analysis of genotypic frequencies is justified even though the inferred ancestral relationships between genotypes are not statistically well supported.

The geographical range of the most frequent genotype (At2) extended throughout the sampled range in the Fraser and Columbia Rivers (Table 2; Fig. 4 and 5). However, the observed frequency of genotypes was significantly different between river systems (log-likelihood ratio chi-square or $G = 36.84$, $d.f. = 5$, $p < 0.0001$). Genotypic frequencies were also distinctive between the upper and lower Columbia River ($G = 21.99$, $d.f. = 5$, $p < 0.001$).

Genotype At2 was found throughout the Columbia River. However, less frequent genotypes were specific to areas above (At3, At6 and At8) or below (At4 and At10) Bonneville Dam (Fig. 4). Two of the rare genotypes (At8 and At10) were unique to the Columbia River.

In the Fraser River, the most common genotypes (At2, At3 and At4) were distributed throughout the middle and lower sections of the river (Fig. 5). In the nine individuals sampled upstream of Hell's Gate Rapids (280 - 320 km from river mouth), only genotypes At2 ($n = 8$) and At4 ($n = 1$) were detected. The Hell's Gate Rapids, located 200 km upstream, have likely isolated upper Fraser River sturgeon populations since 1912. In comparison, eight separate genotypes were detected downstream of the rapids.

Table 2. MtDNA genotypic frequencies of *Acipenser transmontanus* for entire samples from the Fraser and Columbia as well as for various subsamples of the Columbia River. Letter code in parentheses denotes polymorphic DNA fragment patterns for the enzyes *Bcl*I, *Hinc*II, *Ava*I, *Ava*II and *Acc*I.

	Entire Rivers		Columbia		Upper Columbia Pools		
Genotype	Fraser	Columbia	Lower	Upper	Bonneville	The Dalles	John Day
n	105	73	19	54	25	24	5
At 1 (AAABB)	0.12	0.0	0.0	0.0	0.0	0.0	0.0
At 2 (ABBAA)	0.49	0.80	0.68	0.83	0.92	0.75	0.80
At 3 (AABBB)	0.18	0.03	0.0	0.04	0.04	0.04	0.0
At 4 (ABBBB)	0.16	0.07	0.26	0.0	0.0	0.0	0.0
At 6 (ABABB)	0.02	0.07	0.0	0.09	0.04	0.17	0.0
At 7 (ABCAA)	0.01	0.0	0.0	0.0	0.0	0.0	0.0
At 8 (AABAA)	0.0	0.03	0.0	0.04	0.0	0.04	0.20
At 9 (BBABB)	0.01	0.0	0.0	0.0	0.0	0.0	0.0
At10 (BABBB)	0.0	0.01	0.05	0.0	0.0	0.0	0.0
At11 (BBBBB)	0.01	0.0	0.0	0.0	0.0	0.0	0.0

Fig. 4. Uppermost pie charts show mtDNA genotypic frequencies in the lower ($n = 19$) and upper ($n = 54$) Columbia River as delineated by the Bonneville Dam. Bottom pie charts subdivide genotypic frequencies of upper Columbia sample into reservoirs which are Bonneville Pool ($n = 25$), The Dalles Pool ($n = 24$) and John Day Pool ($n = 5$). Solid bars across the river mark dam locations.

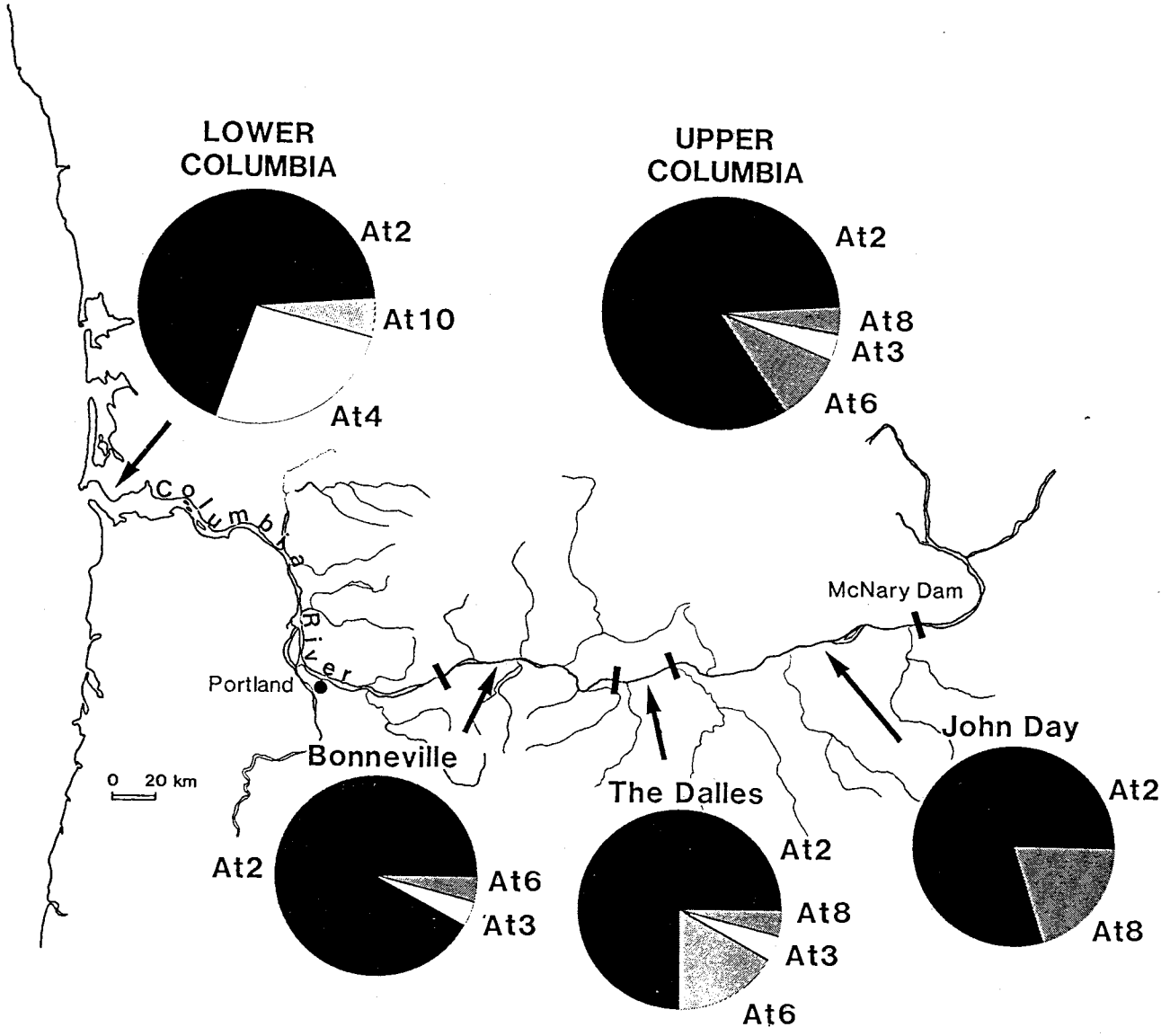
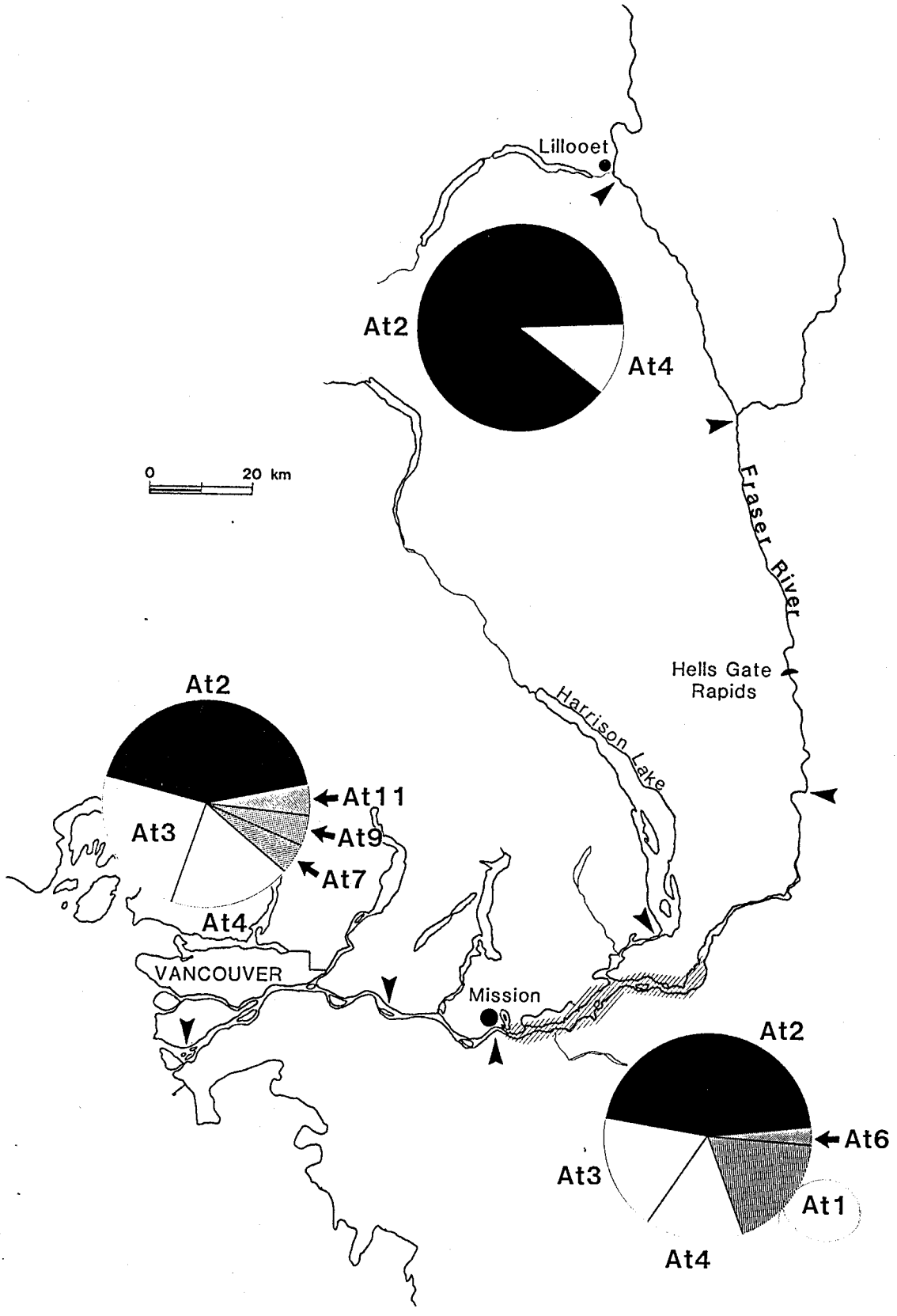

LOWER
COLUMBIA
At2
At10
At4
UPPER
COLUMBIA
At2
At8
At3
At6
Columbia River
McNary Dam
Portland
Bonneville
0 20 km
The Dalles
John Day
At2
At6
At3
At2
At8
At3
At6
At2
At8

Fig. 5. MtDNA genotypic frequencies in the Fraser River. Arrows indicate the edges of sample areas for the lower ($n = 22$), middle ($n = 74$) and upper ($n = 9$) river. Shaded area upstream of the city of Mission covers the range of genotype At1.


Lillooet
At2
At4
0
20 km
Fraser River
Hells Gate
Rapids
Harrison Lake
At2
At11
At3
At9
At7
At4
VANCOUVER
Mission
At2
At6
At3
At1
At4

Only one genotype (At1) had a narrow distribution. This unique genotype was found in 13 individuals caught in a 45 km long section of the Fraser River, 90 km from the river mouth (Fig. 5). This genotype was represented in roughly the same proportions in samples over a three year period.

MtDNA Diversity

In Table 3, both estimators of mtDNA diversity showed that mtDNA diversity of the Fraser River ($h = 0.70$; $\pi = 0.157$) is nearly twice that of the Columbia River ($h = 0.36$; $\pi = 0.076$). In the Columbia River, greater mtDNA diversity was found below ($h = 0.49$; $\pi = 0.103$) than above ($h = 0.30$; $\pi = 0.067$) Bonneville Dam. MtDNA diversity in the upper Columbia River was considerably lower in Bonneville Pool in comparison to The Dalles and John Day Pools. Lower mtDNA diversity ($h = 0.22$) was also estimated for isolated sturgeon populations above Hell's Gate Rapids (Fig. 5) compared to the lower Fraser River ($h = 0.72$).

Diversity estimates for the Fraser River derived from annual subsamples in 1987 ($n = 66$), 1988 ($n = 19$) and 1989 ($n = 20$) were consistently higher than estimates for the Columbia River ($n = 73$), suggesting that low diversity estimates for the Columbia River population were not artifacts of sample size.

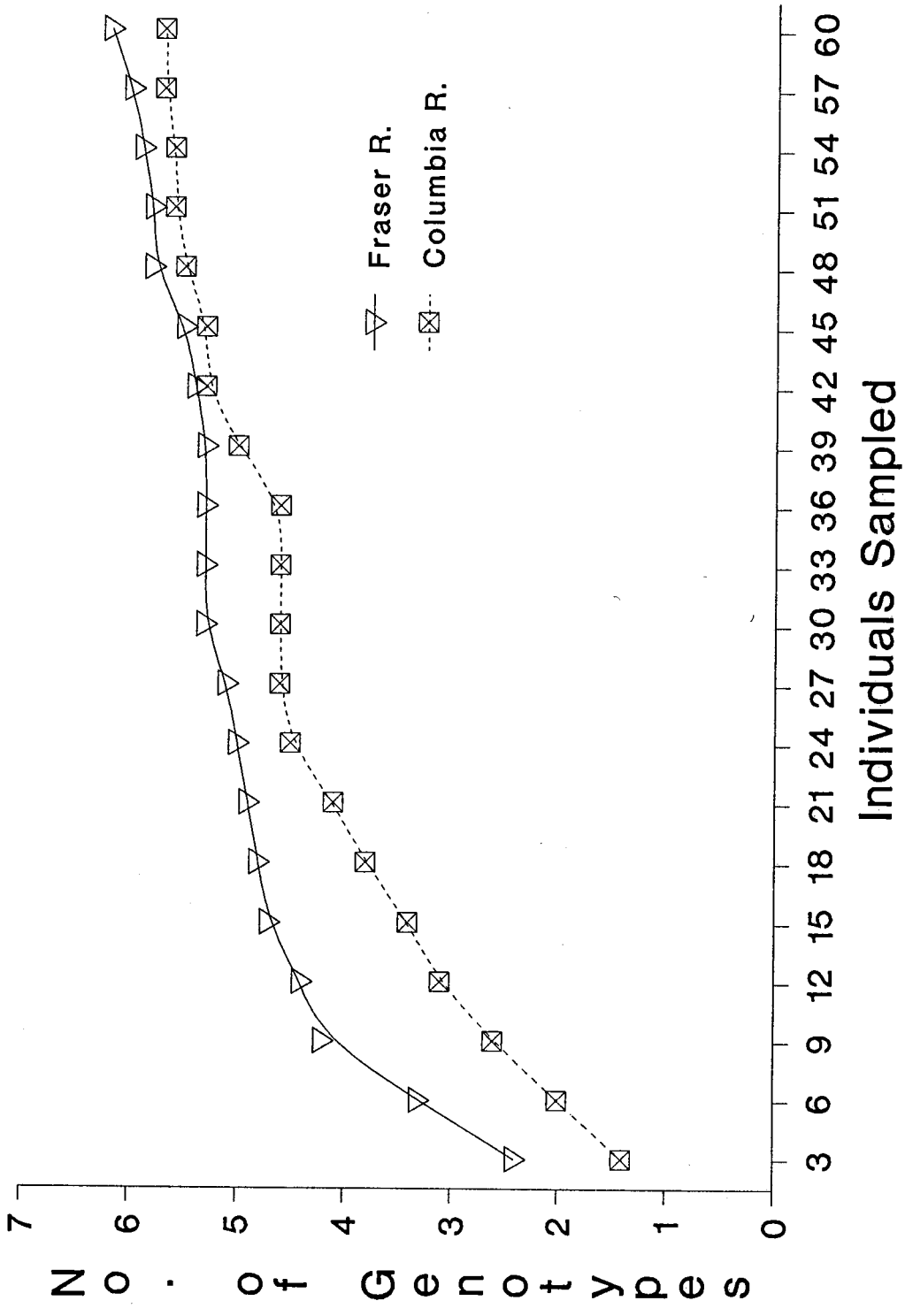
Sample Size and Detection of Genotypes

In the combinatorial analysis, the number of genotypes detected rose with sample size until an asymptote was reached for samples larger than 36 - 39 individuals (Fig. 6). Fewer genotypes were detected in the Columbia River (4 - 5 genotypes) compared to the Fraser River (5 - 6 genotypes).

Table 3. Estimates of mtDNA diversity (h) and nucleotide diversity (π) using the methods of Nei and Tajima (1981) and Nei and Li (1979), respectively.

Location	n	h	π
Fraser and Columbia River	178	0.65	0.135
Fraser River	105	0.70	0.157
1987	66	0.75	0.151
1988	19	0.53	0.134
1989	20	0.51	0.142
Columbia River	73	0.36	0.076
Lower	19	0.49	0.103
Upper	54	0.30	0.067
Bonneville	25	0.16	0.039
The Dalles	24	0.42	0.096
John Day	5	0.40	0.029

Fig. 6. Sample size and detection of *A. transmontanus* genotypes in the Fraser and Columbia Rivers. Each data point represents the mean of 10 random subsamples.


No. of Genotypes
0
1
2
3
4
5
6
7
3 6 9 12 15 18 21 24 27 30 33 36 39 42 45 48 51 54 57 60
Individuals Sampled
Fraser R.
Columbia R.

Discussion

Intraspecific Phylogeny

The mtDNA phylogenetic analysis indicates that *A. transmontanus* populations are little diverged and generally lack significant geographical substructure. Low confidence intervals around intraspecific branching points can be attributed to either a deficiency in the number of nucleotides sampled or historical events which have greatly reduced mtDNA diversity in this species.

An examination of the recognition sites of the polymorphic enzymes, *Acc*I (GT[A/C,G/T]AC), *Ava*I (GPyCGPuG), *Ava*II (GG[A/T]CC) and *Hinc*II (GTPyPuAC), reveals that all have multiple recognition sequences. Of the fixed hexameric enzymes used (*Apa*I, *Bcl*I, *Bgl*II, *Eco*RI, *Hind*III, *Pvu*II and *Xba*I), only *Bcl*I was polymorphic, with 3 individuals displaying a variant fragment pattern. In other studies of intraspecific mtDNA variation, including some of the fish populations studies cited here, fixed hexameric enzymes were polymorphic. DNA sequence comparisons of a hypervariable segment in the control region of *A. transmontanus* mtDNA generally concurs with the low intraspecific divergence estimates obtained from restriction enzyme data (Chapter IV). The low levels of mtDNA sequence variation detected in *A. transmontanus* are probably reflective of the lack of sequence differences among individuals rather than inadequate sampling of nucleotides or populations (see below).

Low genetic variability has been noted for other North American Chondrosteans. Little intraspecific variation was observed in a protein electrophoretic study of American paddlefish (*Polyodon spathula*) populations (Carlson et al. 1982). Two other species of sturgeon, *Scaphirhynchus albus* and *S. platorynchus,* were indistinguishable at 37 protein loci (Phelps and Allendorf 1983).

Restriction enzyme analysis have revealed low intraspecific mtDNA variation in northern populations of the sturgeon species, *A. fulvescens*, *A. medirostris* and *A. oxyrhynchus* (Chapter IV). Populations of *A. oxyrhynchus* in the southeastern United States also show very low mtDNA diversity (Bowen and Avise 1990).

Extensive gene flow between *A. transmontanus* populations would minimize geographical structuring while recent, severe genetic bottleneck effects would reduce genetic diversity. Gene flow between the Fraser and Columbia Rivers might be bidirectional since the fidelity of migratory *A. transmontanus* to natal rivers is unknown. Sturgeon tagged in the Columbia River have been recaptured in marine waters off the coast of northern Washington State (Galbreath 1985). A few migratory, reproductively active females per generation could prevent significant divergence of mtDNA genotypes between river systems (Slatkin 1987).

The most geographically segregated sturgeon mtDNA genotype was At1. Over a three year period, all individuals with this genotype were caught in a 45 km section of the lower Fraser River. This genotype ($n = 13$) comprised about 12 % of the Fraser River sample. The restricted range of At1 and higher divergence (> 1.0 %) between certain conspecific genotypes (Table 1) suggests the existence of moderate isolation processes. Discussion regarding reproductive barriers is highly speculative given our limited understanding of sturgeon behaviour. Potential microgeographic isolation of spawning females could occur in two large lakes draining into the Fraser River. Sturgeon may also be segregated into nonsynchronous spawning cohorts separated by reproductive cycles of several years (Roussow 1957; Semakula and Larkin 1968; Doroshov 1985).

MtDNA Diversity

Previous studies have consistently shown that intraspecific mtDNA diversity (h) is lower in recently founded populations relative to populations in areas of long term ecological stability. As an example, North American whitefish (*Coregonus clupeaformis*) populations in recently (less than 18,000 years ago) deglaciated rivers had much lower diversity ($h = 0.26 \pm 0.25$) than Eurasian populations ($h = 0.90 \pm 0.06$) of a closely related species, *C. lavaretus* (Bernatchez et al. 1989). Genetic bottleneck effects were probably less severe for the Eurasian species because over its contemporary range, glaciation was incomplete and several local refugia had probably existed. Ashley and Wills (1987) concluded that, in the deer mouse (*Peromyscus maniculatus*), recent colonization events were responsible for lower mtDNA heterogeneity of island populations ($h = 0.0$ - 0.44) compared to mainland groups ($h = 0.67$ - 1.0).

The *A. transmontanus* population that colonized the Fraser River ($h = 0.70$) had considerably higher mtDNA diversity than the putative founder population of the Columbia River ($h = 0.36$), contrary to expectation. Further, the upper Columbia River population ($h = 0.30$) had reduced diversity relative to the lower Columbia River population ($h = 0.49$). Log-likelihood ratio chi-square tests confirmed that genotypic frequencies were significantly different in all comparisons between the Fraser River, the lower Columbia River and the upper Columbia River.

It is unlikely that additional samples would have increased the detection of novel mtDNA genotypes in either river system (given the enzymes used in this study). The combinatorial analysis estimated that a minimum sample size of 50 - 60 individuals is sufficient to fully detect mtDNA diversity in both river systems (Fig. 6). Samples for the Columbia ($n = 73$) and Fraser ($n = 105$) were considerably larger than this estimate. The consistently fewer number of genotypes detected over

a range of sample sizes in the Columbia River, relative to the Fraser River, is further evidence of lower genetic diversity in the Columbia River.

The mtDNA diversity of both river samples combined ($h = 0.65$) is near the lower estimates for natural populations of Atlantic salmon (*Salmo salar*; $h = 0.60$) and brown trout (*Salmo trutta*; $h = 0.70$ - 1.0; Gyllensten and Wilson 1987). In these same salmonids, mtDNA diversity levels determined for hatchery populations ($h = 0.0$ - 0.2) founded by small maternal groups, approached the diversity estimate of the entire upper Columbia River sturgeon population ($h = 0.30$) and, in particular, the Bonneville reservoir sample ($h = 0.17$).

There are two possible explanations for higher mtDNA diversity in white sturgeon populations of the recolonized Fraser River, relative to the Columbia River refuge. First, migrants from another southern population may have significantly added to the Fraser River mtDNA gene pool. However, with the exception of the Sacramento River, other Pacific Coast rivers do not support substantial white sturgeon populations. The single individual from the Sacramento River examined in this study had a unique, but closely related genotype, not found in either of the northern rivers.

The second, and more plausible, explanation is that although the Columbia River was the source of present mtDNA lineages in the Fraser River, post-glacial events have lowered mtDNA diversity in the Columbia River population. Recent overexploitation and habitat destruction could have greatly reduced genetic diversity in white sturgeon. Semakula and Larkin (1968) document a precipitous decline in the commercial catch of white sturgeon on the Fraser River from over 2,500,000 kgs. in 1897 to below 44,000 kgs. by 1916. Columbia River stocks suffered a parallel decline over the same time period (Galbreath 1985).

Habitat changes have occurred in both river systems but have been more extensive in the Columbia River basin. Hydroelectric dams built on the Columbia

River in the 1930's and 1950's have probably caused major alterations in sturgeon migration patterns and spawning ecology. The abrupt appearance of these barriers over a period of only a few years may be responsible for the observed segregation of certain genotypes to areas either upstream or downstream of Bonneville Dam.

In 1912, an accidental rock slide blocked the Fraser River at Hell's Gate, about 200 km from the mouth. MtDNA diversity was lower in the sample collected above Hell's Gate ($h = 0.22$) relative to the lower Fraser River ($h = 0.72$). This intra-river difference in diversity estimates is similar to that observed in the Columbia River between white sturgeon populations located above and below Bonneville Dam.

Similar patterns of lower genetic diversity in recently isolated, freshwater populations have been observed in other fish species. A freshwater *Salmo trutta* population trapped above a waterfall, had lower heterozygosity at protein loci compared to an anadromous population found in the same river basin (Hindar et al. 1991). Genetic differentiation between landlocked and anadromous forms was significantly greater than that between sympatric *S. trutta* populations with different life histories traits. Australian populations of *Galaxias ruttaceus* landlocked in lakes, had lower mtDNA diversity than stream populations with access to estuaries and the sea (Ovenden and White 1990). The recent formation of these Australian lakes (3000-7000 years ago) was suggested as a severe but transitory bottleneck which reduced mtDNA diversity.

There is no apparent divergence between mtDNA genotypes found above Bonneville Dam and those occurring in the lower Columbia River and the Fraser River. The probability of detecting nucleotide changes so soon after the occurrence of a genetic bottleneck is extremely low (Wilson et al. 1985). However, a steep reduction in mtDNA or genetic diversity has occurred in the landlocked white sturgeon population.

Genetic drift in isolated demes is unlikely to have caused a reduction in mtDNA diversity in upper Columbia River sturgeon populations because fewer than four consecutive generations have passed since the building of Bonneville Dam. Lower mtDNA diversity in the upper Columbia River is more probably the result of selective pressures for individuals (i.e. maternal lines) which best thrive in the landlocked, freshwater habitat. This is not to suggest that selection is acting at the mtDNA level, but that mtDNA genotypes are neutral markers associated with successful maternal lineages.

Human pressures upon populations can be highly selective and result in a net reduction in genetic diversity (Nelson and Soule 1987). Columbia River dams may have reduced sturgeon spawning success by denying lower river sturgeon access to upstream spawning grounds. Upstream sturgeon populations are similarly excluded from estuaries which have greater food productivity. Also the dams have greatly altered natural current velocities which may be critical to sturgeon spawning success (Doroshov 1985; Brannon et al. 1986).

Brannon et al. (1986) suggested that allozyme frequencies in recently isolated white sturgeon populations in the Columbia River were not at Hardy-Weinberg equilibrium. Bartley et al. (1985) noted higher average heterozygosity values (4.9 - 6.9 %) for three white sturgeon populations with ocean access compared with an isolated freshwater population (1.4 %).

If particular groups of females with synchronous migration patterns are distinguishable by mtDNA genotype, the present-day distribution of mtDNA genotypes could be indicative of the location of these maternal cohorts when the dams were closed. The presence of two genotypes (At3 and At6) in both the Fraser and upper Columbia Rivers, and their absence from the lower Columbia River, indicates that upper Columbia sturgeon once migrated throughout the basin and into marine coastal waters.

Any reduction in genetic diversity in sturgeon populations should be of special concern in the conservation of this species given its long generation time and advanced age of reproductive maturity. The endangered or threatened status of two-thirds of the North American sturgeon exemplifies that this ancient group has not fared well in recent times (Williams et al. 1989).

Previous evolutionary studies of intraspecific mtDNA variation have been mainly concerned with the phylogeographical relationship of mtDNA genotypes as a reflection of the influence of natural geographical events on population structure. This study suggests that human intervention in animal populations can also produce detectable shifts in mtDNA genotypic frequencies over a relatively short time.

Chapter II
Intraspecific Sequence Variation in the Control Region of White Sturgeon (*Acipenser transmontanus*) Mitochondrial DNA

Introduction

Mitochondrial DNA (mtDNA) has been widely used in evolutionary studies of animals because of its generally strict maternal inheritance and rapid evolution. DNA-DNA hybridization and restriction enzyme analyses have estimated the rate of nucleotide change in vertebrate mtDNA to be 5 - 10 times greater than that of single copy nuclear DNA (Brown 1983; Moritz et al. 1987). However, base substitution rates may vary considerably among mtDNA genes (Brown 1983). This intergenic variation in mutation rates is of considerable importance to evolutionary biology. Through the selection of a particular mtDNA gene or genes, the resolution of nucleotide sequence polymorphisms can be adapted to the intended level of taxonomic comparisons. Consequently, there has been recent interest in the reconstruction of animal phylogenies using nucleotide sequences of specific mtDNA genes.

The polymerase chain reaction (PCR) has been particularly useful for the rapid amplification of specific mtDNA gene sequences from minute amounts of target DNA (Saiki et al. 1988). PCR amplified cytochrome *b* gene sequences were used for intra- and intergenera comparisons of fish, birds and rodents (Kocher et al. 1989) and the analysis of phylogenetic relationships among mammalian orders (Irwin et al. 1991). A significant tree topology of the African hominoid trichotomy was derived

from comparisons of the cytochrome oxidase subunit II gene from five primate species (Ruvolo et al. 1991). A much deeper divergence event, the origin of tetrapods, was inferred from comparisons of 12sRNA and cytochrome *b* genes from lungfish, the coelocanth, ray-finned fish and frog (Meyer and Wilson 1990).

At the other end of the evolutionary spectrum are comparisons among populations within a species. Previously, restriction enzyme analysis of mtDNA has been largely used in population studies. However, as shown in the previous chapter and in a reanalysis of several other mtDNA population surveys (Nei et al. 1985, Lynch and Crease 1990), large standard errors usually are associated with divergence estimates based on restriction enzyme analysis.

The origin of heavy strand replication or control region is the most variable area in the mtDNA molecule thus highly suited for detecting intraspecific nucleotide differences (Upholt and Dawid 1977; Wilson et al. 1985). The vertebrate control region, also called the D-loop, does not code for proteins and has only a few sequence blocks which are conserved among taxa (Clayton 1982, Saccone et al. 1987). Nucleotide substitutions and length mutations occur commonly in the control region (reviewed in Harrison 1989). In humans, the nucleotide substitution rate of the control region is three to four times higher than the rate of sequence divergence of the entire mtDNA molecule as estimated from restriction enzyme analysis (Aquadro and Greenberg 1983; Vigilant et al. 1989; Horai and Hayasaka 1990).

Intraspecific variation in the control region has been little studied in species other than humans. Phylogenetic studies of kangaroo rat (*Dipodomys panamintinus*; Thomas et al. 1990) and cichlid fish populations (Meyer et al. 1990) have used partial control region sequences. However, there is little data to suggest whether patterns of intraspecific variation observed in the human mtDNA control region can

be generalized to other vertebrates. It is also unknown whether regions of intraspecific variability and rates of nucleotide change observed in the control region of homeotherms are comparable in poikilotherms.

Sturgeon are a relict group which diverged from the teleosts or bony fishes nearly 200 million years ago (Gardiner 1984). Analysis of mitochondrial DNA sequences in sturgeon is particularly relevant to the study of vertebrate mtDNA evolution because of the significant divergence of sturgeon from the tetrapods.

In collaboration with other workers, I have cloned and sequenced the complete mtDNA control region of two *A. transmontanus* individuals (Buroker et al. 1990). The control region of white sturgeon mtDNA was found to vary in size from 761 to 1007 nucleotides depending upon the number of perfect 82 base pair sequences repeated in tandem after the 3' end of the $tRNA^{Pro}$ gene. In the remainder of the control region occurred single base substitutions and deletions which are more typical of nucleotide polymorphisms found in vertebrate mtDNA.

Restriction enzyme analysis of mtDNA detected limited geographical substructuring of *A. transmontanus* populations in the Fraser and Columbia Rivers (Chapter I). In order to determine whether sturgeon populations are truly panmitic, more extensive genetic variation needs to be detected. Single nucleotide changes in the nonrepeat area of the control region are potentially useful in resolving this phylogenetic problem.

In this chapter, intraspecific variation in the control region of *A. transmontanus* mtDNA was examined 1) to compare the pattern of nucleotide variation in white sturgeon with that of higher vertebrates, 2) to compare the extent of base changes in the control region with the level of nucleotide change for the entire mtDNA molecule detected by restriction enzymes 3) to determine in greater detail

the phylogenetic relationships of white sturgeon populations in the Fraser and Columbia Rivers, 4) to evaluate the topologies of phylogenetic tree derived from restriction enzymes and mtDNA sequence data and 5) to compare estimates of genetic diversity and effective population sizes derived from the two data sets.

Material and Methods

Sample Collection

White sturgeon collections from the Columbia and Fraser Rivers and mtDNA extractions are described in Chapter I. Individuals used in the sequence analysis were chosen on the basis of mtDNA restriction enzyme genotype and river of origin. Individuals from the most abundant genotypes (At1, At2, At3 and At4) were selected for analysis using a random number table.

MtDNA control region sequence from a 3 m long, 82 years old female white sturgeon was also used in the analysis (D. Lane, Malaspina College, Nanaimo, British Columbia; pers. commun.). This individual was found dead in 1989 on the shore of the mid-Fraser River about 90 km from the river mouth. Two previously published control sequences from two sturgeon were also used (Buroker et al. 1990). In total, control region sequences were compared for 11 individuals from the Columbia River and 16 individuals from the Fraser River. All ten mtDNA restriction enzyme genotypes found in *A. transmontanus* were represented in the control region data set.

PCR Amplification and DNA Sequencing

All PCR amplifications were done from CsCl purified white sturgeon mtDNA extracted from liver tissue. Primers were designed from the published sequence of

the control or D-loop region of *A. transmontanus* (Buroker et al. 1990). Sequences of the primers follow, the letters L and H refer to the light and heavy strands and the number refers to the position of the primer's 3' base in the reference 1.6 kb fragment: L185 (5' CATCTACCATTAAATGTTATAC 3'), L317
(5' CTGTAGGGATTCACAACTG 3'), L506
(5' TGACATGTAGAACTCCTTCAGA 3') and H740
(5' GATCAAGGTATGTCGATGACA 3'). The primer combination, L185 and H740 consistently amplified a 597 nucleotide long fragment of the control region. The other two oligonucleotides were used as internal primers for sequencing reactions.

PCR amplifications were done using the GeneAmp DNA Amplification Reagent Kit (Perkin Elmer Cetus). Each amplification was performed in 50 µl final volume with 5 µl of 10x reaction buffer (500 mM KCl, 100mM Tris-HCl (pH 8.3), 15 mM $MgCl_2$, 0.1 % gelatin), each dNTP at 200 mM, each primer at 5-7 µM, mtDNA (5-50 *ng*) and 0.5 units of *Thermus aquaticus* DNA polymerase with 50 µl of mineral oil overlaid. Each PCR reaction cycle consisted of denaturation for 1.5 min at 95°C, primer annealing for 2 min at 50°C and extension for 2-3 min at 72°C. Cycles were repeated 40 times in a programmable cyclic reactor (Ericomp Inc.) and the final cycle included an extension reaction of 10 min.

Amplified double-stranded DNA was phenol extracted from 1.5% low melting point agarose after electrophoresis for 2-3 hr. at 50v in 40mM Tris acetate buffer (ph 8.0) with ethidium bromide added (Sambrook et al. 1989). After precipitation with 95 % ethanol (EtOH) and 3 M NH_4OAc , DNA was rehydrated in 7µl of sterile water.

PCR amplified double-stranded DNA was sequenced directly using the dideoxynucleotide chain termination method (Sanger et al. 1977) and Sequenase T7 polymerase (United States Biochemical) with some modifications. Template DNA, 2 µl of 5x Sequenase reaction buffer, 1 µl of primer and 1 µl of DMSO were heated to 95-98°C for 4 min then quick frozen in a bath of dry ice and 95 % EtOH. The annealed primer and template DNA reactions were placed in a cold plastic block and allowed to equilibrate slowly to room temperature for 20-30 min. The remaining steps followed the Sequenase kit protocol except the label mix was diluted 1:50, label reaction time was reduced to 30 seconds and the termination reaction was completed at 42°C for 2-3 min (S. Palumbi, Department of Zoology, University of Hawaii, Honolulu; pers. commun.). Sequencing reactions were resolved in 6 % polyacrylamide-7 M urea gels using a salt gradient running buffer (Ausubel et al 1989). Gels were vacuum dried at 80 °C and autoradiographed on Kodak X-Omat AR film for 12-72 hr.

DNA Sequence Analysis

A 462 nucleotide portion of the control region was sequenced in all 27 individuals. Initial analysis indicated that nearly every polymorphic site in the control region was located within this sequence. Single nucleotide substitutions in the tandem repeats (located between the hypervariable region and the $tRNA^{Pro}$ gene) were not observed in five individuals with different hypervariable sequences (J.R.Brown, unpublished data). At the other end of the control region were several consensus sequence blocks (CSB) common to most vertebrate mtDNA (reviewed in Clayton 1984; Dunon-Bluteau et al. 1985). To examine nucleotide site variation in the CSB region near the 3' end of the L-strand, the sequence of the entire

nonrepeated region of the D-loop was determined in nine individuals (575 nucleotides).

DNA sequences were aligned using the computer programs Delaney (Delaney Software Ltd.) and ESEE (Cabot and Beckenbach 1989). The PHYLIP 3.3 program package was used for the bootstrapped mixed parsimony analysis (DNABOOT), the maximum likelihood method for DNA sequences (DNAML) and computing pairwise estimates of nucleotide substitution (DNADIST) corrected for multiple substitutions (Kimura 1980). A neighbour joining tree was constructed using corrected distance estimates (Saitou and Nei 1987).

In the parsimony and maximum likelihood analysis, *A. medirostris* (green sturgeon) mtDNA sequence served as an outgroup. Bootstrapped parsimony analysis involved 100 replicates. Tree topology was not affected by changing the order sequences were entered into the analysis. Empirical base frequencies were used in the maximum likelihood analysis (Felsenstein 1990). A transition to transversion ratio of four generated the best likelihood value over a wide range of tested ratios and was used in subsequent maximum likelihood analyses (Felsenstein 1990).

A test was developed to evaluate the similarity between trees based on restriction enzyme analysis of the entire mtDNA genome and trees derived from control region sequence data for the same individuals. The tested null hypothesis states that the topology of a tree based on sequence data is not significantly different from a tree derived from restriction enzyme site data. Consequently, maximum likelihood scores of sequence data fitted to both tree topologies should not be statistically different.

Prior to the test, a maximum likelihood tree was determined for the 10 restriction enzyme genotypes in Chapter I from the presence or absence of sites. The

program used was RESTML of the PHYLIP 3.3 package. These 10 restriction enzyme genotypes occurred in at least one of the 27 individuals for which control region sequence data were available. Control region sequences of individuals representing each of the 10 original restriction enzyme genotypes were used in the test. Since several monotypic restriction enzyme genotypes (At1, At2, At3 and At4) were revealed to be polytypic in the sequence analysis, different combinations of sequences were selected to represent each of the restriction enzyme genotypes. Simulations were performed on a total of ten different combinations of sequences.

In each simulation, two maximum likelihood scores were calculated. The first score was of the ten genotypes fitted to the restriction enzyme tree topology using the defined tree option and the second score was derived from maximum likelihood analysis without a predefined tree. Log likelihood values for restriction enzyme based and undefined trees were compared using the criterion of Kishino and Hasegawa (1989) as implemented in the PHYLIP 3.3 program DNAML. In this test, log likelihood values are significantly different if the difference between two scores exceeds twice the standard deviation.

MtDNA diversity was calculated from DNA sequence data using both indices of nucleotide diversity (π) and nucleon diversity (h) described in Chapter I (Nei and Li 1979; Nei and Tajima 1981).

Results

Nucleotide Change

Nineteen different control region genotypes (Cr1-19) were determined for 27 individuals (Fig. 7). Genotypes Cr9 and Cr13 comprised of 3 and 4 individuals,

respectively. Three other genotypes (Cr2, 3 and 6) were each found in two individuals. The remaining genotypes were unique to single individuals.

In the 462 nucleotide sequence there were 34 polymorphic sites among 27 individuals. Most changes were nucleotide substitutions or deletions occurring at a single site. Four recognition sequences were identified for the restriction enzymes used in the previous population analysis (Chapter I). Two restriction sites (*Hin*cII and *Ava*I) were polymorphic because of a single nucleotide substitution in their respective recognition sequences.

Nucleotide changes were strongly biased towards transitions (Table 4) as found in intraspecific comparisons of the human control region (Aquadro and Greenberg 1983). Transitional changes occurred at 26 of the 34 polymorphic sites. Relative frequencies of nucleotide change showed little preference for one transitional pathway over another (Table 5). The single transversion (position 17 in Fig. 7) was observed in only one genotype, Cr9. Single nucleotide deletion or insertions were more common than transversions. At seven positions, there occurred deletions of the nucleotides G, A or T but not C. Multiple changes involving both a transition and a deletion in different genotypes occurred at two positions.

The conserved sequence blocks at the $tRNA^{Phe}$ end of the control region were invariable in nine individuals and only one polymorphism (in 4 of 9 individuals) was observed in the last 113 nucleotides (Fig. 8). Most of the variable sites occurred at the $tRNA^{Pro}$ end of the control region (Fig. 9). This hypervariable region begins almost immediately after the terminal 3' nucleotide of the last 82 base repeat sequence (Buroker et al. 1990).

Fig. 7. Intraspecific variation in a 462 bp hypervariable area of the control region of *A. transmontanus* mtDNA. Nineteen genotypes (Cr1-19) were found in 27 individuals. After the genotype name, letters indicate whether individuals with that genotype originated from the Fraser (F) or lower Columbia (LC) Rivers or from upper Columbia River reservoirs, Bonneville Pool (BP) or The Dalles Pool (DP). Numbers following the coded river of origin correspond with restriction enzyme genotypes (At1-11). Numbers with an 'x' preceeding the river origin indicates multiple individuals sharing the same restriction enzyme genotype. For example, control region genotype Cr13 was found in one individual from the Fraser River, two from the lower Columbia and one from Dalles Pool. All four individuals shared restriction enzyme genotype At2. Dots below the top sequence indicate invariant nucleotide sites in other genotypes. Locations of recognition sites of restriction enzymes used in the previous population studies are shown.

```
                                                                                                                  60
Cr1 F6                  TGTGCACTAA CATAAAAA-C CTCTGACAAC TTAAAAATGT AGTAAGAGCC GAACATGGAG
Cr2 DP3 F0              .......... .......... .......... ....GG-... .......... ..........
Cr3 F3                  .......... .......... .......... ....GG-... .......... ..........
Cr4 F3                  .......... .......... .......... ....GG-... .......... ..........
Cr5 F7 F2               .......... ...G...... ...C..T... ......-... .......... ..........
Cr6 2xDP8               .......... ...G...... ...C..T... ......-... .......... ..........
Cr7 LC2                 .......... ...G...... ...C...... ......-... .......... ..........
Cr8 LC4                 .......... ...G...... ...C..T... ......-... .......... ..........
Cr9 3xF4                .......... ...G.T..G. ...C..T... ......-... .......... ..........
Cr10 F2                 .......... ...G...... ......T... ......-... .......... ..........
Cr11 F9                 ....T..... .......... .C........ ......-... .......... ......A..A
Cr12 F1                 .......... .......... .......... ....GG-... .......... ..........
Cr13 F2 2xLC2 DP2       .......... ...G...... ...C..T... ......-... .......... ..........
Cr14 F3                 .......... .......... .......... ....GG-... .......... ..........
Cr15 LC10               ....T..... .......... .......... ....G.-... .......... ......A...
Cr16 F11                ....T..... .......... .......... ......-... .......... ......A..A
Cr17 F1                 .......... .......... .......... ....GG-... .......... ..........
Cr18 BP2                .......... ...G...... ...C..T... ......-... .......... ..........
Cr19 BP6                .......... .......... .......... .......... .......... ..........
                              XbaI                                                                        120
Cr1 F6                  ATATGTCTAG AACATAAAGT TAATGAGATG AGGGACAATA ACTGTAGGGA TTCACAACTG
Cr2 DP3 F0              .......... .......G.. .......... .......... .T........ ..........
Cr3 F3                  .......... .......G.. .......... .......... .T........ ..........
Cr4 F3                  .......... .......G.. .......... .......... .T.......-. ..........
Cr5 F7 F2               .......... .......... .......... .......... ......A... ..........
Cr6 2xDP8               .......... .......... .......... .......... ......A... ..........
Cr7 LC2                 .......... .......... .......... .......... ......A... ..........
Cr8 LC4                 .......... .......... .......... .......... ......A... ..........
Cr9 3xF4                .......... .......... .......... .......... .......... ..........
Cr10 F2                 .......... .......... .......... .......... .......... ..........
Cr11 F9                 .......... .......G.. .......... .......... .......... ..........
Cr12 F1                 .......... .......G.. .......... .......... .T....-... ..........
Cr13 F2 2xLC2 DP2       .......... .......... .......... .......... ......A... ..........
Cr14 F3                 .......... .......G.. .......... .......... .T........ ..........
Cr15 LC10               .......... .......G.. .......... .......... .......... ..........
Cr16 F11                .......... .......G.. .......... .......... .......... ..........
Cr17 F1                 .......... .......G.. .......... .......... .T........ ..........
Cr18 BP2                .......... .......... .......... .......... ......A... ..........
Cr19 BP6                .......... .......... .......... .......... .......... ..........
```

```
                                                    AvaII                                 180
Cr1 F6              AACTATTACT GGCATCTGGT TCCTATTTCA GGTCCATTAA CAGTTATTTC CCCATAACTG
Cr2 DP3 F0          .......... .......... .......... .......... ...C...... ..........
Cr3 F3              .......... .......... .......... .......... ...C...... ..........
Cr4 F3              .......... .......... .......... .......... ...C...... ..........
Cr5 F7 F2           .......... .......... .......... .......... .......... ..........
Cr6 2xDP8           .......... .......... .......... .......... .......... ..........
Cr7 LC2             .......... .......... .......... .......... .......... ..........
Cr8 LC4             .......... .......... .......... .......... .......... ..........
Cr9 3xF4            .......... .......... .......... .......... .......... ..........
Cr10 F2             .......... .......... .......... .......... .......... ..........
Cr11 F9             .......... .......... .......... .......... .....GC... ..........
Cr12 F1             .......... .......... .......... .......... ...C...... ..........
Cr13 F2 2xLC2 DP2...........  .......... .......... .......... .......... ..........
Cr14 F3             .......... .......... .......... .......... ...C...... ..........
Cr15 LC10           .......... .......... .......... .......... .....C.... ..........
Cr16 F11            .......... .......... .......... .......... .....GC... ..........
Cr17 F1             .......... .......... .......... .......... ...C...... ..........
Cr18 BP2            .......... .......... .......... .......... .......... ..........
Cr19 BP6            .......... .......... .......... .......... .......... ..........
                                                                                      240
Cr1 F6              AACTATGTCT GGCATCTGAT TAATGTTGGA GGTACTATGA AATCCATGAC CCCACATGCC
Cr2 DP3 F0          .......... .......... .......... .......... G......... ..........
Cr3 F3              .......... .......... .......... .......... G......... ..........
Cr4 F3              .......... .......... .......... .......... G......... ..........
Cr5 F7 F2           .......... .......... .......... .......... G......... ..........
Cr6 2xDP8           .......... .......... .......... .......... G......... ..........
Cr7 LC2             .......... .......... .......... .......... G......... ..........
Cr8 LC4             .......... .......... .......... .......... .......... ..........
Cr9 3xF4            .......... .......... .......... .......... .......... ..........
Cr10 F2             .......... .......... .......... .......... G......... ..........
Cr11 F9             .......... .......... .......... .......... .......... ..........
Cr12 F1             .......... .......... .......... .......... G......... ..........
Cr13 F2 2xLC2 DP2...........  .......... .......... .......... G......... ..........
Cr14 F3             .......... .......... .......... .......... G......... ..........
Cr15 LC10           .......... .......... .......... .......... .......... ..........
Cr16 F11            .......... .......... .......... .......... .......... ..........
Cr17 F1             .......... .......... .......... .......... G......... ..........
Cr18 BP2            .......... .......... C......... .......... G......... ..........
Cr19 BP6            .......... .......... .......... .......... .......... ..........
```

HincII 300

```
Cr1 F6              GAGAATCTTG TCAACATTTG GTATTTTTTA TTCGGGTTTC CATTCACTGA CATGTAGAAC
Cr2 DP3 F0          .......... C......... .......... .......... .......... ..........
Cr3 F3              .......... C......... .......... .......... .......... ..........
Cr4 F3              .......... C......... .......... .......... .......... ..........
Cr5 F7 F2           .......... .......... ....C..... .......... .......... ..........
Cr6 2xDP8           .......... C......... ....C..... .......... .......... ..........
Cr7 LC2             .......... .......... ....C..... .......... .......... ..........
Cr8 LC4             .......... .......... .......... .......... .......... ..........
Cr9 3xF4            .......... .......... .......... .......... .......... ..........
Cr10 F2             .......... .......... ....C..... .......... .......... ..........
Cr11 F9             .......... .......... .......... .......... .......... ..........
Cr12 F1             .......... C......... .......... .......... .......... ..........
Cr13 F2 2xLC2 DP2   .......... .......... ....C..... .......... .......... ..........
Cr14 F3             .......... C......... .......... .......... .......... ..........
Cr15 LC10           .......... .......... .......... .......... .......... ..........
Cr16 F11            .......... .......... .......... .......... .......... ..........
Cr17 F1             .......... C......... .......... .......... .......... ..........
Cr18 BP2            .......... .......... ....C..... .......... .......... ..........
Cr19 BP6            A......... .......... .......... .......... .......... ..........
```

AvaI 360

```
Cr1 F6              TCCTTCAGAG AAGAACAACA AGGTGGAACA TACGTGACTG CCCGAGAGGA TGAATAATGA
Cr2 DP3 F0          .......... .......... .......... .......... T......... ..........
Cr3 F3              .......... .......... .......... ....C..... T......... ..........
Cr4 F3              .......... .......... .......... .......... T......... ..........
Cr5 F7 F2           .......... .......... .......... ....C..... T......... ..........
Cr6 2xDP8           .......... .......... .......... ....C..... T......... ..........
Cr7 LC2             .......... .......... .......... ....C..... T......... ..........
Cr8 LC4             .......... .......... .......... ....C..... T......... ..........
Cr9 3xF4            .......... .......... .......... ....C..... T......... ..........
Cr10 F2             .......... .......... .......... ....C..... T......... ..........
Cr11 F9             .......... .......... .......... .......... .........- ..........
Cr12 F1             .......... .......... .......... .......... .......... ..........
Cr13 F2 2xLC2 DP2   .......... .......... .......... ....C..... T......... ..........
Cr14 F3             .......... .......... .......... .......... T.......A. ..........
Cr15 LC10           .......... .......... .......... .......... T......... ..........
Cr16 F11            .......... .......... .......... .......... T........G ..........
Cr17 F1             .......... .......... .......... .......... .......... ..........
Cr18 BP2            .......... .......... .......... ....C..... T......... ..........
Cr19 BP6            .......... .......... .......... .......... .......... ...G......
```

```
                                                                                                  420
Cr1 F6              ATGGTACAAT GACATATCCC T-GATGTCAC ACATGGCCTG TGCTGTGTAC AGAGAGATGT
Cr2 DP3 F0          .......... .......... .......... .......... .......... ..........
Cr3 F3              .......... .......... .......... .......... .......... ..........
Cr4 F3              .......... .......... .......... .......... .......... ..........
Cr5 F7 F2           .......... .......... .......... .......... .......... ..........
Cr6 2xDP8           .......... .......... .......... .......... .......... ..........
Cr7 LC2             .......... .......... .......... .......... .......... ..........
Cr8 LC4             .......... .......... .......... .......... .......... ..........
Cr9 3xF4            .......... .......... .......... .......... .......... ..........
Cr10 F2             .......... .......... .......... .......... .......... ..........
Cr11 F9             .......... .......... .T........ .......... .....C.... .....A....
Cr12 F1             .......... .......... .......... .......... .......... ..........
Cr13 F2 2xLC2 DP2...........  .......... .......... .......... .......... ..........
Cr14 F3             .......... .......... .......... .......... .......... ..........
Cr15 LC10           .......... .......... .T........ .......... .......... .....A....
Cr16 F11            .......... .......... .T........ .......... .....C.... .....A....
Cr17 F1             .......... .......... .......... .......... .......... ..........
Cr18 BP2            .......... .......... .......... .......... .......... ..........
Cr19 BP6            .......... .......... .......... .......... .......... ..........
                                                                          462
Cr1 F6              TTCACAGAGC CTGGTTTTA- TCTTTTCACA CGACAATCAT GG
Cr2 DP3 F0          .......... .......... .......... .......... ..
Cr3 F3              .......... .......... .......... T......... ..
Cr4 F3              .......... .......... .......... T......... ..
Cr5 F7 F2           .......... .........A .......... T......... ..
Cr6 2xDP8           .......... .......... .......... T......... ..
Cr7 LC2             .......... .........A .......... T......... ..
Cr8 LC4             .......... .......... .......... T......... ..
Cr9 3xF4            .......... .......... .......... T......... ..
Cr10 F2             .......... .........A .......... T......... ..
Cr11 F9             .......... .......... .......... T......... ..
Cr12 F1             .......... .......... .......... T......... ..
Cr13 F2 2xLC2 DP2...........  .......... .......... T......... ..
Cr14 F3             .......... .......... .......... T......... ..
Cr15 LC10           .......... .......... .......... T......... ..
Cr16 F11            .......... .......... .......... T......... ..
Cr17 F1             .......... .......... .......... T......... ..
Cr18 BP2            .......... .......... .......... T......... ..
Cr19 BP6            .......... .......... .......... .......... ..
```

Table 4. Observed single base differences in a 462 base sequence of control region of white sturgeon mtDNA.

Reference Sequence			Nucleotide Introduced					
Nucleotide replaced	No. in sequence	Proportion of sequence	G	A	T	C	Deletion	Total
G	94	0.20	-	6	0	0	1	7
A	145	0.31	8	-	1	0	1	10
T	130	0.28	0	0	-	9	0	9
C	90	0.19	0	0	5	0	0	5
Insertion	3	0.01	1	1	1	0	-	3
Total	462							34

Table 5. Relative base substitution frequencies in white sturgeon mtDNA control region as calculated from Gojobori et al. (1982).

Substitution Type		
H Strand	L Strand	Relative frequency
Transitions		
A -> G	T -> C	0.275
T -> C	A -> G	0.219
G -> A	C -> T	0.211
C -> T	G -> A	0.255
Transversions		
T -> A	A -> T	0.028

Fig. 8. Alignment of the 113 bp sequence at the 3' end of the control region from 9 individuals. Numbering scheme continues from Fig. 7. Consensus sequence blocks (CSB 1, 2 and 3) found in vertebrate control regions are shown. Four control region genotypes (Cr1, Cr4, Cr11 and Cr12) were sequence type A and five genotypes (Cr2, Cr9, Cr10, Cr17 and Cr19) were type B.

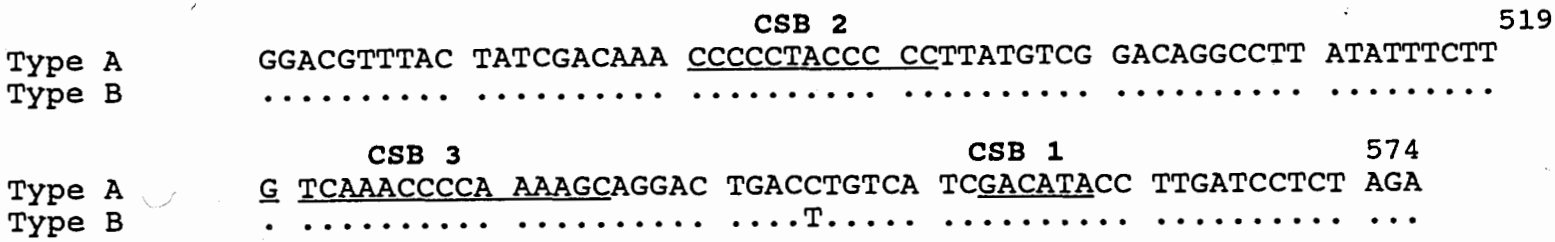
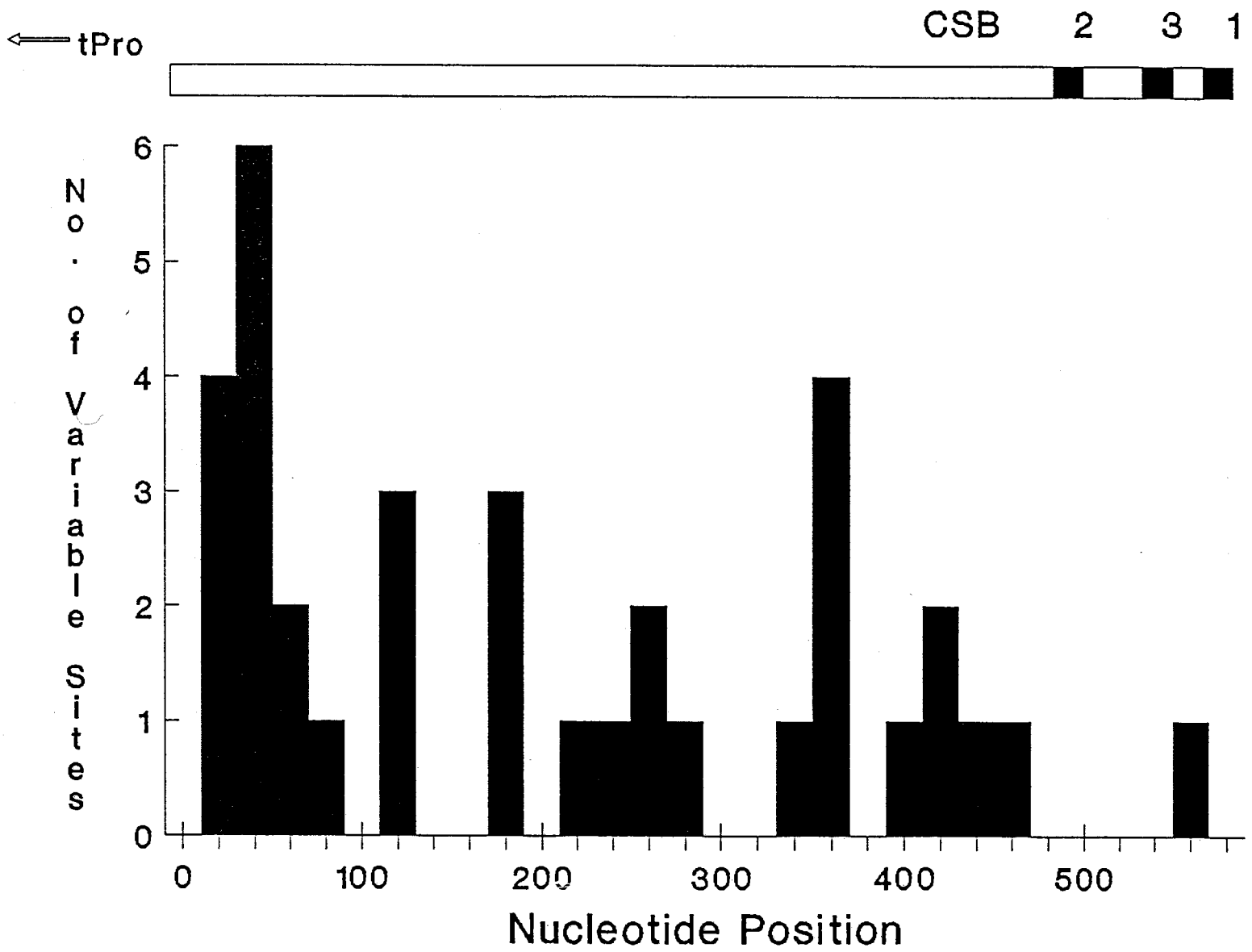

CSB 2
519
Type A GGACGTTTAC TATCGACAAA CCCCCTACCC CCTTATGTCG GACAGGCCTT ATATTTCTT
Type B ....
CSB 3
CSB 1
574
Type A G TCAAACCCCA AAAGCAGGAC TGACCTGTCA TCGACATACC TTGATCCTCT AGA
Type B . ...T..... ....

Fig. 9 Distribution of variable sites in the white sturgeon control region. The number of variable positions in blocks of 20 nucleotides are shown. Schematic drawing at top illustrates the relative positions of consensus sequence blocks (CSB).


tPro
CSB
2
3
1
No. of Variable Sites
0
1
2
3
4
5
6
0
100
200
300
400
500
Nucleotide Position

Sequence Based Phylogeny

The number of nucleotide differences and corrected percent sequence differences (Kimura 1980) between genotypes are shown in Table 6. Genotypic differences varied between 0.01 - 4.07 % over 462 nucleotides.

Three methods of phylogenetic analysis resulted in generally congruent tree toplogies. A neighbor-joining tree was constructed using corrected distance estimates (Fig. 10). *A. medirostris* sequence data was used to root bootstrapped maximum parsimony (Fig. 11) and maximum likelihood trees (Fig.12). The trees based on control sequence data were also generally congruent with the restriction enzyme tree shown in Chapter I. However, several restriction enzyme haplotypes were revealed to be polytypic by the sequence data. Control region sequence analysis revealed 19 genotypes in 27 individuals as opposed to only 10 genotypes in 178 individuals with restriction enzyme analysis. In addition, bootstrapped parsimony and maximum likelihood methods showed that internal nodes of sequence data trees (Fig. 11 and 12) were statistically better supported than internal nodes of the restriction enzyme data trees (Fig. 2 and 3).

Four major clusters of intraspecific genotypes were consistent in all three phylogenies (Fig. 10, 11 and 12). Cluster *A* consisted of six genotypes predominantly from the Fraser River. All individuals with restriction enzyme genotypes At1 and At3 were present in this cluster as well as the 82 year old individual (F 0) found the shore of the Fraser River. The control region of this sturgeon had complete sequence identity with that of an individual from the Columbia River with restriction enzyme genotype At3.

The remaining clusters were a mixture of Fraser and Columbia River fish. A particular restriction enzyme genotype was always exclusive to one of the four

clusters. Cluster *B* was the largest cluster with 8 control region genotypes. Some restriction enzyme genotypes remained monotypic in comparisons of control region sequences. Genotype Cr13 occurred in four individuals with restriction enzyme genotype At2 which were geographically dispersed to the Fraser River and the upper and lower Columbia River. Previously, it was shown that more than 60 % of all sturgeon analyzed had restriction enzyme genotype At2. Control region genotype Cr9 was found in three Fraser River fish with restriction enzyme genotype At4.

Two individuals from the upper Columbia River with restriction enzyme genotype At8 were also indistinguishable in sequence comparisons (Cr6). Control region genotype Cr5 is the only instance where a polytypic relationship detected by restriction enzymes (At2 and At7) was monotypic in sequence data comparisons. However, these genotypes were shown to be closely related in the restriction enzyme analysis (Fig. 2 and 3).

In cluster *C*, the genotypes Cr16, Cr11 and Cr15, represented three rare restriction enzyme genotypes, At11, At9 and At10, respectively. These three genotypes clustered similarly in the restriction enzyme phylogeny because all shared two unique *BclI* sites. Finally, cluster *D* consisted of two individuals from the Fraser and Columbia Rivers with a common restriction enzyme genotype (At6) but slightly divergent control region sequences.

Table 6. Intraspecific comparisons of a 462 nucleotide sequence of the control region of *A. transmontanus*. Number of nucleotide substitutions and deletion/insertions shown in the upper diagonal and percent sequence differences, corrected for multiple substitutions (Kimura 1980), shown in the lower diagonal.

	Cr1	Cr2	Cr3	Cr4	Cr5	Cr6	Cr7	Cr8	Cr9	Cr10	Cr11	Cr12	Cr13	Cr14
Cr1		9	11	11	11	11	10	8	9	9	13	10	10	11
Cr2	2.01		2	2	14	12	13	13	14	12	18	3	13	2
Cr3	2.23	0.22		2	15	10	11	11	12	10	18	3	11	2
Cr4	2.01	0.01	0.22		14	12	13	13	14	12	19	3	13	2
Cr5	2.01	2.69	2.46	2.69		2	1	3	6	2	20	14	3	14
Cr6	2.23	2.46	2.23	2.47	0.22		3	3	6	4	20	12	1	12
Cr7	1.78	2.46	2.23	2.47	0.22	0.44		4	7	3	17	13	2	13
Cr8	1.56	2.69	2.46	2.69	0.44	0.66	0.66		3	5	17	13	1	13
Cr9	1.56	2.69	2.46	2.70	0.89	1.11	1.11	0.44		6	18	15	5	14
Cr10	1.56	2.23	2.01	2.24	0.44	0.66	0.66	0.89	0.89		18	13	3	12
Cr11	2.24	3.38	3.61	3.39	3.84	4.07	3.61	3.38	3.38	3.38		17	19	18
Cr12	1.79	0.22	0.44	0.22	2.92	2.69	2.69	2.92	2.93	2.47	3.16		13	3

Table 6 continued.

	Cr1	Cr2	Cr3	Cr4	Cr5	Cr6	Cr7	Cr8	Cr9	Cr10	Cr11	Cr12	Cr13	Cr14
Cr13	2.01	2.69	2.46	2.69	0.01	0.22	0.22	0.44	0.89	0.44	3.84	2.92		13
Cr14	2.23	0.22	0.44	0.22	2.92	2.69	2.69	2.92	2.92	2.46	3.61	0.44	2.92	
Cr15	1.78	2.01	2.23	2.01	2.92	3.14	2.69	2.46	2.46	2.46	1.33	2.24	2.92	2.23
Cr16	2.46	3.14	3.37	3.15	3.60	3.83	3.37	3.14	3.15	3.14	0.44	3.38	3.60	3.37
Cr17	1.78	0.22	0.44	0.22	2.92	2.69	2.69	2.92	2.92	2.46	3.15	0.01	2.92	0.44
Cr18	2.23	2.92	2.69	2.92	0.22	0.44	0.44	0.66	1.11	0.66	4.07	3.15	0.22	3.14
Cr19	0.66	2.23	2.46	2.24	2.23	2.46	2.01	1.78	1.79	1.78	2.47	2.01	2.23	2.46

Table 6 continued.

	Cr15	Cr16	Cr17	Cr18	Cr19
Cr1	10	13	9	11	2
Cr2	12	17	2	14	11
Cr3	11	16	2	12	13
Cr4	11	16	2	14	13
Cr5	16	18	14	2	13
Cr6	15	18	12	2	13
Cr7	14	17	13	3	12
Cr8	12	15	13	3	10
Cr9	13	16	14	6	11
Cr10	13	16	12	4	11
Cr11	7	2	16	20	15
Cr12	12	17	1	14	12
Cr13	14	17	13	1	12
Cr14	11	16	2	14	13

	Cr15	Cr16	Cr17	Cr18	Cr19
Cr15		5	11	15	12
Cr16	1.11		16	18	15
Cr17	2.23	3.37		14	11
Cr18	3.14	3.83	3.14		13
Cr19	2.01	2.69	2.01	2.46	

Fig. 10. Neighbor-joining tree of *A. transmontanus* control region genotypes using sequence divergence estimates corrected for multiple changes (Kimura 1980). Labels of genotypes, river of origin and restriction enzyme genotype follow Figure 7. Genotypes comprised of more than one individual are marked (*). Estimated number of base changes per nucleotide site are shown along the branches. Major phyletic clusters are marked *A* to *D*.

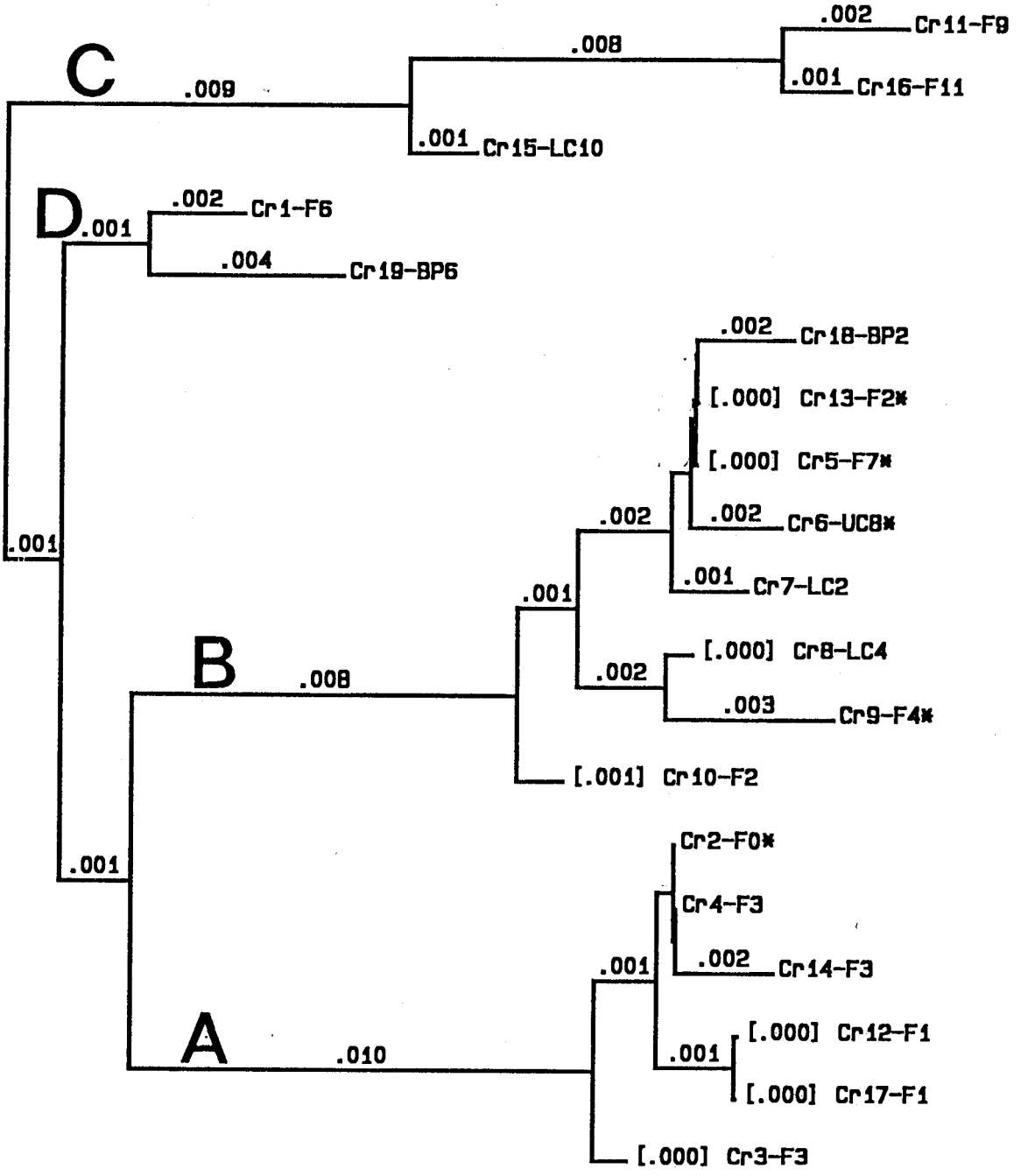

C
.009
.008
.002
Cr11-F9
.001
Cr16-F11
.001
Cr15-LC10
D
.001
.002
Cr1-F6
.004
Cr19-BP6
.001
B
.008
.002
Cr18-BP2
[.000] Cr13-F2*
[.000] Cr5-F7*
.002
.002
Cr6-UC8*
.001
Cr7-LC2
.001
[.000] Cr8-LC4
.002
.003
Cr9-F4*
[.001] Cr10-F2
.001
A
.010
Cr2-F0*
Cr4-F3
.001
.002
Cr14-F3
[.000] Cr12-F1
.001
[.000] Cr17-F1
[.000] Cr3-F3

Fig. 11. Maximum parsimony tree of *A. transmontanus* control genotypes. Labels of genotypes follow Figure 7. River of origin and restriction enzyme genotype are simlarly coded and are underlined. Major phyletic clusters are marked A to D. Since vertical separation does not imply genetic distance, Cluster D genotypes are actually closely related. Numbers along the branches indicate the occurrence of particular nodes in 100 replicates.

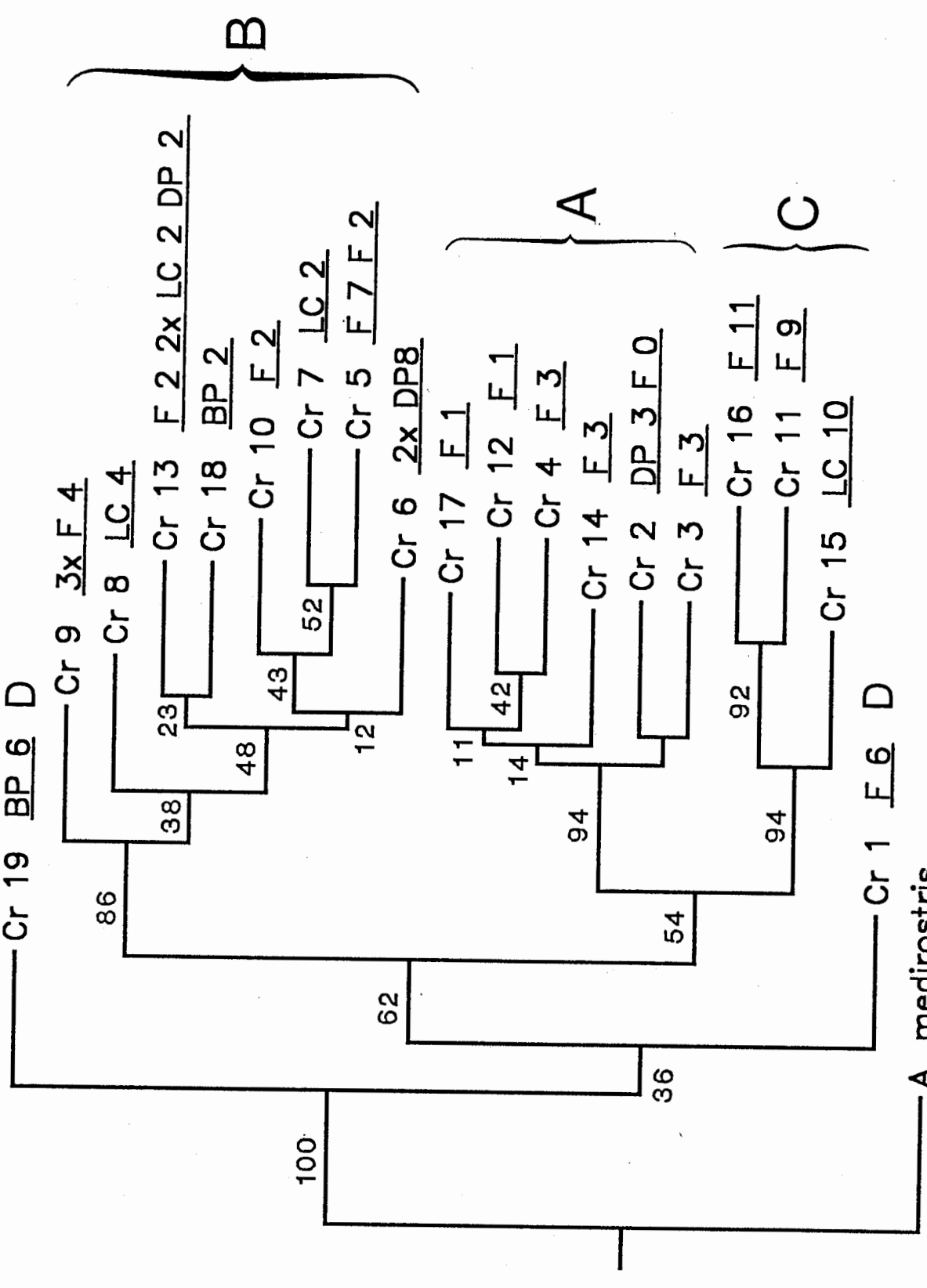
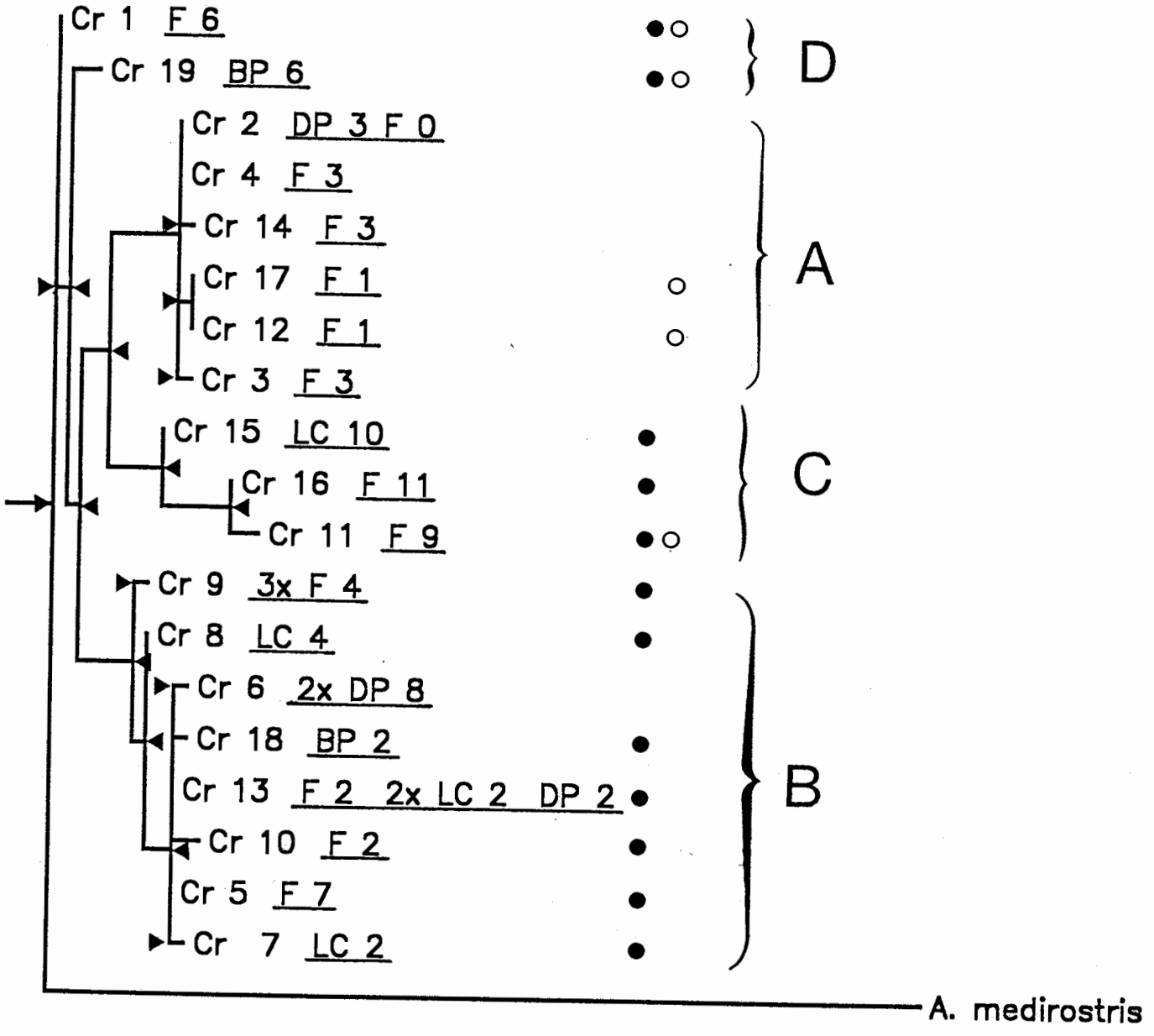

B
A
C
Cr 19 BP 6 D
Cr 9 3x F 4
Cr 8 LC 4
Cr 13 F 2 2x LC 2 DP 2
Cr 18 BP 2
Cr 10 F 2
Cr 7 LC 2
Cr 5 F 7 F 2
Cr 6 2x DP8
Cr 17 F 1
Cr 12 F 1
Cr 4 F 3
Cr 14 F 3
Cr 2 DP 3 F 0
Cr 3 F 3
Cr 16 F 11
Cr 11 F 9
Cr 15 LC 10
Cr 1 F 6 D
A. medirostris
100
36
62
54
94
94
86
38
48
23
12
43
52
14
11
42
92

Fig. 12 Maximum likelihood tree of *A. transmontanus* control region genotypes. Labels of genotypes, river of origin and restriction enzyme genotype follow that of Figure 7. Major phyletic clusters are marked A - D. Significant branching points ($p < 0.05$) are marked with arrows. The *ln* likelihood value for this tree is -1083.04. The presence of polymorphic restriction enzyme sites for *Hinc*II (solid circles) and *Ava*I (open circles) are indicated for each genotype.


Cr 1 F 6
Cr 19 BP 6
Cr 2 DP 3 F 0
Cr 4 F 3
Cr 14 F 3
Cr 17 F 1
Cr 12 F 1
Cr 3 F 3
Cr 15 LC 10
Cr 16 F 11
Cr 11 F 9
Cr 9 3x F 4
Cr 8 LC 4
Cr 6 2x DP 8
Cr 18 BP 2
Cr 13 F 2 2x LC 2 DP 2
Cr 10 F 2
Cr 5 F 7
Cr 7 LC 2
D
A
C
B
A. medirostris

Comparison of Restriction Enzyme and Sequence Methods

Maximum likelihood values were compared for sequence data, first fitted to the restriction enzyme tree followed by analysis without predefined tree toplogy (Table 7). Ten simulations were done with different control region sequences (genotypes) representing the corresponding restriction enzyme genotype. Maximum likelihood values were better (more positive) for the undefined trees in all ten simulations. In nine simulations there was a significant difference in scores.

Maximum likelihood restriction enzyme and unconstrained trees similarly cluster genotypes At2, At4, At7 and At8 (Fig. 13a & b). Rare genotypes At9, At10 and At11 are also grouped together. However, internal nodes of the restriction enzyme tree differ from the consensus unconstrained tree. Genotypes At1 and At3 are nc as closely related in the restriction enzyme toplogy. Two of the unconstrained trees had the wrong topology with the whole tree rooted to the rare restriction enzyme genotype, At9 (Fig. 13c).

Genetic Diversity

Estimates of genetic diversity derived from sequence data show a similar trend as estimates from restriction enzyme data (Table 8). MtDNA nucleotide diversity estimates are higher for white sturgeon sequences from the Fraser River compared with those of individuals from Columbia River. However, sequence based comparisons indicate the lower Columbia population is slightly less genetically diverse than the upper Columbia River population. This is opposite to the conclusion from restriction enzyme estimates perhaps as a result of too small a sample size to compare individuals from the upper and lower Columbia.

Table 7. Pairwise comparisons of *ln* likelihood (*ln*L) values of maximum likelihood trees derived from control region DNA sequence data 1) fitted to the restriction enzyme (R.E.) tree or 2) with the tree undefined (U.D.). Each simulation run involved sequences of a different individuals representing each of the 10 restriction enzyme genotypes (At1 - At11). More positive *ln* likelihood scores indicate a better fit. Standard deviations (S.D.) are shown and significantly different scores (based on the criterion of Kishino and Hasegawa 1989) are indicated (*).

Run No.	R.E. *ln*L	U.D. *ln*L	Difference	S.D.
1	-1065.43	-1030.98	-34.44 *	16.39
2	-1070.56	-1039.80	-30.76	15.87
3	-1078.11	-1038.10	-40.01 *	14.97
4	-1076.82	-1032.38	-44.44 *	15.95
5	-1068.73	-1030.62	-38.11 *	14.72
6	-1061.91	-1020.36	-41.55 *	15.72
7	-1066.76	-1034.11	-32.66 *	15.94
8	-1059.15	-1019.13	-40.02 *	14.84
9	-1059.59	-1019.58	-40.02 *	14.83
10	-1068.73	-1039.03	-29.70 *	14.71

Fig. 13 Maximum likelihood trees based on control region sequences fitted to the restriction enzyme defined tree (A) and not fitted to a predefined tree (B & C). Tree B is the consensus undefined tree (8 of 10 simulations) and Tree C is the variant undefined tree (2 of 10 simulations). Restriction enzyme genotypes are shown and major rearrangements in the unconstrained trees relative to the restriction enzyme defined tree are marked with arrows.

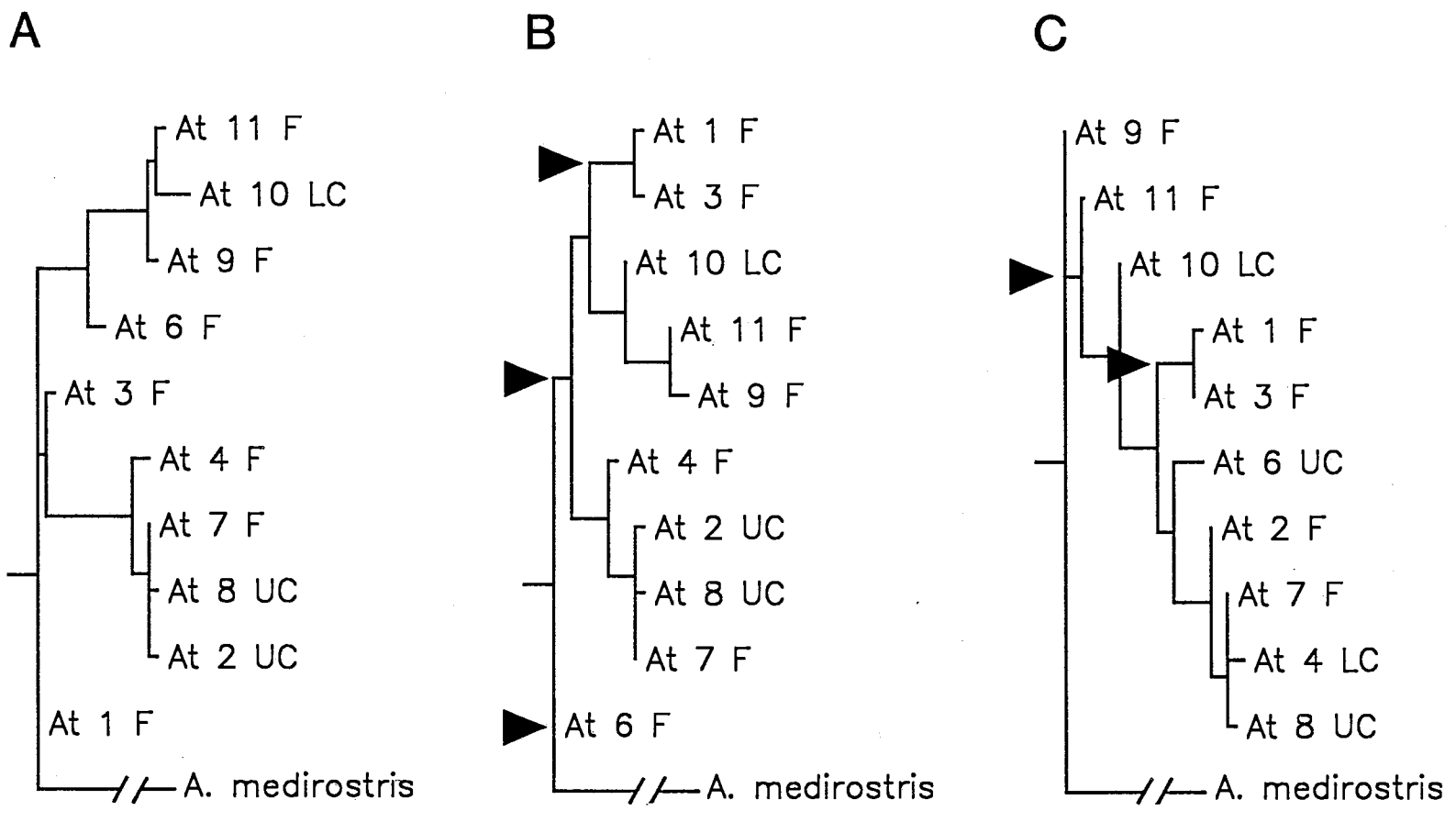

A
At 11 F
At 10 LC
At 9 F
At 6 F
At 3 F
At 4 F
At 7 F
At 8 UC
At 2 UC
At 1 F
A. medirostris
B
At 1 F
At 3 F
At 10 LC
At 11 F
At 9 F
At 4 F
At 2 UC
At 8 UC
At 7 F
At 6 F
A. medirostris
C
At 9 F
At 11 F
At 10 LC
At 1 F
At 3 F
At 6 UC
At 2 F
At 7 F
At 4 LC
At 8 UC
A. medirostris

Table 8. Estimates of mtDNA diversity (h) and nucleotide diversity (π) using the methods of Nei and Tajima (1981) and Nei and Li (1979), respectively, from control region sequence.

Location	n	h	π
Fraser and Columbia River	27	0.97	1.001
Fraser River	16	0.97	0.390
Columbia River	8	0.88	0.116
Lower	4	0.90	0.019
Upper	5	0.93	0.036

Discussion

Nucleotide Variation

The high transition to transversion ratio (28:1) observed in the control region of white sturgeon mtDNA is typical of vertebrate mtDNA. The human control region has a 32-fold bias favouring transitions over transversions (Aquadro and Greenberg 1983). A similar bias for transitions has been observed in the cytochrome *b* gene from a wide variety of animal species (Kocher et al. 1989) and in four other protein coding genes from six species of salmonids (Thomas and Beckenbach 1989). The L-strand in the sturgeon control region is slightly A-T rich but there is little bias favouring one transitional change over another (Table 5). Relative frequencies of transitional pathways were also found to vary little in the human control region (Aquadro and Greenberg 1983). In white sturgeon, single nucleotide deletions and insertions are more frequent than transversions. Postreplicative repair defiencies in the mtDNA polymerase has been suggested as the cause of the large bias towards transitions and deletion or insertions (Tomkinson and Linn 1986; Thomas and Beckenbach 1989).

In white sturgeon mtDNA, the greatest variation occurred near the $tRNA^{Pro}$ end of the control region which is adjacent to the repeated sequences (Fig 9). The opposite $tRNA^{Phe}$ end of the control region, which included the consensus sequence blocks (CSB's), lacked any significant polymorphisms. Nucleotide variation in the human control region has a similar distribution suggesting comparable selective constraints on this region across the vertebrates (Horai and Hayasaka 1990). The short consensus blocks in the sturgeon control have near complete sequence identity and position to those of higher vertebrates (Buroker et al. 1990). While nucleotide

polymorphism in the vertebrate control region is extensive, there also appears to be significant conservation of the general structural organization of the region (Clayton 1982, Saccone et al. 1987).

Rate of Nucleotide Change

Previously, concurrent intraspecific comparisons of variation in the control region and the entire mtDNA molecule were only available for humans. The average rate of sequence divergence in the human control region is 8.4 % per million years as estimated from the fossil record (Viligant et al. 1989). Restriction enzyme analysis suggests that the average rate of divergence of the entire mtDNA molecule is about 2 % per million years (Cann et al. 1987). Therefore, the rate of sequence divergence for the control region in human mtDNA is about 4.2 times greater than that of the entire mtDNA molecule. A fourfold higher rate for the control region in human mtDNA has been confirmed elsewhere (Aquadro and Greenberg 1983, Horai and Hayasaka 1990).

In *A. transmontanus*, mean pairwise percent divergence of control region sequences and restriction enzyme sites in the entire mtDNA are 2.3 % and 0.54 %, respectively. In sturgeon, it appears that the rate of nucleotide change in the control region is about 4.3 times higher than the remaining mtDNA molecule, essentially the same relative rate as observed in human mtDNA. Restriction sites are widely distributed throughout white sturgeon mtDNA thus it is unlikely that sequence divergence estimates are biased due to nonrandom clumping of restriction sites (Fig. 1). Divergence estimates for the control region should also be unbiased because the entire hypervariable region was sequenced.

Morphological and fossil evidence suggest that sturgeon evolved nearly 200 million years ago, making sturgeon one of the most primitive extant fish species (Gardiner 1984). If the relative rates of nucleotide change between the control region and mtDNA coding genes are identical between very primitive and advanced vertebrates then these same relative rates of evolutionary change in mtDNA may be constant over most vertebrate species. If true, the ratio of sequence divergences for intraspecific pairwise comparisons of mtDNA genes (eg. control region vs cytochrome *b*) are likely to be the same across different vertebrate taxa.

Gene content and order has been the same for all vertebrate mtDNA (reviewed in De Giorgi and Saccone 1989) examined with the exception of minor rearrangements in gallinaceous birds (Desjardins and Morais 1990). Therefore, constraints on variation in the organization of the mtDNA genome must be considerable. While nucleotide substitutions and length variation are extensive in the white sturgeon control region, CSB and TAS (termination-association sequences) blocks are arranged like the mammalian control region (Clayton 1984). Such selective constraints on mtDNA functional organization may contribute towards the possible maintainence of intramolecular relative rates of nucleotide change across a wide range of taxa.

Selective constraints may act to remove mtDNA molecules with one or more genes that have too high of a mutation rate. Each gene may have its own allowable rate of nucleotide substitution set by functional constraints on its coded protein. Selection at the cellular level could fix intergenic differences in sequence divergence rates. The lack of recombination between mtDNA molecules would also serve to maintain these intergenic rate differences.

Alternatively, control region versus whole mtDNA substitution rates may be the same for sturgeon and humans because the two species have similar generation times. Sturgeon reach sexual maturity between 11 and 22 years of age and probably reproduce in cycles of two to five years (Roussow 1957; Semakula and Larkin 1968; Doroshov 1985). Differences between man and rodents in the rate of nucleotide substitutions in nuclear genes has been suggested to be linked to shorter generation time in rodents (Wu and Li 1985). If generation time is a major factor in setting the molecular clock, then little difference should be expected in the rate of nucleotide change between human and sturgeon mtDNA.

Unfortunately, the incomplete fossil record for the genus *Acipenser* prevents further testing of the rate-constancy hypothesis (Zuckerkandl and Pauling 1965; Wilson et al. 1985). The mean rate of divergence of 2 % per million years for the entire mtDNA molecule is an estimate based on primate data but evidence suggests that mtDNA in other mammalian species (Wilson 1985 et al.) and birds (Schields and Wilson 1987) evolves at the same rate. However, the rate of amino acid replacement in the cytochrome *b* gene, appears to be slower in fish (Kocher at al. 1989).

The relative consistency of the ratio of the rate of nucleotide change between the control region and entire mtDNA among sturgeon and humans is insufficient evidence to support the universality of the vertebrate molecular clock. Intraspecific control region and restriction enzyme comparisons on species of fish with better fossil records are clearly required.

Intraspecific Phylogeny

Intraspecific trees produced by neighbour joining, parsimony and maximum likelihood methods have similar topologies. All three phylogenetic methods determined four major clusters of control region genotypes. Alternative parsimony trees examined also had these clusters and differences between trees involved only minor rearrangements of terminal branches.

Several factors may be responsible for the lack of significant geographical structuring of white sturgeon populations. Migration behaviour of *A. transmontanus* is enigmatic and conventional tagging and allozyme studies have contributed little to our understanding of population structure in this species (reviewed in Binkowski and Doroshov 1985). Individual longevity, high fecundity, multiple spawnings and marine migratory behaviour, are factors which could facilitate a high level of gene flow between white sturgeon populations. In addition, the mtDNA diversity of white sturgeon populations has been greatly reduced through bottlenecks resulting from Pleistocene glacial events and human exploitation (Wilson et al. 1985; also Chapter I).

The Columbia River basin was the post-glacial source of founders for many Northwest Pacific fish populations, including white sturgeon (MacPhail and Lindsay 1986). Restriction enzyme analysis of a 178 individuals revealed that sturgeon populations in the Fraser and Columbia Rivers share the majority of mtDNA genotypes (Chapter I). That analysis suggested migrants from the Columbia River colonized the Fraser River after the last glacial retreat (about 12,000 years ago) and contemporary gene flow may still be significant. The sequence phylogeny generally supports that conclusion by clustering together similar restriction enzyme genotypes. However, control regions comparisons detected at least an order of magnitude greater

sequence variation than restriction enzyme analysis. This greater resolution found some significant but previously undetected population structures.

A distinctive grouping (cluster *A*) of control region genotypes, predominantly comprised of individuals from the Fraser River, was detected in the phylogenetic analyses (Fig. 10, 11 and 12). This cluster included all individuals with restriction enzyme genotypes At1 and At3 as well as the very old, large sturgeon (designated genotype F 0 in Figs. 10, 11 and 12) found on the shore of the Fraser River. Restriction enzyme analysis had shown genotype At1 to be unique to sturgeon caught in a localized area of the mid Fraser River. Restriction enzyme genotype At3 was mainly found in Fraser River sturgeon ($n = 19$) but two individuals with same genotype were caught in the upper Columbia River. The UPGMA phylogeny of restriction enzyme data did not clearly show the geographical distinction of genotypes At1 and At3 (Fig. 2). However, these genotypes were clustered together in parsimony analysis of the same data (Fig. 3).

The estimated rate of nucleotide substitutions in the human control region is 8.4 % per million years (Viligant et al. 1989). Application of this rate to white sturgeon mtDNA is highly provisional although the relative rate of sequence divergence between the control region and the rest of the sturgeon mtDNA genome appears to be consistent with humans. Mean pairwise distances between genotypes in cluster *A* and cluster *B* (with the greatest abundance of Columbia River genotypes) is 2.68 % ($\pm$0.26 %). Assuming an equal rate of nucleotide substitutions in both lineages, the two clusters diverged about 160,000 years ago. However, mean pairwise distances among genotypes within Cluster *A* is 0.29 % ($\pm$0.11 %) and the estimated time of divergence is about 17,300 years ago. The maternal ancestor of contemporary cluster *A* genotypes likely diverged from other Columbia River lineages

prior to the last glacial excursion. Divergence of intracluster genotypes was more recent and closer to the time of colonization of the Fraser River. These time estimates are speculative because recent bottleneck effects could also reduce intraspecific divergence estimates (Wilson et al. 1985; Chapter I).

The phylogeny of mtDNA genotypes suggest that sturgeon have a low fidelity to their natal rivers. For example, genotype At3 occurred in two individuals caught in the Columbia River above Bonneville Dam. The control region of one of these Columbia River individuals was sequenced and found to have complete nucleotide identity with the control region (genotype Cr2) of the large, female sturgeon found on the shore of the Fraser River (genotype F 0).

Although the present analysis can not precisely determine very recent maternal relationships between individuals, it is suggested that sturgeon with genotype Cr2 are closely related on the basis of age cohort analysis. The estimated ages of Fraser and Columbia River individuals with this genotype are 82 years and 20 years, respectively (D. Lane pers. commun.). Therefore, the older individual likely hatched around 1907 in either the Fraser or Columbia River and reached reproductive maturity some 11 to 22 years later between 1918 to 1929. Sturgeon in the upper Columbia River had access to marine coastal waters up to the completion of the Bonneville Dam in the early 1930's. Therefore, this individual could have spawned in the upper Columbia and returned to the lower river or the ocean before the dam closure. It is unknown from mtDNA analysis if this female subsequently spawned in the Fraser or the lower Columbia Rivers. Cohorts of the 1920 to 1930 year class should have been sexual mature by the early 1950's. The upper Columbia River sturgeon with the same genotype, Cr2, likely hatched around 1968 thus was possibly the progeny of a female from the 1920 to 1930 year class.

The remaining three phyletic clusters are comprised of geographically dispersed genotypes. The most abundant grouping of control regions genotypes, cluster *B*, included 14 of the 27 individuals analyzed. This cluster includes the restriction enzyme genotype At2 which had been previously found in 61 % of sturgeon sampled from the Fraser and Columbia Rivers. Four individuals captured in the Fraser and Columbia Rivers had complete identity of both control region sequence (Cr 13) and restriction enzyme sites (At2). Two other control region genotypes (Cr6 and Cr9) in this cluster had two or more individuals which also shared restriction enzyme genotypes.

Further sequencing of the control region probably would not have revealed any new polymorphic sites in these individuals since the entire hypervariable region was used in the analysis. Individuals that share both control region and restriction enzyme site genotypes are identical at nearly 780 nucleotide positions or about 5 % of the mtDNA genome. This low level of sequence divergence among genotypes is consistent with the conclusion of a recent genetic bottleneck from the restriction enzyme analysis.

The generally low concordance of both restriction enzyme and control region genotypes with geographical regions suggests that significant mtDNA gene flow occurs in this species. The broad geographical distribution of both common and rare mtDNA genotypes is typical of species with high levels of gene flow such as American eels (*Anguilla rostrata*; Avise et al. 1986) and red-winged blackbirds (*Agelaius phoeniceus*; Ball et al. 1988). However, relatively few female migrants per generation will limit geographical differentiation of two populations (Slatkin 1987). The close genetic relationship of the old sturgeon found in the Fraser River to some upper Columbia River sturgeon, shows that is highly probable that a single female may

change spawning areas more than once over her life cycle. Mixing of nuclear genes is probably even more extensive because white sturgeon are iteroparous (repeat), broadcast spawners and often breed in large groups (Doroshov 1985).

If any differentiation of subpopulations does occur, such as the four clades of genotypes, it is probably the result of temporal rather spatial isolation. Very little is known about the white sturgeon reproductive cycle. Reproduction does not appear to occur annually and intervals between spawnings could last 4 to 9 years (Roussow 1957; Semakula and Larkin 1968). The age of first reproduction and the duration of spawning cycles could vary between cohorts and lead to the formation of temporal demes.

Genetic Diversity

The extensive level of variation detected in direct sequence comparisons usually results in high estimates of genetic diversity (Li and Graur 1991). Therefore, only relative comparisons between diversity estimates from restriction enzyme and sequence data are valid. Estimates of mtDNA (h) and nucleotide (π) diversity based on control region sequence were higher for the Fraser River compared to the Columbia River (Table 8). Diversity estimates from restriction enzyme data also clearly showed lower genetic diversity in the Columbia River population (Chapter I; Table 3). Lower nucleotide sequence diversity of white sturgeon from the Columbia River refuge compared to the recently founded Fraser River population, supports the theory that a postglacial genetic bottleneck has affected the Columbia River population.

In Chapter I, recent dam construction was suggested as the cause of lower genetic diversity in the Columbia River. Lower genetic diversity observed in

populations above Bonneville Dam compared with populations in the lower river was consistent with this hypothesis. Similar measures of genetic diversity from control region sequence data show the upper Columbia to have higher diversity than the lower Columbia. This inconsistency between restriction enzyme and sequence data diversity indices is probably due to the small sample size of control region sequences from above ($n = 6$) and below ($n = 5$) Bonneville Dam.

Effective Population Size

The longterm effective population size (N_e) for mtDNA is equal to the number of reproductive females and can be calculated from estimates of intraspecific divergence and mutation rate (Wilson et al. 1985). The time (t) since two randomly picked individuals shared a common mother is equal to the inverse of the mutation rate (% per million years) multiplied by the mean pairwise divergence (δ) within a species (Tajima 1983). N_e is equal to t (in years) divided by the mean number of years per generation.

For *A. transmontanus*, separate calculations of N_e can be made from restriction enzyme and control region sequence divergence estimates. Since mtDNA gene flow is probably significant in white sturgeon, N_e is calculated from mean pairwise divergence estimates of all mtDNA genotypes in the Fraser and Columbia Rivers. For restriction enzymes, δ is equal to 0.54 %. Assuming the generalized mutation rate for mtDNA of 2 % per million years is applicable to sturgeon then t equals 0.27 million years ago. If the average generation time for sturgeon is assumed to be 30 years (Semukula and Larkin 1968), then N_e equals 9000 females.

For sturgeon control region sequences δ is equal to 2.3 %. The rate of nucleotide change in the human control region is estimated to be 8.4 % per million years (Viligant et al. 1989). Accordingly, the time (t) since all Pacific Northwest sturgeon shared a common ancestor was again estimated to be 0.27 million years ago. Therefore, N_e estimated from nucleotide sequence data is equal to N_e calculated from restriction enzyme site data. The equality of the two estimates is not surprising given the similarlity between relative rates of control region evolution in white sturgeon and humans.

The estimate of a total of 9000 reproductive females in the Fraser and Columbia Rivers is higher than that suggested by census data. Semukula and Larkin (1968) estimated the average spawning population of sturgeon in the Fraser River to be about 300 to 600 females. Independent estimates of the number of reproductive females in the Columbia River are not available but data on annual catch (Galbreath 1985) and mtDNA diversity estimates indicate a small number of maternal lines.

One reason for this discrepancy between population size estimates is that sturgeon populations may have undergone a recent bottleneck. The divergence events postulated in the present analysis occurred no earlier than 10,000 years ago. Therefore, the N_e estimate reflects historical rather than current effective population size. Reduction in the size of sturgeon populations due to human exploitation would not be reflected in this estimate of N_e.

Avise et al. (1988) found that N_e estimated from mtDNA restriction enzyme data for three highly abundant species were actually much lower than contemporary census population sizes. While these species have likely expanded their geographical ranges from a few founders, the sum of the present mtDNA analysis suggests a recent constriction in the size and range of sturgeon populations.

Control Region and Restriction Enzyme Phylogenies

Intraspecific phylogenetic trees derived from restriction enzyme and control region data were qualitatively similar. In the control region trees, clustering of individuals was generally concordant to their respective restriction enzyme genotype. Similar general concordances have been found between mtDNA phylogenies derived from sequence and restriction enzyme data for other species. The geographical distribution of Australian babblers (*Pomatostomus temporalis*) mtDNA genotypes were similar in trees derived from cytochrome *b* nucleotide sequence and restriction enzyme site data (Edwards and Wilson 1990). Phylogenies based on mtDNA restriction enzyme site data and sequence comparisons of a 2214 bp protein coding region were nearly identical for six Pacific salmonid species (Thomas and Beckenbach 1989).

However, for salmonids and white sturgeon, phylogenies derived from direct sequence data are quantitatively superior to the restriction enzyme phylogenies. The increased amount of genetic variation found in the control region contributed towards greater statistical support for internal nodes and the detection of new population substructuring among the Fraser River white sturgeon.

The maximum likelihood method of phylogenetic inference is very useful because it provides a test for the statistical significance of internal branching points as well as a criteria for evaluating alternative tree toplogies (Felsenstein 1988). Maximum likelihood scores for unconstrained analysis of white sturgeon control region sequences were usually significantly higher than those of sequence data fitted to the restriction enzyme defined tree (Table 7). The consensus unconstrained tree differed from the fitted tree in the clustering of genotypes At1 and At3 as well as the rooting of the cluster with genotypes At9, At10 and At11 (Fig. 13).

In species with low levels of genetic variation, restriction enzyme changes which converge in different lineages may be wrongly considered informative. For example, in the control sequence, there are polymorphic sites for the restriction enzymes, *Ava*I and *Hinc*II. In the control sequence phylogeny, the *Hinc*II site was absent from all genotypes in cluster *A* and from genotype Cr6 in cluster *B* (Fig. 12). The loss of the *Hinc*II site in genotype Cr6 must be interpreted as a convergent event rather than as an indication of immediate common ancestory with individuals of cluster *B*. Similarly, the restriction site for the enzyme *Ava*I occurs in some but not all genotypes of clusters *A*, *C* and *D*. This blinking on and off of sites is a potentially recurring problem in phylogenetic reconstruction from restriction enzyme data (Swofford and Olsen 1990).

The unconstrainted analysis produced the wrong tree topology for two subsets of control region sequences (Fig. 13c). In this toplogy, all genotypes were rooted to a rare genotype, At9. Thus a random combination of sequences produced, by chance, the wrong tree topology. Sequence data from a larger number of indivduals is definitely required in this instance. In intraspecific studies, the number of sequences or individuals sampled should be given equal consideration as the number of nucleotide positions examined within each individual (Martin et al. 1990).

Accurate phylogenetic reconstruction from nucleotide sequences rests upon comparisons of homologous regions of the genome which have a sufficient amount of genetic variation to resolve taxonomic relationships. Intraspecific comparisons of control regions in humans (Viligant et al. 1989, Horai and Hayasaka 1990), kangaroo rat (*Dipodomys panamintinus*; Thomas et al. 1990) and white sturgeon have yielded statistically reliable phylogenies at the population level. The widespread use of PCR amplification will certainly facilitate the greater use of direct sequence comparisons in

population studies. Current technical difficulties in sequencing the control region, such as highly stable secondary structures and the design of specific oligonucleotide primers, will likely be overcome as more data becomes available on control region structure in different taxonomic groups.

Chapter III

Mitochondrial DNA Length Variation and Heteroplasmy in Populations of White Sturgeon (*Acipenser transmontanus*)

Introduction

The principal characteristics of animal mitochondrial DNA (mtDNA) are its compact organization, maternal inheritance and rapid evolution. Types of variation which have been observed in mtDNA include rearrangements of gene order, point nucleotide changes and size differences (reviewed in Moritz et al. 1987). Of the latter, length variation among mtDNA can occur as homopolymer runs of the same nucleotide (Brown and DesRosiers 1983; Hauswirth et al. 1984), tandem duplication or deletion of large (1 - 8 kb) regions of the genome (Moritz and Brown 1987; Snyder et al. 1987; Wallis 1987) or varying numbers of shorter (<300 bp) tandemly repeated sequences (Boyce et al. 1989, Rand and Harrison 1989, Buroker et al. 1990).

Species with mtDNA size variation are often heteroplasmic with different mtDNA length variants existing within a single individual. Instances of sequence length heteroplasmy have been reported in several species of invertebrates and lower vertebrates including *Drosophila* species (Fauron and Wolstenholme 1976; Solignac et al. 1983), crickets (*Gryllus* sp.; Harrison et al. 1985), scallops (*Placopecten magellanicus*; Synder et al. 1987), frogs (*Rana esculenta*; Monnerot et al. 1984), Cnemidophorus lizards (Densmore et al. 1985) and the bowfin (*Amia calva*; Bermingham et al. 1986) and American shad (*Alosa sapidissima*; Bentzen et al. 1988). Length variation in mammalian mtDNA is considerably less frequent (Boursot et al. 1987) and deletions of long mtDNA sequences are nearly always pathogenic in humans (Johns et al. 1989, Schon et al. 1989).

Length heteroplasmy is heritable and persistent across generations. In heteroplasmic crickets (*Gryllus* sp.; Rand and Harrison 1986) and fruit flies (*Drosophila mauritiana*; Solignac et al. 1983), it has been estimated that fixation of a particular mtDNA size class may take several hundred generations. In contrast, heteroplasmy for nucleotide substitutions has been rarely observed. Heteroplasmic restriction enzyme site lineages in a cow were rapidly sorted out after five or fewer generations (Hauswirth and Laipis 1982).

The mechanisms which generate size variation are poorly understood (reviewed in Moritz et al. 1987). Intermolecular recombination has not been observed in vertebrate mtDNA (Hayshi et al. 1985) although recent evidence of partial paternal inheritance of mtDNA in Drosophila (Kondo et al. 1990) and mussels (Hoeh et al. 1991) allows for the possibility of recombination between different mtDNA's in some invertebrate species. Intra- and intermolecular recombination have been proposed for maintaining length variation in crickets (Rand and Harrison 1989). Other models for the generation of length variation include polymerase stalling in homopolymer regions (Hauswirth et al. 1984) and replicative misalignments (Efstratiadis et al. 1980).

Sequence length heteroplasmy is extensive in white sturgeon (*Acipenser transmontanus*) and its molecular basis has been well characterized (Buroker et al. 1990). Restriction enzyme analyses have shown that *A. transmontanus* mtDNA ranges in size from 16.0 to 16.5 kb with several discrete length variants. Sequence analysis of a small (1.6 kb) and a large (1.9 kb) mtDNA fragment revealed that size variation was caused by perfect repeated copies of a 82 bp sequence in the D-loop or control region. The smallest size mtDNA had one perfect copy while the larger mtDNA had five copies. These repeats begin near the 3' end of the $tRNA^{Pro}$ and also share partial sequence identity with that gene.

The perfect repeat also contained a 14 bp sequence which had close similarity with the terminal associated sequence (TAS) found in the mouse control region (Doda et al. 1981). The TAS block is believed to have functional importance in determining the length of nascent DNA fragments involved in priming mtDNA replication. In addition, one or more repeats appear capable of forming stable secondary structures. This observation suggested a model in which secondary structure facilitates frequent misalignment of the repeat region prior to heavy strand elongation thus causing repeat sequence units to be added or deleted.

According to the model, high recurrent mutation would maintain size heteroplasmy despite the tendency of segregation, and perhaps selection, to eliminate it. These processes can be studied further by examining the distribution of length variation and heteroplasmy in natural populations of *A. transmontanus*.

Rand and Harrison (1989) examined the molecular basis and population genetics of mtDNA size variation in two cricket species (*Gryllus firmis* and *G. pennsylvanicus*). DNA sequence analysis revealed that extensive length variation and heteroplasmy in these species was the result of tandemly repeated sequences, 220 bp in length. Frequencies of mtDNA size classes in cricket populations were determined from Southern blot analysis and the degree of partitioning of genetic variation was quantified using hierarchial statistics adapted from Birky et al. (1983, 1989).

Birky et al. (1983) defined genetic diversity (K) as the probability of sampling two distinct genes from different levels of the population such as the zygote, cell, individual, population, regions and species (Birky et al. 1983). In crickets, 35 % of the total diversity for length variation occurred within individuals while over 50 % of the diversity was found between individuals but within populations (Rand and Harrison 1989). There was little difference in mtDNA length variation between cricket populations.

The population genetics of mtDNA length variation in vertebrates has not been as thoroughly examined. Most studies to date have only reported the fraction of heteroplasmic individuals found in the total sample (reviewed in Birky et al. 1989). The distribution of length variation in natural populations can provide some insight into the effects of various evolutionary forces upon mtDNA genome structure.

In this chapter, the occurrence of mtDNA size variation and heteroplasmy were quantified for 174 sturgeon from the Fraser and Columbia Rivers. Hierarchial and conventional statistics were used to analyse the genetic diversity of length variation among river populations of sturgeon as well as among previously defined mtDNA restriction enzyme genotypes.

Methods and Materials

Sample Collection

Sturgeon collections from the Columbia and Fraser Rivers and mtDNA extractions are described in Chapter I. Of the 178 individuals used in the restriction enzyme analysis in Chapter I, four did not produce readable bands in autoradiographs. Therefore, the Southern blot analysis included a total of 174 individuals from the Fraser ($n = 102$) and Columbia ($n = 72$) Rivers. Sample sizes from the lower and upper areas of the Columbia River were 19 and 53 individuals, respectively.

Probe Construction

A *BclI* digested fragment (1.6 kb in length) of *A. transmontanus* mtDNA had been previously subcloned into a *BamHI* site in the vector pUC19 (Buroker et al. 1990). DNA sequence analysis revealed that this fragment contained the entire control region, the genes for $tRNA^{Pro}$ and $tRNA^{Thr}$ and part of the cytochrome *b*

gene (Gilbert et al. 1988, Brown et al. 1989). The control region was flanked by *Eco*RI sites located in the pUC19 polylinker and in the middle of the $tRNA^{Pro}$ (Fig. 14). In an *EcoR*I digest, an 800 bp long fragment containing the control region was cut from the recombinant plasmid. The probe fragment was isolated in a 1.5% low melting agarose gel and extracted with phenol (Sambrook et al. 1989).

For hybridizations, the control region probe fragment was digested with ExoIII nuclease then end labelled with [5'-α-^{32}P]dATP using the klenow fragment of DNA polymerase I (Sambrook et al. 1989).

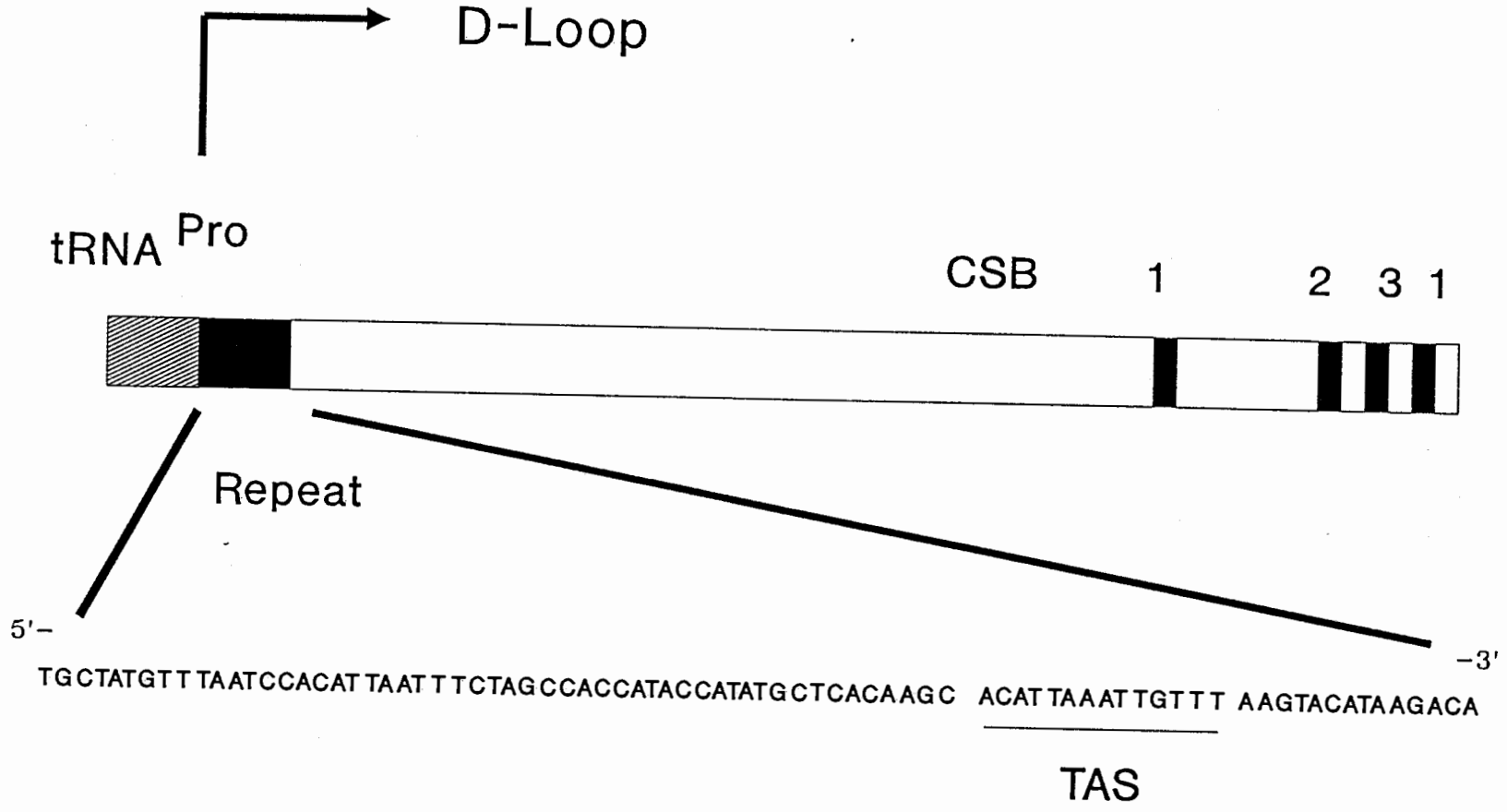
Southern Blotting and Hybridization

About 500 *ng* of mtDNA from each sample was digested with the restriction enzyme *Bcl*I. The digest resulted in variable sized fragments ranging from 1.6 to 2.0 kb which spanned the D-loop region (Buroker et al. 1990). The fragment patterns of up to 17 individuals were compared in 1.2 % agarose gels. After electrophoresis, gels were photographed and transferred to nitrocellulose filters (Southern 1975).

Hybridizations were done under stringent conditions, 68°C overnight followed by three washes in 0.5x SSPE (20x SSPE stock; 3.6M NaCl, 200mM NaH_2PO_4 pH 7.4, 20mM EDTA pH 7.4). Wet filters were wrapped in plastic wrap for autoradiography. Kodak O-Mat AR film was pre-exposed to ensure a linear response for microdensitometry readings (Laskey 1980). Several exposures (0.5 - 14 days) were made to ensure the best contrast between bands for each heteroplasmic individual.

The amount of DNA detected in the present analysis was determined by running on a gel, known quantities (0.05-100 *ng*) of the recombinant control region plasmid which had been linearized by digestion with *Hind*III. The gel was blotted, hybridized and exposed to film as described above.

Figure 14. Schematic diagram of the control region or D-loop probe. The D-loop fragment contained one perfect 82 bp repeat which can occur as multiple copies within an individual (Buroker et al. 1990). The sequence of the repeat is shown with the location of the termination association sequence (TAS) underlined. Also shown are the hypervariable area and several consensus sequence blocks (CSB's) which have similarity to mtDNA sequences found in the frog *Xenopus laevis* (Dunon-Bluteau et al. 1985).


D-Loop
tRNA Pro
CSB
1
2
3
1
Repeat
5'–
–3'
TGCTATGTTTAATCCACATTAATTTCTAGCCACCATACCATATGCTCACAAGC ACATTAAATTGTTT AAGTACATAAGACA
TAS

Densitometry

In the autoradiographs, different mtDNA size classes were visible as distinct bands. Homoplasmic individuals had only one visible band while in heteroplasmic individuals, as many as seven distinct bands were present in a ladder-like arrangement. Densitometer scans were taken of heteroplasmic individuals in order to quantify the relative frequencies of distinct size classes within that individual. Individuals showing only a single band were not scanned and the frequency for that mtDNA size variant was recorded as 1.0.

Chart printouts were made for densitometry scans of heteroplasmic individuals. The points of greatest width and height in tracings for each band were measured and the area was calculated as $1/2 width \cdot height$. For an individual sturgeon, the areas under all the peaks were summed and the frequency of a mtDNA size class was proportional to the area a particular band contributed to the total area.

MtDNA of two individuals, for which the control region sequence had been determined (Buroker et al. 1990), were Southern blotted and hybridized along with several highly heteroplasmic individuals. The bands in the heteroplasmic individuals were found to correspond with the number of tandem perfect repeats found in the DNA sequence analysis. One heteroplasmic individual, with five mtDNA size classes, was used as a size standard in all subsequent gels.

In the 174 individuals analyzed, no mtDNA was smaller than the reference sequence with one perfect repeat. The maximum number of observed repeats was six. Therefore, mtDNA size variants were labelled RPT1, RPT2...RPT6 which directly corresponds with the presence of 1, 2...6 perfect 82 bp repeats. For example, a heteroplasmic indivdual with 10 % of its mtDNA size variants with one perfect repeat, 30 % with two repeats and 70 % with three repeats would contribute the

frequencies 0.10, 0.30 and 0.70 to the size classes RPT1, RPT2 and RPT3, respectively.

Statistical Analysis

Several different transformations were applied to the data including $\log_{10}$, $\log_e$ and arcsin$\sqrt{}$. However, tests showed that untransformed data best conformed to expectations of normality and critical values of skewness (Zar 1984). For chi-square tests, it was necessary to represent the distribution of mtDNA length variants within heteroplasmic individuals as single value. A composite variable (CRPT), essentially the weighted sum of the frequency of size classes, was calculated:

$$\text{CRPT} = \text{RPT1} + 2{\cdot}\text{RPT2} + 3{\cdot}\text{RPT3} + 4{\cdot}\text{RPT4} + 5{\cdot}\text{RPT5} + 6{\cdot}\text{RPT6} \qquad 3.1$$

As an example, if size classes RPT3 and RPT4 were found in respective frequencies of 0.3 and 0.7 in a heteroplasmic individual, then the CRPT equaled 3.7. Similarly, CRPT would equal 2.0 for an individual homoplasic for size class RPT2.

Frequency distributions, *t*-tests and log likelihood chi-square (G test) analyses were done using programs P2D, P3D and P4F, respectively, of the BMDP statistical software package (Dixon 1983).

Hierarchial Statistical Analysis

Birky et al. (1983) proposed that genetic diversity (K) represents the operation of drawing two distinct genes from a population. The K statistics permits the subdivision of populations or species into hierarchial levels where i) Kz, is the sampling of two different genes from a single zygote; ii) Ka is the sampling of two distinct genes from a single cell; iii) Kb is the sampling of genes from an individual;

iv) Kc is the sampling from different individuals within a population and v) Kd is the sampling of two different genes from different populations within a species.

The diversity indices Kz and Ka can not be measured directly and can only be estimated from the data. For the other levels, gene diversity is calculated as : $K = 1 - \Sigma\ x_i^2$ where i is the frequency of the ith allele or mtDNA size class in that population level.

Rand and Harrison (1989) used these diversity indices to quantify the degree of population subdivision, within individuals:

$$C_I = \overline{Kb}/Kd \qquad 3.2$$

Among individuals within populations:

$$C_{IP} = (\overline{Kc} - \overline{Kb})/Kd \qquad 3.3$$

Among individuals within a species:

$$C_{IL} = (Kd - \overline{Kb})/Kd \qquad 3.4$$

Among populations within a species:

$$C_{PL} = (Kd - \overline{Kc})/Kd \qquad 3.5$$

where $\overline{Kb}$ is the mean Kb among all sampled individuals in the species, $\overline{Kc}$ is the mean Kc among all sampled populations in a species and Kd is the probability that two different size mtDNA molecules were selected from different populations in the

species. The sum of levels C_I, C_{IP} and C_{PL} equal 1.0. The level C_{IL} is excluded because it is not calculated from adjacent hierarchial levels.

In separate analyses, white sturgeon populations were defined by geographical location (i.e. Fraser, Columbia and upper and lower Columbia Rivers) and restriction enzyme genotype (At1 - 11).

More recently, Birky et al. (1989) proposed two more K values. The first is Kc^* which considers the probability of sampling two distinct genes from different individuals within a population *and* the probability that the genes occurred in the same individual:

$$Kc^* = [\overline{Kb} + Kc(n - 1)]/n \qquad 3.6$$

where n is the total number of males and females in the population. The probability (Kt) of selecting different genes from anywhere among L populations:

$$Kt = [\overline{Kc}^* + Kd(L\text{-}1)]/L \qquad 3.7$$

The K^* statistics can also be used to calculate a new set of hierarchial levels using equations 3.2 - 3.5 by replacing Kc and Kd with Kc^* *and* Kt, respectively. In the nomenclature of Rand and Harrison (1989) these new statistics are C_I^*, C_{IP}^*, C_{IL}^* and C_{PL}^*. Therefore, C and C^* statistics refer to the probabilities of sampling of distinct genes between different populations and between or within different populations, respectively.

The K statistics can also be related to another measure of population subdivision, GST (Nei 1973), where:

$$GST = (Kt - \overline{Kc}^*)/Kt \qquad 3.8$$

According to Birky et al. (1989), *GST* represents the fraction of local gene diversity reduced by population subdivision.

Rand and Harrison (1989) extended their hierarchial analysis to consider genetic diversity both within and among eight populations of two species of crickets. However, since genotype and river subdivisions of mtDNA size variation in sturgeon overlap, a combined hierarchial analysis of the two data sets would be redundant.

Results

Detection of mtDNA Size Classes

In hybridizations of the test blot, the lane with 100 *pg* of control region DNA is clearly visible after a 5 day exposure (Fig. 15). A faint band is just visible in the lane loaded with 50 *pg* of DNA. In hybridizations of Southern blots of whole mtDNA, very intense bands can overshadow nearby weaker bands in heteroplasmic individuals. Several exposures (0.5 - 14 days) were usually made to obtain the best contrast among bands for densitometry. Therefore, a conservative estimate of routine detection levels would be around 100 *pg* of DNA.

MtDNA Length Variation

A. transmontanus is highly heteroplasmic for mtDNA size variation (Fig. 16). Among the 174 individuals analyzed, 42 % were heteroplasmic for at least two mtDNA size classes. The smallest observed size variant corresponded with a fragment known to have only one perfect repeat from DNA sequence analysis (Buroker et al. 1990). The largest fragment consistently observed was equivalent to a control region with six repeated sequences.

Fig. 15. Hybridization of a test Southern blot of known amounts of linearized pUC19/*A. transmontanus* control region DNA. Numbers are *ng* of DNA loaded in each lane. The film was pre-exposed and then exposed to the blot for 5 days.

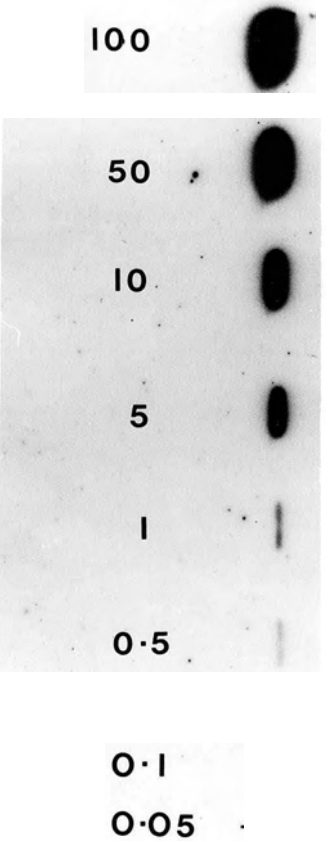

100
50
10
5
1
0·5
0·1
0·05

Fig. 16. An example autoradiograph of mtDNA size variation in *A. transmontanus*. MtDNA was digested with the enzyme *BclI* was probed with the entire mtDNA control region cloned from white sturgeon. Bands (bottom to the top) in lane 1 correspond with mtDNA with 1 - 5 copies of a perfect 82 bp tandemly repeated sequence (Buroker et al. 1990).

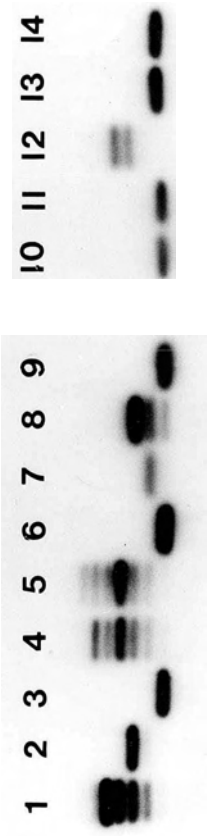
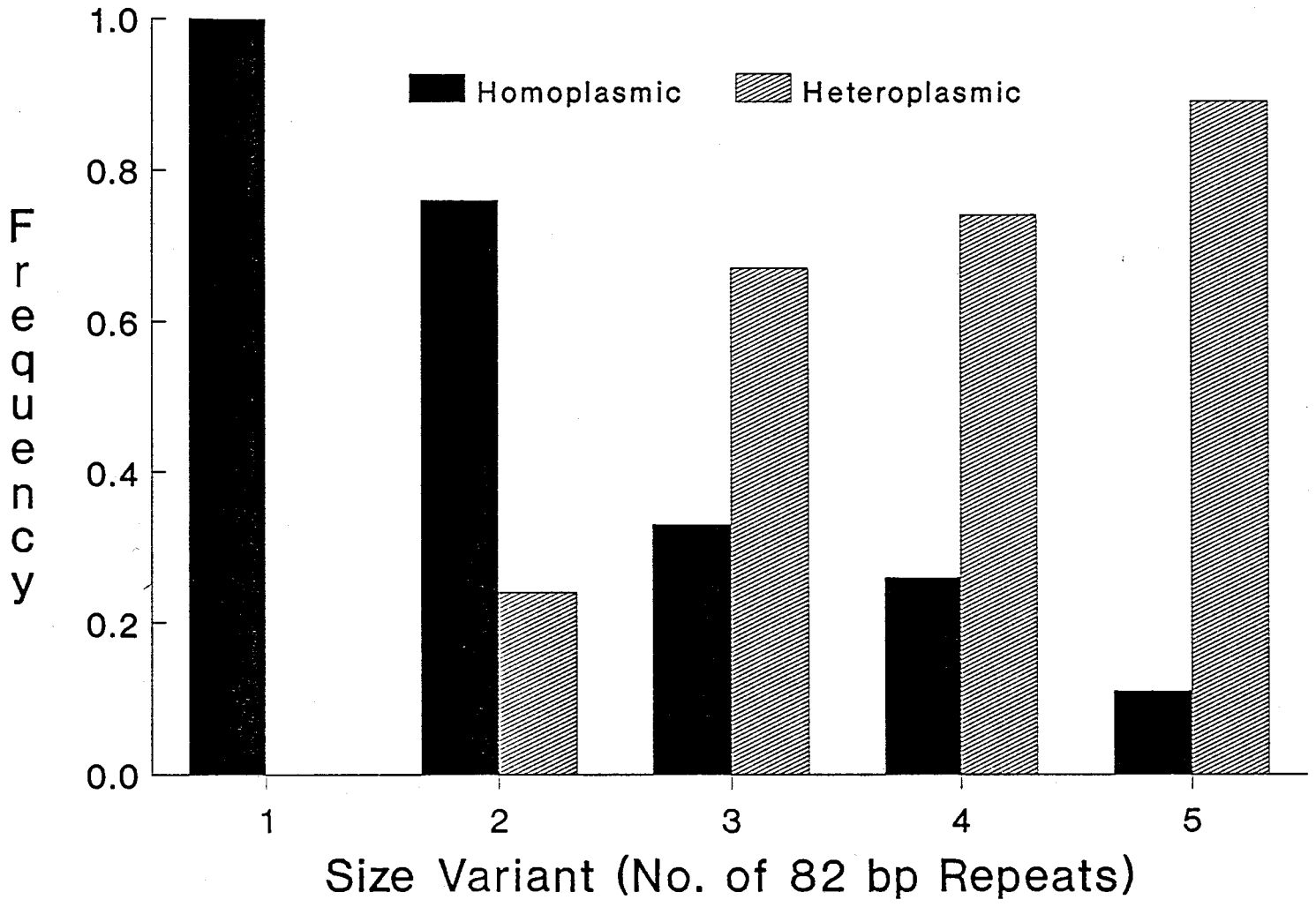

10
11
12
13
14
1
2
3
4
5
6
7
8
9

Fig. 17. Frequency of heteroplasmy associated with different mtDNA size classes. The size variant (along the x axis) represents the size class in the highest frequency within an individual. Frequencies of heteroplasmy among individuals with predominant size variants are shown. For size variants 1, 2, 3, 4 and 5, the sample sizes were $n = 45$, 41, 51, 27 and 9, respectively. One individual with two size classes occuring with equal frequency was excluded from this analysis.


1.0
0.8
0.6
0.4
0.2
0.0
Frequency
Homoplasmic
Heteroplasmic
1
2
3
4
5
Size Variant (No. of 82 bp Repeats)

One individual was heteroplasmic for seven mtDNA size classes but the largest fragment was faint and at a low frequency (Lane 5 in Fig. 16). In the analysis, the area under this band was pooled with the size class for six repeats (RPT6).

Figure 17 considers the incidence of heteroplasmy among individuals with different dominant mtDNA size classes. The dominant mtDNA size class within an individual is defined as size variant with the highest frequency. As an example, mtDNA size class RPT5 is the most frequent size class in the first individual in Figure 16. Figure 17 shows a trend of increased heteroplasmy among individuals with large mtDNA (RPT2 - 6). Individuals with the smallest size class as the dominant type (i.e. Lanes 3, 6, 9, 10, 11, 13 and 14 in Fig. 16) had no visible heteroplasmy. Size classes always occurred sequentially in heteroplasmic individuals. For example, a jump in size classes from one to three or more repeats was never observed in an individual. However, near bimodal distributions of size variants occurred in some individuals (i.e. in Fig. 16, the brightest bands in lane 4 correspond to size classes with four and six repeats).

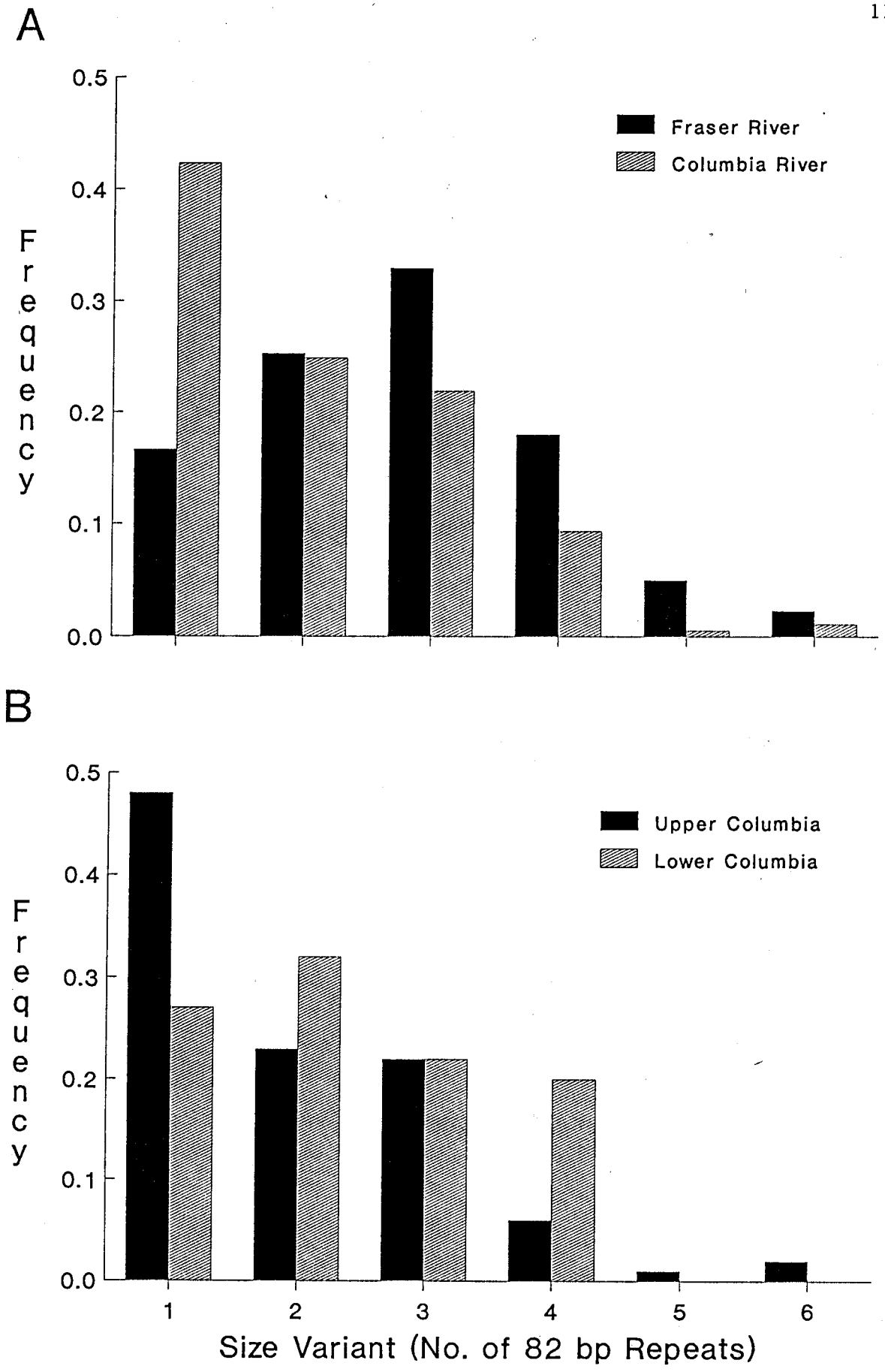
Size Variation Among Rivers

The frequencies of mtDNA size classes (Table 9) and the incidence of heteroplasmy were significantly different between sturgeon sampled from the Fraser and Columbia Rivers. Heteroplasmy occurred in 53 % of the sturgeon ($n = 102$) from the Fraser River and in only 25 % of individuals ($n = 72$) from the Columbia River. The frequency distribution of mtDNA size classes (based on CRPT values) were significantly different between the Fraser and Columbia Rivers ($G = 20.73$; $d.f. = 5$; $p = 0.0009$). For the Columbia River, the frequency distribution was skewed to mtDNA with one repeat (RPT1) while size classes of the Fraser River had an approximately normal distribution with a mean of about three perfect repeats (RPT3; Fig. 18a).

Table 9. Mean frequencies of mtDNA size variants (RPT1 - 6) and diversity indices for river populations. Size variants are defined by the number (1 to 6) of perfect 82 bp repeats in the control region. $\overline{K_b}$ is the mean diversity of within individuals in a particular river. K_c is the diversity between individuals in a river and K_c^* is the diversity between and within individuals. The subdivision of variation among individuals within a river is $G_{ip} = (K_c - \overline{K_b})/K_c$.

River	n	RPT1	RPT2	RPT3	RPT4	RPT5	RPT6	$\overline{K_b}$	K_c	K_c^*	G_{ip}
Fraser and Columbia	174	0.273	0.251	0.283	0.144	0.031	0.018	0.183	0.760	0.756	0.759
Fraser	102	0.166	0.253	0.329	0.180	0.050	0.023	0.222	0.765	0.760	0.710
Columbia	72	0.423	0.249	0.219	0.093	0.005	0.011	0.127	0.702	0.694	0.819
Upper	19	0.265	0.317	0.222	0.196	0.0	0.0	0.120	0.742	0.709	0.838
Lower	53	0.480	0.225	0.217	0.056	0.006	0.015	0.129	0.668	0.658	0.807

Fig. 18. Frequency distributions of mtDNA size classes in white sturgeon populations in the A) Fraser ($n = 102$) and Columbia Rivers ($n = 72$) and B) the Columbia River subdivided into upper ($n = 53$) and lower portions ($n = 19$), as delineated by Bonneville Dam. The numbering of size variants on the x-axis correspond with mtDNA size classes, RPT1 - 6.


A
0.5
0.4
0.3
0.2
0.1
0.0
Frequency
Fraser River
Columbia River
B
0.5
0.4
0.3
0.2
0.1
0.0
Frequency
Upper Columbia
Lower Columbia
1
2
3
4
5
6
Size Variant (No. of 82 bp Repeats)

Within the Columbia River basin, the distributions of size variants in the upper and lower portions of the river were not significantly different ($G = 2.994$; $d.f. = 3$; $p = 0.393$). However, the distribution was slightly more skewed to size class RPT1 in the upper Columbia River compared to the lower Columbia (Fig. 18b). The mean diversity within individuals ($\overline{Kb}$) in the total sample was 0.183 (Table 9). The higher diversity within Fraser River individuals in comparsion to individuals from the Columbia, reflects the higher incidence of heteroplasmy in Fraser River sturgeon. In both rivers, the greatest proportion of genetic diversity (Kc, Kc^* and G_{IP}) is attributed to variation among individuals in the population rather than within individuals.

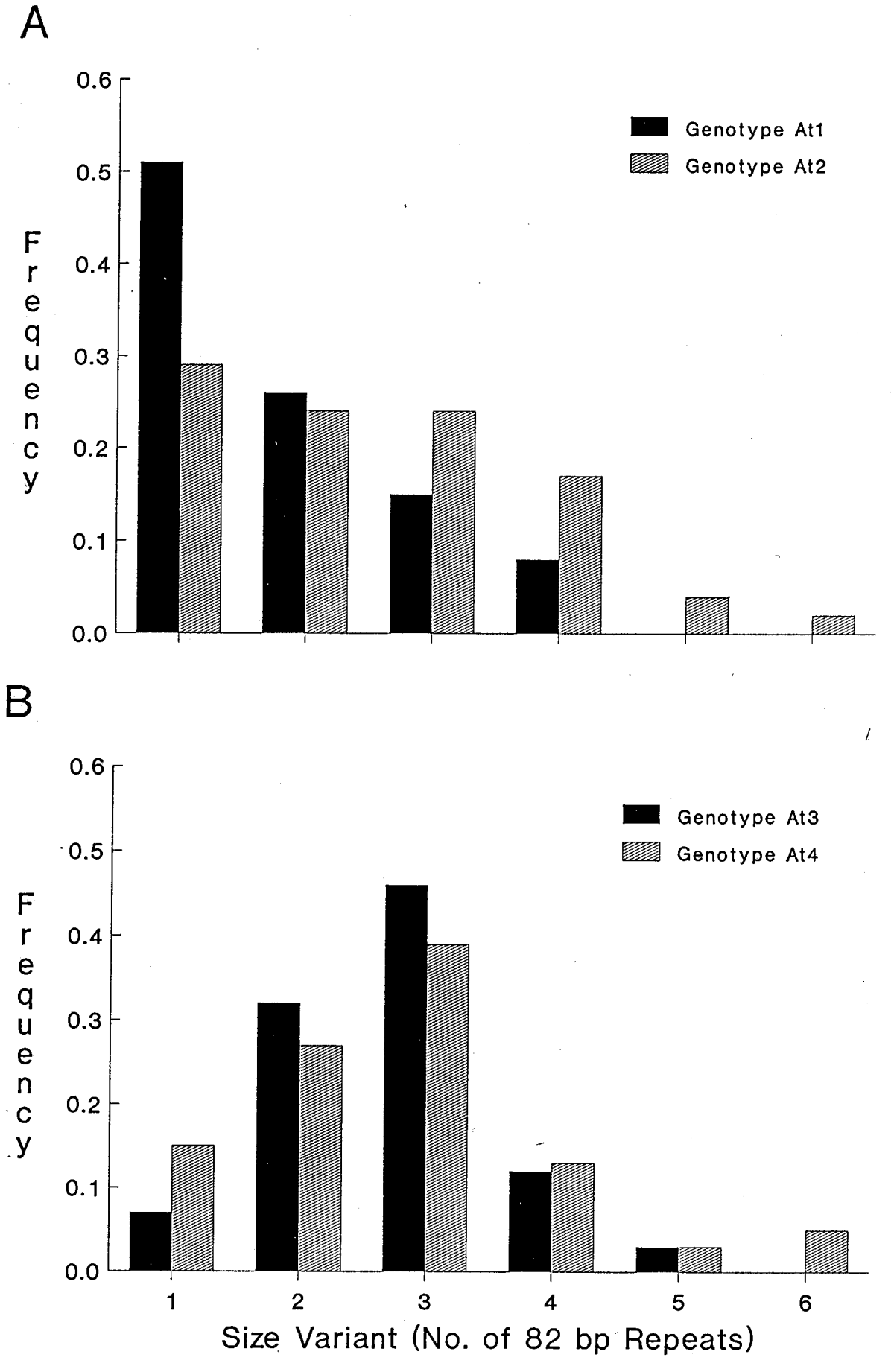
Size Variation and Restriction Enzyme Genotypes

A total of ten restriction enzyme genotypes were detected in Northwest Pacific *A. transmontanus* populations (Chapter I). Six of those genotypes were rare and seldom found in more than one individual. Therefore, frequency distributions of mtDNA size classes could only be analyzed in the four most abundant genotypes, At1 ($n = 13$), At2 ($n = 107$), At3 ($n = 21$) and At4 ($n = 20$). The frequency distribution was heavily skewed to size class RPT1 in genotype At1 (Fig. 19a). In genotype At2, there was a more equal distribution of the first four size classes (RPT1-4). Genotypes At3 and At4 have a normal-like distribution about RPT3 (Fig. 19b). In pairwise comparisons, the size class frequency distribution of genotype At1 was significantly different (t-tests; $p < 0.02$) from those of genotypes At2, At3 and At4. However, there were no significant differences between the latter three genotypes. The percentage of heteroplasmic individuals was fairly consistent among genotypes At1 (38 %), At2 (41 %) and At4 (35 %). However, heteroplasmy (67 %) was more frequent in individuals with genotype At3.

Table 10. Mean frequencies of mtDNA size variants (RPT1 - 6) and diversity indices for restriction enzyme genotypes (At1 - 11). Size variants are based on the number (1 to 6) of perfect 82 bp repeats in the control region. $\overline{K_b}$ is the mean diversity within individuals with a particular restriction enzyme genotype. K_c is the diversity among individuals within a genotype and K_c^* is the diversity within and among individuals. The subdivision of variation among individuals in a genotype is $G_{ip} = (K_c - K_b)/K_c$.

Genotype	*n*	RPT1	RPT2	RPT3	RPT4	RPT5	RPT6	$\overline{K_b}$	K_c	K_c^*	G_{ip}
At1	13	0.505	0.260	0.154	0.082	0.0	0.0	0.161	0.647	0.610	0.751
At2	107	0.294	0.237	0.243	0.168	0.038	0.020	0.175	0.768	0.762	0.762
At3	20	0.067	0.320	0.464	0.116	0.032	0.0	0.305	0.663	0.639	0.540
At4	21	0.147	0.267	0.385	0.125	0.027	0.050	0.147	0.740	0.712	0.801
At6	7	0.429	0.240	0.331	0.0	0.0	0.0	0.062	0.649	0.562	0.904
At7	1	0.0	0.0	0.540	0.460	0.0	0.0	0.497	0.497	0.497	0.0
At8	2	1.000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
At9	1	0.0	0.210	0.790	0.0	0.0	0.0	0.332	0.332	0.332	1.000
At10	1	0.0	1.000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
At11	1	0.0	0.080	0.220	0.620	0.07	0.0	0.556	0.556	0.556	1.000

Fig. 19. Frequency distributions of mtDNA size classes in the restriction enzyme genotypes A) At1 ($n = 13$) and At2 ($n = 107$) as well as B) At3 ($n = 21$) and At4 ($n = 20$). The numbering of size variants on the x-axis correspond with mtDNA size classes, RPT1 - 6.


A
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Frequency
Genotype At1
Genotype At2
B
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Frequency
Genotype At3
Genotype At4
1
2
3
4
5
6
Size Variant (No. of 82 bp Repeats)

Within individual diversity estimates (Kb) were distorted upwards in the rare genotypes with only a few individuals (Table 10). In the more abundant genotypes, At1, At2 and At4, mean Kb values (0.147 - 0.175) were similar to those calculated for the river subdivisions. $\overline{Kb}$ was higher in genotype At3 (0.305) which also had a high proportion of heteroplasmic individuals. Variation among individuals within a genotype (Kc, Kc^* and G_{IP}) was considerable higher than variation within individuals.

Hierarchial Statistics

In the hierarchial analysis, diversity indices for variation within individuals (C_I and C_I^*), within populations (C_{IP} and C_{IP}^*) and among populations (C_{PL} and C_{PL}^*) sum to 1.0 (Table 11). Intra-individual diversity (C_I and C_I^*) for mtDNA size classes is about the same for genotype and river subdivisions. About 24 % of total diversity is due to length variation within individuals (i.e heteroplasmy).

The largest component of variation is diversity among individuals within a river or genotype (C_{IP} and C_{IP}^*). Nearly 72 % and 40 % of mtDNA size variation occurs within river and genotype groups, respectively.

Variation among genotypes is greater than variation among rivers (C_{PL} and C_{PL}^*). Nearly 36 % of the variation in mtDNA size occurs between restriction enzyme genotypes. However, in the river subdivision only about 3 % of mtDNA length is due to between river differences. The index GST also shows greater population subdivision among genotypes ($GST = 0.388$) than among rivers ($GST = 0.067$).

There are few differences between the C and C^* statistics. Rand and Harrison (1989) had a similar result for crickets. It would appear that there are the same probabilities of sampling two distinct genes from among population levels as for sampling different genes from either among populations or within those populations.

Table 11. Hierarchial diversity statistics for restriction enzyme genotypes and river populations. Diversity indices calculated for rivers were subdivided into i) the Fraser and Columbia Rivers (F/C) and ii) the Fraser River and the upper and lower Columbia River (F/LC/UC).

		River	
Statistic	Genotype	F/C	F/LC/UC
$\overline{K_c}$	0.485	0.734	0.705
$\overline{K_c}^*$	0.467	0.727	0.709
K_d	0.763	0.760	0.760
K_t	0.731	0.744	0.743
Among Diversity			
C_i	0.247	0.241	0.241
C_{ip}	0.397	0.725	0.713
C_{il}	0.759	0.759	0.759
C_{pl}	0.362	0.034	0.046
Within and Among Diversity			
C_i^*	0.250	0.246	0.246
C_{ip}^*	0.389	0.731	0.708
C_{il}^*	0.750	0.754	0.754
C_{pl}^*	0.361	0.023	0.046

Table 12. Hierarchial diversity statistics for mtDNA length variation within rivers and genotypes.

River/Genotype	c_i	c_{ip}	c_{il}
Fraser	0.292	0.714	0.709
Columbia	0.167	0.757	0.833
Lower Columbia	0.158	0.818	0.842
Upper Columbia	0.170	0.709	0.830
At1	0.212	0.639	0.788
At2	0.230	0.780	0.764
At3	0.401	0.471	0.609
At4	0.193	0.780	0.807

In addition, subdividing the Columbia River into the upper and lower regions did not greatly alter diversity indices for the Fraser and Columbia Rivers.

The diversity indices C_I and C_{IP} were also calculated separately for rivers and genotypes (Table 12). Within individual variation was highest for the Fraser River and genotype At3. The lowest intra-individual variation occurred in the Columbia River. Variation among individuals (C_{IP}) is greater than within individual variation for river subdivisions and most genotypes. The genotype At3 is the exception where variation within and among individuals is nearly equal.

Estimates of Cellular Diversity and Mutation Rate

The genetic diversity of zygotes (Kz) and adult cells (Ka) were not measured directly but can be estimated using the equations in Birky et al. (1983). Rearranging their Equation 3:

$$Kz = \overline{Kb}[(N_e + 1)/ (N_e - 1)] \qquad 3.9$$

where N_e is the effective population size or, in the case of mtDNA, the number of females. Little census data exists for Northwest Pacific sturgeon population but the number of reproductive females probably does not exceed 1000 for the combined Fraser and Columbia Rivers (Semakula and Larkin 1968; Chapter II). Mean within individual diversity or Kb was estimated to be 0.183 from Table 9. Therefore, Kz was calculated to be 0.183 for sturgeon.

In Birky et al. (1983) the probability (Ka) of sampling distinct genes from the same cell can be estimated from Equation 2:

$$Ka = (1 - 1/N_{eo})^{c} \cdot Kz \qquad 3.10$$

where c is the number of germline cell divisions in a sexual generation and N_{eo} is the

effective number of organelle genes. For mtDNA, N_{eo} is equal to the number of effective females in the population. The true value of the parameter c is not known for *A. transmontanus* but for vertebrates, its approximate range is from 10 to 50 (Birky et al. 1983, 1989). Therefore, if N_e equals 1000, the range for Ka is 0.174 - 0.181.

Rand and Harrison (1989) showed that the mutation rate (u) for length variation can be approximated from $\overline{Kc}$ by rearranging Equation 25 of Birky et al. (1989):

$$u \approx 1/(1/\overline{Kc} - 1)(2N_e) \qquad 3.11$$

The mean diversity ($\overline{Kc}$) among individuals within a population varied from 0.485 for genotypes to 0.734 for rivers (Table 11). If N_e is equal to 1000, then u is estimated to range from 1.33 x10^{-4} to 1.70 x10^{-4} mutations per gene per sexual generation.

For the calculation of u, Rand and Harrison (1989) discussed several potential sources of error including estimates of population size and cell divisions and measurements in the densitometry. Birky et al. (1989) added that Kc may not be at equilibrium for some populations which can also affect the value of u.

Discussion

Heteroplasmy in Sturgeon MtDNA

Over 42 % of *A. transmontanus* individuals were heteroplasmic for up to seven different mtDNA sizes. This level of variation is extensive in comparison to other fish species with mtDNA heteroplasmy. In bowfin (*Amia calva*), 8 % of individuals were heteroplasmic for two mtDNA size variants (Bermingham et al. 1986). Thirty of 244 (12 %) American shad (*Alosa sapidissima*) were heteroplasmic

for two different mtDNA types (Bentzen et al. 1988). In other cold-blooded vertebrates, including Cnemidophorous lizards (Densmore et al. 1985), the newt (*Triturus cristatins*; Wallis 1987) and *Hyla* treefrogs (Bermingham et al. 1986), the frequency of intraspecific heteroplasmy does not exceed 25 %.

However, in the frog, *Rana esculenta*, each individual was heteroplasmic for different size mtDNA (Mommerot et al. 1984). Mulligan and Chapman (1989) report a similar high occurrence of heteroplasmy in white perch (*Morone americana*) involving multiple copies of a 110 bp fragment.

Certain insects also have a high degree of mtDNA size heteroplasmy. In two species of gryllus crickets, 46 % of 319 individuals were heteroplasmic (Rand and Harrison 1989). Extensive mtDNA heteroplasmy has also been reported for species of bark weevils (Boyce et al. 1989) and *Drosophila* (Solignac et al. 1983; Hale and Singh 1986).

Inferences from the Hierarchial Analysis

Rand and Harrison (1989) suggested that heteroplasmy represents a dynamic balance between mutation (which increases variation) and genetic drift (to eliminate variation). In support of their argument, the results of their hierarchial analysis show that 35 % of mtDNA length diversity is within individuals and over 50 % occurs among individuals within populations. In the absence of counteracting forces, mutation would tend to increase within individual diversity until all the genetic diversity for length variation in the population lay within the individual.

The present hierarchial analysis suggests that similar mechanisms cause the generation and reduction of length variation in cricket and sturgeon mtDNA. Diversity indices for within individual sturgeon ($C_I = 0.241$ to 0.247) and among individuals within populations ($C_{IP} = 0.397$ to 0.725) are close to values calculated

for crickets (Rand and Harrison 1989). Population level subdivision (C_{PL} and Gst) for mtDNA length variation is also low in both species.

In addition, within individual diversity ($\overline{Kb}$ for crickets = 0.2 - 0.3; sturgeon = 0.183) and estimates of cellular diversity (Ka for crickets = 0.18 - 0.29; sturgeon = 0.174 - 0.181) were similar. Ka for sturgeon was at least twice as high as estimates for other heteroplasmic species reported by Birky et al. (1989). Since Birky et al. (1989) used similar values for the parameters c and N_e ($c = 10$; $N_e = 500$ and $c = 50$ and $N_e = 50$), the higher degree of heteroplasmy in sturgeon relative to other vertebrates is the principal factor behind differences in estimates of Ka. It should be noted that estimates of GST, Kc^* and Kt could be in error by 25 % for a range of Ka values between 10^{-2} to 10^{-1} (Birky et al. 1989).

Given the similarity of the K parameters, it follows that calculated estimates of mutation rates (i.e. mutations per gene per sexual generation) were nearly equal for sturgeon (1.3 x10^{-4} to 1.7 x10^{-4}) and crickets (1.25 x10^{-4} to 2.14 x10^{-4}; Rand and Harrison 1989). This similiarity is not surprising from a molecular perspective because, in both species, mtDNA length variation is the result of varying numbers of tandemly repeated sequences occurring near the origin of replication. However, the physiology and metabolism of crickets and sturgeon are obviously distinct, especially with respect to lifespan and reproductive generation time. These factors suggest that the period of mtDNA segregation to germline cells following formation of the zygote is fundamental to the fixation of mtDNA size variants.

In *Xenopus laevis* and *X. mulleri*, the maternal mtDNA segregated to primary germ cells in the embryo is the mtDNA inherited by the subsequent generation (Dawid and Blackler 1972). The same mtDNA size variant, RPT2, was detected in separate mtDNA isolations from liver, brain, muscle, gill, heart and gonadal tissues from two individual sturgeon (J.R. Brown unpublished data). These data suggest that mutations for mtDNA length variation occur in germ cells rather than somatic

cells although it would be desirable to confirm this observation for a heteroplasmic individual.

The parameters affecting estimates of u and Ka (Equations 3.10 and 3.11) are the diversity (Kb) within individuals, effective population size (N_{eo} or N_e) and the number of germ cell generations per sexual generation. Kb was shown empirically to be similar in sturgeon and crickets. Rand and Harrison (1986) estimated N_{eo} for *Gryllus firmus* to be in the range of 87 to 395 which is not greatly different from N_e (300 to 1000 females) estimated for *A. transmontanus* (Semakula and Larkin 1968). Finally, the parameter c equaled 10 for crickets but is unknown for sturgeon. However, estimates for vertebrates were in the range of 10 to 50 germ cell generations per sexual generation (Birky et al. 1989). Based on these assumptions, similar estimates of u and Ka were derived for crickets and white sturgeon. In both species, the period of germ cell segregation in the developing embryo probably provides the principal opportunity for the generation or elimination of mtDNA size variants.

The Maintenance of Heteroplasmy

Studies on crickets (Rand and Harrison 1986) and *Drosophila* sp. (Solignac et al. 1983; Matsuura et al. 1989) suggest that the heteroplasmic condition can persist for many generations. The question remains on how multiple size classes of mtDNA are maintained within an individual in light of cell line segregation and, possibly selection, acting to cause homoplasy.

Heteroplasmy is a transitory and dynamic evolutionary state where mutation is balanced by genetic drift and/or selection. The higher frequency of heteroplasmy among large mtDNA size variants (Fig. 17) is indicative of a dynamic state where mutation increases the number of repeat sequences while counter-forces attempt to reduce the copy number. The increasing proportion of heteroplasmic individuals as copy numbers increase, reflects lower fixation rates of larger mtDNA. Replication

errors and costs associated with too large of a control region could be one of the selective pressures opposing fixation of mtDNA with a large number of tandem repeat sequences.

Skewed distributions in favour of smaller mtDNA's have reported for other heteroplasmic species including *Drosophila melanogaster* (Hale and Singh 1986), crickets (Rand and Harrison 1989), lizards (Moritz and Brown 1987) and American shad (Bentzen et al. 1988). The lack of any size variants smaller than one perfect repeat and the absence of heteroplasmy among individuals which predominantly have this size of mtDNA, suggests that small mtDNA are also selected for in white sturgeon.

Buroker et al. (1990) suggested that the ability of repeat sequences to fold into complex secondary structures prior to mtDNA replication is crucial for the generation of new repeats. The internal pairing capability of repeat sequences, hence thermodynamic stability, greatly increases with copy number. The calculated stability of a single 82 bp sequence is -14.1 kcal/mole versus -34.4 kcal/mole for two sequences. This circumstantial evidence suggests that the mutation rate to increase repeat copy number from one is lower than those rates among size variants with two or more repeats.

In the Columbia River and restriction enzyme genotype At1, mtDNA size frequency distributions were skewed to size variants with only one repeat. There is both molecular and historical evidence to suggest that both these populations have lower effective population sizes. Diversity indices for genetic heterogeneity (h and π) calculated from mtDNA restriction enzyme site and control region sequence data show that Columbia River white sturgeon populations have markedly lower genetic diversity than Fraser River populations (Chapter II and III). Genotype At1, is geographically localized to a small area of the Fraser River which suggests a small group of reproductive individuals. Fewer mothers (low N_e) in these populations

would act as genetic bottlenecks and reduce the number of mutation events which produce larger mtDNAs in a population. The result would be a skewed distribution to smaller mtDNA.

The estimated mutation rate for length variation is considerably higher than the estimated rate of nucleotide substutions in the control region (Viligant et al. 1989; Chapter II). The uncoupling of mutation rates for length variations from nucleotide substitutions is also suggested by the phylogenetic relationship between sturgeon genotypes At1 and At3. Both restriction enzyme analysis (Fig. 3) and control sequences (Fig. 10, 11 and 12) show low sequence divergence between individuals with these genotypes. However, genotype At3 had nearly twice (67 %) the incidence of heteroplasmy as At1 (38 %) and there was a significant difference between the size class distributions of these genotypes. Therefore, the majority of the mutation events responsible for existing patterns of length variation in these genotypes must have occurred since the time of their divergence.

The long term persistence of tandem repeat sequences in sturgeon mtDNA is intriguing. While multiple copies of repeats are not required for a functional mtDNA, the absence of any smaller size variants suggests it is necessary to have at least one copy of the repeat. The existence of a TAS sequence block may partially explain the functional importance of a single 82 bp repeat. The TAS block is implicated as the termination site for the synthesis of a short piece of DNA important in the initiation of mtDNA replication (Doda et al. 1981). The acquisition of a functional role by "selfish DNA" has been noted before in the nuclear genome (Doolittle and Sapienza 1980; Orgel and Crick 1980). Unlike the nuclear genomic environment, there are few areas in the compact gene organization of mtDNA which can tolerate the insertion of "selfish DNA" sequences. The control region is probably the only area in the mtDNA genome with a sufficiently long noncoding region to accept any novel length changes induced by self-proliferating DNA.

Chapter IV
Mitochondrial DNA Sequence Variation and Length Heteroplasmy in Four Sturgeon Species (Genus *Acipenser*)

Introduction

Restriction enzyme studies have shown mitochondrial DNA (mtDNA) to be a sensitive genetic marker for analyzing the geographical distributions of populations and species (reviewed in Avise 1989). The duration and permanence of geological barriers to gene flow is the principal factor which determines the degree of concordance between mtDNA genotypes and geography. As an example, mtDNA genotypes of freshwater fish populations were correlated with river drainages in an area of the southeastern United States which had been ecologically and geologically stable for several million years (Bermingham and Avise 1986). In contrast, geographical substructuring of mtDNA genotypes is considerably less in northern populations of fish species which had been founded within the last 20,000 years, following the last glacial excursion (Bentzen et al. 1988; Bernatchez et al. 1989).

The effective separation of major North American river systems by the Continental Divide since the Miocene, provides another opportunity to examine species evolution using mtDNA. The emergence of western mountain ranges about 5 to 10 million years ago divided North America into eastern and western regions and has lead to the evolution of two highly distinctive assemblages of fish species (Smith 1981). Comparative mtDNA analysis of closely related species found to the east and the west of the Continental Divide can suggest processes and patterns of speciation.

In addition, comparisons of DNA sequences between species are important in understanding genomic patterns of spatial and temporal variation. It is known with some certainty that fish species on opposite sides of the Continental Divide have

been separated for at least 5 million years. The longterm evolutionary significance of nucleotide change, length variation and heteroplasmy in animal mtDNA can be examined by studying closely related but reproductively isolated species with such a geographical distribution.

Sturgeon and their closest living relatives, paddlefish, are members of the Infraclass Chondrostei, a pre-Jurassic diversion from the teleost lineage (Lauder and Liem 1983; Gardiner 1984). In North America, five of the eight indigenous sturgeon species belong to the genus *Acipenser* (Doroshov 1985). The phylogeny of Chondrostean species has received little attention. In this chapter, mtDNA variation was examined in four sturgeon species of the genus *Acipenser*. The sturgeon species examined include white sturgeon (*A. transmontanus*), green sturgeon (*A. medirostris*) lake sturgeon (*A. fulvescens*) and Atlantic sturgeon (*A. oxyrhynchus*). All these species share the characteristics of individual longevity ($>$ 50 years), late age of maturity (5 to 30 years), high fecundity (5 x10^4 to 4 x10^6 eggs per female) and extensive, but largely unknown, migratory behaviour (Doroshov 1985, Conte et al. 1988).

The contemporary distributions of these four species span the North American continent from east to west (Fig. 20). *A. fulvescens* and *A. oxyrhynchus* occur only on the eastern side of the Continental Divide. *A. oxyrhynchus* is mainly found in marine coastal areas but it returns to freshwater to spawn. There are two subspecies of Atlantic sturgeon; *A. oxyrhynchus oxyrhynchus* occurs from Labrador to northeastern Florida and *A. oxyrhynchus desotoi* inhabits the Gulf of Mexico. Individuals of the former subspecies were examined in this study. *A. fulvescens* is predominantly a freshwater species and inhabits the continental interior.

The species *A. transmontanus* and *A. medirostris*, are found only on the western side of the Continental Divide. The ranges of both species overlap in coastal waters and estuaries. However, *A. transmontanus* migrates further upstream in

freshwater compared to *A. medirostris*. *A. medirostris* usually inhabits marine coastal waters and its range also includes northern Asia.

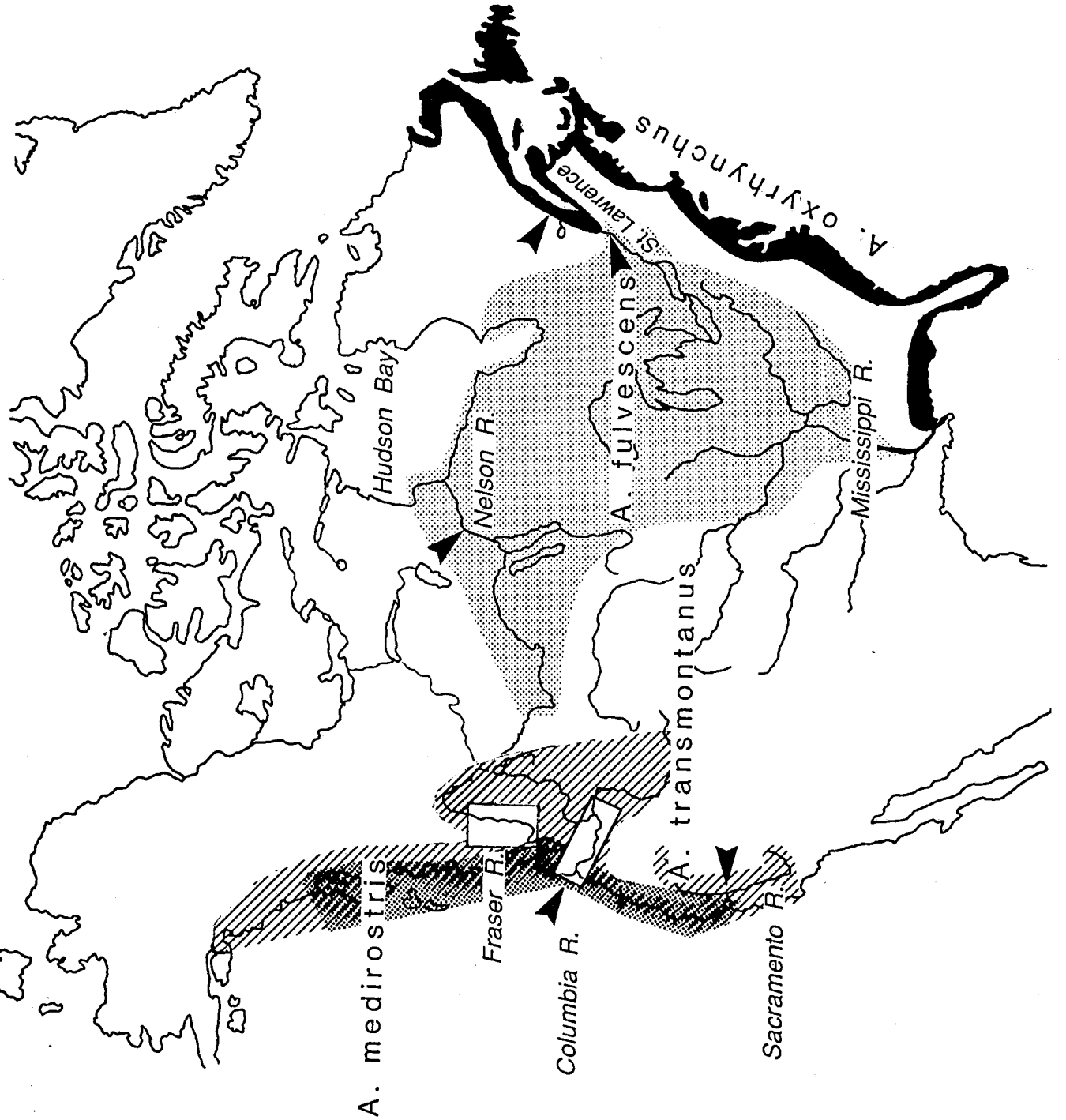
In this study, the mtDNA of these four sturgeon species were examined for nucleotide variation in the whole mtDNA using restriction enzymes and in the control region by direct DNA sequence analysis. Levels of mtDNA length variation and heteroplasmy was also determined in the four species.

Materials and Methods

Sample Collection and mtDNA Isolation

A. transmontanus specimens ($n = 174$) were caught in the Columbia and Fraser Rivers as part of the previous population analysis (Fig. 20). *A. medirostris* ($n = 10$) were obtained from the lower Columbia River. *A. oxyrhynchus* ($n = 19$) were caught in the lower St. Lawrence River, Quebec. *A. fulvescens* ($n = 21$) specimens orginated from two river systems; the Nelson River and Lake Winnipeg in Manitoba ($n = 14$) and the upper St. Lawrence River at Lac St. Pierre, Quebec ($n = 7$). In all species, mitochondrial DNA (mtDNA) was extracted from fresh liver tissue (as described in Chapter I) usually within 6 to 72 hours after dissection in the field.

Fig. 20. Distributions of the four sturgeon species in this study (Scott and Crossman, 1973). Arrows indicate specific collection sites. Boxed areas cover *Acipenser transmontanus* collection ranges on the Columbia and Fraser Rivers which are shown in detail in Figures 4 and 5, respectively.


A. oxyrhynchus
St. Lawrence
A. fulvescens
Hudson Bay
Nelson R.
Mississippi R.
A. transmontanus
A. medirostris
Fraser R.
Columbia R.
Sacramento R.

Restriction Enzyme Analysis

Interspecific comparisons were based on digests with 12 restriction enzymes with specificities for six nucleotide (*Apa*I, *Bcl*I, *Bgl*II, *Eco*RI, *Hind*III, *Pvu*II and *Xba*I), multiple six nucleotide (*Acc*I, *Ava*I, *Hae*II and *Hinc*II) and multiple five nucleotide (*Ava*II) restriction sites. Enzyme digest conditions were as specified by the vendor (BRL and Pharmacia). DNA fragments were visualized under ultra-violet light after electrophoresis in ethidium-bromide stained agarose gels.

Since only small amounts of tissue were obtained for some of the species, the extracted mtDNA was preferentially used in the Southern blot analysis. Therefore, only a limited survey was made of intraspecific restriction enzyme site variation in the species, *A. fulvescens* ($n = 12$), *A. medirostris* ($n = 9$) and *A. oxyrhynchus* ($n = 9$). Several of the restriction enzymes (*Acc*I, *Ava*I, *Ava*II and *Hinc*II) found to be polymorphic for *A. transmontanus* were used to analysis the mtDNA of the other species.

Restriction Enzyme Data Analysis

Interspecific differences in DNA fragment patterns were due to relatively few restriction enzyme site changes. Therefore, it was possible to derive divergence estimates from maps of restriction enzyme sites. Nucleotide substitutions per site (δ) between pairs of species were estimated using the maximum likelihood method of Nei and Tajima (1983).

The UPGMA method was used to construct a phylogenetic tree from estimates (d) of percent pairwise sequence divergence. In the tree, each interior node was plotted as $1/2d$ (Nei 1987; p. 293-296) and standard errors about the nodes were calculated (Nei et al. 1985). The presence or absence of restriction sites between

species were also analyzed using the Wagner parsimony method. The bootstrapped mixed parsimony algorithm (BOOT) of the PHYLIP 3.3 program package was replicated 100 times (Felsenstein 1990).

Southern Blotting and Hybridization

It was discovered early in the restriction enzyme analysis that digests with the restriction enzyme *Hae*II resulted in a fragment of varying length, from 1.5 to 2.0 kb, in all four species. Several individuals from each species were surveyed for length variation by probing Southern blots of *Hae*II digested mtDNA of individuals, with the cloned control region of *A. transmontanus* mtDNA (Chapter III).

Hybridizations were done under moderately stringent conditions of 60°C overnight followed by three washes in 1.0x SSPE. Filters were left wet and wrapped in plastic wrap for autoradiography. The film (Kodak O-Mat AR) was pre-exposed to ensure a linear response for microdensitometry readings (Laskey 1980). Several exposures (0.5 - 14 days) were made to ensure the best contrast between bands for heteroplasmic individuals.

Analysis of D-loop Hybridizations

The densitometry and analysis of autoradiographs were described in Chapter III. An *A. transmontanus* individual, heteroplasmic for five different mtDNA size classes, was included as a size standard in interspecific Southern blots. MtDNA size classes observed in *A. medirostris*, *A. fulvescens* and *A. oxyrhynchus* were recorded relative to size variants seen in *A. transmontanus*.

The small sample sizes of three species made it inappropriate to apply the hierarchial statistics used in Chapter III. However, the within individual diversity (K_b) was calculated as a relative measure of heteroplasmy (Birky et al. 1983; Rand and Harrison 1989). Pairwise comparisons were made of size frequency distributions

among the species using programs P2D and P3D of the BMDP statistical software package (Dixon 1983).

PCR Amplification and DNA Sequencing

All PCR amplifications were done from CsCl purified sturgeon mtDNA extracted from liver tissue (as described in Chapter II). Primers were designed from the published sequence of the control or D-loop region of *A. transmontanus* (Buroker et al. 1990). Sequences of the primers follow, the letters L and H refer to the light and heavy strands and the number refers to the position of the primer's 3' base in the reference 1.6 kb fragment: L185 (5' CATCTACCATTAAATGTTATAC 3'), L317 (5' CTGTAGGGATTCACAACTG 3'), L506 (5' TGACATGTAGAACTCCTTCAGA 3'), H164 (5' GTATAACATTTAATGGTAGATG 3'; the complement of L185), H245 (5' ATGTTCGGCTCTTACTACA 3') and H740 (5' GATCAAGGTATGTCGATGACA 3'). Another primer was designed to anneal just outside the D-loop in the $tRNA^{Pro}$ gene (Gilbert et al. 1988). According to the published sequence for that gene, it was labelled tPro123 (5' ACCCTTAACTCCCAAAGC 3').

The primer combination, tPro123 and H740 consistently amplified an approximately 800 nucleotide long fragment of the control region in all species. The other oligonucleotides were used as internal primers for sequencing either the hypervariable area (L185, L317 and L506) or the area with repeat sequences (H164 and H245).

PCR amplifications were done using the GeneAmp DNA Amplification Reagent Kit (Perkin Elmer Cetus). Each amplification was performed in 50 µl final volume with 5 µl of 10x reaction buffer (500 mM KCl, 100mM Tris-HCl (pH 8.3), 15 mM $MgCl_2$, 0.1 % gelatin), each dNTP at 200 mM, each primer at 5-7 µM, mtDNA

(5-50 ng) and 0.5 units of *Thermus aquaticus* DNA polymerase with 50 µl of mineral oil overlaid. Each PCR reaction cycle consisted of denaturation for 1.5 min at 95°C, primer annealing for 2 min at 42°C and extension for 2-3 min at 72°C. Cycles were repeated 40 times in a programmable cyclic reactor (Ericomp Inc.) and the final cycle included an extension reaction of 10 min. A negative control amplification reaction (all the above reagents but no DNA) was performed in all PCR series to check for possible contamination.

Amplified double-stranded DNA was phenol extracted from 1.5 % low melting point agarose after electrophoresis for 2 to 3 hr. at 50v in 40mM Tris acetate buffer (pH 8.0) with ethidium bromide added (Sambrook et al. 1989). After precipitation with 95 % EtOH and 3 M NH_4OAc, DNA was rehydrated in 7 µl of sterile water.

PCR amplified double-stranded DNA was sequenced directly using the dideoxynucleotide chain termination method (Sanger et al. 1977) and Sequenase T7 polymerase (United States Biochemical) with some modifications. Template DNA, 2 µl of 5x Sequenase reaction buffer, 1 µl of primer and 1 µl of DMSO were heated to 95-98°C for 4 min then quick frozen in a bath of dry ice and 95 % EtOH bath. The annealed primer and template DNA reactions were allowed to warm slowly to room temperature for 20-30 min.

The remaining steps followed the Sequenase kit protocol except the label mix was diluted 1:50, label reaction time was reduced to 30 seconds and the termination reaction was completed at 42°C for 2-3 min (S. Palumbi, Department of Zoology, University of Hawaii, Honolulu; pers. commun.). Sequencing reactions were resolved in 6 % polyacrylamide-7 M urea gels using a salt gradient running buffer (Ausubel et al. 1989). Gels were vacuum dried at 80 °C and autoradiographed on Kodak X-Omat AR film for 12-72 hr.

DNA Sequence Analysis

D-loop nucleotide sequence data were analysed separately for the hypervariable and tandemly repeat sequence regions. Alignment of highly variable DNA sequences is problematic (Swofford and Olsen 1990). Existing computer algorithms will not reliably produce the best alignment but can help guide the visual alignment of sequences. Accordingly, preliminary sequence alignments were obtained using the computer programs Delaney (Delaney Software Ltd.) and Multialign. These alignments were improved upon by visually aligning sequences with the program ESEE (Cabot and Beckenbach 1989). An attempt was made to properly align long blocks of conserved sequences using a minimum number of introduced gaps.

The PHYLIP 3.3 program package was used for the bootstrapped mixed parsimony analysis (DNABOOT) with 100 replicates. The parsimony tree topology was not affected by changing the ordering of sequences entered into the analysis. The program DNADIST was used to calculate pairwise estimates of sequence divergence with a correction for multiple substitutions (Kimura 1980). The corrected distance estimates were used to construct a tree using the neighbour joining method (Saitou and Nei 1987).

Results

Restriction Enzyme Phylogeny

The 12 restriction enzymes detected between 40 (*A. oxyrhynchus*) to 59 (*A. fulvescens*) sites per mtDNA genome (Fig. 21). DNA fragment lengths for each enzyme are presented in Appendix A.

Sequence divergence estimates among the four sturgeon species (Table 13) were at least fourfold greater than intraspecific variation observed in *A. transmontanus* (see Table 1). The two most closely related species, *A. transmontanus* and *A. medirostris*, are about 4.3 $\pm$ 0.96 % divergent. *A. fulvescens* is 7.1 $\pm$ 1.24 % diverged from these two species. Sequence divergence of *A. oxyrhynchus* is nearly 8.6 $\pm$ 1.42 % from the cluster of the other 3 species. In the UPGMA tree, the *A. transmontanus* / *A. medirostris* node was significantly different from the *A. fulvescens* branching point because standard error (SE) bars did not overlap (Fig. 22).

The phylogeny derived by UPGMA analysis was well supported by bootstrapped Wagner parsimony analysis (Fig. 22). In 100 replicates, the *A. oxyrhynchus* / *A. fulvescens* and *A. transmontanus* / *A. medirostris* branching points occurred 100 and 89 times, respectively.

Intra-specific Variation

Fewer restriction site polymorphisms were found in *A. fulvescens*, *A. medirostris* and *A. oxyrhynchus* relative to *A. transmontanus* (Chapter I), although this is a preliminary finding given the small sample sizes. Nine *A. oxyrhnychus* individuals were monomorphic for all restriction enzymes.

Fig. 21. An ethidium bromide stained agarose gel with a *Hinc*II restriction digest of mtDNA from *A. transmontanus* (At), *A. medirostris* (Am), *A. fulvescens* (Af), and *A. oxyrhynchus* (Ao). *A. transmontanus* individuals are also polymorphic a single restriction site. The size marker lambda phage DNA double digested with the enzymes *Eco*RI and *Hind*III.

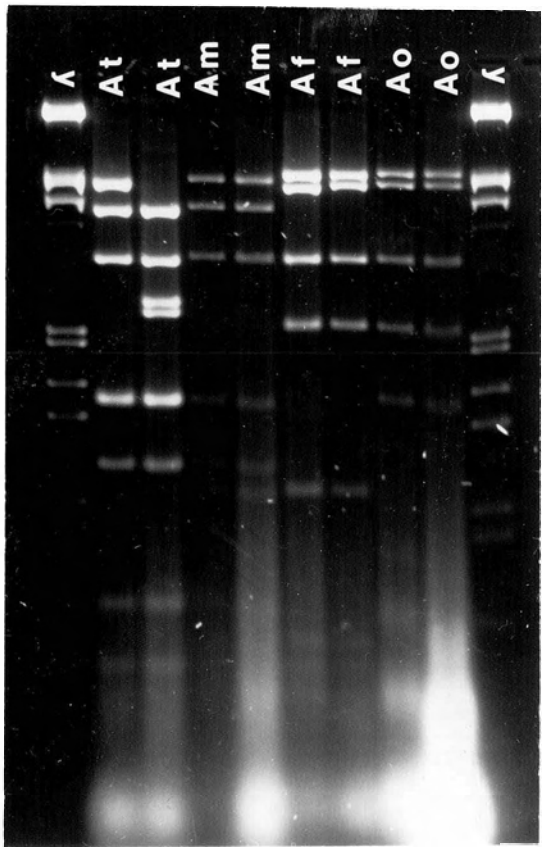
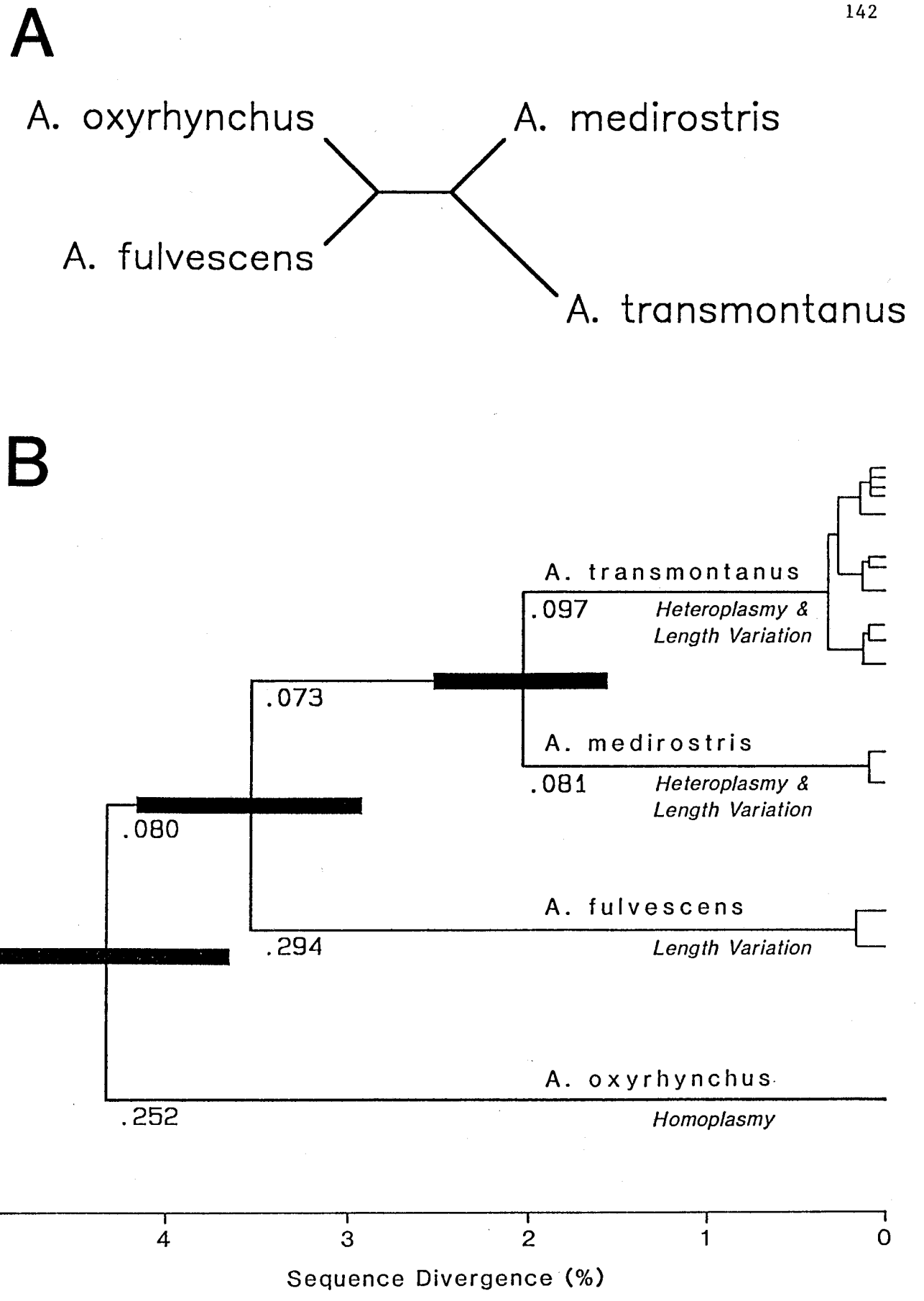

λ
At
At
Am
Am
Af
Af
Ao
Ao
λ

Table 13. Pairwise divergence estimates for four species of sturgeon of the genus *Acipenser*. Percent sequence divergence from differences in restriction enzyme sites and standard deviations (in parenthesis) are given in the lower diagonal. In the upper diagonal are shown percent differences from nucleotide sequence comparisons of the hypervariable region and the repeated region (in parenthesis) of the D-loop. Percent sequence differences for nucleotide comparisons were corrected for multiple substitutions (Kimura 1980).

Species	*A. transmontanus*	*A. medirostris*	*A. fulvescens*	*A.oxyrhynchus*
A. transmontanus	--	17.92	46.99	50.00
	--	(14.82)	(18.54)	(25.54)
A. medirostris	4.14	--	44.47	49.27
	(0.96)	--	(16.22)	(15.60)
A. fulvescens	5.91	8.27	--	62.72
	(1.18)	(1.50)	--	(27.63)
A. oxyrhynchus	7.82	8.49	9.07	--
	(1.59)	(1.69)	(1.72)	--

Fig. 22. Phylogenetic relationships of sturgeon species of the genus *Acipenser*. A) The most parsimonious tree of the four sturgeon species. In 100 bootstrap replicates of data on 1) the loss or gain of restriction sites over the entire mtDNA molecule and 2) 425 nucleotides of the D-loop hypervariable region, this tree was supported 89 times and 100 times, respectively.

B) Phylogenetic relationships according to distance methods. A UPGMA tree based on percent pairwise divergence from restriction enzyme site data is shown. Solid bars representing one standard error on either side of each node. Numbers along the branchs are branch lengths estimated by the neighbor joining method from the pairwise divergence of 425 nucleotides in the hypervariable region. Divergence estimates for DNA sequences were corrected for multiple changes (Kimura 1980). The occurrence of mtDNA length variation and heteroplasmy determined from Southern blot analysis is shown for each species.


A
A. oxyrhynchus
A. medirostris
A. fulvescens
A. transmontanus
B
A. transmontanus
.097
Heteroplasmy &
Length Variation
.073
A. medirostris
.081
Heteroplasmy &
Length Variation
.080
A. fulvescens
.294
Length Variation
A. oxyrhynchus
.252
Homoplasmy
5
4
3
2
1
0
Sequence Divergence (%)

One of nine *A. medirostris* individuals had an additional *Ava*II site (genotype A.m.-*B* in Appendix A) and intra-specific divergence in this species was estimated to be 0.174 %.

Among 12 *A. fulvescens* individuals, two genotypes were distinguished by single polymorphic sites in *Hinc*II and *Ava*II digests (Appendix A). Divergence between these two genotypes was 0.314 %. Six individuals from the Nelson River, Manitoba and four individuals from Lac St. Pierre, Quebec, shared the same genotype (A.f.-*A*). The second genotype was found in two individuals from the Nelson River (A.f.-*B*).

Length Polymorphisms and Heteroplasmy

In all four species, *Hae*II digests produced a fragment varying in length from 1.5 to 2.0 kb. Southern blot analysis revealed that the size range of this fragment was consistent among species and within the range of variation observed for *A. transmontanus* (Fig. 23).

It was shown earlier for *A. transmontanus* that about 42 % of 174 individuals were heteroplasmic for more than one mtDNA size class (Chapter II). In *A. medirostris* ($n = 10$), 50 % of individuals appeared to be heteroplasmic for a size range of mtDNA similar to that found in *A. transmontanus* (Table 14). Diversity (Kb) within individuals was slightly higher for *A. medirostris*. In terms of *A. transmontanus* mtDNA size classes, the size of mtDNA types in *A. medirostris* ranged from one to four perfect repeats.

Different length mtDNA, but no apparent heteroplasmy, were observed in *A. fulvescens* individuals ($n = 21$). In one individual, there was a slight blurring around the dominant mtDNA size class which may be a second size variant with a very low frequency.

Fig. 23. Autoradiograph of *Hae*II digested mtDNA from *A. transmontanus* (At), *A. medirostris* (Am), *A. fulvescens* (Af) and *A. oxyrhynchus* (Ao) probed with the control region or D-loop of *A. transmontanus*. The bands (bottom to top) in *A. transmontanus* correspond with mtDNA size classes with 1 to 5 tandemly repeated 82 bp sequences.

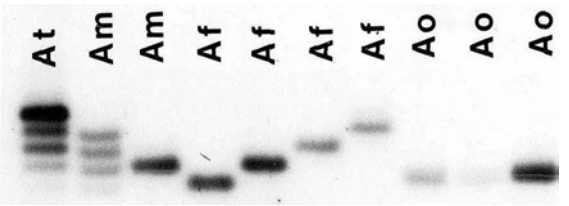

At
Am
Am
Af
Af
Af
Af
Ao
Ao
Ao

Table 14. Mean frequencies of mtDNA size variants (RPT1 - 6) and heteroplasmy in four species of *Acipenser* sturgeon. In all species, size variants were defined on the basis of equivalent gel migration to mtDNA size classes in *A. transmontanus*. MtDNA size classes (RPT1 to RPT6) in that species are caused by a varying number (1 to 6) of perfect 82 bp repeats in the control region. $\overline{K_b}$ is the mean diversity of within individuals of the species (Birky et al. 1983). *HP* is the proportion of heteroplasmic individuals detected in the sample.

Species	*n*	RPT1	RPT2	RPT3	RPT4	RPT5	RPT6	$\overline{K_b}$	*HP*
A. transmontanus	174	0.273	0.251	0.283	0.144	0.031	0.018	0.183	0.42
A. medirostris	10	0.027	0.306	0.476	0.192	0.0	0.0	0.247	0.50
A. fulvescens	21	0.381	0.143	0.238	0.190	0.048	0.0	0.0	<0.05
A. oxyrhynchus	19	0.947	0.053	0.0	0.0	0.0	0.0	0.0	0.0

However, the shadow band was faint and indistinct compared to very low frequency (< 0.05) mtDNA size classes found in *A. transmontanus*. The range of mtDNA sizes in *A. fulvescens* corresponded with one to five perfect repeat units in *A. transmontanus*.

No heteroplasmy and very little length variation was observed in mtDNA of *A. oxyrhynchus* ($n = 19$). The common size of mtDNA in Atlantic sturgeon was equivalent to the smallest size variant (one repeat) found in the other three species. One Atlantic sturgeon had a slightly larger mtDNA which corresponded to *A. transmontanus* mtDNA with two repeats.

In pairwise comparisons, there were no significant differences between the frequency distributions of mtDNA size classes of *A. transmontanus*, *A. medirostris* and *A. fulvescens*. *A. oxyrhynchus*, having little size variation in its mtDNA, was significantly different (t-tests; $p < 0.001$) from these three species. Since *A. fulvescens* and *A. oxyrhynchus* were not heteroplasmic, *Kb* equals zero for these species.

Hypervariable Sequences

A 541 nucleotide sequence of the hypervariable portion of the control region was determined for *A. transmontanus*, *A. medirostris* and *A. oxryhnychus* (Fig. 24). PCR amplifications and direct sequencing were not as effective for *A. fulvescens* mtDNA. The compositions of a 52 nucleotide and 78 nucleotide sequence in the middle and at the end of the control region, respectively, have yet to be determined in this species.

Sequence variation had a spatially heterogeneous distribution in the D-loop region (Fig. 25). There were some highly conserved areas of the D-loop, most notably the consensus sequence blocks (CSB) found in control regions of vertebrate mtDNA (reviewed in Clayton 1982).

Fig. 24. Aligned sequences of the hypervariable region in the D-loop or control region of four sturgeon species. Dots indicate nucleotide positions unchanged from the reference sequence. Dashes are gaps introduced to optimize the sequence alignment. Blank areas in the sequence of *A. fulvescens* D-loop indicates regions yet to be sequenced in this species. Underlined are the locations of consensus sequence blocks (CSB) found in the control regions of other vertebrate species (reviewed in Clayton 1982).

```
A. transmontanus TTTATGTGCACTAACATGAAAACCTCTGATAACTT-AAAATGTAGTAAGAGCCGAACATG 59
A. medirostris   .C..............ACG.TTTC.C....-...AG.T...................A
A. fulvescens    .C....C.AC........CC..TTC....G.-.CAT..T...................T
A. oxryhnychus   .C............GCA..CTT-C.C..C...A.T...G...................-

A. transmontanus GAGATATGTCTAGAACATAAAGTTAATGAGATGAGGGACAATAACTGTAGGGATTCACAA 119
A. medirostris   A..G.C............T.....................C..................
A. fulvescens    CTAT.C....C........GG.............AA.G..G...T.....AAT.A..T..
A. oxryhnychus   TCAT.C.T.T....-......C.............AA....G...TGCG.AAT.GCATA..

A. transmontanus CTGAACTATTACTGGCATCTGGTTCCTATTTCAGGTCCATTAACAGTTATTTCCCCATAA 179
A. medirostris   .....................................G.T..C.............
A. fulvescens    .....................................A.C.GT.GCTACA........
A. oxryhnychus   T.....................TA...C.....CA...-.....C.C.C.........

A. transmontanus CTGAA-CTATGTC-TGGCATCTGATTAATGTTGGAGGTACTATGAGATCCATGACCCCAC 237
A. medirostris   ..............................A.......AT-A.AG.............
A. fulvescens    ..AG.T.-G.TA.....................A..A....C.-ACAG...G...GGT.C.
A. oxryhnychus   T..G.T..G....C......T............GAA...AT-ACA-.........A.CA

A. transmontanus ATGCCGAGAATCTTGTCAAC-ATTTGGTATCTTTTATTCGGGTTT-CCATTCACTGACAT 295
A. medirostris   .....A.........C..............T.......T......A.............
A. fulvescens    .....A....C.CCA..............ATC....A.TTTAG--
A. oxryhnychus   T...G.GC...T..AC...GC.......TAT.......G....--C.............

A. transmontanus GTAGAACTCCTTCAGAGAAGAACAACAAGGTGGAACATACGCGA-CTGTCCGAGAGGATG 354
A. medirostris   ..................G..................TTATA.C...C...G...A...
A. fulvescens                                         .TGTGT-TTGC..GA..TTA..T
A. oxryhnychus   ---AC...T.......AT...TATCT.....A.....-GTATATCA.TCGTTCA....CA
```

CSB 1

```
A. transmontanus AATAATGAATGGTACAATGACATATCCCTGATGTCAC-ACATGGCC--TGTGCTGTGTAC 411
A. medirostris   ....G.....................T...C.CT.T......A......CA..AC.C..
A. fulvescens    ....G......ACTT............-...ATATCAC....A.T......A.CA..C..
A. oxryhnychus   ..C.G.......C............T.-...-TAT.TC...C.A..CC..CTGC..AC..

A. transmontanus AGAGAGA--T-GTTTCACAGAGCCTGGTTTTAATCTTTT----CACATGACAA-TCATGG 463
A. medirostris   .T.T...GG.........G..A.........GGC.C..ACCCT........TCAAA..AA
A. fulvescens    .TGT..TGAGC....AC.GAG....AAG.C.T.C.CCCACATAGTA..C.A.TGC..CAA
A. oxryhnychus   .A.ACATTT.AC.A.A..CCTAAG.A-...GC.CAAACACTAAATA.CC.A..TA.T.AA
```

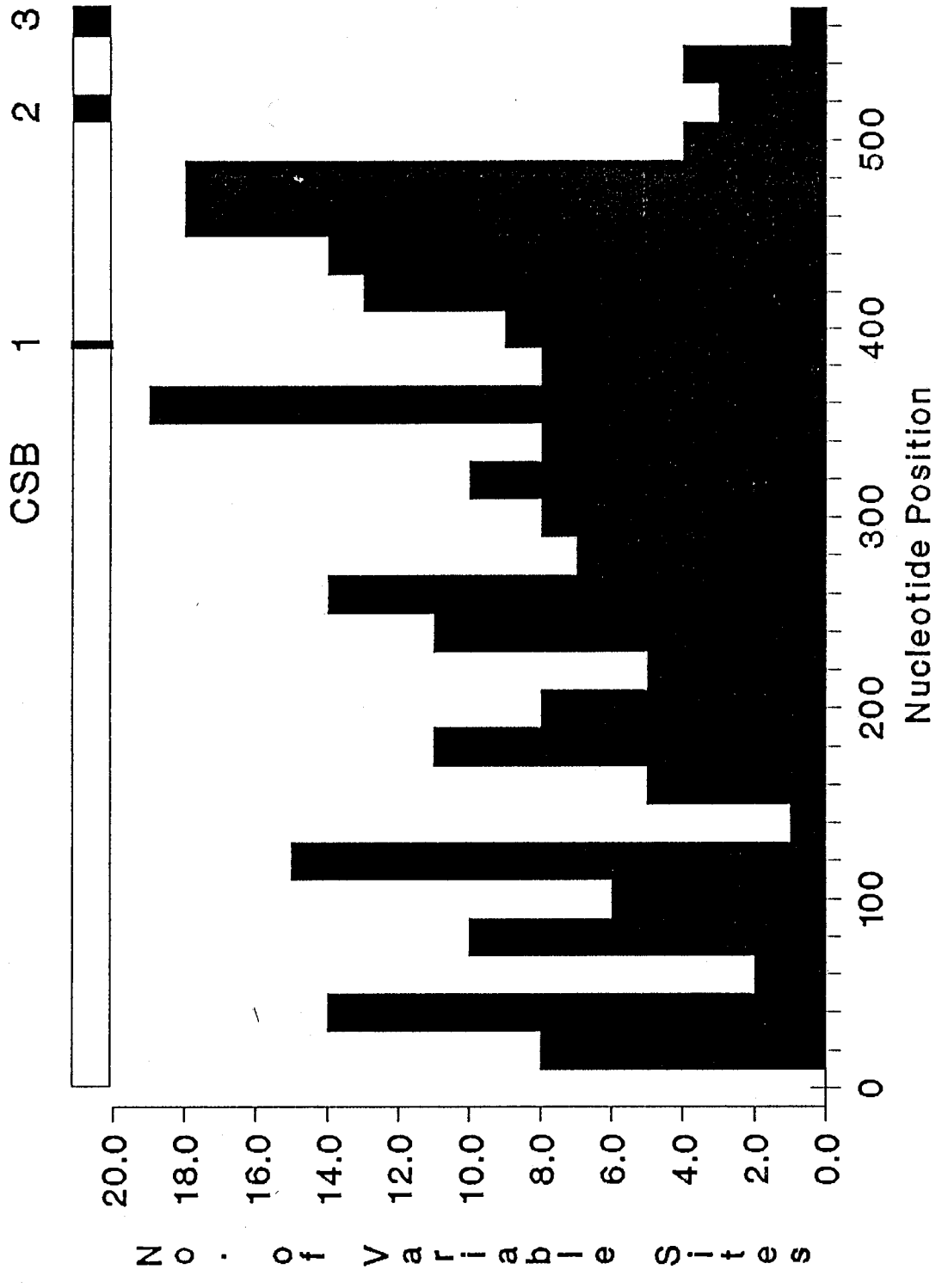
CSB 2

```
A. transmontanus ACGTTTACTATCGACAAACCCCC-TACCCCCTTATGTCGGACAGGCCTTA-TATTTCTT 520
A. medirostris   .......T..............CC.............................A.
A. fulvescens
A. oxryhnychus   ...C...T..............TAC....-...C........AA.....A......C.
```

CSB 3

```
A. transmontanus GTCAAACCCCAAAAGCAGGA                                      541
A. medirostris   ....................
A. fulvescens
A. oxryhnychus   .C..................
```

Fig. 25. Number of variable nucleotide sites (in blocks of 20 nucleotides) in the hypervariable region of the D-loop. Figure at the top showns the relative location of the CSB blocks in the sequence.


CSB
1
2
3
No. of Variable Sites
20.0
18.0
16.0
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
0
100
200
300
400
500
Nucleotide Position

However, overall sequence divergence was high and ranged from 17.9 % to 62.7 % (Table 13). Pairwise percent divergence estimates were four- to sevenfold higher than estimates from restriction enzyme data.

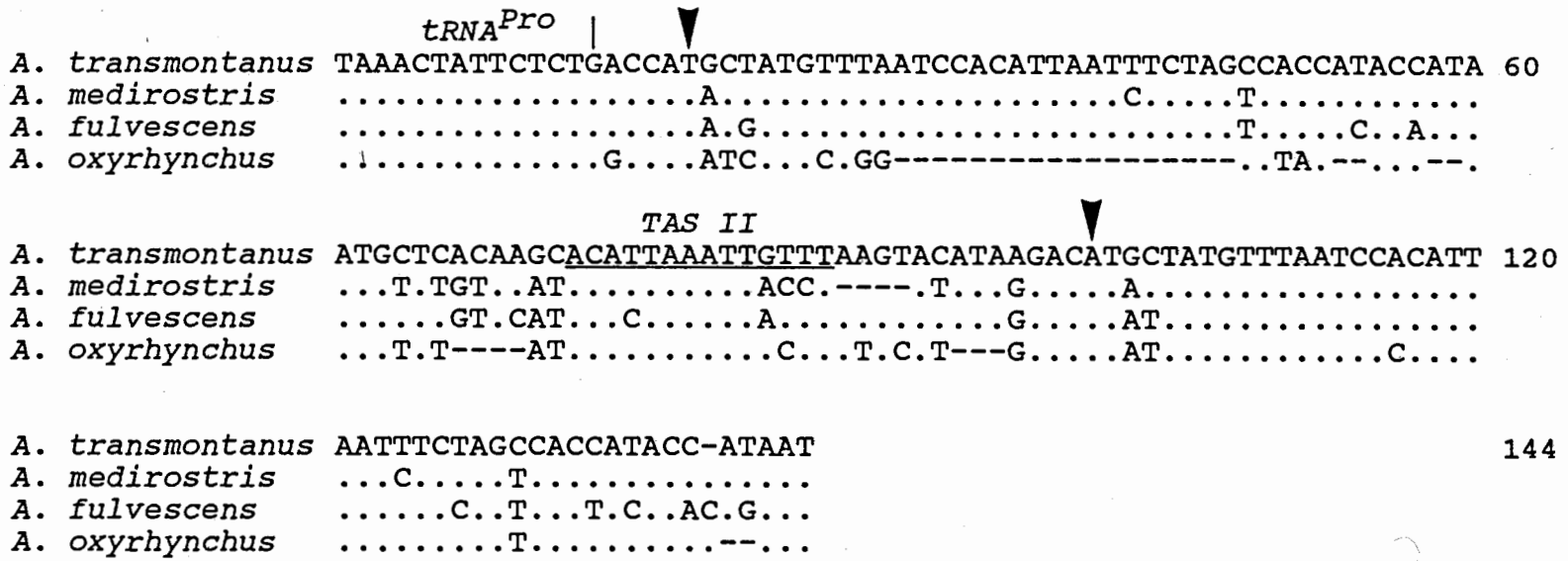
Since complete sequences of all four species were required for phyletic analysis, regions not sequenced in *A. fulvescens* were also excluded for the other species. In total, 425 nucleotides of homologous sequence were used to construct the phylogenetic trees. This analysis resulted in neighbour joining and parsimony trees which were congrugent with trees derived from restriction enzyme data (Fig. 22). The divergence of western species, *A. transmontanus* and *A. medirostris* from the eastern species, *A. fulvescens* and *A. oxyrhynchus*, was highly significant in the bootstrap parsimony analysis.

Divergence of the Repeat Region

In all four species, nucleotide sequence data was obtained for the region of the D-loop known to have multiple 82 bp repeat sequences in *A. transmontanus* (Buroker et al. 1990). Partial sequences of the $tRNA^{Pro}$ gene were totally conserved among the four species. In contrast, several nucleotide differences were found in the D-loop region within a few base pairs of the junction with the 3' end of $tRNA^{Pro}$ gene.

The tentative ends of the 82 bp repeat sequence in *A. transmontanus* are shown in Fig. 26. Partial sequence of a second copy of the repeat begins immediately after the 3' end of the first repeat. This region is also repeated in *A. medirostris* with the partial sequence of the second copy of the repeat showing complete homology with the first repeat. However, there were several nucleotide differences between the repeat sequences in *A. medirostris* and *A.transmontanus*. The most notable change was the deletion of 4 nucleotides in the *A. medirostris* sequence thus the length of each repeat is 78 nucleotides in this species compared with an 82 nucleotide long repeat seen in *A. transmontanus*.

Fig. 26. Aligned sequences of the repeated region of the D-loop in four sturgeon species. The junction between the D-loop and $tRNA^{Pro}$ is indicated. Arrows indicate the boundary of the 82 bp repeat sequence in *A. transmontanus*. A similar repeat sequence, 78 bp long, occurs in *A. medirostris*. The location of a putative termination associated sequence (TAS II) is also indicated.


tRNA^Pro
TAS II
A. transmontanus TAAACTATTCTCTGACCATGCTATGTTTAATCCACATTAATTTCTAGCCACCATACCATA 60
A. medirostris .....A.....C.....T.....
A. fulvescens .....A.G.....T.....C..A...
A. oxyrhynchus .....G....ATC...C.GG----------..TA.--...--.
A. transmontanus ATGCTCACAAGCACATTAAATTGTTTAAGTACATAAGACATGCTATGTTTAATCCACATT 120
A. medirostris ...T.TGT..AT.....ACC.----.T...G.....A.....
A. fulvescens .....GT.CAT...C.....A.....G.....AT.....
A. oxyrhynchus ...T.T----AT.....C...T.C.T---G.....AT.....C....
A. transmontanus AATTTCTAGCCACCATACC-ATAAT 144
A. medirostris ...C.....T.....
A. fulvescens .....C..T...T.C..AC.G...
A. oxyrhynchus .....T.....--...

A repeated sequence arrangement was not evident in *A. fulvescens* or *A. oxyrhynchus*. The "repeated" region was highly diverged in *A. oxyrhynchus* where the best probable alignment required the introduction of several gaps, including a gap 18 bp long beginning at nucleotide position 29. In *A. fulvescens*, no deletions occurred but a single C residue was inserted at position 139. Numerous nucleotide substitutions, mostly transitions, occurred among species including several in the putative termination associated sequence or TAS II (Doda et al. 1981; Buroker et al. 1990).

Overall sequence divergence in this region is lower than the hypervariable portion of the D-loop but higher than restriction enzyme estimates (Table 13). *A. medirostris* was almost equally diverged from *A. transmontanus* (14.8 %), *A. fulvescens* (16.2 %) and *A. oxyrhynchus* (15.6 %). The greatest sequence divergence was *A. oxyrhynchus* from *A. transmontanus* (25.5 %) and *A. fulvescens* (27.6 %). The divergence (18.5 %) between among repeated regions of *A. transmontanus* and *A. fulvescens* was intermediate among the species.

Relative Rates of Nucleotide Change

In Table 15 are shown the interspecific relative rates of nucleotide change of whole mtDNA and the repeated and hypervariable regions of the D-loop. These relative rates are ratios of the estimates of pairwise sequence divergence derived from restriction enzyme and sequence data in Table 13. Rates of change for the hypervariable region relative to the entire mtDNA ranged from about 4.3 for the comparison of *A. transmontanus* / *A. medirostris* to 8.0 for *A. transmontanus* / *A. fulvescens*.

Table 15. Interspecific relative rates of nucleotide substitutions of entire mitochondrial DNA (mtDNA) and the hypervariable (Hyper) and repeated (Rept) regions of the D-loop in the genus *Acipenser*. Relative rates are the ratios of percent pairwise divergence estimates in Table 13. Restriction enzyme divergence estimates were used as the rate for the entire mtDNA molecule.

Species Comparisons	Hyper:mtDNA	Rept:mtDNA	Hyper:Rept
A. transmontanus / *A. medirostris*	4.33	3.56	1.21
A. transmontanus / *A. fulvescens*	7.95	3.18	2.53
A. transmontanus / *A. oxyrhynchus*	6.39	3.27	1.96
A. medirostris / *A. fulvescens*	5.38	1.96	2.74
A. medirostris / *A. oxyrhynchus*	5.80	3.25	3.16
A. oxyrhynchus / *A. fulvescens*	6.92	3.04	2.25

The rate of nucleotide change for the repeated region of the D-loop was generally threefold greater than rates for the entire mtDNA. However, the relative rate of change for this region in the comparison of *A. medirostris* / *A. fulvescens* was only twofold higher than the rate for the entire mtDNA.

The interspecific relative rates of nucleotide change in the hypervariable region to the repeated D-loop region varied among species comparisons. In the comparison of *A.transmontanus* / *A. medirostris*, rates for the two areas of the D-loop were nearly equal while for *A. oxyrhynchus* / *A. medirostris* nucleotide changes in the hypervariable region occurred at a threefold higher rate than the repeated area.

Discussion

Interspecific Phylogeny

The interspecific mtDNA phylogeny of the genus *Acipenser,* as reconstructed from restriction enzyme and hypervariable sequence data, clearly shows the evolution of distinctive species assembleages on opposite sides of the Continental Divide. MtDNA analyses revealed that the Northwest Pacific species, *A. transmontanus* and *A. medirostris*, are significantly more closely related to each other than to either species found east of the Continental Divide, *A. fulvescens* and *A. oxyrhnychus.* In distance and parsimony analysis, the nodes of the two species groups were significantly different for both restriction enzyme and D-loop sequence data.

The general rate of nucleotide substitutions in vertebrate mtDNA estimated from restriction enzyme analysis is 2 % per million years (reviewed in Brown 1985; Wilson et al. 1985). This rate estimate has not been confirmed for fish mtDNA and should be regarded as being highly provisional for Chondrostreans. Accordingly, *A. transmontanus* and *A. medirostris* diverged from *A. fulvescens* about 3.5 $\pm$0.62 MYA which is nearly concordant with Miocene western block faulting and uplifting which

occurred between 5 to 10 MYA (Cavender 1986). However, this time of divergence better corresponds with a proposed Pliocene connection between the Columbia River and eastern drainages in central Montana and South Dakota (McPhail and Lindsay 1986). The existence of this route was suggested by the presence of fossil remains of several eastern Teleost genera (*Esox*, *Ictalurus* and *Myoxocephalus*) in Cenozoic deposits in the Columbia River. A Pliocene connection may have facilitated the dispersal of *A. fulvescens* west of the Continental Divide. The closure of this east-west corridor would have isolated western sturgeon populations and led to their subsequent divergence from eastern fauna.

MtDNA analyses suggest the divergence of *A. transmontanus* from *A. medirostris* occurred about 2.1 $\pm$0.48 MYA. These species are sympatric with only slight differences in their contemporary distributions. *A. medirostris* appears to prefer estuaries and marine coastal waters and migrates only short distances inland to spawn in freshwater (Scott and Crossman 1973). *A. transmontanus* has a similar marine distribution but can be found much further inland in lakes and rivers along the western slope of the Continental Divide.

Separation of *A. oxyrhynchus* from the other sturgeon species was the result of an ancient isolation event (at least 4.3 $\pm$0.71 MYA), perhaps in the Mississippi drainage. The contemporary distributions of *A. fulvescens* includes the middle to upper areas of the Mississippi drainage while *A. oxyrhynchus* is limited to the lower river and coastal waters (Swift et al. 1986). The ecological stability of the Mississippi drainage throughout the Cenozoic allowed for a rich accumulation of species diversity which may have also included the sturgeons (Smith 1981; Robison 1986).

Actual divergence times in the genus *Acipenser* may be greater than those estimated from the nucleotide substitution rate of 2 % per million years. The earliest fossil remains of *Acipenser* were found in Upper Cretaceous beds but the emergence of this genus probably significantly preceeded this period (Lauder and Liem 1983;

Gardiner 1984). Some inconsistences between clock estimates based on restriction enzyme analysis of mtDNA and geographical changes associated with speciation have been previously noted in the mtDNA evolution of other lower vertebrates such as tortoises (*Xerobates* sp.; Lamb et al. 1989) and newts (*Triturus* sp.; Wallis and Arntzen 1989).

A highly significant interspecific phylogeny was established using the DNA sequence of the hypervariable region of the D-loop. However, the level of interspecific variation in this region may be too great for establishing reliable time estimates of specific divergence events. Possible sources of error in estimating divergence times from hypervariable sequences include 1) unknown levels of intraspecific variation in species other than *A. transmontanus*, 2) sequence alignment ambiguities, 3) saturation of sites for nucleotide change (Moritz et al. 1987), 4) unequal rates of nucleotide change between species.

Interspecific pairwise divergence estimates were four to thirty times greater than intraspecific comparisons for *A. transmontanus* (Table 6; Chapter II). The complete hypervariable regions of two individual *A. medirostris* were sequenced and found to share complete sequence identity. Therefore, intraspecific variation is probably only a small component of the sequence differences observed between species.

The analysis of relative nucleotide substitution rates (Table 15) suggests that the hypervariable region is saturated for single base changes between species. Rates of nucleotide substitutions in the hypervariable region relatve to the whole mtDNA are more variable than relative rates estimated for the repeat region. Relative rates of the hypervariable region to the repeat region also vary greatly. However, different evolutionary constraints on the repeat region between the species must also be considered (see below).

The estimated rate of nucleotide divergence of the control region in human mtDNA is 8 % per million years (Viligant et al. 1989). Application of this estimated rate to sturgeon mtDNA sequences is highly provisional, particularly in light of probable saturation for nucleotide site changes in the hypervariable region. Therefore, the divergence of *A. transmontanus* and *A. medirostris* from *A. fulvescens* is very tentatively estimated to have occurred about 5.7 MYA according to differences between hypervariable regions. Similarly, the bifurcation of *A. transmontanus* and *A. medirostris* is estimated to have occurred about 2.2 MYA. These estimated times of divergence are relatively close to those derived from restriction enzyme data.

The evolutionary relationships of three of the four sturgeon species in the present study are in general agreement with protein electrophoresis studies. Bartley (1987) determined that *A. transmontanus* and *A. medirostris* had the greatest genetic identity while *A. fulvescens* was more closely related to *A. brevirostrum*. *A. brevirostrum* inhabits Atlantic coastal waters and its range overlaps that of *A. oxyrhynchus* (which was not represented in Bartley's study).

MtDNA Variation in Sturgeon

Recent dispersals from only one or two refuges probably accounts for the apparent lack of intraspecific geographical segregation of mtDNA genotypes in both *A. transmontanus* and *A. fulvescens*. MtDNA genotypes were similar among *A. fulvescens* populations in the Nelson River and the St. Lawrence River although these drainages have been isolated for at least 8,000 years (Crossman and McAllister 1986). Approximately 12,000 yr ago, these two drainages were interconnected with each other and the Mississippian refugium by a series of large lakes formed during the northward retreat of the ice-sheet. These temporary routes may have facilitated the colonization of the Hudson Bay and St. Lawrence River drainages by a common

group of founders. However, postglacial dispersal patterns of fish from southern refuges probably varied greatly between species. Structuring of mtDNA variation in walleye (*Stizostedion vitreum*) caught in the Great Lakes and northern Manitoba suggested that founders of these populations possibly originated from not one but three separate refuges (Ward et al. 1989).

Intraspecific mtDNA variation was low in *A. medirostris* and *A. oxyrhynchus*. Although sample sizes ($n = 9$) of these species were small, general application of the combinatorial analysis of Fraser River populations of *A. transmontanus* (Chapter I; Fig. 6) reveal that a minimum of 3-4 genotypes, if present, should have been detected in each species rather than the observed 1-2 genotypes.

A. medirostris and *A. oxyrhynchus* spend a significant part of their life in marine coastal waters. Other species with marine life histories have been known to be homogeneous for mtDNA genotypes over large distances (Avise et al. 1986; Avise et al. 1987; Palumbi and Wilson 1990). In addition, effective population sizes of sturgeon species may have gone through several periods of constriction and expansion as a consequence of their ancient ancestry, with the net effect of a reduction in mtDNA diversity.

Low genetic variability has been noted for other North American Chondrosteans. Little intraspecific variation was observed in a protein electrophoretic study of American paddlefish (*Polyodon spathula*) populations (Carlson et al. 1982). Two other species of sturgeon, *Scaphirhynchus albus* and *S. platorynchus*, were indistinguishable at 37 loci (Phelps and Allendorf 1983). Restriction enzyme analysis have shown low levels of mtDNA diversity in populations of Atlantic sturgeon in the southern United States (Bowen and Avise 1990) and lake sturgeon in Quebec (S. Guenette, University of Quebec, Montreal; pers. commun.).

The Evolution of mtDNA Size Variation in Sturgeon

The occurence of mtDNA size variation and heteroplasmy among species of the genus Acipenser appears to be highly concordant with the proposed phylogeny of this group (Fig. 22). Heteroplasmy was absent in species found to the east of the Continental Divide. In *A. oxyrhynchus*, nearly all individuals were homoplasmic for a single, small mtDNA size variant. *A. fulvescens* individuals were fixed for a particular size variant and showed no evidence of heteroplasmy.

Given that, for samples of *A. medirostris* ($n = 10$) and *A. transmontanus* ($n = 174$) between 42 % and 50 % of individuals were heteroplasmic, respectively, sample sizes for *A. fulvescens* ($n = 21$) and *A. oxyrhynchus* ($n = 19$) were probably sufficiently large to detect even a low level of heteroplasmy in those species. In addition, the possiblity that heteroplasmy occurs in only certain populations of these species is unlikely since *A. fulvescens* individuals were obtained from two separate drainage systems. No mtDNA length variation or heteroplasmy was reported in *A. oxyrhnychus* collected from populations in the southern United States (Bowen and Avise 1990).

Sequence analysis strongly suggests that length variation and heteroplasmy in *A. medirostris* and *A. transmontanus* is the result of a varying number of tandemly repeated sequences. The structural arrangement of these repeated sequences is similar between the two species but the actual nucleotide composition of the repeat unit is specific to each species. The specific nucleotide sequence of the repeated region appears to be perfectly conserved among individuals of both *A. transmontanus* ($n = 5$) and *A. medirostris* ($n = 3$). The lack of any similarly repeated sequence may explain the uniform size of *A. oxyrhynchus* mtDNA and the fixation of length variation in the absence of heteroplasmy in *A. fulvescens* mtDNA.

A model based on the D-loop sequence in *A. transmontanus* suggested that frequent misalignment of the repeat region prior to elongation, facilitated by stable secondary structures, generated the observed mtDNA length variation and heteroplasmy (Buroker et al. 1990). For *A. medirostris*, it would desirable to obtain further sequence data an individual with multiple copies of the repeat. However, preliminary analysis suggests that a 78 bp repeat sequence in *A. medirostris* is capable of forming a thermodynamically stable (-10.5 kcal/mole) complex secondary structure comparable to that proposed for an 82 bp repeat (-14.1 kcal/mole) in *A. transmontanus* (Buroker et al. 1990).

Another prediction of the elongation model was that the central repeats should evolve in a concerted manner. The nucleotide sequence of the repeats may be different between distantly related genotypes (or species) but within a genotype, the sequence would be perfectly conserved. Such is the case for repeated regions of *A. medirostris* and *A. transmontanus*. There is considerable interspecific variation between repeats but complete nucleotide conservation within each species.

The widespread occurrence of length variation and heteroplasmy throughout a genus has been observed before in Cnemidophorus lizards (Densmore et al. 1985). In restriction enzyme analysis, species of lizards were highly variable for different size mtDNA but showed little sequence divergence, unlike sturgeon. With the rapid mutation rate for generating size variation in mtDNA (Densmore et al. 1985: Rand and Harrison 1989), is length variation and heteroplasmy in *Acipenser* the result of convergent or divergent evolution?

The evolutionary pattern of mtDNA size variation and heteroplasmy on the phylogenetic tree of *Acipenser* provides more support to the existence of a heteroplasmic ancestor. The existence of fixed but different mtDNA size variants in *A. fulvescens* individuals would require some heteroplasmic ancestor. The degree of sequence degenerency in "repeats" found *A. fulvescens* and *A. oxyrhynchus* relative

those in *A. transmontanus* suggests that short sequences cannot self-proliferate in the D-loop regions of the two eastern species.

In the absence of recurring mutation for length variation, selection and/or genetic drift may gradually reduce the degree of heterogeneity for mtDNA size. In the previous chapter, it was suggested that there might be selection for smaller mtDNA in *A. transmontanus* populations, especially if the effective population size is small. The sturgeon are an ancient vertebrate group and have probably witnessed many of the tectonic upheavals of the North Amercian landmass. These geographical changes have probably caused sturgeon populations to go through several periods of expansion and reduction which could have been associated with genetic bottlenecks for mtDNA size variation.

Conclusion

Restriction enzyme and DNA sequence analysis revealed that white sturgeon populations generally have low geographical subdivision for mtDNA genotypes. However, genotypic frequencies are significantly different between river systems which seems paradoxical. Shifts in mtDNA genotypic frequencies are likely to occur more rapidly than mutations leading to new mtDNA genotypes. The observed differences in mtDNA genotypic frequencies could be the result of recent genetic bottlenecks as a consequence of the Pleistocene glaciations and human intervention. The accumulation of DNA sequence variation is more gradual, particularly in sturgeon which have a long generation time.

The intraspecific phylogeny suggests that mtDNA gene flow is high while geographical differences in genotypic frequencies indicates that the actual number of migrant individuals is only a small proportion of the total population. Population theory indicates that relative few migrants per generation could result in low geographical subdivision for organelle genotypes (Birky et al. 1983, 1989; Slatkin 1987). If the number of females in a population is N and the average rate of immigration is m then genetic drift will not result in substantial local differentiation of mtDNA genotypes if Nm is greater than one (Slatkin 1987). The effective populations of female sturgeon in the Fraser and Columbia Rivers are probably small (Chapter II). However, there is the potential for a single female to be highly migratory and to reproduce in different rivers over her life span. As an example, the 82 year old female sturgeon found in the Fraser River could have contributed to the gene pool of the upper Columbia River population several years ago. Thus, in white sturgeon populations, N is probably the smaller and m the larger component of Nm.

Restriction enzymes were shown to be less sensitive in detecting genetic variation than direct DNA sequence analysis (Chapter II). While direct sequencing

of DNA has become less complex, the trade-off remains between the number of individuals examined and the number of nucleotide positions examined in each individual. It can be concluded for population studies that restriction enzyme methods are best suited for estimating genetic diversity (provided the number of individuals examined is large) while comparisons of control region sequences produce statistically better intraspecific phylogenies (provided a large number of nucleotides are examined).

Hypervariable sequences of the D-loop are the most useful region of the mtDNA molecule for intraspecific comparisons. In rainbow trout (*Oncorhynchus mykiss*), only 13 variable sites were found in a 2214 bp fragment of mtDNA encoding four proteins and two tRNAs (Beckenbach et al. 1990). Low intraspecific variation was found in the cytochrome *b* genes of Australian songbird species (Edwards and Wilson 1990). Intraspecific changes in the same gene in Atlantic cod (*Gadus morhua*; Carr and Marshall 1991) and tuna (*Thunnus* species; Bartlett and Davidson 1991) were also too few for phylogenetic analysis.

The diversity of life histories and dispersal patterns in marine and freshwater species provides a wealth of opportunities to study microevolutionary processes using molecular markers (Quinn and Brodeur 1991). Restriction enzyme analyses of mtDNA have added to knowledge of migration patterns in humpback whales (*Megaptera novaeangliae*; Baker et al. 1990) and green turtles (*Chelonia mydas*; Meylan et al. 1990). High resolution intraspecific analysis afforded by control region sequence comparisons can now be applied to a wide range of species for which little is known about their life histories.

Hypervariable sequences may also reveal new population structuring in species, such as skipjack tuna (*Katsuwonus pelamis*; Graves et al. 1984) and albacore (*Thunnus alalunga*; Graves and Dizon 1989), which had little demographic variation in restriction enzyme surveys. The analysis of hypervariable sequences in the

mtDNA control region should be especially revelant in future studies of microevolutionary processes because of its high level of genetic variation and maternal line of descent.

The rapid saturation of the D-loop region for nucleotide substitutions limits its utility in examining the relationships between higher orders of taxa. In Chapter IV, interspecies alignments between certain segments of the D-loop were ambiguous. However, the sequence derived phylogeny did support the restriction enzyme analysis and revealed a significant geographical pattern of mtDNA variation in the four *Acipenser* species. The inclusion of other species of the order Acipenseriformes would be a logical extension of the phylogenetic analysis. Sequences of conserved mtDNA protein coding genes are probably more suitable for interspecific comparisons. Data on different gene sequences would also provide additional estimates of nucleotide substitution rates in sturgeon mtDNA. Finally, comparisons of mtDNA sequences from other species across the Continental Divide would be useful in the study of speciation.

Chapters III and IV examined the nature of mtDNA size variation within and between sturgeon species. Length variation appears to arise through rapid mutation which is balanced by genetic drift, and possibly selection. The rapid mutation rate and apparent non-neutrality of this polymorphism makes mtDNA length variation unsuitable as a genetic marker in population analysis. Harrison et al. (1987) reached a similar conclusion regarding mtDNA length variants in crickets. However, mtDNA size variation provides an interesting model system to study genome evolution, particularly with regards to the role of microevolutionary processes.

Population processes, suggested as possible influences in the distribution of mtDNA size variation in *A. transmontanus* populations (Chapter II), may have caused some of the interspecific differences in heteroplasmy (Chapter IV). Additional nucleotide sequence is required of large mtDNA size variants found in *A. medirostris*

and *A. fulvescens*. The developmental basis of mtDNA length variation requires investigation either by examining segregation of germ cells in embryos or sampling oocytes and progeny of a heteroplasmic female for size variation.

In this study, mtDNA analysis was shown to be useful in determining the structure of *A. transmontanus* populations and the systematic relationships of the genus *Acipenser*. In addition, intra- and interspecific mtDNA comparisons of these species have provided broader insights into the evolution of vertebrate mtDNA.

References

Aquadro, C.F. and B.D. Greenberg. 1983. Human mitochondrial DNA variation and evolution: Analysis of nucleotide sequences from seven individuals. Genetics 103:287-312.

Ashley, M. and C. Wills. 1987. Analysis of mitochondrial DNA polymorphisms among Channel Island deer mice. Evolution 41:854-863.

Avise, J.C. 1989. Gene trees and organismal histories: A phylogenetic approach to population biology. Evolution 43:1192-1208.

Avise, J.C., J. Arnold, R.M. Ball, E. Bermingham, T. Lamb, J.E. Neigel, C.A. Reeb and N.C. Saunders. 1987. Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. Ann. Rev. Ecol. Syst. 18:489-522.

Avise, J.C., R.M. Ball and J. Arnold. 1988. Current versus historical population sizes in vertebrate species with high gene flow: A comparison based on mitochondrial DNA lineages and inbreeding theory for neutral mutations. Mol. Biol. Evol. 5:331-344.

Avise, J.C., G.S. Helfman, N.C. Saunders and L.S. Hales. 1986. Mitochondrial DNA differentiation in North Atlantic eels: Population genetic consequences of an unusual life history pattern. Proc. Nat. Acad. Sci. USA 83:4350-4354.

Avise, J.C., C.A. Reeb and N.C. Saunders. 1987. Geographic population structure and species differences in mitochondrial DNA of mouthbrooding marine catfishes (Ariidae) and demersal spawning toadfish (Batrachoididae). Evolution 41:991-1002.

Ausubel, F.M., R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith and K. Struhl. 1989. Current protocols in molecular biology. John Wiley and Sons. Toronto.

Baker, C.S., S.R. Palumbi, R.H. Lambertsen, M.T. Weinrich, J. Calambokidis and S.J. O'Brien. 1990. Influence of seasonal migration on geographic distribution of mitochondrial DNA halotypes in humpback whales. Nature 344:238-240.

Ball, R.M., Jr., S. Freeman, F.C. James, F. Bermingham and J.C. Avise. 1988. Phylogeographic population structure of red-wing blackbirds assessed by mitochondrial DNA. Proc Nat. Acad. Sci. USA 85:1558-1562.

Bartley, D.M. 1987. The genetic structure of chinook and coho salmon populations in California, with a note on the genetic variability in sturgeon (Acipenseridae). PhD. dissertation. University of California, Davis.

Bartley, D.M., G.A.E. Gall and B. Bentley. 1985. Preliminary description of the genetic structure of white sturgeon, *Acipenser transmontanus*, in the Pacific Northwest. In: F.P. Binkowski and S.I. Doroshov (eds.). 1985. North American sturgeons: Biology and aquaculture potential. Dr. W. Junk, Dordrecht, The Netherlands.

Bartlett, S.E. and W.S. Davidson. 1991. Identification of *Thunnus* tuna species by the polymerase chain reaction and direct sequence analysis of their mitochondrial cytochrome *b* genes. Can. J. Fish. Aquat. Sci. 48:309-317.

Beckenbach, A.T., W.K. Thomas and H. Sohrabi. 1990. Intraspecific sequence variation in the mitochondrial genome of rainbow trout (*Oncorhynchus mykiss*). Genome 33:13-15.

Bentzen, P., G.G. Brown and W.C. Leggett. 1989. Mitochondrial DNA polymorphism, population structure, and life history variation in American shad (*Alosa sapidissima*). Can. J. Fish. Aquat. Sci. 46:1446-1454.

Bentzen, P., W.C. Leggett and G.G. Brown. 1988. Length and restriction site heteroplasmy in the mitochondrial DNA of the American shad (*Alosa sapidissima*). Genetics 118:509-518.

Bermingham, E. and J.C. Avise. 1986. Molecular zoogeography of freshwater fishes in the southeastern United States. Genetics 113:939-965.

Bermingham, E., T. Lamb and J.C. Avise. 1986. Size polymorphism and heteroplasmy in the mitochondrial DNA of lower vertebrates. J. Hered. 77: 249-252.

Bernatchez, L., J.J. Dodson and S. Boivin. 1989. Population bottlenecks: Influence on mitochondrial DNA diversity and its effect in coregonine stock discrimination. J. Fish. Biol. 233-244.

Binkowski, F.P. and S.I. Doroshov (ed.). 1985. North American sturgeons: Biology and aquaculture potential. Dr. W. Junk, Dordrecht, The Netherlands.

Birky, C.W., P. Fuerst and T. Maruyama. 1989. Organelle gene diversity under migration, mutation, and drift: Equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. Genetics 121:613-627.

Birky, C.W., T. Maruyama and P. Fuerst. 1983. An approach to population and evolutionary genetic theory in mitochondria and chloroplasts, and some results. Genetics 103:513-527.

Boursot, P., H. Yonekawa and F. Bonhomme. 1987. Heteroplasmy in mice with deletion of a large coding region of mitochondrial DNA. Mol. Biol. Evol. 4:46-55.

Bowen, B.W. and J.C. Avise. 1990. Genetic structure of Atlantic and Gulf of Mexico populations of sea bass, menhaden, and sturgeon: influence of zoographical factors and life-history patterns. Marine Biology 107:371-381.

Boyce, T.M., M.E. Zwick and C.F. Aquadro. 1989. Mitochondrial DNA in the bark weevils: Size, structure and heteroplasmy. Genetics 123:825-836.

Brannon, E., A. Setter, M. Miller, S. Brewer, G. Winans, F. Utter, L. Carperter and W. Hershberger. 1986. Columbia River white sturgeon, *Acipenser transmontanus*, population genetics and early life history study. Bonneville Power Administration, Division of Fish and Wildlife. Portland, Ore.

Brown, J.R., T.L. Gilbert, D.J. Kowbel, P.J. O'Hara, N.E. Buroker, A.T. Beckenbach and M.J. Smith. 1989. Nucleotide sequence of the apocytochrome *b* gene in white sturgeon mitochondrial DNA. Nucleic Acids Res. 17:4389.

Brown, W.M. 1983. Evolution of animal mitochondrial DNA. In M. Nei and R.K. Koehn (eds.). Evolution of Genes and Proteins. Sinauer. Sunderland, MA.

Brown, W.M. 1985. The mitochondrial genome of animals. In R.J. MacIntyre (ed.). Molecular Evolutionary Genetics. Plenum Press. NewYork.

Brown, G.G. and L.J. DesRosiers. 1983. Rat mitochondrial DNA polymorphisms: Sequence analysis of a hypervariable site for insertions/deletions. Nucleic Acids Res. 11:6699-6708.

Brown, W.M., M. George, Jr. and A.C. Wilson. 1979 Rapid evolution of animal mitochondrial DNA. Proc. Natl. Acad. Sci. USA 76: 1967-1971.

Buroker, N.E., J.R. Brown, T.A. Gilbert, P.J. O'Hara, A.T. Beckenbach, W.K. Thomas and M.J.Smith. 1990. Length heteroplasmy of sturgeon mitochondrial DNA: An illegitimate elongation model. Genetics 124:157-163.

Cabot, E.L. and A.T. Beckenbach. 1989. Simultaneous editing of multiple nucleic acid and protein sequences with ESEE. Computer Applications in the Biosciences. 5:233-234.

Cann, R.L., M. Stoneking and A.C. Wilson. 1987. Mitochondrial DNA and human evolution. Nature 325:31-36.

Carlson, D.M., M.K. Kettler, S.E. Fisher and G.S. Whitt. 1982. Low genetic variability in paddlefish populations. Copeia 1982:721- 725.

Carr, S.M. and H.D. Marshall. 1991. Detection of intraspecific DNA sequence variation in the mitochondrial cytochrome *b* gene of Atlantic cod (*Gadus morhua*) by the polymerase chain reaction. Can. J. Fish. Aquat. Sci. 48:48-52.

Cavender, T.M. 1986. Review of the fossil history of North American freshwater fishes. In C.H. Hocutt and E.O. Wiley (eds.) The Zoogeography of North American Freshwater Fishes. John Wiley and Sons. New York.

Clague, J.J. and J.L. Luternauer. 1983. Late quaternary geology of southwestern British Columbia. Field Trip 6. Geological Association of Canada, Victoria Section.

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

Clayton, D.A., 1984. Transcription of the mammalian mitochondrial genome. Ann. Rev. Biochem. 53: 573-594.

Crossman, E.J. and D.E. McAllister. 1986. Zoogeography of freshwater fishes of the Hudson Bay Drainage, Ungava Bay and the Arctic Archipelago. In C.H. Hocutt and E.O. Wiley (eds.) The Zoogeography of North American Freshwater Fishes. John Wiley and Sons. New York.

Conte, F.S., S.I. Doroshov, P.B. Lutes and E.M. Strange. 1988. Hatchery manual for the white sturgeon, *Acipenser transmontanus* Richardson, with application to other North American Acipenseridae. Division of Agriculture and Natural Resources. University of California. Oakland CA

Dawid, I.B. and A.W. Blackler. 1972. Maternal and cytoplasmic inheritance of mitochondrial DNA in Xenopus. Develop. Biol. 29:152-161.

De Giorgi. C. and C. Saccone. 1989. Mitochondrial Genome in Animal Cells: Structure, organization and evolution. Cell Biophysics 14:67-78.

Densmore, L.D., J.W. Wright and W.M. Brown. 1985. Length variation and heteroplasmy in mitochondrial DNA from parthenogenetic and bisexual lizards (Genus Cnemidophorus). Genetics 110: 689-707.

Desjardins, P. and R. Morais. 1991. Nucleotide sequence and evolution of coding and noncoding regions of a quail mitochondrial genome. J. Mol. Evol. 32:153-161.

Dixon, B.N. 1986. Age, growth and migration of white sturgeon in the Nechako and upper Fraser Rivers of British Columbia. British Columbia Ministry of Environment. Fisheries Technical Circ. 70.

Dixon, W.J. 1983. BMDP Statistical Software. University of California Press. Berkeley CA.

Doda, J.N., C.T. Wright and D.A. Clayton. 1981. Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. Proc. Natl. Acad. Sci. USA 78: 6116-6120.

Doolittle, W.F. and C. Sapienza. 1980. Selfish genes, the phenotype paradigm and genome evolution. Nature 284:601-603.

Doroshov, S.I. 1985. Biology and culture of sturgeon Acipenseriformes. In J.F. Muir and R.J. Roberts (eds.) Recent Advances in Aquaculture. Croom Helm. London.

Dunon-Bluteau, D., M. Volovitch and G. Brun. 1985. Nucleotide sequence of a *Xenopus laevis* mitochondrial DNA fragment containing the D-loop, flanking tRNA genes and apocytochrome *b* gene. Gene 36:65-78.

Edwards, S.V. and A.C. Wilson. 1990. Phylogenetically informative length polymorphism and sequence variability in mitochondrial DNA of Australian songbirds (Pomatostomus). Genetics 126:695-711.

Efstratiadis, A., J.W. Posakony, T. Maniatis, R.W. Lawn, C. O'Connell, R.A. Spritz, J.K. Deriel, B.G. Forget, S.M. Weissman, J.L. Slightom, A.E. Blechi, O. Smithies, F.E. Baralle, C.C. Shoulders and N.J. Proudfoot. 1980. The structure and evolution of the human -globin gene family. Cell 21:653-668.

Fauron, C.M.-R. and D.R. Wolstenholme. 1976. Structural heterogeneity of mitochondrial DNA molecules within the genus *Drosophila*. Proc. Natl. Acad. Sci. USA 73: 3623-3627.

Felsenstein, J. 1990. Phylogeny Inference Package (PHYLIP) 3.3. University of Washington. Seattle WA.

Felsenstein, J. 1988. Phylogenies from molecular sequences: Inference and reliability. Annu. Rev. Genet. 22:521-565.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the boostrap. Evolution 39:783-791.

Galbreath, J.L. 1985. Status, life history and management of Columbia River white sturgeon, *Acipenser transmontanus*. In F.P. Binkowski and S.I. Doroshov (eds.). North American sturgeons: Biology and aquaculture potential. Dr. W. Junk, Dordrecht, The Netherlands.

Gardiner, B.G. 1984. Sturgeon as living fossils. In N. Eldredge and S.M. Stanley (eds.). Living Fossils. Springer Verlag. New York.

Gilbert, T.L., J.R. Brown, P.J. O'Hara, N.E. Buroker, A.T. Beckenbach and M.J. Smith. 1988. Sequence of $tRNA^{Thr}$ and $tRNA^{Pro}$ from white sturgeon (*Acipenser transmontanus*) mitochondria. Nucleic Acids Res. 16:11825.

Gojobori, T., W.-H. Li and D. Graur. 1982. Patterns of nucleotide substitution in pseudogenes and functional genes. J. Mol. Evol. 18:360-369.

Graves, J.E. and A.E. Dizon. 1989. Mitochondrial DNA sequence similarity of Atlantic and Pacific albacore tuna (*Thunnus alalunga*). Can. J. Fish. Aquat. Sci. 46:870-873.

Graves, J.E., S.D. Ferris and A.E. Dizon. 1984. Close genetic similarity of Atlantic and Pacific skipjack tuna (*Katsuwonus pelamis*) demonstrated with restriction endonuclease analysis of mitochondrial DNA. Marine Biology 79:315-319.

Gyllensten, U. and A.C. Wilson. 1987. Mitochondrial DNA of salmonids: Inter- and intraspecific variability detected with restriction enzymes. In N. Ryman and F. Utter (eds.). Population genetics and fishery management. University of Washington Press. Seattle.

Hale, L.R. and R.S. Singh. 1986. Extensive variation and heteroplasmy in size of mitochondrial DNA among geographical populations of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA 83: 8813-8817.

Harrison, R.G. 1989. Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. Trends in Ecology and Evolution. 4:6-11.

Harrison, R.G., D.M. Rand and W.C. Wheeler. 1987. Mitochondrial DNA size variation in field crickets across a narrow hybrid zone. Mol. Biol. Evol. 4:144-158.

Harrison, R.G., D.M. Rand and W.C. Wheeler. 1985. Mitochondrial DNA size variation within individual crickets. Science 228: 1446-1448.

Hauswirth, W.W. and P.J. Laipis. 1982. Mitochondrial DNA polymorphism in a maternal lineage of Holstein cows. Proc. Natl. Acad. Sci. USA 79: 4686-4690.

Hauswirth, W.W., M.J. Van de Walle, P.J. Laipis and P.D. Olivo. 1984 Heterogeneous mitochondrial DNA D-loop sequences in bovine tissue. Cell 37:1001-1007.

Haynes, J.M. and R.H. Gray. 1981. Diel and seasonal movements of white sturgeon, *Acipenser transmontanus*, in the mid-Columbia River. Fishery Bulletin 79:367-370.

Haynes, J.M., R.H. Gray and J.C. Montgomery. 1978. Seasonal movements of white sturgeon (*Acipenser transmontanus*) in the mid-Columbia River. Trans. Am. Fish. Soc. 107:275-280.

Hayshi, J.-I., Y. Tagashira and M.C. Yoshida. 1985. Absence of extensive recombination between inter- and intraspecies mitochondrial DNA in mammalian cells. Expt. Cell Res. 160: 387-395.

Hebert, P.D.N., R.D. Ward and L.J. Weider. 1988. Clonal- diversity patterns and breeding-system variation in *Daphnia pulex*, an asexual-sexual complex. Evolution 42:147-159.

Hillis, D.M. and C. Moritz. 1990. Molecular systematics. Sinauer Associates, Inc. Sunderland MA.

Hindar, K., B. Jonsson, N. Ryman and G. Stahl. 1991. Genetic relationships among landlocked, resident and anadromous brown trout, *Salmo trutta* L. Heredity 66:83-91.

Hoeh, W.R., K.H. Blakley and W.M. Brown. 1991. Heteroplasmy suggests paternal leakage of *Mytilus* mitochondrial DNA. Science 251:1488-1490.

Horai, S. and K. Hayasaka. 1990. Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. Am. J. Hum. Genet. 46:828-842.

Irwin, D.M., T.D. Kocher and A.C. Wilson. 1991. Evolution of the cytochrome *b* gene of mammals. J. Mol. Evol. 32:128-144.

Jukes, T.H. and C.R. Cantor. 1969. Evolution of protein molecules. In H.N. Munro (ed.). Mammalian protein metabolism. Academic Press. New York.

Johns, D.R., S.L. Rutledge, O.C. Stine and O. Hurko. 1989. Directly repeated sequences associated with pathogenic mitochondrial DNA deletions. Proc. Natl. Acad. Sci. USA 86:8559-8062.

Kimura, M. 1968. Evolutionary rate at the molecular level. Nature 217:624-626.

Kimura, M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16:111-120.

Kishino, H. and M. Hasegawa. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data and the branching order in Hominoidea. J. of Mol. Evol. 29:170-179.

Kocher, T.D., W.K. Thomas, A. Meyer, S.V. Edwards, S. Paabo, F.X. Villablanca and A.C. Wilson. 1989. Dynamics of mitochondrial DNA evolution in animals: Amplification and sequencing with conserved primers. Proc. Natl. Acad. Sci. USA 86:6196-6220.

Kondo, R., Y. Sotta, E.T. Matsuura, H. Ishiwa, N. Takahata and S.I. Chigusa. 1990. Incomplete maternal transmission of mitochondrial DNA in Drosophila. Genetics 126:657-663.

Lake, J.A. 1987. A rate-independent technique for analysis of nucleic acid sequences: operator metrics. Mol. Biol. Evol. 4:167-191.

Lamb, T., J.C. Avise and J. W. Gibbons. 1989. Phylogeographic patterns in mitochondrial DNA of the desert tortoise (*Xerobates agassizi*) and evolutionary relationships among the North American gopher tortoises. Evolution 43:76-87.

Lansman, R.A., R.O. Shade, J.F. Shapira and J.C. Avise. 1981. The use of restriction endonucleases to measure mitochondrial DNA sequence relatedness in natural populations: III. Techniques and potential applications. J. Mol. Evol. 17:214-226.

Laskey, R.A. 1980. The use of intensifying screens or organic scintillators for visualizing radioactive molecules resolved by gel electrophoresis. Meth. Enzymol. 65:363-371.

Lauder, G.V. and K.F. Liem. 1983. The evolution and interrelationships of the Actinopterygian fishes. Bull. Mus. Comp. Zool. 150:95-197.

Li, W.-H. and D. Graur. 1991. Fundamentals of molecular evolution. Sinauer Associates, Inc. Sunderland MA.

Lynch, M. and T.J. Crease. 1990. The analysis of population survey data on DNA sequence variation. Mol. Biol. Evol. 7:377-394.

Martin, A.P., B.D. Kessing and S.R. Palumbi. 1990. Accuracy of estimating genetic distance between species from short sequences of mitochondrial DNA. Mol. Biol. Evol. 7:485-488.

Matsuura, E.T., S.I. Chigusa and Y. Niki. 1989. Induction of mitochondrial DNA heteroplasmy by intra- and interspecific transplantation of germ plasm in Drosophila. Genetics 122:663-667.

Meyer, A., T.D. Kocher, P. Basasibwaki and A.C. Wilson. 1990. Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. Nature 347:550-553.

Meyer, A. and A.C. Wilson. 1990. Origin of tetrapods inferred from their mitochondrial DNA affiliation to lungfish. J. Mol. Evol. 31:359-364.

Meylan, A.B., B.W. Bowen and J.C. Avise. 1990. A genetic test of the natal homing versus social facilitation models for green turtle migration. Science 248:724-727.

McPhail, J.D. and C.C. Lindsay. 1986. Zoogeography of the freshwater fishes of Cascadia (The Columbia system and rivers north of the Stikine). In C.H. Hocutt and E.O. Wiley (eds.) The Zoogeography of North American Freshwater Fishes. John Wiley and Sons. New York.

Miller, L.W. 1972a. White sturgeon population characteristics in the Sacramento-San Joaquin estuary as measured by tagging. Calif. Fish and Game 58:94-101.

Miller, L.W. 1972b. Migrations of sturgeon tagged in the Sacramento-San Joaquin estuary. Calif. Fish and Game 58:102- 106.

Monnerot, M., J.-C. Mounolou and M. Solignac. 1984. Intra-individual length heterogeneity of *Rana esculenta* mitochondrial DNA. Biol. Cell 52: 213-218.

Moritz, C. and W.M. Brown. 1987. Tandem duplications in animal mitochondrial DNAs: Variation in incidence and gene content among lizards. Proc. Natl. Acad. Sci. USA 84:7183-7187.

Mortiz, C., T.E. Dowling and W.M. Brown. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Ann. Rev. Ecol. Syst. 18: 269-292.

Mulligan, T.J. and R.W. Chapman. 1989. Mitochondrial DNA analysis of Chesapeake Bay white perch, *Morone americana*. Copeia 1989:679-688.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. USA 70:3321-3323.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press. New York.

Nei, M. and W.-H. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA 76:5269-5273.

Nei, M., J.C. Stephens and N. Saitou. 1985. Methods for computing the standard errors for branching points in an evolutionary tree and their application to molecular data from human and apes. Mol. Biol. Evol. 2:66-85.

Nei, M. and F. Tajima. 1981. DNA polymorphisms detectable by restriction endonucleases. Genetics 97:145-163.

Nei, M. and F. Tajima. 1983. Maximum likelihood estimation of the number of nucleotide substitutions from restriction sites data. Genetics 105:207-217.

Nelson, K. and M. Soule. 1987. Genetical conservation of exploited fishes. In N. Ryman and F. Utter (eds.). Population genetics and fishery management. University of Washington Press. Seattle.

Orgel, L.E. and F.H.C. Crick. 1980. Selfish DNA: The ultimate parasite. Nature 284:604-607.

Ovenden, J.R. and R.W.G. White. 1990. Mitochondrial and allozyme genetics of incipient speciation in a landlocked population of *Galaxias truttaceus* (Pisces: Galaxiidae). Genetics 124:701-716.

Palumbi, S.R. and A.C. Wilson. 1990. Mitochondrian DNA diversity in the sea urchin *Strongylocentrotus purpuratus* and *S. droebachiensis*. Evolution 44:403-415.

Phelps, S.R. and F.W. Allendorf. 1983. Genetic identity of pallid and shovelnose sturgeon (*Scaphirhynchus albus* and *S. platorynchus*). Copeia 1983:696-700.

Quin, T.P. and R.D. Brodeur. 1991. Intra-specific variations in the movement patterns of marine animals. Amer. Zool. 31:231-241.

Rand, D.M. and R.G. Harrison. 1989. Molecular population genetics of mtDNA size variation in crickets. Genetics 121: 551-569.

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

Robison, H.W. 1986. Zoogeographic implications of the Mississippi River basin. In C.H. Hocutt and E.O. Wiley (eds.) The Zoogeography of North American Freshwater Fishes. John Wiley and Sons. New York.

Roussow, G. 1957. Some considerations concerning sturgeon spawning periodicity. J. Fish. Res. Bd. Canada 14:553-572.

Ruvolo, M., T.R. Disotell, M.W. Allard, W.M. Brown and R.L. Honeycutt. 1991. Resolution of the African hominoid trichotomy by use of a mitochondrial gene sequence. Proc. Natl. Acad. Sci. USA 88:1570-1574.

Saccone, C., M. Attimonelli and E. Sbisa. 1987. Structural elments highly preserved during the evolution of the D-loop containing region in vertebrate mitochondrial DNA. J. Mol. Evol. 26:205-211.

Saiki, R.K., D.H. Gelfand, S.Stoffel, S.J. Scharf, R. Higuchi, G.T. Horn, K.B. Mullis and H.A. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487-491.

Saitou, N. and M. Nei. 1987. The neighbor-joining method: A new method for constructing phylogenetic trees. Mol. Biol. Evol. 4:406-425.

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. Molecular cloning: A laboratory manual. 2nd edition. Cold Spring Harbor Press. Cold Spring Harbor NY.

Sanger, F., S. Nicklen and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74: 5463-5467.

Schields, G.F. and A.C. Wilson. 1987. Calibration of mitochondrial DNA evolution in geese. J. Mol. Evol. 24:212-217.

Schon, E.A., R. Rizzuto, C.T. Moraes, H. Nakase, M. Zeviani and S. DiMauro. 1989. A direct repeat is a hotspot for large-scale deletion of human mitochondrial DNA. Science 244:346-349.

Scott, W.B. and E.J. Crossman. 1973. Freshwater fishes of Canada. Fish. Res. Bd. of Canada Bull. 184.

Semakula, S.N. and P.A. Larkin. 1968. Age, growth, food and yield of the white sturgeon (*Acipenser transmontanus*) of the Fraser River, British Columbia. J. Fish. Res. Bd. Canada 25:2589- 2602.

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

Smith, G.R. 1981. Late cenozoic freshwater fishes of North America. Ann. Rev. Ecol. Syst. 12:163-193.

Smith, M.J., D.K. Banfield, K. Doteval, S. Gorski and D.J. Kowbel. 1989. Gene arrangement in sea star mitochondrial DNA demonstrates a major inversion event during echinoderm evolution. Gene 76:181-185.

Snyder, M., A.R. Fraser, J. LaRoche, K.E. Gartner-Kepkay and E. Zouros. 1987. Atypical mitochondrial DNA from the deep-sea scallop *Placopecten magellanicus*. Proc. Natl. Acad Sci. USA 84: 7595-7599.

Soliganc, M., J. Genermont, M. Mommerot and J.-C. Mounolou. 1987. Drosophila mitochondrial genetics: Evolution of heteroplasmy through germ line cell divisions. Genetics 117: 687-696.

Solignac, M., M. Monnerot and J.-C. Mounolou. 1983. Mitochondrial DNA heteroplasmy in *Drosophila mauritiana*. Proc. Natl. Acad. Sci. USA 80: 6942-6946.

Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503-517.

Swift, C.C., C.R. Gilbert, S.A. Bortone, G.H. Burgess and R.W. Yerger. 1986. Zoogeography of the freshwater fishes of the southeastern United States: Savannah River to Lake Pontchartrain. In C.H. Hocutt and E.O. Wiley (eds.) The Zoogeography of North American Freshwater Fishes. John Wiley and Sons. New York.

Swofford, D.L. and G.L. Olsen. 1990. Phylogeny reconstruction. In D. M. Hillis and C. Moritz (eds.). Molecular Systematics. Sinauer Associates, Inc. Sunderland MA.

Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics 105:437-460.

Thomas, W.K. and A.T. Beckenbach. 1989. Variation in salmonid mitochondrial DNA: Evolutionary constraints and mechanisms of substitution. J. Mol. Evol. 29:233-245.

Thomas, W.K., S. Paabo, F.X. Villablanca and A.C. Wilson. 1990. Spatial and temporal continuity of kangaroo rat populations shown by sequencing mitochondrial DNA from museum specimens. J. Mol. Evol. 31:101-112.

Tomkinson, A.E. and S. Linn. 1986. Purificiation and properties of a strand-specific endonuclease from mouse cell mitochondria. Nucleic Acids Res. 14:9579-9593.

Upholt, W.B. and I.B. Dawid. 1977. Mapping of mitochondrial DNA of individual sheep and goats: Rapid evolution in the D-loop region. Cell 11: 571-583.

Vigilant, L., R. Pennington, H. Harpending, T.D. Kocher and A.C. Wilson. 1989. Mitochondrial DNA sequences in single hairs froma southern African population. Proc. Natl. Acad. Sci. USA 86:9350-9354.

Wallis, G.P. 1987. Mitochondrial DNA insertion polymorphism and germ line heteroplasmy in the *Triturus cristatus* complex. Heredity 58: 229-238.

Wallis, G.P. and J.W. Arntzen. 1989. Mitochondrial DNA variation in the crested newt superspecies: Limited cytoplasmic flow among species. Evolution 43:88-104.

Ward, R.D., N. Billington and P.D.N. Hebert. 1989. Comparison of allozyme and mitochondrial DNA variation in populations of walleye, *Stizostedion vitreum.* Can. J. Fish. Aquat. 46:2074-2084.

Williams, J.E., J.E. Johnson, D.A. Hendrickson, S. Contreras- Balderas, J.D. Williams, M. Navarro-Mendoza, D.E. McAllister and J.E. Deacon. 1989. Fishes of North America endangered, threatened or of special concern: 1989. Fisheries. 14:2-20.

Wilson, A.C., R.L. Cann, S.M. Carr, M. George, U.B. Gyllensten, K.M. Helm-Bychowski, R.G. Higuchi, S.R. Palumbi, E.M. Prager, R.D. Sage and M. Stoneking. 1985. Mitochondrial DNA and two perspectives on evolutionary genetics. Biol. J. Linn. Soc. 26:375-400.

Wilson, G.M., W.K. Thomas and A.T. Beckenbach. 1987. Mitochondrial DNA analysis of Pacific Northwest populations of *Oncorhynchus tshawytscha.* Can. J. Fish. Aquat. Sci. 44:1301-1305.

Wu, C.-I. and W.-H. Li. 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. Proc. Natl. Acad. Sci. USA 82:1741-1745.

Zar, J.H. 1984. Biostatistical analysis. Prentice-Hall, Inc. Englewood Cliffs NJ.

Zuckerkandl, E. and L. Pauling. 1962. Molecular disease, evolution and genic heterogeneity. In M. Kash and B. Pullman (eds.). Evolving Genes and Proteins. Academic Press. New York.

Appendix A

DNA fragment lengths (in base pairs) for the restriction enzymes used in this study. Species are *Acipenser transmontanus* (A.t.), *A. medirostris* (A.m.), *A. fulvescens* (A.f) and *A. oxyrhynchus* (A.o.). Intra-specific polymorphic DNA fragment patterns are denoted by the letter code as Table 3. Variable length fragments which are known to span the origin of heavy strand replication are marked with " * ".

ApaI				*EcoRI*				*HindIII*			
A.t.	A.m.	A.f.	A.o.	A.t.	A.m.	A.f.	A.o.	A.t.	A.m.	A.f.	A.o.
7200	7200	5400	16500	8400	8600	8400	16800	12000	12000	12000	12000
5400	5400	4300		8200	7600	8200		3000	3000	2200	3000
4300	4300	4000						1100	1100	1100	1100
		1700								800	
		1300									

Appendix A: continued

AccI					*AvaII*						
A.t.-A	A.t.-B	A.m.	A.f.	A.o.	A.t.-A	A.t.-B	A.m.-A	A.m.-B	A.f.-A	A.f.-B	A.o.
*4600	*4600	*4600	6400	8200	6200	4000	*10000	*8300	*3800	*3400	*3300
3600	2800	2500	2800	2800	*2700	*2700	1600	1700	2200	2800	3100
1900	1900	1900	2700	2300	2100	2200	1100	1600	1800	1800	2500
1750	1750	1750	*1420	1800	1400	2100	1000	1100	1500	1500	1800
1550	1550	1550	800		1000	1400	760	1000	1400	1400	1100
1300	1300	1300	560		760	1000	620	760	1350	1350	1000
1100	1100	1100	500		620	760	560	620	1100	1100	620
560	800	1100			560	620	520	560	620	620	560
500	560	560			520	560	400	520	620	620	520
	500	500			400	520	340	400	560	560	400
					340	400		340	520	520	340
						340			400	400	
									340	340	

Appendix A: continued

*Bgl*II				*Pvu*II				*Xba*I			
A.t.	A.m.	A.f.	A.o.	A.t.	A.m.	A.f.	A.o.	A.t.	A.m.	A.f.	A.o.
16500	16500	16500		16500	10000	14000		11000	8200	10700	9000
					6400	2200		4700	4700	4700	4700
								1100	2800	1450	2800
									1100		

Appendix A: continued

HaeII				*BclI*				
A.t.	A.m.	A.f.	A.o.	A.t.-A	A.t.-B	A.m.	A.f.	A.o.
15000	15000	6800	8100	8000	8000	4800	5500	*6230
*1800	*1800	4600	4200	2500	2500	*4800	*5000	4800
		3600	2500	2300	2300	2500	2300	2500
		*1800	*1500	*1800	1370	2300	1500	2300
				750	*900	750	1000	
				680	820	680	750	

Appendix A: continued

AvaI						*HincII*					
A.t.-A	A.t.-B	A.t.-C	A.m.	A.f.	A.o.	A.t.-A	A.t.-B	A.m.	A.f.-A	A.f.-B	A.o.
6600	11000	11000	6600	5400	8400	3700	*5000	*5000	*5000	*5000	*4700
5400	5400	3400	5400	5400	4200	2700	3700	3700	4800	3400	4800
4600		2000	4600	4200	1700	2500	2700	2700	2700	2700	2700
				1500	1500	*2500	1400	1400	1900	1900	1900
						1400	1400	1100	940	1400	1400
						1400	1100	940	480	940	
						1100	580	580	450	480	
						580	450	480		450	
						450		450			