EVOLUTION OF MITOCHONDRIAL DNA IN THE GENUS *SALMO*

by

Milly Sin Yan So
B.Sc., University of British Columbia, 2001

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

In the Department of
Molecular Biology and Biochemistry

SIMON FRASER UNIVERSITY

Spring 2006

APPROVAL

Name: **Milly So**

Degree: **Master of Science**

Title of Thesis: **Evolution of mitochondrial DNA in the genus *Salmo***

Examining Committee:

Chair: **Dr. Christopher T. Beh**
Professor, Department of Molecular Biology and Biochemistry

_____

Dr. William S. Davidson
Senior Supervisor
Professor, Department of Molecular Biology and Biochemistry

_____

Dr. Barry M. Honda
Supervisor
Professor, Department of Molecular Biology and Biochemistry

_____

Dr. Nancy C. Hawkins
Supervisor
Assistant Professor, Department of Molecular Biology and Biochemistry

_____

Dr. Andrew T. Beckenbach
Internal Examiner
Professor, Department of Biological Sciences

Date Defended/Approved: Apr. 12/06

ABSTRACT

The mitochondrial genome (mtDNA) is a valuable source of data for evolutionary studies because of its small size, lack of recombination and its higher rate of accepted mutations than nuclear coding sequences. All salmonid mitochondrial genomes are ~16.7 Kb in size and identical in their genome organization. PCR amplification with thirty-three conserved primer pairs and subsequent direct sequencing was used to obtain whole mitochondrial genome sequences from fourteen Atlantic salmon (*Salmo salar*) samples. The nucleotide and amino acid sequences were aligned and compared with those of a sister species, brown trout (*Salmo trutta*) to study the mode and tempo of mtDNA evolution. Varying percent sequence divergence was observed in different parts of the genome suggesting that different constraints operate across the genome. Further, by measuring the amount of variation in the Atlantic salmon from different geographical locations, previous hypotheses regarding the structuring of Atlantic salmon populations were confirmed.

Keywords

Atlantic salmon, Brown trout, Complete mitochondrial genome, Salmonid mtDNA primers, Phylogeography.

DEDICATION

To mom, dad, Danny and Siemon.

ACKNOWLEDGEMENTS

I sincerely thank my supervisor Dr. William Davidson for giving me the opportunity to work in his lab. I am grateful for his continuous support and encouragement along each step of way. I also want to thank all the Davidson team members, Evelyn, Colin, Krzysztof, Siemon, Leslie, Amit, Carlo, Yanli, Christine and Jay who made the lab fun everyday. Thanks also to Jonathan McCarthy who helped me with primer designing and Evelyn Davidson for assisting me with sequencing. Many thanks to my family and friends, especially mom, dad, Siemon, thanks for believing in me all through this journey.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 - INTRODUCTION

1.1 Vertebrate mitochondrial genomes

The mitochondrion is the energy production center of all respiring eukaryotic cells. Its main function is to produce energy in the form of ATP for cellular functions (Cann *et al.*, 1984). Inside each mammalian cell, there are more than 1,000 mitochondria with a total weight of 0.017 pg, which is about 1% of the weight of the 1.7 pg nucleus. Each small mitochondrion contains many copies of circular, double-stranded DNA known as the mitochondrial genome. Mitochondrial DNA (mtDNA) is replicated, transcribed and its mRNAs are translated independently from the nuclear genome. In addition, there is no exchange of information between the mitochondrial genome and the nuclear genome. Unlike the nuclear genome, the mitochondrial genome is uniparentally descended through the maternal lineage (MacIntyre, 1985). The characteristic single parent inheritance allows geneticists to use data from mtDNA sequences to draw conclusions about evolution.

Mitochondrial genomes are much simpler than their corresponding nuclear genomes. Thus, the mitochondrial genome serves as an excellent model for studying evolution. It has several characteristics that the nuclear genome does not possess. First, the size of the vertebrate mitochondrial genome (less than 17,000 bp) is much smaller compared to the nuclear genome. Animal mitochondrial genomes are ~25,000 times smaller than the smallest nuclear genome (MacIntyre, 1985). The small genome shortens

the time required to collect data. Secondly, mtDNA has been shown to evolve 2-10 times faster than nuclear DNA (Brown *et al.*, 1979). Several reasons for the faster evolutionary rate include: (1) the absence of proof-reading ability of the mitochondrial DNA polymerase for mtDNA replication; (2) the more frequent self-multiplication of mtDNA increases the number of errors than occur during replication process; (3) the lack of recombination within the mitochondrial genome results in a lower tendency for genome rearrangements to occur (Stepien and Kocher, 1997). These advantageous features of the mitochondrial genome attracted many researchers to use this simpler genome as a model to study molecular evolution in animals. The recent availability of sequence data makes it possible to learn about the properties of the mitochondrial genome by comparing various mtDNAs. The study of mitochondrial genome evolution has impacted the field of anthropology, human pathology, population genetics and human evolution, and will continue to provide information about the evolution of various species.

The complete mitochondrial genome contains thirteen coding genes, twenty-two transfer RNAs (tRNAs), two ribosomal RNAs (rRNAs), and one non-coding region commonly known as the displacement loop (D-loop) in vertebrates. The mitochondrial genome is extremely compact in gene organization; >90% of mtDNA is made up of coding genes with no intergenic regions. The organization of the mitochondrial genome is believed to have been established very early in the vertebrate lineage (Lee and Kocher, 1995), and the vertebrate consensus order has been maintained throughout most of the vertebrate lineages. Molecular data such as gene sequences and gene order can be used to infer phylogenetic relationships within and between species, even between distantly

diverged lineages. However, most studies have used mtDNA to investigate the relationships of individuals and populations or closely related species groups.

1.2 History of techniques used to study mtDNA

A major technological breakthrough occurred when Type II restriction endonucleases were applied to mtDNA isolated from three mammalian cell lines (Brown and Vinograd, 1974). This process resulted in the first restriction maps of mtDNA, and paved the way for assessing relationships between species and populations that diverged relatively recently. The use of mtDNA for evolutionary studies really took off in the late 1970s (Brown *et al.*, 1979), mainly because of the characteristics mentioned above.

In early studies, restriction digestion mapping was the technique used to estimate the percentage of genome sequence differences between pairs of organisms (Brown *et al.*, 1979). The first step required the preparation of mtDNA and made use of differential centrifugation to isolate a sub-cellular fraction enriched for mitochondria and relatively free from nuclei, followed by the purification of the covalently-closed, circular form of mtDNA free of contaminating nuclear DNA by equilibrium sedimentation in caesium chloride density gradients in the presence of an intercalating dye such as propidium iodide or ethidium bromide. The second part involved the development of procedures that enhanced the separation and detection of fragments of DNA. These consisted of electrophoresis through agarose or polyacrylamide gels which resolved fragments, generated by restriction endonuclease digestions, over the range of 10,000 bp to 100 bp. The DNA fragments were visualized directly by staining with ethidium bromide when agarose gels were used. For polyacrylamide gels it was necessary to radioactively end-label the restriction fragments using the Klenow fragment of DNA polymerase and the

incorporation of α^{32}P dNTPs. Based on the fragment sizes, cleavage maps were created giving the positions of the endonuclease restriction sites relative to one another. Pairwise comparisons of cleavage maps of different samples (from different species) yielded a score denoting the fraction of sites shared by each pair. The percentage of unshared sites could be interpreted as the minimum number of base substitutions per site compared, which could then be converted to the estimated percentage sequence difference between the two samples. These techniques provided the tools and the impetus for many population biologists and systematists to switch from the traditional allozyme analysis to a DNA-based system for their studies, or at least to incorporate both where possible.

The procedures outlined above emphasized the use of purified mtDNA, which usually required a large amount of tissue. However, large tissue samples are not always available from samples of natural populations commonly targeted for population biology and evolutionary studies. Therefore, novel techniques were developed by Lansman *et al.* (1981), which combined the use of restriction endonuclease digestion and DNA-DNA hybridization to increase the sensitivity of visualizing digestion patterns on agarose gels, and by so doing it overcame the need for a large amount of pure mtDNA. The DNA-DNA hybridization technique was beneficial in its ability to detect a minimal amount of DNA from crude tissue samples. In addition, a less pure mtDNA sample could be substituted as the starting material as long as the restriction enzyme was able to cleave the DNA. Cytoplasmic nucleic acids (CNA), an enriched mtDNA-containing tissue fraction substituted the need for pure mtDNA. CNA was isolated quickly, subsequently digested, and the resulting fragments were electrophoresed on an agarose gel, before being subjected to Southern blot analysis. The filters were incubated with radioactively-

labelled pure mtDNA probe sequences and following autoradiography, revealed the mtDNA digestion profile with the pattern of hybridized restriction fragments. The intensity of the bands no longer depended on the amount of fragmented DNA, but rather on the specific binding of the hybridized probe, thereby alleviating the requirement of large amount of tissue sample for extraction and restriction analysis. The mtDNA digestion profile is the gel phenotype produced by one restriction endonuclease. Each complete digestion of one CNA sample would yield a pattern of bands that has a sum of ~17 Kb. The proportion of fragments shared in their mtDNA digestion profiles was rated as an index of relative genetic similarity between pairs of organisms, and further used to estimate the number of base substitutions per nucleotide. Several digestion profiles can be combined to create a composite profile created from multiple digestion profiles. The organisms that share most similarity in composite profile are most likely to have arisen from a common ancestor. The geographical distribution and movement of these organisms can be deduced according to their shared patterns. A most parsimonious phylogenetic network created by interconnecting composite phenotypes provides a representation of the estimated matriarchal phylogeny (Lansman *et al.*, 1981).

The time and expense required to conduct restriction analysis on mtDNA meant that the number of samples examined was often orders of magnitude lower than had been routinely possible using allozymes. Moreover, the indirect assessment of DNA sequence variation obtained by restriction analysis had drawbacks. The preparation and alignment of restriction maps are tedious and results from different labs were often difficult to compare because of variation in the gel separation techniques and the suite of restriction enzymes used. It was certainly possible to overcome these problems by cloning specific

fragments of mtDNA and determining their nucleotide sequences so that intra- and inter-specific variation could be measured, but this was prohibitively expensive for large-scale surveys. The end of the 1980s set the scene for the next major technological advance. The procedure involved the direct sequencing of segments of the mitochondrial genome amplified using the polymerase chain reaction (PCR) and sets of conserved primer pairs (Kocher *et al.*, 1989).

The first mitochondrial genome to be sequenced was that of human (Anderson *et al.*, 1981), and this was quickly followed by those of cow (Anderson *et al.*, 1982), mouse (Bibb *et al.*, 1981) and a frog (Roe *et al.*, 1985). Comparisons of these sequences revealed some highly conserved regions. This enabled the design of three pairs of oligonucleotide primers that could be used in conjunction with PCR (Saiki *et al.*, 1988) to amplify specific segments of mtDNA from a rather crude preparation of total DNA extracted from vertebrates (Kocher *et al.*, 1989). Furthermore, the same protocol was able to amplify mtDNA from preserved archival specimens. In both cases, the amplified products were in a state that sequencing could be performed directly without the need of cloning.

Although it was initially quite expensive to contemplate establishing a facility to carry out high-throughput sequence analysis for population studies, the development of DNA sequencing technology as a result of the human genome project now makes this quite realistic. In the interim period, researchers developed a compromise; they used sets of primer pairs to amplify defined regions of the mitochondrial genome, and then analyzed the amplified fragments with a series of restriction endonucleases that usually had four base pair recognition sites. This latter technique has proved particularly useful

for large-scale population surveys. Today, sequencing a mitochondrial genome is usually done by PCR amplification of overlapping segments of the entire genome and subsequent assembly of the sequences of these segments using standard bioinformatic tools.

1.3 Fish mtDNA

MtDNA has been used extensively to study fish at various levels of organization. These include analysis of variation at the population level, which provides useful information for the conservation of endangered species and the identification of stocks for making management decisions, and comparisons between different species to determine phylogenetic relationships (Wilson *et al.*, 1985). For example, measuring the amount and distribution of variation in the mtDNA of the highly endangered cyprinid, *Anaecypris hispanica*, allowed the geographical structuring of this species to be deduced. The results suggested to conservationists ways to manage these populations while maintaining the genetic diversity of the species (Alves *et al.*, 2001). Concerns with aquaculture hatcheries regarding the hybridization and introgression of wild and cultivated populations motivated mtDNA studies on several fish species such as the Southern Steelhead (Nielsen *et al.*, 1994), ayu (Iguchi *et al.*, 1999) and other migratory freshwater fishes (Sivasundar *et al.*, 2001). Furthermore, the amount of genetic variation within and among populations assisted in selecting broodstock for *L. elongates*, an endangered fishery resource, to preserve the genetic diversity within this aquaculture species (Martins *et al.*,, 2003). Single nucleotide polymorphisms in the mitochondrial genome are now being used to monitor gene flow among populations of species such as cod (Carr *et al.*, 1995). MtDNA variation has provided markers for stock identification and the management of walleye stocks (Billington and Hebert, 1988; Billington *et al.*,

1988). The amount of genetic differentiation among populations of freshwater fish tends to be lower than what is observed in anadromous fish, which in turn is less than that observed in marine species. This is thought to be because there are fewer isolating barriers to gene flow, in the continuous realm of the oceans. It may also reflect differences in population sizes. The amount of mtDNA sequence divergence was greater between Arctic and coastal cod, *Gadus morhua*, (1.8–5.6%), than was observed between populations in coastal localities in the North East Atlantic Ocean (<1%) (Dahle, 1991).

Aside from stock identification and population genetic studies, mtDNA is a useful tool for inferring phylogenetic relationships. With the availability of procedures for analysing mtDNA, phylogenetic relationships that were previously based on morphological characteristics could be re-examined from the point of view of genetic differences. A combination of molecular and morphological data appears to give the best approach for systematics and the taxonomy assignment in fishes. The molecular systematics of gadoid fishes demonstrated the success of mtDNA sequence in determining phylogenetic and biogeographic relationships (Carr *et al.*, 1999). The investigation of two partial mtDNA genes (896 bp in total length) from fourteen cod fish species clarified relationships between these species. As well, the taxonomy of the Gadidae based on mtDNA is consistent with time of speciation events. The study suggested that morphological data are not as conclusive as molecular evidence for determining the phylogenetic relationships of gadoids.

The phylogeny of divergent lineages has been found difficult to resolve with only morphological data and limited mtDNA sequences. Because the understanding of deep level relationships is important for investigating the evolutionary history of vertebrates,

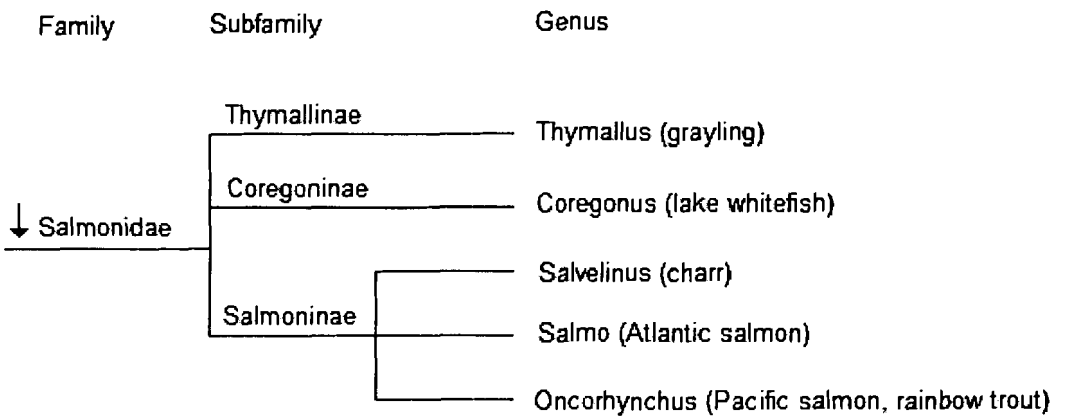
researchers turned to sequencing complete mtDNA genomes to elucidate more information. Comparisons of full mitochondrial genomes have proven successful in resolving controversial groupings. For instance, in the determination of the closest sister group to tetrapods, based on paleontological and morphological data, the ray-finned fishes is accepted as the outgroup to the lobe-finned fishes and tetrapods. However, the distance between the two groups of lobe-finned fish (lungfish and coelancanth) to tetrapod was uncertain. Using the complete mtDNA sequence of lungfish (Zardoya and Meyer, 1996), it was possible to produce a statistically confident tree that groups coelacanth as a sister clade to tetrapods. This result demonstrated the benefit of whole mitochondrial genome sequencing to resolve discrepancies in deeper level phylogenies. A separate study of the phylogeny of basal euteleostei (protacanthopterygian), a diversified subgroup of teleost fishes, examined the controversial relationships within the euteleostean lineage (Ishiguro *et al.*, 2003). 8,130 bp of homologous mtDNA sequence from thirty-four species was used to infer the phylogeny. A well resolved tree found that the basal euteleostie is not one monophyletic group; rather there are five monophyletic groups. The topology showed the following relationships: {(Esociformes, Salmoniformes), (Osmerioids, Argentinoids), Alepocephaloidea}. The ability of mitogenomic (whole mitochondrial genome) data to resolve persistent controversies in teleost phylogeny again revealed the benefit of whole mitochondrial genome sequences for phylogenetic analysis.

1.4 The salmonids

Salmonids are members of the Salmonidae family, and include the whitefishes and ciscos (subfamily Coregoninae), graylings (subfamily Thymallinae) and trout,

salmon and charr (subfamily Salmoninae) (Figure 1). These fish have been further classified into nine genera and roughly sixty-eight species (Nelson, 1994). The genus *Salmo* lies in the Salmoninae, and comprises two main species: *Salmo salar* (Atlantic salmon) and *Salmo trutta* (brown trout). All salmonid species breed in freshwater; however, many of them are migratory and spend a period of their adult life in the sea. There are some "landlocked" salmonid species, which forego the ocean stage and complete their entire lifecycle in freshwater.

Figure 1 Salmonid taxonomy showing the relationships of the three subfamilies in the Salmonidae.



The salmonids comprise several species of substantial importance for aquaculture, wild stock fisheries and recreational sport fisheries. Besides their great economic and societal importance, the salmonids are also of considerable scientific importance in such fields as physiology, genetics, immunology, toxicology, nutritional and environmental science. There is no other species group that receives such a comprehensive combined

commercial and scientific human attention. More is known about the biology, life history, population dynamics, biogeography and evolution of salmonids than any other fish family (Brown and Thorgaard, 2002). The salmonids are especially useful as model systems for ecological and evolutionary studies addressing questions regarding the effects of population structure, genetic and morphological divergence, gene flow and Pleistocene glaciations (Brunner *et al.*, 2001).

1.4.1 MtDNA studies on salmonids

The complete sequence of a mitochondrial genome has the greatest utility for molecular phylogenetic studies, because it can provide insights in all levels of phylogenetic problems (Boore and Brown, 1998). Salmonids whose mtDNAs have been completely sequenced include: whitefish, *Coregonus lavaretus* (Miya and Nishida, 2000); rainbow trout, *Oncorhynchus mykiss* (Zardoya *et al.*, 1995); chinook salmon, *Oncorhynchus tshawytscha* (Wilhelm *et al.*, 2003); cutthroat trout, *Oncorhynchus clarki* (Hickey *et al.*, Unpublished), Arctic char, *Salvelinus alpinus* (Doiron *et al.*, 2002); brook trout, *Salvelinus fontinalis* (Doiron *et al.*, 2002) and Atlantic salmon, *Salmo salar* (Hurst *et al.*, 1999). The organization of all salmonid mitochondrial genomes is the same, and they are all approximately the same size. MtDNA has been used to characterize salmonid phylogenies, both at the inter-specific and the intra-specific level.

1.4.1.1 Inter-specific relationships

Inter-specific relationships among salmonids have been studied using many different procedures. In the early days, experimental hybridization between species pairs was used to evaluate the relationship between *Oncorhynchus* and *Parasalmo* (Chevassus,

1979). Viable offspring were used as the indicator of a close relationship between the two genera, which later collapsed into a single genus. Advances in fish biology allowed morphological and life-history data to replace hybridization crosses, and these results suggested that rainbow trout should belong in a genus similar to *Salmo* (Neave, 1958). Thomas *et al.* (1986) used cleavage sites for thirteen restriction enzymes in the mitochondrial genome to examine inter-specific relationships among the five Pacific salmon (coho, chinook, sockeye, pink and chum) and rainbow trout. By examining 48 restriction sites, ~1.7% of the mitochondrial genome from each individual was sampled. The results showed that rainbow trout and coho and chinook salmon displayed low levels of nucleotide sequence divergence compared to the 10% observed between rainbow trout and other *Salmo* species (Gyllensten and Wilson, 1987). Subsequent confirmation of this result contributed to rainbow trout being placed in the genus *Oncorhynchus* along with the five Pacific salmon.

In the genus *Oncorhynchus*, restriction analysis of mtDNA using 13 restriction enzymes confirmed that the Pacific salmon are monophyletic. Inter-specific variation in mitochondrial DNA sequences revealed the relationships between pink, chum and sockeye salmon as well as their relationship to masu salmon. Within the Pacific salmon, three distinct groups were formed; one group contains pink and chum salmon, another group includes coho and chinook and the third group contains sockeye salmon alone. Their mtDNA sequence divergence ranges from 2.5 – 6.9% between species (Thomas *et al.*, 1986). Examination of a 2,214 bp *Hind*III fragment of mtDNA sequence containing four protein genes and two tRNAs from five Pacific salmon and one rainbow trout revealed 4.74-8.76% in sequence difference (Thomas and Beckenbach, 1989). This level

is higher than that reported from restriction analysis which shows that restriction site data often underestimate the level of variation. A lower inter-species difference was observed in the genus *Salmo*, with only 5.4 % between Atlantic salmon and brown trout based on a 295 bp segment of the cytochrome *b* gene (Mcveigh and Davidson, 1991) and 6.2-6.5% calculated using RFLP analysis of total mtDNA (Gyllensten and Wilson, 1987).

Salmonidae systematics has been inferred using a variety of techniques including morphological (Neave, 1958), ecological (Vladykov, 1963), cytological (Philips *et al.*, 1989), DNA restriction analysis (Thomas *et al.*, 1986) and sequencing data (Thomas and Beckenbach, 1989). Based on 295 bp of the cytochrome *b* gene, a closer relationship between *Salmo* and *Salvelinus* than of *Salmo* and *Oncorhynchus* was observed (Mcveigh and Davidson, 1991). However mtDNA RFLP data (Gyllensten and Wilson, 1987) and allozyme data found a different grouping whereby *Salmo* and *Oncorhynchus* group together with *Salvelinus* being more distant. The controversies could be due to the difference in the level of resolution from different techniques. Direct sequence comparisons should be clear, consistent and comparable between different laboratories. Crespi and Fulton (2004) examined the phylogeny within Salmonidae using all sequence data from GenBank. It provided a thorough picture of inter-genus divergence, and found that *Salmo* and *Oncorhynchus* are not sister taxa. Instead there was stronger evidence for a sister-taxon relationship between *Oncorhynchus* and *Salvelinus*. The same relationship was well-supported in a previous study on the ND3 region of the mitochondrial genome (Oakley and Phillips, 1999). The success of molecular data in resolving phylogenetic relationship shows that mtDNA sequences are valuable resources.

1.4.1.2 Intra-specific relationships

Intra-specific variation in mtDNA is, as expected, much lower than that observed in inter-specific comparisons. A 640 bp region of the D-loop was sequenced from individuals representing 24 populations of brown trout (S. *trutta*) from different locations in Europe, and 12 mtDNA genotypes were identified (Bernatchez *et al.*, 1992). These genotypes clustered into five phylogeographic groups with a mean sequence difference of 0.96-1.44%. Similarly, a 552 bp fragment of the mtDNA control region was sequenced from 159 Arctic charr from 83 different populations (Brunner *et al.*, 2001). The sequence differences split the Arctic charr into five main complexes: Siberia, Atlantic, Acadia, Arctic and Bering. The mean sequence divergence between the groups was 0.8%.

1.4.1.3 MtDNA studies on Atlantic salmon populations

Studies on Atlantic salmon using mtDNA markers changed tremendously over the past 20 years. As technology developed, they advanced from restriction mapping of mtDNA (Birt *et al.*, 1986) to PCR-based direct sequencing of a small region of the mitochondrial genome (McVeigh *et al.*, 1991) and more recently, to a combination of PCR amplification of specific genes followed by restriction endonuclease analysis (Verspoor *et al.*, 1999). As will be described in this thesis, the rapid sequencing of complete mitochondrial genomes enables the identification of new mtDNA markers to study salmonid population dynamics and to resolve deeper phylogenies.

Birt *et al.* (1986) used restriction endonuclease analysis to conduct the first investigation of mtDNA variation in Atlantic salmon when they examined allopatric anadromous and non-anadromous populations in Newfoundland. Similar experiments were carried out in Atlantic salmon from Sweden (Gyllensten and Wilson, 1987) and

Finland (Palva *et al.*, 1989). These experiments found some variable restriction endonuclease cutsites in the mtDNA, and in some instances they could be used to differentiate the Atlantic salmon populations. Bermingham *et al.* (1991) discovered that the restriction enzyme *Bgl*II produced a different cleavage pattern when sampling "European-origin" and "North American-origin" Atlantic salmon. The presence or absence of this *Bgl*II site was used to predict the continent of origin of Atlantic salmon in the mixed-stock fishery off the coast of Greenland. The main research interest in the late 1980s and 1990s involved looking for molecular markers that could be used to survey the mitochondrial genome, with a goal to determining population substructures and the evolutionary relationships among populations of Atlantic salmon (Birt *et al.*, 1991; Hovey *et al.*, 1989; Knox and Verspoor, 1991; Nielsen *et al.*, 1996; Oconnell *et al.*, 1996; Verspoor *et al.*, 1999). A limitation of restriction enzyme analysis is that it can only sample a limited region of the mitochondrial genome. It is also difficult to interpret restriction enzyme data of mtDNA fully without some knowledge of the location of the mutations and their possible consequences. These drawbacks made some researchers turn to direct sequencing analysis when the technology became available (McVeigh *et al.*, 1991; Palsson and Arnason, 1994), but it never became popular because of concerns over the high cost of DNA sequencing and the lack of high-throughput methods in the 1990s. As will be seen from this thesis, DNA sequencing is no longer an issue for analyzing many samples at a reasonable cost.

1.5 Analyses of fish mitochondrial sequences

A 307 bp fragment of the cytochrome *b* gene was sequenced from 100 organisms using conserved primers and PCR amplification (Kocher *et al.*, 1989). The number of

amino acid replacement in three lineages was investigated based on branch lengths of a phylogenetic tree. There were five times more amino acid changes in the mammalian and bird lineages compared to fish lineages within the same evolutionary length of time. Therefore, this suggested that there was a five-fold higher rate of amino acid substitution in the bird and mammalian lineages. The fish species used in this study included sharks, cichlids, salmonids and coryphenids. The same observation was made in a separate analysis of full cytochrome *b* sequences in some *Perciforms* (Cantatore *et al.*, 1994). They found that fish mitochondrial DNA had a nucleotide substitution rate three to five times lower than that of mammals. Furthermore, in another study that examined cytochrome *b* protein evolution in sharks, the rate of shark cytochrome *b* protein was found to evolve six times slower than in mammals (Martin and Palumbi, 1993). This confirmed that not all taxa operate under the same molecular consistent clock, and that there are different molecular evolutionary rates between the mammalian lineage and the fish lineage, with fish mtDNA appearing to evolve more slowly than in other vertebrates. The consistency of results showing a slower rate of amino acid substitution among cold-blooded vertebrates compared to mammals was tested in Pacific salmon (Thomas and Beckenbach, 1989). The mtDNA sequences from six Pacific salmonids showed a high occurrence of silent substitutions compared to replacement substitutions.

The non-coding region of the mitochondrial genome is thought to possess different characteristics from the rest of the mitochondrial genome. In many vertebrate studies, this segment was found to be the most rapidly evolving region of the mitochondrial genome (Upholt and Dawid, 1977). This region contains the D-loop, which forms a displacement loop during replication of the heavy strand of the circular

mitochondrial genome in mammals. The size of this region varies among different animals, ranging from 200 – 1,400 bp (Brown *et al.*, 1982). In the human D-loop, the substitution rate is about 2.8 to 5 times faster than that of the rest of the mitochondrial genome (Aquadro and Greenberg, 1983; Cann *et al.*, 1984; Meyer, 1993). Despite the high tendency to accept mutations, there are several sequence blocks that are conserved in all mammalian and vertebrate mtDNA (Chang and Clayton, 1985). A number of reports suggested that fish D-loops do not behave the same way as those of mammals. Bernatchez and Danzmann (1993) and Shedlock *et al.* (1992) found that the D-loop region evolves more slowly in salmonid fishes. Moreover, in freshwater rainbow fishes (genus *Melanotaenia*), the overall rate of divergence between the cytochrome *b* gene and the D-loop was similar (Zhu *et al.*, 1994).

1.6 Purpose of thesis

The first goal of my thesis was to establish methods that could rapidly determine the sequence of the entire mitochondrial genomes of Atlantic salmon from different populations throughout their range. The second goal was to compare the complete mitochondrial genome sequences from individual Atlantic salmon with one another and with the corresponding sequence from a brown trout to reveal the mode and tempo of nucleotide and amino acid substitutions in various regions of the mitochondrial genomes in the genus *Salmo*. Finally, the third goal was to test hypotheses, developed from previous studies using restriction endonuclease analysis of Atlantic salmon mtDNA, concerning the population structure of Atlantic salmon.

CHAPTER 2 - MATERIALS AND METHODS

2.1 DNA extraction

Samples of fin tissue from Atlantic salmon specimens were used to isolate total DNA. The geographical origins of the samples are listed in Table 1. The extractions were performed using the Puregene DNA Isolation Kit by Genetra System. DNA was isolated according to the "mouse tail tissue protocol" with 300 μg of fin tissue.

2.2 Primer development

Thirty-three pairs of mitochondrial specific primers were designed using a template generated from the consensus sequence of four different complete salmonid mitochondrial sequences downloaded from the NCBI database (GenBank Accession Nos. NC_001960, *Salmo salar*; NC_00861, *Salvelinus alpinus*; NC_00860, *Salvelinus fontelinus*; NC_001717, *Oncorhynchus mykiss*). Conserved regions were selected to place overlapping forward and reverse primers (Figure 2). Primer design was carried out using Primer3, a web based primer design program (Rozen and Skaletsky, 2000; Whitehead Institute for Biomedical Research; http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). The primers were mostly 20 to 25 nucleotides in length, with some sites being degenerate to enable them to amplify mtDNA from a range of salmonids (Table 2). PCR products varying from ~500 bp to ~1000 bp were acquired which suits the subsequent direct sequencing of these PCR products. The combination of thirty-three overlapping PCR products yields the full mitochondrial genome sequence.

Table 1 **Origin and size of Atlantic salmon and brown trout mitochondrial genomes used in current analysis**

Species	Sample Name	Geographical origin of sample	Full MtDNA available	Whole mtDNA genome size (bp)
Salmo salar	Alm	Scotland	No	N/A
Salmo salar	Nfld	Newfoundland	Yes	16,674
Salmo salar	ConRiv	Newfoundland	Yes	16,674
Salmo salar	K1	Baltic Sea	Yes	16,671
Salmo salar	K28	Baltic Sea	Yes	16,671
Salmo salar	K31	Baltic Sea	No	N/A
Salmo salar	K36	Baltic Sea	Yes	16,671
Salmo salar	K42	Baltic Sea	Yes	16,671
Salmo salar	LoirV	France	Yes	16,671
Salmo salar	Mork1	Baltic Sea	Yes	16,671
Salmo salar	NB	New Brunswick	Yes	16,674
Salmo salar	Swed	Sweden	Yes	16,671
Salmo salar	UngBay	Labrador	Yes	16,674
Salmo salar	WGreen1	West Greenland	Yes	16,672
Salmo trutta	BT89	Newfoundland	Yes	16,679

Figure 2 **The organization of the linearized mitochondrial genomes of *Salmo salar* and *Salmo trutta*.**

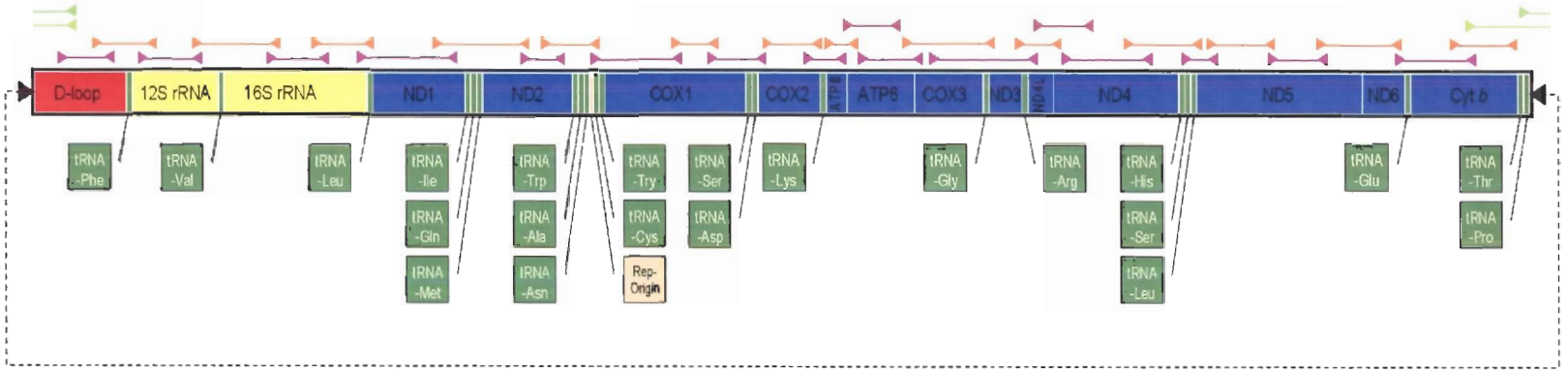


Table 2 **Thirty-three pairs of mitochondrial specific PCR and sequencing primers for species in the genus *Salmo*.**

Note: Y = C/T; R = A/G; M = A/C; W = A/T; S = C/G; K = G/T; H = A/C/T

Pair #	Site	Primer Sequence (5'-> 3')	Size (bp)	Annealing Temp(°C)
1	277F 818R	CTCCGTCTTTACCCACCAAC TGGTTTAGGGGTTRACAGGAA	541	60.9
2	629F 1305R	TGAATTCCAGAGAACCCATGT AGTCAAGCTTTCGCTTATGG	676	60.9
3	1142F 1842R	AAGTCTCCGCAYTCCTGTGA GTAGCCCATTTCTTCCCACC	700	56.2
4	1772F 2708R	CCCGCCTATATACCACCGTC TAGCCGCCTGTWAAGGTTGT	936	56.2
5	2552F 3251R	GCTGGTTGCTTAGGAAATGAA CGTGATGCCATTCATACAGG	699	60.9
6	3082F 3731R	ACACAAGCCTCGCCTGTTTA CCTGGATTACTCCGGTCTGA	649	56.2
7	3568F 4624R	CCTAGGGATAACAGCGCAAT GGATGTGGGATGCKCCTA	1056	60.9
8	1149F 1904R	CTGTGAGGATGCCCTTAATC CAGAGCCAGTTTCAAGAGAAC	755	51.4
9	4401F 5482R	ACTMGTACCAGCCTGACCCC TGAATTATAAGTGCGAAAGGTG	1081	56.2
10	5352F 5861R	GCCCTTGCACTTAAACTTGG GTTCTTGYAAAATAAGTCATTTWGG	509	51.4
11	5602F 6250R	TCTTCAATYGCCCACYTAGG GAGGCCTTCCCACCTAGAAA	648	56.2
12	6137F 7161R	CAAAGCTCTAARCGGGRGTG CCGGGTCAAAGAAAGTGGT	1024	56.2
13	7042F 7574R	CCCRCTTTTTGTTTGAGCTGT CAAGGACAATGCCYGTRAGT	532	51.4
14	7418F 8099R	YACCATRATCATCGCCATCC GTTATGCGGTTGGCTTGAAA	681	60.9
15	8055F 8702R	GAGGAATTGAACCCCCATGT ACGGCTCAGGAGTGAAGGAC	647	60.9
16	8514F 8993R	ATGGGCCAYCAATGATAYTG AGTTGGGGCATGTCACTAGG	479	60.9

Pair #	Site	Primer Sequence (5'-> 3')	Size (bp)	Annealing Temp(°C)
17	8731F 9132R	GTYCCAGGACGATTAAACCA GGTCAGTTTCAGGGTTCAGG	401	56.2
18	8966F 9586R	CCTAGTGACATGCCCCAACT TAAAAAGGCTAATTGTYTCGAT	620	56.2
19	9127F 9821R	CTGACCATGACACTAAGCTTCT TGCCATTARACGTTTTCTTG	694	56.2
20	9602F 10636R	GCCCTTGGYGTACGACTYAC GACCGGGTGATTGGAAGTC	1034	56.2
21	9900F 11092R	TCAGGCACTGCAGTCTGATT CATAAGGCGGTCATGGACTT	1192	56.2
22	10861F 11357R	CCTRCGCTTCTTTCTAATCG GTGWGTTCGKGCAGTTGCTA	496	51.4
23	11043F 11718R	CCTTGATTTCGGCTCAAAAG GGCCCCGAATGCTAAKAY	678	56.2
24	11352F 12359R	ACCGCCTCCAAAGCYTAAA AGTGCTGAGGAGGCRAGRC	1007	60.9
25	12042F 12911R	GTAGAAGCCCCAATCGCMG TCCTTTAGAAGCACGAGTGAA	869	60.9
26	12697F 13093R	CCCGSGAACACCTACTTATY GAGTGAGGGCTCAGTTTTYWTG	396	56.2
27	12973F 13722R	ATGCACCCGACYACACTCATC ACCGGCGTAGGACCYTCTAT	749	64.5
28	13642F 14303R	CTAGCCGCCACYGGYAAAT GTAAAGCGGGGGTGTCCYAT	661	64.5
29	14194F 15128R	TCTCACCTTAACGCCTGAGC GCTTATTCAGCGGCTTTGG	934	60.9
30	15062F 15928R	ACCCTAAAACCGAACGATCC TTAGGGTGGCGTTGTCTACAG	866	60.9
31	15841F 428RO	ACAAACCTCCTCTCHGCTGT TAGGAACCAAATGCCAGGAA	1252	56.2
32	15684F 16413R	CGCTAACGGAGCATCTTTCT TATGTCYGCTACYAGGGTTC	729	60.9
33	16431F 428RO	GAACCCTRGTAGCRGACATA TAGGAACCAAATGCCAGGAA	662	60.9

2.3 PCR conditions

PCR amplification were performed with 50 ng of total genomic DNA in a 25 μL reaction including 0.05U Taq DNA polymerase, 12.5 pmoles of each specific primer, 2.5 μL of 10X PCR Buffer, 12.5 μmole of dNTPs (deoxy-nucleotide-tri-phosphate) and autoclaved water to bring the final volume to 25 μL. The PCR amplification cycle began with an initial denaturation step of 4 minutes at 95°C, followed by 35 cycles consisting of a denaturation step at 95°C for 45 seconds, annealing at the optimal annealing temperature specific to each pair of primers for 45 seconds, extension at 72°C for 1 minute, and ending with a final extension step at 72°C for 10 minutes. Following mtDNA amplification, 5 μL of PCR products were electrophoresed on a 1% agarose gel made with 1X TBE. The size of each product was verified using a 1 Kb standard ladder. The remaining 20 μL of PCR products were purified using the QIAGEN QIAquick PCR product purification kit, and quantified with a spectrophotometer. These purified products were directly applied to sequence analysis using the florescent dideoxy terminator technique.

2.4 Sequencing

Sequencing reactions were done using the Amersham DYEnamic ET Terminator sequencing kit (dye labelled dideoxy-nucleotides, ddNTPs). 10 ng of purified PCR product was added for every 100 bp of the PCR product, 0.5 μL primer (10 μM stock), 8 μL sequencing premix, and autoclaved water) to bring the volume to 20 μL. Sequencing reactions were carried out with the following conditions: 94°C 30 seconds, Tm (55°C) 20 seconds, 60°C 1 minute for 30 cycles using a thermocycler (Biometra) for

amplification. Unincorporated ddNTPs were removed from each sequencing mixture by ethanol precipitation, and the reactions were resuspended in 2 μL of formamide loading dye. Sequence analysis was carried out on an ABI 377 DNA Sequencer (Applied Biosystems).

2.5 Assembling the genome

The sequences were analyzed using Sequencher 4.1 (Gene Codes Corporation), and manually edited. Overlapping sequence reads were assembled into a contig using a previously sequenced Atlantic salmon mtDNA genome from GenBank (GenBank Accession No. NC_001960) as a template.

2.6 Bioinformatics programs for analysis

A multiple sequence alignment of the fourteen Atlantic salmon mtDNA sequences was generated with ClustalW (v.1.83; Thompson *et al.*, 1994) using default parameters for gap opening, penalty of 10.0, and gap extension, penalty of 5.0. RevSeq (Emboss) was used to find the reverse complement sequence of ND6 that is located on the non-coding strand opposite from the sequenced template strand. To assess the characteristics of the *Salmo* mtDNAs, the base composition profiles, the number of nucleotide differences and amino acid differences were examined using MEGA version 3.1 (Kumar *et al.*, 2004). The degree of pairwise sequence divergence and the number of transitions and transversions were also investigated. The location of sites that are variable amongst the fourteen Atlantic salmon sequences were recorded to study the relative rate of evolution at different regions of the mtDNA genome of *Salmo*. Furthermore, the differentiation among individuals was studied using UPGMA analysis to infer phylogeny

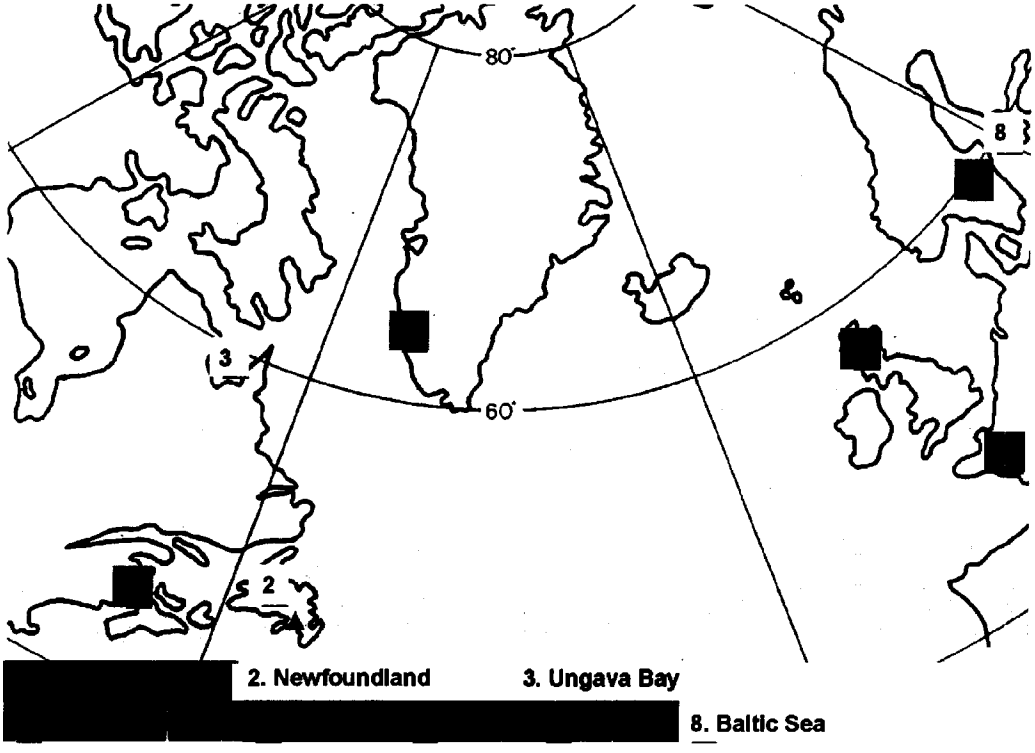
based on pairwise genetic distances between individuals. A neighbour-joining tree, a maximum parsimony (MP) tree and a UPGMA-based tree were constructed using MEGA version 3.1 (Kumar *et al.*, 2004). These trees showed the same topology suggesting the validity of the relationship among these sequences.

CHAPTER 3 - RESULTS

3.1 Atlantic salmon mitochondrial genomes

Twelve complete and two near-complete Atlantic salmon (*Salmo salar*) mitochondrial genomes were sequenced. Individuals were sampled from eight geographic regions as shown in Figure 3 and described further in Table 1. In addition, the complete mtDNA sequence was determined for a brown trout (*Salmo trutta*) from Rennies River, St. John's, Newfoundland.

Figure 3 Geographical origins of the fourteen Atlantic salmon whose mtDNAs were sequenced. The origin of the brown trout sampled is marked by ▲.



3.2 Characteristics of *Salmo* mitochondrial genomes

3.2.1 Genome content

The mitochondrial genomes from both members of the *Salmo* genus are identical in their genome organization. The complete circular genome is 16671 – 16679 bp in size. Protein coding genes and ribosomal RNA genes were identified by their similarity to the published *Salmo salar* mitochondrial DNA sequence (Hurst *et al.*, 1999). The *Salmo* mitochondrial genome contains one non-coding control region (D-loop), two genes for ribosomal RNA, 22 genes encoding tRNAs and thirteen protein-coding genes (Figure 2). The proteins it encodes include seven subunits of the NADH ubiquinone oxidoreductase complex (ND), and one subunit of the ubiquinol cytochrome c oxidoreductase complex (cytochrome *b*), three subunits of the cytochrome c oxidase complex (COX), and two subunits of ATP synthase (ATPase). Twelve of the thirteen protein genes are located on the "+" strand, with NADH ubiquinone oxidoreductase subunit 6 (*ND6*) being the only protein gene positioned on the "-" strand. The relative positions and orientation of the genes in the Atlantic salmon and brown trout mitochondrial genomes are consistent with those observed in other salmonids such as rainbow trout, chinook salmon and Arctic charr.

3.2.2 Mitochondrial genome organization

The Atlantic salmon mitochondrial genome is organized in an economical way, with genes either abutting or overlapping. The twenty-two genes for tRNAs are interspersed throughout the circular genome, connecting the thirteen protein genes. Few short spacer regions are present between some genes, but they are mostly one to two bp

in size. The longest spacer is 14 bp located between *ND1* and tRNA-isoleucine. The overlapping arrangements in three sets of genes *ATP8/ATP6*, *ND4L/ND4* and *ND5/ND6* were 10, 7, and 4 bp, respectively. The overlaps are identical to that found in other Salmonidae species. The similarity in the organization of mitochondrial genomes across other fish species (such as cod, carp, pike and zebrafish) verified the conserved nature of mtDNA organization in teleosts.

The size of the fourteen Atlantic salmon mitochondrial genomes ranges from 16671-16674 bp. The differences are contributed by a two bp insertion in the D-loop region and the gene for tRNA-tyrosine. The brown trout mitochondrial genome sequence is 16679 bp in length. Small single or double base pair insertions in the D-loop and 1 bp in the 16S ribosomal RNA gene demonstrated that the control region is the most flexible region to accept indel mutations.

3.2.3 Base composition

The base composition of the coding strand in the Atlantic salmon mitochondrial genome illustrated a bias against the use of guanine nucleotides with only 15.3% of the nucleotides being guanine. The average distribution of the other bases was; T, 26.3%; C, 30.1%; A, 28.3% (Table 3), which was similar to that observed in the rainbow trout mitochondrial genome, which also had a sight preference for cytosine. An examination of the nucleotide composition profile in each coding gene showed a similar pattern in all coding genes where guanine is not preferred in the third position of codons. However, this trend was not observed in *ND6*. Instead, the higher occurrence of guanine was observed at the first position of codons in this gene. The ND6 gene is distinct from other coding genes in that it is located on the "-" strand of the mitochondrial genome. Cytosine

and adenine were observed at a lower percentage on the minus strand, in particular at the first and third codon position: T = 37.9%, C = 12.3%, A = 15.5% and G = 34.4%. These results indicate that the coding strand (+) and non-coding strand (-) have different preferences for nucleotide usage.

Table 3 The nucleotide composition profile for the *Salmo salar* mitochondrial genome

14 *Salmo salar* (Atlantic salmon)																
Position	1				2				3				Total			
Nucleotide	T	C	A	G	T	C	A	G	T	C	A	G	T	C	A	G
D-loop													31.7	22.1	31.6	
12S rRNA													19.9	27.3	30.4	
16S rRNA													19.5	25.5	32.8	
ND1	18.2	30.5	29.0	22.4	36.3	30.8	17.3	15.7	27.5	34.0	29.6		27.3	31.7	25.3	
ND2	18.6	31.1	29.9	20.4	39.4	34.3	15.1	11.1	20.9	26.1	39.4		26.3	33.9	28.1	
ND3	19.7	32.5	21.4	26.5	40.6	32.7	14.7	12.1	22.6	37.7	34.5		27.6	34.3	23.5	
ND4	22.3	28.9	28.7	20.1	41.3	27.0	16.1	15.7	20.4	34.1	38.4		28.0	30.0	27.7	
ND4L	19.2	34.3	19.2	27.3	41.4	28.3	14.1	16.2	18.1	35.6	41.9		26.2	32.7	25.1	
ND5	18.9	25.6	33.9	21.5	39.2	28.9	20.2	11.7	24.0	37.4	34.6		27.4	30.6	29.6	
ND6	35.6	8.6	10.9	44.8	43.7	21.3	12.6	22.4	34.5	6.9	23.0	35.6	37.9	12.3	15.5	34.3
COX1	22.4	21.5	22.5	30.5	40.4	26.3	18.0	15.3	28.8	33.8	30.0		30.6	27.2	24.5	
COX2	18.2	26.8	22.6	32.3	37.8	23.5	27.8	10.9	28.7	33.4	32.4		28.2	27.9	27.6	
COX3	26.7	25.2	21.0	27.1	36.3	25.2	21.8	16.8	23.6	37.2	26.6		28.9	29.2	26.5	
ATPase8	23.2	30.4	26.8	19.6	28.6	33.9	25.0	12.5	26.8	30.4	37.5		26.2	31.5	29.8	
ATPase6	13.6	36.4	29.4	20.6	47.8	26.8	14.0	11.4	24.5	35.6	36.5		28.6	32.9	26.6	
Cyt *b*	23.1	26.3	23.6	27.0	40.0	26.3	19.7	14.0	25.0	39.7	31.3		29.4	30.8	24.9	

3.3 Sequence variation within Atlantic salmon

3.3.1 Amount of sequence variation

The fourteen Atlantic salmon mitochondrial genomes were compared with one another to study the amount of variation between Atlantic salmon from different populations. A multiple sequence alignment of these sequences revealed 215 variable sites among the fourteen individuals (Table 4), and gave a consensus sequence of 16,674 bp (Figure 4). The CLUSTALW alignment requires 3 bp of indels, with 2 bp located in the D-loop and 1 bp in the tRNA-tyrosine gene. The total number of substitutions observed ranges from 7 to 144 bp indicating that about 0.04-0.86% of the mtDNA genome is variable within *Salmo salar* in the individuals studied (Table 5).

Indels were observed in the D-loop (two base pairs within *Salmo salar* samples), and the gene for tRNA-tyrosine. It was not unexpected that there were no indels seen in the protein genes as any would cause a frame-shift and this would prove lethal. The lack of indels indicates that the structures of these molecules are tightly regulated, and suggests that an insertion or deletion would be detrimental to the organism.

Table 4 Table of variable sites in Atlantic salmon mtDNA from 14 individuals

Position	K1	K31	K28	Swed	Mork1	K42	K36	WGreen	LoirV	Alm	Nfld	UngBay	ConRiv	NB
20	C	C	C	C	C	C	C	C	C	C	T	T	T	T
176	T	T	T	T	T	T	T	T	C	T	T	T	T	T
303	A	A	A	A	A	A	A	A	A	A	G	G	G	G
390	C	C	C	C	C	C	C	C	T	C	T	T	T	T
449	C	C	C	C	C	C	C	C	C	C	T	T	T	T
473	T	T	T	T	T	T	T	T	G	T	G	G	G	G
504	T	T	T	T	T	T	T	T	T	T	C	C	C	C
530	C	C	C	C	C	C	C	C	C	C	T	T	T	T
563	T	C	T	T	T	T	T	T	T	T	T	T	T	T
571	G	-	G	G	G	G	G	G	G	G	G	G	A	A
703	G	G	G	G	G	G	G	G	G	G	A	A	G	G
711	T	T	T	T	T	T	T	T	T	T	C	C	C	C
750	C	C	C	C	C	C	C	C	C	N	C	C	C	T
921	G	G	G	G	G	G	A	G	G	N	G	G	G	G
943	A	A	A	A	A	A	A	A	A	N	C	C	C	C
964	-	-	-	-	-	-	-	-	-	N	C	C	C	C
965	-	-	-	-	-	-	-	-	-	N	T	T	T	T
971	G	G	G	G	G	G	G	G	G	N	C	C	C	C
972	T	T	T	T	T	T	T	T	T	N	A	A	A	A
973	C	C	C	C	C	C	C	C	C	N	G	G	G	G
974	A	A	A	A	A	A	A	A	A	N	T	T	T	T
976	T	T	T	T	T	T	T	T	T	N	A	A	A	A
978	G	G	G	G	G	G	G	G	G	N	A	A	A	A
1220	C	C	C	C	C	C	C	C	C	C	C	C	C	G
1336	G	A	G	G	G	G	G	G	G	G	G	G	G	G
1432	G	G	G	G	G	G	G	G	G	G	A	A	A	A
1441	T	T	T	T	T	T	T	T	T	T	C	C	C	C
1465	A	A	A	A	A	A	A	A	A	A	G	G	G	G
1539	G	G	G	G	G	G	G	G	G	G	A	A	A	A
1566	A	G	G	G	G	G	G	G	G	G	G	G	G	G
1760	G	G	G	A	G	G	G	G	G	G	G	G	G	G
1843	A	A	A	A	A	A	A	A	A	A	G	G	G	G
2218	G	G	G	G	G	G	G	G	G	G	G	A	G	G
2850	G	G	G	G	A	G	G	G	A	G	A	A	A	A
2903	T	T	T	T	T	T	T	T	C	T	T	T	T	T

Position	K1	K31	K28	Swed	Mork1	K42	K36	WGreen	LoirV	Alm	Nfld	UngBay	ConRiv	NB
3444	C	C	C	C	C	C	C	C	C	C	G	G	G	G
3711	T	T	T	T	T	T	T	T	T	T	C	C	C	T
3741	A	A	A	A	A	A	A	A	A	A	C	C	C	A
3828	A	A	A	A	A	A	A	G	A	A	A	A	A	A
3892	G	G	G	G	G	G	G	A	G	G	G	G	G	G
3933	C	C	C	C	C	C	C	T	C	C	C	C	C	C
3945	A	A	A	A	A	A	A	A	G	A	A	A	A	A
3978	G	G	G	G	G	G	G	G	G	G	A	A	A	G
3996	G	G	G	G	G	G	A	G	G	G	G	G	G	G
4086	C	C	C	C	C	C	C	C	C	C	T	T	T	C
4089	T	T	T	T	T	T	T	T	T	T	C	C	C	T
4137	G	G	G	G	G	G	G	G	A	G	G	G	G	G
4173	G	G	G	G	A	G	G	G	G	A	G	G	G	A
4251	T	T	T	T	T	T	T	T	T	T	T	T	C	T
4431	C	C	C	C	C	C	C	C	C	C	T	T	T	C
4524	A	A	A	A	A	G	G	G	G	A	G	G	G	G
4572	G	G	G	G	G	G	G	G	G	G	A	A	A	A
4575	C	C	C	C	C	C	C	C	C	C	T	T	T	T
4657	G	A	G	G	G	G	G	G	G	G	G	G	G	G
4673	G	G	G	G	A	G	G	G	G	G	G	G	G	G
4759	G	G	G	G	G	G	G	G	G	G	A	A	A	A
4789	A	A	A	A	A	A	A	A	A	A	A	A	G	G
4829	A	A	A	A	A	A	A	A	A	A	G	G	G	G
4918	G	N	A	A	A	A	A	A	A	A	A	A	A	A
5091	C	N	C	C	C	C	C	C	T	C	C	C	C	C
5172	C	N	C	C	C	C	C	C	C	C	T	T	C	C
5187	C	N	C	C	C	C	C	C	T	C	C	C	C	C
5298	A	N	A	A	A	A	A	G	A	A	A	A	A	A
5320	C	N	C	C	C	C	C	T	C	C	C	C	C	C
5335	G	N	G	G	G	G	G	G	G	G	G	A	G	G
5373	C	C	C	C	C	T	C	C	C	C	C	C	C	C
5388	A	A	A	A	A	A	A	G	A	A	A	A	A	A
5424	A	A	A	A	A	A	A	A	A	G	G	G	G	G
5448	G	G	G	G	G	G	G	A	G	G	G	G	G	G
5451	C	C	C	C	C	C	C	C	C	C	C	A	C	C
5502	A	A	A	A	A	A	A	A	A	G	G	G	G	A
5547	A	A	A	A	A	A	A	G	A	A	A	A	A	A
5571	T	T	T	T	T	T	T	T	T	C	T	T	C	T

Position	K1	K31	K28	Swed	Mork1	K42	K36	WGreen	LoirV	Alm	Nfld	UngBay	ConRiv	NB
5760	T	T	T	T	T	T	T	T	T	Y	T	T	C	C
5767	G	G	G	G	G	G	G	G	G	R	A	A	A	A
5776	G	G	G	G	G	G	G	G	G	R	A	A	A	A
5820	C	C	C	C	C	C	C	C	C	M	A	A	A	A
5844	A	A	A	A	A	A	A	A	A	A	A	A	G	G
6024	G	G	G	G	G	G	G	G	G	G	G	G	G	A
6072	G	G	G	G	G	G	G	G	G	G	A	A	A	A
6080	T	T	T	T	T	T	T	T	T	T	T	C	T	T
6361	G	G	G	G	G	G	G	G	G	G	A	A	A	A
6397	G	G	G	G	G	G	G	G	A	G	G	G	G	G
6409	A	A	A	A	A	A	A	A	G	A	A	A	A	A
6413	-	-	-	-	-	-	-	T	-	T	T	T	T	T
6951	C	C	C	C	C	C	C	A	C	C	C	C	C	C
6961	T	T	T	T	T	T	T	T	T	T	G	G	G	G
6980	T	T	T	T	C	T	T	C	T	T	C	C	C	C
7069	G	G	G	G	G	G	G	G	G	A	A	A	A	G
7162	A	A	A	A	A	A	A	A	A	G	G	G	G	A
7177	C	C	C	C	C	C	C	C	T	C	C	C	C	C
7219	T	T	T	T	T	T	T	T	T	C	C	C	C	T
7294	C	C	C	C	C	C	C	C	C	T	T	T	C	C
7372	G	G	G	G	G	G	G	G	A	G	G	G	G	G
7492	G	G	G	G	G	G	G	A	A	A	A	A	A	R
7546	G	G	G	G	G	G	G	G	G	G	G	G	A	A
7564	T	T	T	T	T	T	T	T	T	T	C	C	C	C
7579	C	C	C	C	C	C	C	C	C	C	T	T	T	T
7580	G	G	G	G	G	G	G	G	G	G	G	G	A	G
7636	A	A	A	A	A	A	A	A	A	A	G	A	A	A
7693	T	T	T	T	T	T	T	T	T	T	C	C	C	C
7852	C	C	C	C	C	C	C	T	C	C	C	C	C	C
8005	A	A	A	A	A	A	A	A	A	A	G	G	G	G
8066	C	T	C	C	C	C	C	C	C	C	C	C	C	C
8388	G	G	G	G	G	G	G	G	G	G	A	A	A	A
8411	T	T	T	T	T	T	T	T	T	T	C	C	C	C
8426	T	T	T	T	T	T	T	T	C	T	T	T	T	T
8501	T	T	T	T	T	T	T	T	T	T	C	C	C	C
8642	T	T	T	T	T	T	T	T	T	T	C	C	C	C
8801	G	G	A	A	A	G	G	G	G	G	G	G	G	G
8840	G	G	G	G	G	G	G	G	A	G	A	A	G	G

Position	K1	K31	K28	Swed	Mork1	K42	K36	WGreen	LoirV	Alm	Nfld	UngBay	ConRiv	NB
8953	A	A	A	A	A	A	A	G	A	A	A	A	A	A
9218	A	A	A	A	A	A	A	A	A	A	A	A	G	A
9272	C	C	C	C	C	C	C	C	C	C	T	T	C	C
9290	A	A	A	A	A	A	A	G	A	A	A	A	A	A
9317	A	A	A	A	A	A	A	G	A	A	A	A	A	A
9323	T	T	T	T	T	T	T	T	C	T	T	T	T	T
9365	T	T	T	T	T	T	T	T	T	T	T	C	C	C
9533	A	A	A	A	A	A	A	A	C	A	A	A	A	A
9560	G	G	G	G	G	G	G	G	G	G	A	A	A	A
9587	C	C	C	C	C	C	C	C	C	C	C	C	A	A
9595	G	G	G	G	G	G	G	G	G	G	G	G	C	G
9677	C	C	C	C	C	C	C	C	C	C	T	T	T	T
9788	T	C	T	T	T	T	T	T	T	T	C	C	C	C
9880	T	T	T	T	T	T	T	T	T	T	C	C	C	C
10000	C	C	C	C	C	C	C	C	C	C	T	T	T	T
10126	T	T	T	T	T	T	T	T	T	T	C	C	C	C
10291	C	C	C	C	C	C	C	C	C	C	T	T	C	C
10402	A	A	A	A	A	A	A	A	A	A	C	C	C	C
10429	G	G	G	G	G	G	G	G	G	G	A	A	A	A
10606	-	A	A	A	A	A	A	A	A	A	A	A	A	A
10644	G	G	G	G	G	G	G	G	G	G	A	A	A	A
10774	C	C	C	C	C	C	C	C	C	C	T	T	T	T
10837	T	T	T	T	T	T	T	T	T	C	C	C	C	C
10882	C	C	C	C	C	C	C	C	C	T	T	T	T	T
10888	T	T	T	T	T	C	T	T	T	T	T	T	T	T
10908	C	T	C	C	C	C	C	C	C	C	C	C	C	C
11163	G	G	G	G	G	G	G	G	G	A	A	A	A	G
11205	T	T	T	T	T	T	T	T	T	T	C	C	C	T
11214	C	C	C	C	C	C	C	C	T	C	C	C	C	C
11262	G	G	G	G	G	G	G	G	G	A	A	A	A	G
11325	A	A	A	A	A	A	A	A	A	M	A	A	A	M
11381	C	C	C	C	C	C	C	T	C	C	C	C	C	C
11390	A	A	A	A	A	A	A	A	G	A	A	A	A	A
11498	G	G	G	G	G	G	G	G	G	G	A	A	A	A
11639	A	A	A	A	A	A	A	G	A	A	A	A	A	A
11762	T	T	T	T	T	T	T	T	T	T	C	C	T	T
11900	G	G	G	G	G	G	G	G	G	G	A	A	A	A
11915	C	C	C	C	C	C	C	C	C	C	T	T	T	T

Position	K1	K31	K28	Swed	Mork1	K42	K36	WGreen	LoirV	Alm	Nfld	UngBay	ConRiv	NB
11939	G	G	G	G	G	G	G	G	G	G	A	A	G	G
11943	G	G	G	G	G	G	G	G	G	G	G	G	A	A
11946	G	G	G	G	G	G	A	G	G	A	G	G	G	G
11960	A	A	A	A	A	A	A	A	A	A	G	G	G	G
12002	T	T	T	T	T	T	T	T	T	T	C	C	C	C
12089	G	G	G	G	G	G	G	G	G	G	A	A	A	A
12176	A	A	A	A	A	A	A	A	G	A	G	G	G	G
12203	G	G	G	G	G	G	A	G	G	G	A	A	A	A
12212	A	A	A	A	A	A	A	A	A	A	G	G	G	G
12216	C	C	C	C	C	C	C	C	C	C	T	T	T	T
12266	A	A	A	A	A	A	A	G	A	A	A	A	A	A
12269	G	G	A	G	G	G	G	G	G	G	G	G	G	G
12287	G	G	G	G	G	G	G	G	G	G	A	A	A	A
12329	C	C	C	C	C	C	C	C	C	C	T	T	T	T
12350	G	G	G	G	G	G	G	G	G	G	A	A	A	A
12518	T	T	T	T	T	T	T	T	T	T	C	C	C	C
12656	G	G	G	G	G	G	G	G	G	G	A	A	A	A
12728	A	A	A	A	A	A	A	A	A	A	G	G	G	G
12879	G	G	G	G	G	G	G	G	A	G	A	A	A	A
13151	A	-	A	A	A	A	A	A	A	A	G	G	G	G
13235	C	C	C	C	C	C	C	C	C	C	T	T	T	T
13259	C	C	C	C	C	C	C	C	C	C	T	T	T	T
13283	A	A	A	A	A	A	A	A	A	A	C	C	C	C
13514	A	A	A	A	A	A	A	A	A	A	G	G	G	G
13695	G	A	G	G	G	G	G	G	G	G	G	G	G	G
13706	G	T	G	G	G	G	G	G	G	G	G	G	G	G
13838	A	A	A	A	A	A	A	A	A	A	G	G	G	G
13937	A	A	A	A	A	A	A	A	A	A	G	G	G	G
13997	A	A	A	A	A	A	A	A	A	A	G	G	G	G
14258	G	G	G	G	G	G	G	G	G	G	A	A	A	A
14321	C	C	C	C	C	C	C	C	C	C	T	T	T	T
14339	G	G	T	T	T	T	T	T	T	T	T	T	T	T
14417	T	T	T	T	T	T	T	T	T	T	C	C	C	C
14582	C	C	C	C	C	C	C	C	C	C	T	T	T	T
14615	C	C	C	A	C	C	C	C	C	C	C	C	C	C
14657	G	G	G	G	G	G	G	G	G	G	A	A	A	A
14689	C	C	C	A	C	C	C	C	C	C	C	C	C	C
14762	C	C	C	C	C	C	C	C	C	C	T	T	T	T

Position	K1	K31	K28	Swed	Mork1	K42	K36	WGreen	LoirV	Alm	Nfld	UngBay	ConRiv	NB
14830	A	A	G	A	A	A	A	A	A	A	A	A	A	A
14848	A	A	A	A	A	A	A	A	G	A	A	A	A	A
14882	G	G	G	G	G	G	G	G	G	G	G	A	G	G
14906	G	G	A	A	A	A	A	A	A	A	A	A	A	A
14924	T	T	T	T	T	T	T	T	C	T	T	T	T	T
14933	C	C	C	C	G	C	C	C	C	C	C	C	C	C
14934	G	G	G	G	G	G	A	G	G	G	G	G	G	G
14966	C	C	C	C	C	C	C	C	C	C	T	T	T	T
15029	C	C	C	C	C	C	C	C	C	C	T	T	T	T
15032	A	A	A	C	A	A	A	A	A	A	A	A	A	A
15229	A	A	A	A	A	A	A	G	A	A	A	A	A	A
15422	G	G	G	G	G	G	G	A	G	G	G	G	G	G
15497	C	C	C	C	C	C	C	C	T	C	C	C	C	C
15503	C	C	C	C	C	C	C	C	C	C	A	A	A	A
15527	A	A	A	A	A	A	A	A	A	A	A	A	G	A
15560	C	C	C	C	C	C	C	C	C	C	T	T	C	C
15836	A	A	A	A	A	A	A	G	A	A	A	A	A	A
15953	A	A	A	A	A	A	A	A	A	A	G	G	A	A
16178	C	C	C	C	C	C	C	C	C	C	A	A	A	C
16190	A	A	A	A	A	A	G	A	A	G	A	A	A	A
16292	G	G	G	G	G	G	G	G	G	G	C	C	C	G
16306	T	G	T	T	T	T	T	T	T	T	T	T	T	T
16340	G	A	G	G	G	G	G	G	G	G	G	G	G	G
16373	G	G	G	A	G	G	A	G	G	G	A	A	A	A
16433	A	A	A	A	A	A	A	A	A	A	G	G	C	C
16442	C	C	C	C	C	C	C	C	C	C	T	T	C	C
16551	G	G	G	G	G	G	G	G	G	G	A	A	A	A
16561	G	G	G	G	G	G	G	A	G	G	G	G	G	G
Total Mis-Matched Residues					**215**									

NOTE: Position # relative to the beginning of D-Loop (see Figure 4).

Figure 4 **Atlantic salmon mtDNA consensus sequence from the multiple alignment of fourteen individuals**

Note: Base #1 corresponds to the beginning of the D-loop. The locations of genes on the "+" strand are shown above the sequence, with the beginning and ends marked in yellow. The locations of genes on the "-" strand are shown below the sequence, with the beginning and ends marked in blue. The areas of overlap in genes are marked in green. Start codons (ATG and GTG) are marked in bold, and stop codons are indicated by italics. All tRNAs are underlined.

```
      D-loop →
   1  ACGTTTCAGCTATGTACAACAATAAATGTTATATCTAGCTAACCCAATGTTATACTACATCTATGTATAACATTACATATTATGTATTTACCCATATATA
 101  TAATATCGCATGTGAGTAGTACATTATATGTATTATCAACATAAGTGGATTTAACCCCTCATACATCAGCACTAATCCAAGGTTTACATAAAGCAAAACA
 201  CGTGATAATAACCAACTAAGTTGTTTTAACCCGATTAATTGCTATATCAATAAAACTCCAACTAACACGGGCTCCGTCTTTACCCACCAACTATTTAGCA
 301  TCAGTCCTATTTAATGTAGTAAGAACCGACCAACGATTTAATAGTAGGCATACTCTTATTGATGGTCAGGGACAGATATCGTATTAGGTCGCATCTCGTG
 401  AACTATTCCTGGCATTTGGTTCCTATATCAAGGGCTATCCTTAAGAAACCACCCCCTGAAAGCCGAATGTAATGCATCTGGTTAATGGTGTCAACCTTAT
 501  TGCTCGTTACCCACCAAGCCGGGCGTTCTCTTATATGCATAGGGTTCTCTTTTTTTTTTTTTCCTTTCAGCTTGCATATACAAGTGCAAGCAAAGAAGT
 601  CTAACAAGGTCGAACTAGATCTTGAATTCCAGAGAACCCATGTATCATGGTGGAAGATATTCTATAAAGAATCACATACTTGGATATCAAGTGCATAAGG
 701  TTGATTTTTTTCTTCATATATATCTAAGATCCCCCCGGCTTCCGCGCGGCAAACCCCCCTACCCCCCTACGCTGAAGGATCCTTATATTCCTGTCAAACC
 801  CCTAAACCAGGAAGTCTCAAATCAGCGCTAATCTTTTTTATATACATTAATAAACTTTTTGCCAATTTTATAGCATTCGGCACCGACTACACTGTCATTG
 901  GTACCACTTTTATAATTAAAGTATACATTAATGAACTTTTCAATAAATTTATAGCATCTAGCACTAACTAGTCACTAGTACCCGCTCAGTTAAATATATA
        --|tRNA-Phe →                                                          --|12S-rRNA→
1001  AAGGCCTAGTTGGCGTAGCTTAACTAAAGCATAACACTGAAGCTGTTAAGATGGACCCTAGAAAGTCCCGCGAACACAAAGGCTTGGTCCTGACTTTACT
1101  ATCAGCTCTAACTGAACTTACACATGCAAGTCTCCGCATTCCTGTGAGGATGCCCTTAATCCCCTGCCCGGGGACGAGGAGCCGGCATCAGGCACGCCCA
1201  AGCAGCCCAAGACGCCTTGCTAAGCCACACCCCCAAGGAAACCCAGCAGTGATAAATATTAAGCCATAAGCGAAAGCTTGACTTAGTTAAAGTTAAGAGG
1301  GCCGGTAAAACTCGTGCCAGCCACCGCGGTTATACGAGAGGCCCTAGTTGATAACTACCGGCGTAAAGAGTGGTTACGGAAAAATATTTAATAAAGCCGA
1401  ACACCCCCTCAGCCGTCATACGCACCTGGGGGCACGAAGATCTACTACGAAAGCAGCTTTAATTATACCTGAACCCACGACAGCTACGACACAAACTGGG
1501  ATTAGATACCCCACTATGCCTAGCCGTAAACTTTGATGGAAACATACAACTGACATCCGCCAGGGGACTATAAGCGCCAGCTTAAAACCCAAAGGACTTG
1601  GCGGTGCCTCAGACCCACCTAGAGGAGCCTGTTCTAGAACCGATAACCCCCGTTCAACCTCACCACCTCTTGTTTTCCCCGCCTATATACCACCGTCGTC
1701  AGCTTACCCTGTGAAGGCCTTATAGTAAGCAAAATGGGCAAAACCCAAAACGTCAGGTCGAGGTGTAGCGCATGGGGTGGGAAGAAATGGGCTACATTCT
1801  CTAAATTAGAGCACTACGAACCACGCTGTGAAACCAGCGTCCAAAGGTGGATTTAGCAGTAAATAGAAAATAGAGAGTTCTCTTGAAACTGGCTCTGAGG
1901  CGCGCACACACCGCCCGTCACTCTCCCCAAGTTCAATTAACCCTTCTAACTAAGAAATTAACCGAACAAAGGGGAGGCAAGTCGTAACATGGTAAGTGTA
                        --|tRNA-Val →                                                          --|16S-
2001  CCGGAAGGTGCACTTGGAATAACCAGAGTGTAGCTAAAATAGGAAAGCACCTCCCTTACACCGAGAAGACATCCGTGCAAATCGGGTCACCCTGAGCTGA
      rRNA →
2101  CTAGCTAGCCAACACACTTGGTCTAACACCACAACATACATACCCCTATAAAACTTAAAACTAAGTCAACAAACCATTTTTCCACCTTAGTACGGGCGAC
```


--|tRNA-Leu
--|ND1
*|
tRNA-Ile →
--|
tRNA-MeT→
|--
← tRNA-Gln


--| ND2 →
5001 GGCCCATACCCCGAATATGTTGGTTAAAATCCTTCCCTCACTAATGAACCCCTACGTACTTACCATCTTACTTTCTAGCTTAGGCCTAGGCACAATCCTC
5101 ACCTTCGCCAGCTCCCACTGACTCCTTGCATGAATAGGCCTAGAAATCAATACTCTCGCTATTATCCCAATCATAGCACAACAATCCCACCCACGAGCAA
5201 TTGAAGCTACAACCAAATATTTTTTAACACAAGCAACAGCCGCAGCAATAATCCTCTTCGCCAGCACTACCAACGCCTGACTAGTAGGAGAATGAGAAAT
5301 CCACCAGCTATCCCACCCTCTAGCAACCACCACAGTTATACTAGCCCTTGCACTTAAACTTGGCCTAGCACCCGTTCACTTCTGATTACCAGAAGTCCTT
5401 CAAGGACTTGAACTCACCACAGGACTAATCCTCTCAACTTGACAAAAGCTCGCACCTTTCGCACTTATAATTCAAGTAGCCCCAACTATCAACTCTTCTC
5501 TACTCATTGCGATAGGCCTTCTATCAACACTTGTAGGAGGCTGAGGAGGACTCAATCAAACCCAACTACGTAAAATCCTAGCTTATTCTTCAATCGCCCA
5601 CTTAGGGTGGATAGTACTAATTTTACAATACGCACCCTCCCTAACACTCCTCAGTCTCTTTCTCTACATCATCATAACATCCTCAGCATTCCTCACACTA
5701 AAAACCAACAATTCCCTCACCATTAATACTCTAGCGACTTCATGAACTAAATCCCCAACTCTTGCCGCATTAACCGCTCTCGTATTACTATCCCTAGGAG
5801 GCCTCCCCCCTCTCTCAGGCTTTATACCTAAATGACTTATTTTACAAGAACTCACAAAACAGGGACTCCCACTATCTGCCACACTAGCTGCTATAACAGC
5901 CCTACTCAGTCTTTACTTTTACCTACGCCTCTGTTATGCCATAACCCTCACTATCTACCCCAACACTCTAACTGCCACCGCCCCCTGACGCCTCAACTTC
*| tRNA-
6001 ACCCTCATTACCCTCCCTCTTTCGATCATTACTATTTTAGCCCTAGGCTTGCTTCCCCTCACTCCAGCTGTGACCACGTTACTAACCTTATaaCAAGGGC
-Trp →
--|
6101 TTAGGATAGCACTAAGACCAAGAGCCTTCAAAGCTCTAAGCGGGAGTGAAAATCTCCCAGCCCTTGTTTAAGACCTGCAGGACTTTATCCCACATCTTCT
|--
6201 GAATGCAACCCAGACACTTTAATTAAGCTAAAGCCTTTCTAGGTGGGAAGGCCTCGATCCTACAAACTCTTAGTTAACAGCTAAGCGCTCTATCCAGCGA
← tRNA Ala |--
6301 GCATCCATCTACTTTCCCCCGCCACCGGGGTGGCGAGGCGGGGGAAAGCCCCGGCAGGCTGTTAGCCTACTTCTTTAGATTTGCAATCTAACATGTGGTA
←tRNA-Asn||-Replication origin of L-strand-||--
COX1 →
6401 CACCACAGAGCTTGATAAGGAGAGGACTTAAACCTCTGTCCATGGAGCTACAATCCACCGCTTAGGCGCTCAGCCACCCTACCTGTGGCAATCACACGAT
← tRNA-Cys|-- ← tRNA-Tyr
6501 GATTTTCTCAACCAACCACAAAGACATTGGCACCCTCTATTTAGTATTTGGTGCCTGAGCCGGAATAGTCGGCACCGCCCTAAGTCTCTTGATTCGAGC
6601 AGAACTCAGCCAGCCTGGCGCCCTTCTGGGAGATGACCAAATTTATAACGTAATTGTTACAGCCCATGCCTTCGTCATAATTTTCTTTATAGTCATACCG
6701 ATTATGATCGGCGGCTTTGGAAACTGATTAATTCCTCTTATAATCGGGGCCCCCGACATAGCATTCCCCCGAATGAATAACATAAGTTTTTGACTTCTCC
6801 CTCCCTCCTTTCTTCTCCTCCTGGCCTCATCTGGAGTTGAAGCCGGCGCTGGCACCGGATGAACAGTCTACCCCCCTCTAGCAGGTAATCTTGCCCACGC
6901 AGGAGCTTCCGTTGACTTAACTATTTTTTCCCTCCATTTGGCTGGTATTTCTTCAATTCTTGGGGCCATTAATTTTATTACAACCATTATTAATATAAAA
7001 CCCCCAGCTATCTCTCAGTATCAAACCCCACTTTTTGTTTGAGCTGTATTAGTCACTGCCGTCCTTTTGTTACTCTCCCTCCCTGTTCTAGCAGCAGGCA
7101 TTACCATACTACTTACAGACCGAAATCTAAATACCACTTTCTTTGACCCGGCAGGCGGAGGAGACCCAATCTTGTACCAACATCTCTTTTGGTTCTTTGG
7201 CCATCCAGAAGTCTATATTCTCATTCTCCCAGGCTTTGGTATAATTTCACACATCGTTGCATACTACTCTGGCAAAAAAGAACCTTTCGGGTACATAGGA
7301 ATAGTCTGAGCTATGATAGCCATCGGACTCTTAGGTTTTATCGTTTGAGCCCACCATATGTTTACTGTCGGGATAGATGTAGACACTCGTGCCTACTTCA
7401 CATCTGCCACCATAATCATCGCCATCCCAACTGGAGTAAAAGTGTTTAGTTGACTAGCCACACTGCACGGCGGCTCAATCAAATGAGAAACGCCACTTCT


tRNA-Asp →
tRNA-
-Ser
COX2 →
tRNA-Lys →
ATP8 →
ATP6 →
COX3 →


*|tRNA-Gly
--|ND3 →
--|tRNA-ArG→
--|ND4L →
ND4→ *|
*|tRNA-His →


--|tRNA-Ser →
--| tRNA-Leu→
12801 TTAAAACCCTCTTATCCACCGAGAGTTATCTGTTGATGATAGAGACTGCTAATCTTCTACCCCCTCGGTTAAATTCCGGGGTTCACTCGTGCTTCTAAAG
--|ND5 →
12901 GATAACAGCTCATCCATTGGTCTTAGGAACCAAAAACTCTTGGTGCAAATCCAAGTAGCAGCTATGCACCCTGACTACACTCATCCTAAGCTCAACCCTT
13001 TTAATAAGTCTTCGCACTTCTCCTCTATCCTCTTATTACTACCCTTAACCCAACCCCCCAACAAGAAAACTGAGCCCTCACTCACGTAAAAACTGCTATC
13101 AAAATAGCTTTCTTAGTAAGTCTACTCCCCCTTTTTATTTTTCTAGACCAAGGGACCGAAACAATTGTTACTAATTGACAATGAATAAATACTACAACCT
13201 TCGACATTAACCTCAGCTTTAAATTTGACCACTACTCCATCATTTTTACCCCAATTGCCCTATACGTAACTTGATCTATTCTAGAATTCGCATCCTGATA
13301 CATGCATGCCGACCCCAATATGAACCGATTCTTTAAATACCTCCTCCTCTTCTTAATTGCCATAATTATCTTAGTAACCGCCAACAACATGTTCCAACTA
13401 TTTATTGGCTGAGAAGGTGTTGGCATTATATCGTTCCTCCTCATCGGGTGATGGTACGGCCGAGCTGACGCTAACACAGCTGCCATACAAGCTGTAATTT
13501 ATAACCGAGTCGGAGATATCGGACTTATCTTAAGCATAGCCTGATTCGCAACAAACCTGAACTCCTGAGAAATTCAACAAATATTTGCCTCCTCTAAAGA
13601 ACTCGACCTCACACTCCCCCTCATAGGCCTCATCCTAGCCGCCACCGGTAAATCAGCACAATTTGGACTTCATCCCTGACTTCCTTCAGCGATAGAAGGT
13701 CCTACGCCGGTATCTGCCCTACTTCATTCTAGCACCATAGTAGTTGCGGGTATCTTCCTACTAATTCGACTCCACCCCCTTATAGAAAATAACCAAACAG
13801 CCCTAACTACCTGCTTATGCCTAGGAGCACTAACCACACTATTCACCGCTACTTGTGCCTTAACACAAAATGATATCAAAAAAATCGTTGCATTCTCTAC
13901 ATCCAGTCAACTAGGACTGATAATAGTCACAATCGGACTTAACCAACCACAACTAGCCTTCCTCCATATCTGCACCCACGCATTCTTTAAAGCCATACTC
14001 TTCCTGTGCTCCGGATCAATTATTCATAGCTTAAACGATGAACAAGACATTCGAAAAATAGGAGGCATACATAACCTCACTCCTTTCACTTCCTCCTGTC
14101 TTACAATTGGAAGCCTAGCACTTACTGGCACTCCATTCTTAGCGGGATTCTTTTCCAAAGATGCTATCATTGAAGCCTTAAATACCTCTCACCTTAACGC
14201 CTGAGCCCTTACTCTTACCTTACTAGCTACCTCCTTCACTGCCGTCTATAGCCTCCGGGTCGTCTTTTTCGTCTCGATAGGACACCCCCGCTTTACAGCT
14301 ACAGCCCCTATTAATGAAAACAACCCCTCCGTTATTAATCCCATCAAACGACTAGCTTGAGGAAGCATTATTGCAGGACTTCTAATCACCTCAAACTTTC
14401 TACCCTCCAAAACCCCTATCATAACCATACCCCTCCCATTAAAATTAGCCGCCCTCCTAGTTACCATCTCAGGCCTTCTCATCGCACTAGAACTTGCATC
14501 CCTAACTAATAAACAATTCAAAACTACACCTAATCTCATTACCCACAACTTCTCCAACATGCTAGGATTCTTCCCAGCCATCATCCACCGATTAGCCCCC
14601 AAACTAAACTTGACCCTAGGACAAACCATTGCCAGCCAAATAGTAGATCAAACATGGTTTGAAAAAATCGGCCCAAAGGGAGTTGTATCCACTCACCTAC
14701 CTATGGTCACAACAACAAGCAATATCCAACAAGGCATAATTAAAACATACCTCACTCTATTCTTCCTTTCAACAGCCCTAGCCGTTCTACTCACACTAAC
*|
14801 CtaaACTGCTCGAAGCGCTCCCCGACTCAACCCCCGTGTTAATTCCAATACCACAAAAAGTGTTAACAAAAGCACTCACGCGCACACAATTAATATTCCA
|*
14901 CCCCCATAAGAGTACATTAAAGCTACCCCGCTCGTATCGCCTCGTAAAACAGAAAACTCCTTAAACTCATCCACTACTACCCACGAGGTTTCGTATCACC
15001 CACTTCAAAATCAACCCGCCACCAACATCACACCCACCGTATACACCACTACATACCCTAAAACCGAACGATCCCCCCAAGACTCTGGAAAAGGTTCGGC
15101 AGCCAAAGCCGCTGAATAAGCAAACACTACCAGCATCCCCCCCAAATAAATTAAAAATAATACCAAAGACAAGAAAGACCCCCCGTAACCCACCAAAACT
15201 CCGCAACCTACACCTGCTGCCACAACCAAACCCAAAGCAGCAAAGTAAGGCGCAGGGTTAGATGCAACAGCCACAAGCCCTAAAACCAACCCTAAAAGAA
tRNA-Glu →
Cytb→
15301 ATAAAGACACAATATAAGTCATTAATTCCTGCTCGGACTTTAACCGAAACTAATGACTTGAAAAACCACCGTTGTTATTCAACTACAAGAACCTTAATGGC
← ND6
15401 CAACCTCCGAAAAACTCACCCGCTCCTAAAAATTGCTAATGACGCACTAGTCGATCTCCCAGCACCATCTAACATCTCAGTTTGATGAAACTTTGGCTCA
15501 CTCTTAGGCCTATGTCTAGCCACCCAAATCCTTACCGGGCTCTTCCTAGCCATACACTACACCTCCGATATCTCAACAGCTTTTTCCTCTGTTTGCCACA

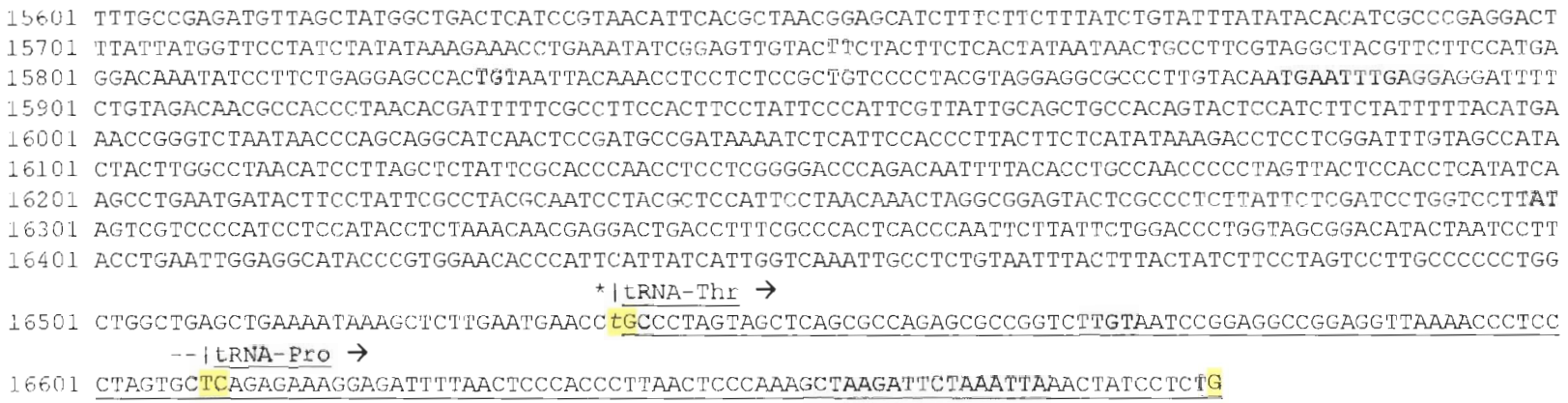

15601 TTTGCCGAGATGTTAGCTATGGCTGACTCATCCGTAACATTCACGCTAACGGAGCATCTTTCTTCTTTATCTGTATTTATATACACATCGCCCGAGGACT
15701 TTATTATGGTTCCTATCTATATAAAGAAACCTGAAATATCGGAGTTGTACTTCTACTTCTCACTATAATAACTGCCTTCGTAGGCTACGTTCTTCCATGA
15801 GGACAAATATCCTTCTGAGGAGCCACTGTAATTACAAACCTCCTCTCCGCTGTCCCCTACGTAGGAGGCGCCCTTGTACAATGAATTTGAGGAGGATTTT
15901 CTGTAGACAACGCCACCCTAACACGATTTTTCGCCTTCCACTTCCTATTCCCATTCGTTATTGCAGCTGCCACAGTACTCCATCTTCTATTTTTACATGA
16001 AACCGGGTCTAATAACCCAGCAGGCATCAACTCCGATGCCGATAAAATCTCATTCCACCCTTACTTCTCATATAAAGACCTCCTCGGATTTGTAGCCATA
16101 CTACTTGGCCTAACATCCTTAGCTCTATTCGCACCCAACCTCCTCGGGGACCCAGACAATTTTACACCTGCCAACCCCCTAGTTACTCCACCTCATATCA
16201 AGCCTGAATGATACTTCCTATTCGCCTACGCAATCCTACGCTCCATTCCTAACAAACTAGGCGGAGTACTCGCCCTCTTATTCTCGATCCTGGTCCTTAT
16301 AGTCGTCCCCATCCTCCATACCTCTAAACAACGAGGACTGACCTTTCGCCCACTCACCCAATTCTTATTCTGGACCCTGGTAGCGGACATACTAATCCTT
16401 ACCTGAATTGGAGGCATACCCGTGGAACACCCATTCATTATCATTGGTCAAATTGCCTCTGTAATTTACTTTACTATCTTCCTAGTCCTTGCCCCCCTGG
*|tRNA-Thr →
16501 CTGGCTGAGCTGAAAATAAAGCTCTTGAATGAACCtGCCCTAGTAGCTCAGCGCCAGAGCGCCGGTCTTGTAATCCGGAGGCCGGAGGTTAAAACCCTCC
--|tRNA-Pro →
16601 CTAGTGCTCAGAGAAAGGAGATTTTAACTCCCACCCTTAACTCCCAAAGCTAAGATTCTAAATTAAACTATCCTCTG

Table 5 **Pair-wise sequence differences in the mtDNAs from fourteen Atlantic salmon and brown trout.**

(Above diagonal: Transition/Transversion; Below diagonal: total number of nucleotide differences)

	K1	K28	K31	K36	K42	Swed	Mork	WGre	LoirV	Alm	Nfld	UngBay	ConRiv	NB	Brown T
K1	----	6/2	9/9	11/2	6/3	6/5	8/3	27/3	27/5	19/2	107/18	111/17	106/21	91/18	754/134
K28	8	----	12/9	11/0	6/1	4/3	6/1	27/1	27/3	19/0	107/16	111/17	106/19	91/16	758/132
K31	18	21	----	17/9	12/10	12/12	14/10	31/10	31/12	25/8	109/25	112/26	108/28	93/25	722/136
K36	13	11	26	----	9/1	9/3	13/1	30/1	30/3	19/0	106/16	110/17	105/19	90/16	757/132
K42	9	7	22	10	----	6/4	8/2	25/2	25/4	19/1	104/17	108/18	103/20	88/17	758/133
Swed	11	7	24	12	10	----	6/4	27/4	27/6	19/3	105/19	109/20	104/22	89/19	759/133
Mork	11	7	24	14	10	10	----	27/2	27/4	19/1	105/17	109/18	104/20	87/17	757/133
WGre	30	28	41	31	27	31	29	----	44/4	38/1	122/17	126/18	121/20	107/17	768/133
LoirV	32	30	43	33	29	33	31	48	----	38/3	114/17	118/18	115/20	101/17	757/131
Alm	21	19	33	19	20	22	20	39	41	----	95/9	99/10	93/12	90/9	740/127
Nfld	125	123	134	122	121	124	122	139	131	104	----	6/1	24/3	34/6	756/131
UngBay	130	128	138	127	126	129	127	144	136	109	7	----	26/4	36/7	759/137
ConRiv	127	125	136	124	123	126	124	141	135	105	27	30	----	20/5	764/139
NB	109	107	118	106	105	108	104	124	118	99	40	43	25	----	758/138
Brown T	888	890	858	889	891	892	890	901	888	867	887	896	903	896	----

3.3.2 Mode of evolution

There are two main types of nucleotide substitutions: transitions and transversions. Transition substitutions are purine to purine or pyrimidine to pyrimidine mutations (A <->G or C <-> T). This type of substitution is generally less detrimental because it rarely causes a change at the amino acid level at third positions of codons. Transversion substitutions are purine to pyrimidine or pyrimidine to purine nucleotide changes that more often alter the amino acid sequence. Between the Atlantic salmon sequences, the transition substitutions were more prominent than transversions (Table 5, upper diagonal). The average ratio of transitions to transversions is 7.3.

An in-depth examination of the 215 nucleotide variable sites in the genome showed that 14 sites were located in tRNA genes, 178 in protein coding genes, and 23 were found in the D-loop. In the thirteen protein genes, the majority of the substitutions were located at the third positions of a codon, and therefore did not translate into amino acid changes. Surveying the whole genome, only fourteen non-synonymous nucleotide substitutions were observed in all pairwise comparisons, whereas 149 synonymous nucleotide substitutions were found (Table 6).

Table 6 **The comparison of how susceptible different regions of the Atlantic salmon mitochondrial genomes are in accepting nucleotide substitutions.**

Also examined are the positions of codons and the number of amino acid changes. The genome is divided into 16 segments, thirteen protein genes, two rRNA genes and the D-loop.

Gene	Size (bp)	# of Trans-itions	# of Trans-versions	Variable sites	# of Syn. Codon change	# of Non-Syn. Codon Change	Position		
							1	2	3
ND6	522	9	2	11	10	1	3	1	7
ND2	1050	20	2	22	19	3	2	1	19
D loop	1006	14	7	21	---	---	---	---	---
ND1	975	18	0	18	16	2	2	0	16
ND4	1380	25	0	25	23	2	3	0	22
ATP6	684	9	3	12	11	1	0	1	11
ND4L	297	4	1	5	5	0	0	0	5
ND3	348	5	0	5	5	0	1	0	4
CYTB	1140	11	4	15	15	0	0	0	15
COX1	1551	16	2	18	16	2	1	1	16
ND5	1839	14	5	19	17	2	1	1	17
COX2	690	7	0	7	6	1	1	0	6
12S	947	8	1	9	---	---	---	---	---
COX3	786	5	1	6	6	0	0	0	6
16S	1680	4	3	7	---	---	---	---	---
ATP8	168	0	0	0	0	0	0	0	0
Total	**15063**	**169**	**31**	**200**	**149**	**14**	**14**	**5**	**144**

It was expected from the wobble hypothesis of the mitochondrial genetic code that most third position codon substitutions will not affect the protein sequence since the tRNA anticodons are able to recognize all four nucleotides at the third position of most codons. Therefore, the vast majority of mutations at this position will not translate into a change at the amino acid level. Fourteen substitutions occurred at the first codon position, four of which did not result in a change of amino acid (all giving leucine). The remaining substitutions resulted in four conservative replacements and six non-conservative replacements. Five mutations occurred at the second codon position, resulting in three conservative non-synonymous amino acid changes, and one non-conservative amino acid replacement, one mutation did not cause a change in amino acid (Table 7).

Table 7 **Location and details of the amino acid changes observed in Atlantic salmon mtDNA.**

Replacement types abbreviations are UP = uncharged polar amino acid, PCP = positively charged polar amino acid, NCP = Negatively charged polar amino acid, NP = non-polar amino acid.

Gene	Position in codon	Replacement Type	Conservative Replacement	Amino Acid (Before)	Amino Acid (After)
ND1	1	NP -> UP	No	Alanine	Threonine
ND1	1	NP -> NP	Yes	Valine	Isoleucine
ND2	1	NCP -> UP	No	Aspartic acid	Threonine
ND2	1	NP -> UP	No	Alanine	Threonine
COXI	1	NP -> NP	Yes	Valine	Isoleucine
COX2	1	NP -> NP	Yes	Valine	Isoleucine
ND4	1	NP -> UP	No	Glycine	Serine
ND4	1	NCP -> UP	No	Aspartic acid	Asparagine
ND5	1	NCP -> PCP	No	Glutamic acid	Lysine
ND6	1	NP -> NP	Yes	Tryptophan	Methionine
ND2	2	NP -> UP	No	Leucine	Serine
COX1	2	UP -> UP	Yes	Serine	Tyrosine
ATP6	2	NP -> NP	Yes	Valine	Leucine
ND5	2	UP -> UP	Yes	Serine	Tyrosine

3.3.3 Tempo of evolution

Previous work on the mtDNA genome in fish had shown that not all regions of the mtDNA genome accumulate substitutions at the same rate (Shedlock *et al.*, 1992). Regions that accumulate more substitutions, especially substitutions that cause amino acid changes, are said to evolve relatively faster than the rest of the genome. A pairwise comparison was made between one sample from Sweden and one from Newfoundland (Table 8). Various genomic regions were examined for the amount and the type of substitution. The regions include the thirteen protein coding genes, two ribosomal RNA genes, and the D-loop. Interestingly, the D-loop possesses the highest amount of sequence variation, and evolves three times faster than the cytochrome *b* region. The gene coding for ATPase 8 showed no variation between the two individuals, indicative of its important function in ATP production.

Four of the sixteen genomic regions evolve relatively slower than the cytochrome *b* gene; they are *ATP6*, *ND6*, *16S* rRNA and *ATP8*. There was no correlation found between sequence divergence rate and the fragment size.

Table 8 **Tempo of evolution in different regions of the Atlantic salmon mitochondrial genome.**

Comparison made between one European individual from Sweden and one North American individual from Newfoundland. (Ts = Transition nucleotide substitution; Tv = Transversion nucleotide substitution; Syn. = synonymous; Non-syn. = non-synonymous)

Genomic Region	Size (bp)	Total # of Nt. Sub.	# of Ts.	# of Tv.	% Seq. Diff.	# of Indels	Codon Change		Codon Position			Relative Rate compared to cytochrome b
							Syn.	Non-Syn.	1	2	3	
D loop	1008	16	9	7	1.79	2	--	--	--	--	--	2.93
ND4	1380	18	18	0	1.30	0	18	0	1	0	17	2.13
ND4L	297	3	3	0	1.01	0	3	0	0	0	3	1.66
ND5	1839	16	13	3	0.87	0	15	1	0	1	15	1.43
COX2	690	6	6	0	0.87	0	5	1	1	0	5	1.43
ND3	348	3	3	0	0.86	0	3	0	0	0	3	1.41
ND1	975	8	8	0	0.82	0	7	1	1	0	7	1.34
COX1	1551	12	11	1	0.77	0	12	0	0	0	12	1.26
COX3	786	6	5	1	0.76	0	6	0	0	0	6	1.25
ND2	1050	7	6	1	0.67	0	5	2	2	0	5	1.10
12S	947	6	6	0	0.63	0	--	--	--	--	--	1.03
CYTB	**1140**	**7**	**4**	**3**	**0.61**	**0**	**7**	**0**	**0**	**0**	**7**	**1.00**
ATP6	684	4	4	0	0.58	0	4	0	0	0	4	0.95
ND6	522	3	2	1	0.57	0	3	0	0	0	3	0.94
16S	1680	4	2	2	0.30	1	--	--	--	--	--	0.49
ATP8	168	0	0	0	0.00	0	0	0	0	0	0	0.00

3.4 Sequence variation in the genus *Salmo*

3.4.1 Amount of sequence variation

Salmo trutta (brown trout) is a sister species of the Atlantic salmon (*Salmo salar*). The sequence of the mitochondrial genome of a Newfoundland brown trout was compared against that of an Atlantic salmon from Newfoundland. By selecting both fish from the same geographical location, environmental factors were minimized. The mitochondrial genome size of the brown trout is 5 bp longer than that of Atlantic salmon. There are 8 bp of insertion in the brown trout (7 bp in the D-loop, 1 bp in the 16S rRNA) and 3 bp in the Atlantic salmon (1 bp in the D-Loop and 2 bp in the16S rRNA). Overall, there are 887 variable sites between the two genomes. The percent sequence difference increased five fold (5.3%) when the mtDNAs of *Salmo salar* and *Salmo trutta* were compared relative to within *Salmo salar* (0.04% - 0.90%). The number and the locations of the variable sites are listed in Table 9.

Table 9 **Percent sequence divergence between *Salmo salar* and *Salmo trutta* for thirteen mitochondrial protein-coding genes, two ribosomal RNA genes and D-loop.**

Gene	Size (bp)	# nt. diff.	% Seq Diff.	# Ts	#Tv	# of indels	Codon Change		Position			Rate of Evolution Relative to Cyt *b*
							Syn.	Non-syn.	1	2	3	
ND3	348	32	9.20%	27	5	0	29	3	4	0	28	1.46
ND6	522	43	8.26%	34	9	0	39	4	6	1	36	1.30
ND2	1050	80	7.62%	72	8	0	68	12	15	3	62	1.21
ND5	1839	130	7.07%	109	21	0	115	15	15	4	111	1.12
ATP6	684	48	7.02%	36	12	0	46	2	4	0	44	1.11
ND1	975	68	6.97%	61	7	0	60	8	7	2	59	1.10
COX1	1551	108	6.96%	98	10	0	105	3	7	0	101	1.10
D-loop	1014	67	6.61%	44	15	8	---	---	--	--	--	1.05
ND4	1380	91	6.59%	86	5	0	86	5	11	1	79	1.04
Cytb	**1140**	**72**	**6.32%**	**54**	**18**	**0**	**71**	**1**	**3**	**0**	**69**	**1.00**
COX3	786	42	5.34%	36	6	0	39	3	6	0	36	0.84
COX2	690	27	3.91%	24	3	0	25	2	2	1	24	0.62
ND4L	297	10	3.37%	10	0	0	9	1	1	1	8	0.53
16S	1683	39	2.32%	30	6	3	---	---	--	--	--	0.37
12S	947	19	2.01%	18	1	0	---	---	--	--	--	0.32
ATP8	168	2	1.20%	2	0	0	0	0	1	0	1	0.19

3.4.2 Mode of evolution

Of the 887 substitutions, 756 were transition-type substitutions and 131 were transversion substitutions. A ratio of one transversion per 5.6 transitions was observed, which is lower than that in the intra-specific comparison. In the inter-specific comparisons, more mutations at the second position of codons were observed (Table 9), and they translated into an increased occurrence of non-synonymous amino acid substitutions than was seen between Atlantic salmon individuals. There are a total of 55 non-synonymous substitutions; an almost three-fold increase from the 19 observed in the intra-specific comparison.

3.4.3 Tempo: relative rates of evolution

The protein-coding genes, the two ribosomal RNA genes and the non-coding control regions from *Salmo salar* (Newfoundland strain) and *Salmo trutta* were compared in order to detect differences in substitution patterns between different regions of the mitochondrial genome that are under different functional constraints (Table 9). As might be expected, the percent differences were highly variable among genes in the *Salmo* mitochondrial genomes, ranging from 1.5% - 9.2%. The rates were subsequently standardized to one gene, cytochrome *b*, to give relative rates of sequence divergence. The cytochrome *b* gene was chosen as the standard as it has been studied extensively in mitochondrial genomes from a wide variety of organisms.

Nine of the sixteen genomic regions in *Salmo* mtDNA evolved faster than cytochrome *b*, with the gene encoding NADH complex 3 (*ND3*) being the fastest and the ATPase 8 gene the slowest (Table 9). There is no apparent relationship between the rate

of evolution and the length of coding region. The two ribosomal RNA genes accept mutations at about one third the rate of the cytochrome *b* gene. It was somewhat surprising to observe that the D-loop was not the fastest evolving region of the mitochondrial genome, as this has been noted for mammalian mtDNAs. Rather, this region evolves at approximately the same rate as the cytochrome *b* gene between two *Salmo* species. A comparison of the genes that are evolving relatively faster than cytochrome *b* in the intra-specific and inter-specific comparisons gave almost the same list of genes suggesting that the pattern of evolution that affects the rate of evolution is affecting the whole genus of *Salmo* (Table 10).

Table 10 **The list of genes that evolve relatively faster and slower than cytochome *b* in the intra-specific and inter-specific comparisons.**

Genes are listed in decreasing order of rates, those underlined were found to differ.

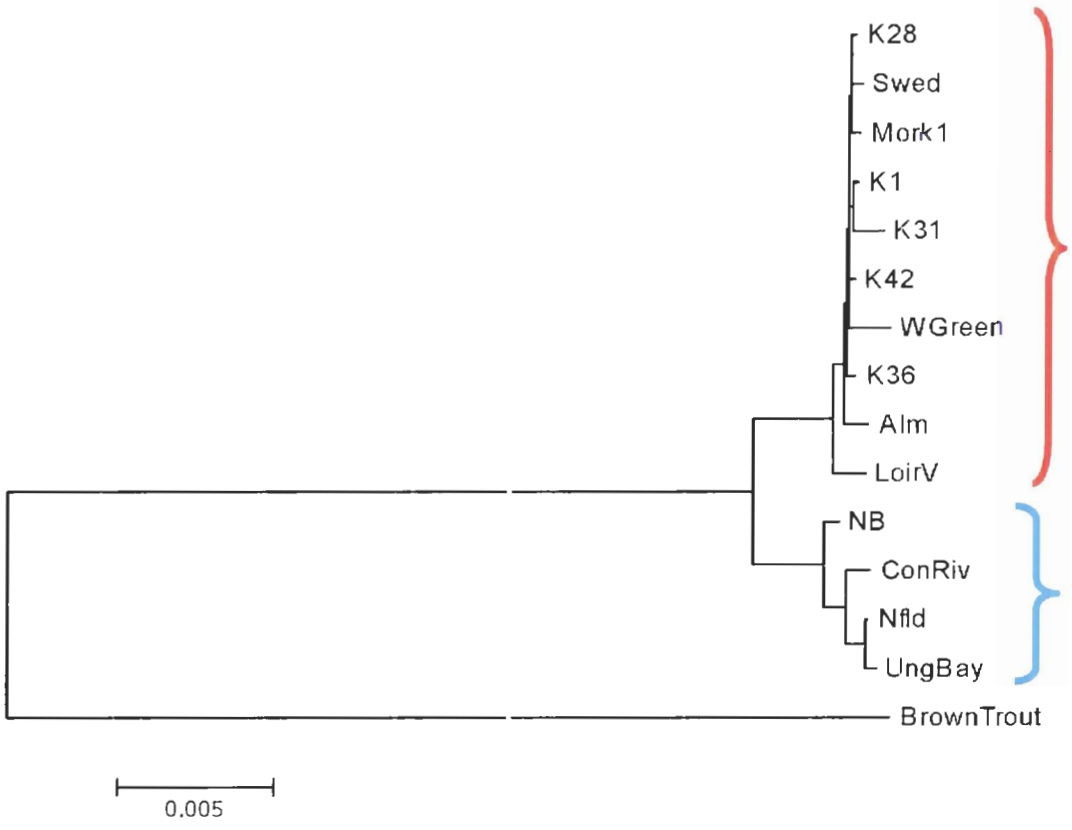
	Intra-specific comparison	Inter-specific comparison
Genes evolving **faster** than cytochrome *b*	*D-loop*, *ND4*, *ND4L*, *ND5*, *COX2*, *ND3*, *ND1*, *COX1*, *COX3*, *ND2*, *12S*	*ND3*, *ND6*, *ND2*, *ND5*, *ATP6*, *ND1*, *COX1*, *D-loop*, *ND4*
Genes evolving **slower** than cytochrome *b*	*ATP6*, *ND6*, *16S*, *ATP8*	*COX2*, *COX3*, *ND4L*, *12S*, *16S*, *ATP8*

3.5 Phylogeography of Atlantic salmon populations

Phylogeography correlates sequence similarity with geographic distribution. A multiple sequence alignment was performed on the fifteen mtDNA sequences to examine the relationships between the Atlantic salmon from various regions of the world. Based on the number of nucleotide differences, a neighbour joining phylogenetic tree was constructed using brown trout as the outgroup. The topology of the tree showed a clear separation of the fourteen samples into two main clusters corresponding to Atlantic salmon originating from Europe and North America. These are the two main haplotypes of Atlantic salmon in the world today.

Figure 4 **Neighbour-joining tree based on fourteen mitochondrial genomes of Atlantic salmon with brown trout as the out-group.**

(Red bracket are samples from Europe, Blue brackets are samples from North America).



CHAPTER 4 - DISCUSSION

Fourteen Atlantic salmon and one brown trout mitochondrial genomes were sequenced and compared against each other to study the variation among these sequences. The goal of sequencing whole mitochondrial genomes of Atlantic salmon from various populations was to gain a better understanding of how mtDNA evolves over a relatively short period of time. The comparison of the amount and distribution of mtDNA variation could also shed light on the evolution of Atlantic salmon populations. Furthermore, it was anticipated that specific markers would be discovered which could be useful for stock identification and help in making management decisions.

4.1 Conserved primers for sequencing whole mtDNA genome

A PCR-based approach for sequencing whole mitochondrial genomes is a more rapid and economical way to obtain mtDNA sequences than the traditional method of cloning mtDNA and using a combination of sequencing the hybrid recombinant clones and primer walking. The ability to establish "universal primers" by Kocher *et al.* in 1989 allow the rapid amplification of specific sequences from a large number of samples, which is beneficial for large scale population studies (Stepien and Kocher, 1997). This technique can be used to accurately determine complete mitochondrial genome sequences that have wide phylogenetic utility such as in population genetics and examining deep phylogenetic relationships (Curole and Kocher, 1999).

PCR-based sequencing has been used to sequence whole mitochondrial genomes from various species such as finches and cuckoos (Sorenson *et al.*, 1999), cockroach and dragonfly (Yamauchi *et al.*, 2004), moth (Coates *et al.*, 2005) and the deep-sea fish, *Gonostoma gracile* (Miya and Nishida, 1999). The current study provided primer sequences that were successful in amplifying mtDNAs from a range of salmonid species: *Salmo salar*, *Salmo trutta*, *Salvelinus alpinus* and *Salvelinus namaycush* (unpublished data). The set of thirty-three primer pairs yields overlapping fragments of mtDNA sequences that, when combined, produce a whole genome sequence. Some of these primers were degenerate primers, which have a number of nucleotides at several positions in the sequence that enable them to anneal to and amplify a variety of related sequences (Avise, 2004). This set of primers will be beneficial for future sequencing of mitochondrial genomes from broader taxonomic categories in Salmoniformes. The similarity in the primer sequences used for amplifying avian mitochondrial genomes and vertebrate mtDNA sequences indicates these primers could have a broad taxonomic utility (Sorenson *et al.*, 1999). The high level of conservation observed in vertebrate mitochondrial genome organization (Lee and Kocher, 1995) also suggests the same trend should be true about the mtDNA sequence. The access to primer sequences for salmonid mitochondrial genomes will facilitate future whole mitochondrial sequencing of other teleost species.

4.2 Advantages of whole mtDNA sequence

The comparison of complete animal mitochondrial sequences was shown to be more informative than shorter sequences of individual genes for discerning evolutionary relationships (Boore *et al.*, 2005). This is because the whole genome sequence can act as

a model for genome evolution, which yields genome-level characters such as the mode and tempo of nucleotide evolution. Previously, short regions of Atlantic salmon mtDNA were examined, of which, cytochrome *b* has been heavily characterized in many salmonids (Mcveigh and Davidson, 1991; Nielsen *et al.*, 1996). A complete mitochondrial genome for Atlantic salmon was published by Hurst *et al.* in 1999 (GenBank Accession No. NC_001960). A combination of restriction enzyme digests, sub-cloning of fragments, sequencing of cloned DNA and PCR amplification was used to obtain the sequence. The comparison of the whole mtDNA sequence from a Newfoundland Atlantic salmon to the published sequence revealed small indels at conserved locations such as in the 16S rRNA gene. This suggested that the published sequence contains sequencing errors. Hence, that sample was not included in the analysis. Furthermore, the examination of sites that were variable amongst the two main genotypes (European and North American), showed that the published sequence was inconsistently identified as either European or North American in origin. These observations indicated that the published sequence might be a hybrid of two genotypes. This would have occurred if mtDNA from different fish samples was used.

The advances in molecular technology made it possible to sequence whole mitochondrial genomes, meaning that more accurate data could be obtained and more sensitive analysis carried out. The range of sequence differences between two whole mitochondrial sequences from Atlantic salmon (0.04% - 0.9%) is similar to that predicted from restriction analysis (0.12%, Gyllensten and Wilson, 1987; 0.2%, Birt *et al.*, 1991). The upper range of whole genome sequence differences between any two individuals is 0.9%, which is also comparable to the 1.0% predicted by (Birt *et al.*, 1991) using

restriction endonucleases. Both experiments sampled a wide range of Atlantic salmon samples from both sides of the Atlantic Ocean. The sample area ranged from Sweden to France in Europe, and throughout Atlantic Canada and West Greenland located on the west side of the Atlantic Ocean. The parallel results demonstrated that a suitable set of restriction enzymes can yield estimates of variation that are comparable to molecular sequence data.

4.3 Characteristics of *Salmo* mtDNA

MtDNA sequence data have inherent advantages over other kinds of systematic data such as morphological data, restriction maps, nuclear data, etc. Each gene, as well as individual sites within a gene, might evolve at a unique rate because of different levels of functional constraint. Slowly evolving genes may be useful for deducing relationships among highly divergent groups while more rapidly evolving areas, such as the vertebrate mtDNA control regions, may be useful for inferring lower-level systematic relationships such as among populations (Faber and Stepien, 1997). The availability of these data has allowed the examination intra-generic phylogenetics at a higher resolution than had been previously possible. At the lowest taxonomic levels such as among populations, molecular studies have been markedly successful (Avise *et al.*, 1987).

MtDNA regions have been well studied in fishes in the past ten years (Martin and Palumbi, 1993; Miya and Nishida, 1999). Base substitutions events were found to occur relatively rapidly compared to changes in gene order. Most studies were done on the cytochrome *b* gene in fishes (e.g., Carr and Marshall, 1991; Carr *et al.*, 1995; Kocher *et al.*, 1989; Meyer and Wilson, 1990; Zhu *et al.*, 1994).

4.3.1 Base composition bias

Base composition bias against guanine was observed in the mtDNA from Atlantic salmon and brown trout. This bias has been previously observed when analyzing whole mitochondrial genomes of a few other salmonid species (rainbow trout, Arctic char, and Pacific salmon). The bias against the use of guanine is particularly noticeable at the third codon positions in most vertebrates: mammals (Irwin *et al.*, 1991), cod (Carr *et al.*, 1995), Atlantic Alcidae (Friesen *et al.*, 1993), coelacanth (Zardoya and Meyer, 1996) and zebrafish (Broughton *et al.*, 2001). There is no explanation yet for the variation in base composition in vertebrate mitochondrial genome except evidence that showed a connection between the strand asymmetric replication mechanism (Clayton, 1982; Shadel and Clayton, 1997) and base composition (Gibson *et al.*, 2005). Mitochondrial genomes have an asymmetric replication mechanism. During replication, the H-strand (heavy with protein coding genes) is replicated first, displacing the parental H-Strand and leaving it single-stranded until replication of the L-strand begins approximately two-thirds of the way around the genome. Several studies have correlated the base composition of a gene on the H-strand with the relative amount of time that the gene is single-stranded and have found a relationship between the two (Bielawski and Gold, 2002; Faith and Pollock, 2003; Reyes *et al.*, 1998; Tanaka and Ozawa, 1994). Others have attributed the base composition bias to the energetics of codon-anticodon interactions (Friesen *et al.*, 1993).

4.3.2 Transition bias

The two types of nucleotide substitutions were not distributed uniformly; a bias for transition substitutions was observed in both intra-specific and inter-specific

comparisons. This tendency was also observed in the comparison of other salmonid mtDNA (Thomas and Beckenbach, 1989).

A transition bias in nucleotide substitution is common in all animal mtDNA studied to date. This phenomenon has been reported in both vertebrate species (Beckenbach *et al.*, 1990; Brown *et al.*, 1982; Irwin *et al.*, 1991) and invertebrate species (Thomas and Wilson, 1991). Two factors have been suggested to contribute to this: (1) most spontaneous mutations are transitional type substitutions (Li *et al.*, 1984); (2) purifying selection in protein-coding genes can tolerate transition substitutions which tend to yield synonymous mutations rather than transversional mutations, that more frequently lead to amino acid changes. Strong purifying selection against transversions have been observed at two-fold degenerate sites in the cytochrome *b* sequences of pocket gophers (Xia *et al.*, 1996).

The ratio of transitions to transversions (ts/tv) also illustrated the bias for transitions. When comparing the nucleotide substitutions in intra-specific and inter-specific comparisons, a higher percent sequence divergence is related to a lower transition to transversion ratio. This occurrence matches with previous observations that closer related species tend to have more transitions than transversions (Blouin *et al.*, 1998). This was observed in the mtDNA evolution study of *Caenorhabditis* where a noticeable drop in ts/tv ratio was observed between intra- to inter-specific comparisons (Thomas and Wilson, 1991).

Caution must be taken when using ts/tv ratio data for phylogenetic inference; a saturation effect of transitions is common for pairs with greater divergence time. It can be observed that the proportion of transitions declines because multiple substitutions

might have occurred at the same site when the divergence time is long (Wilson *et al.*, 1985). This effect would obscure the record of transitions. A linear relationship between the number of transitions versus total number of substitutions would show that multiple substitutions are not present. Since the species within the genus *Salmo* have diverged only recently, the increase in substitutions was not high enough to reach saturation of transition substitutions. Further, a linear relationship showed that multiple hits did not occur (data not shown). In the case of nematodes, Blouin *et al.* (1998) recommended the use of more slowly evolving mtDNA genes to avoid the saturation effect. It has been suggested that mtDNA for phylogenetics should only be used for closely related species. Because taxa often differ in composition bias, few number of sites can be applied to infer relationships at deeper parts of the phylogenetic tree (Hasegawa and Hashimoto, 1993).

4.3.3 Accepted mutations occur more frequently in the third position of codons

The position of nucleotide mutation also differs in intra-specific and inter-specific comparisons. There were only fourteen first position and five second position substitutions when pairwise comparisons of the 14 complete Atlantic salmon sequences were made. However, there was an increase to 76 and 12, first and second positions substitutions, respectively, when comparing Atlantic salmon to brown trout from Newfoundland. The effect is an increase in non-synonymous amino acid changes. A four-fold increase in the number of replacement substitutions was observed between intra- and inter-specific comparisons. In both cases third position transitions have not yet reached saturation. It has been known that third position transitions usually saturate at the level of congeneric species, and at higher taxonomic levels, sometimes even

replacement substitutions begin to encounter saturation (Blouin *et al.*, 1998). These findings echoed the limit of using mtDNA data for inferring deeper phylogenetic relationships.

4.3.4 Relative rate of change in the mitochondrial genome is conserved in the genus *Salmo*

A close examination of the relative rates of evolution in each genomic region of the mitochondrial genome found a consistency in the tempo of evolution at certain regions of the genome. The ATPase 8 and 16S ribosomal RNA genes are consistently evolving the slowest among all the regions. This could reflect the importance of ATP8 in ATP energy production and 16S rRNA in translating the genes in the mitochondrial genome, which prevents them from accepting mutations as quickly. The D-loop and ND3 gene were the fastest evolving regions in intra-specific and inter-specific comparisons, respectively.

In previous Salmonid inter-specific comparisons, the ND3 gene was used for phylogenetic studies in pink, sockeye, and chum salmon (Domanico and Phillips, 1995; Thomas and Beckenbach, 1989) and confirmed its relatively higher rate of evolution. Another region that has been known to evolve relatively faster is *ATP6*. It has been found that *ATP6* evolves faster than the D-loop, at a rate 1.57 times faster in *Salvelinus*, 1.93 times faster in *Oncorhynchus* and 2.46 times faster in *Thymallus* (Froufe *et al.*, 2005). This pattern can also be observed in the current data, but at a lesser extent as *ATP6* is only evolving slightly faster than the D-loop (1.11 times). A reason for the differences could be that the range of relatedness of the species within the different genera is not the same. Among human mitochondrial genes, *ATP6* has been found to

contain the highest number of variable restriction sites (Cann *et al.*, 1984). *ATP6* can be observed as one of the genes that evolve relatively faster than cytochrome *b*, showing that some salmonid mtDNA features can also be found in humans.

Between Atlantic salmon, the genomic region with the overall highest percent sequence difference is D-loop. However, upon a closer examination of the type of substitutions, it was found to possess a series of five transversion substitutions and a 2 bp insertion within a short 15 bp "hot spot" region close to the end of the D-loop which contributes to its faster rate. The rest of the 993 bp D-loop only accumulated nine transitions and two transversions. Also, the three conserved sequence blocks (CSB-1, CSB-2 and CSB-3) are well-conserved among all Atlantic salmon and brown trout. The relative faster rate of evolution was not observed in the D-loop region of brown trout; instead, the D-loop was found to accept mutations at about the same rate as cytochrome *b*. Previous experiments on mtDNA sequence variation in Greek brown trout saw that the D-loop evolves at a lower rate than other regions of the mitochondrial genome (Apostolidis *et al.*, 1997). Other evidence also showed that the D-loop evolves slower in Salmonids than in mammals (Bernatchez and Danzmann, 1993; Shedlock *et al.*, 1992). Since the distribution of mutations is not uniform, it is more likely that the fast rate of D-loop evolution is an independent case. Further examination of the D-loop regions must be performed to identify the cause of the higher accumulation of transversions at the "hot spot".

4.4 Atlantic salmon population genetics

Intra-specific phylogeography uses molecular data to reconstruct population histories in relation to geography. Avise *et al.* (2004) suggest that patterns of concordance between genealogy and geography should reflect historical events that give rise to the current distribution of individuals. Analysis of mtDNA sequence variation has become a well-established tool for studying fish evolution, especially the use of the cytochrome *b* gene (Carr *et al.*, 1995; Krajewski and King, 1996; Mcveigh and Davidson, 1991; Zhu *et al.*, 1994). The relationships between Atlantic salmon populations have been of interest for a long time, mostly because of the economic importance of the species. Further, due to the interesting life cycle of Atlantic salmon, ecological studies on the relatedness of anadromous and non-anadromous populations have also been carried out (Birt *et al.*, 1991). The conservation of Atlantic salmon in various areas with over-fishing is also of interest to environmentalists and fisheries managers who hope to preserve the genetic diversity and increase the survivability of the species as wild populations. Advances in technology greatly transformed the type and amount of information that can be obtained from the mitochondrial genome of this species, and this has contributed to studying the population genetics of Atlantic salmon.

4.4.1 Allozyme data displayed population splits

In the early 1980s, researchers began to apply biochemical genetics to perform population genetics studies (Utter, 1987). Early experiments focused on blood proteins such as hemoglobin and transferrin (Payne *et al.*, 1971), then later transferred to enzymatic proteins, allozymes (Stahl, 1987). Both methods aim to measure protein polymorphisms at different genetic loci between individuals from various populations.

Using a combination of starch gel electrophoresis, and the application of histochemical stains, the enzyme or other protein products can be visualized (Avise, 2004). This method made use of proteins migrating differently under a charged environment when a mutation had altered the protein's overall charge. The frequencies of the alleles at several loci are then used to calculate the genetic distance (Nei, 1978) between groups of individuals. A protein that has been studied heavily is serum transferrin (Payne *et al.*, 1971). Allozyme studies found a consistent differentiation of Atlantic salmon into two groups, one group corresponds to fishes originating from North America, while the second includes fishes of European origin (Verspoor *et al.*, 2005).

The investigation of refugial areas for salmon from Europe and North America led to the hypothesis about the pattern of post-glacial colonization. In parts of Europe, North American-specific variants were detected in Northern Russia (Verspoor *et al.*, 1999) suggesting that there had to be differentiation of Northern Russian populations from those from elsewhere of Europe. These results can be converted into a hypothesis concerning the possible route taken by the fish during post-glacial colonization. These variants also showed the presence of a subdivision of groups within European populations. The Baltic drainage region in particular showed a distinct genotype from the rest of the European genotypes (Koljonen *et al.*, 1999). This was later confirmed by results found by performing restriction enzyme analysis on mtDNA (Bermingham *et al.*, 1991).

4.4.2 Restriction enzyme analysis saw the same split

The first application of restriction endonuclease analysis of mtDNA from Atlantic salmon is a study of allopatric anadromous and non-anadromous populations from

Newfoundland (Birt *et al.*, 1986) Eleven restriction enzymes were used to digest mtDNA from 15 individuals from the two populations. One enzyme, *Xba*I, revealed a polymorphism in one individual. This was the first evidence that variation existed in the mitochondrial genome of Atlantic salmon. Subsequently, the use of a few six base pair recognition enzymes found that *Bgl*II and *BstE*II were able to produce digestion patterns that were different between Atlantic salmon from Newfoundland and Scandinavia. These results suggested that these enzymes would be useful for determining the continent of origin of an Atlantic salmon (Davidson *et al.*, 1989a; Bermingham *et al.*, 1991) employed these enzymes to distinguish Atlantic salmon of North America and European/Icelandic origin. The survey included 328 Atlantic salmon in the West Greenland fishery, and showed consistent results with salmon tagged in their rivers of birth.

Based on the observed number of nucleotide differences, the relationships between the fourteen Atlantic salmon from different populations were constructed. This revealed the separation of the samples into two main genotypes, corresponding to the origin of the fishes, European or North American. By creating a restriction map using enzymes that previously proved successful in discriminating Atlantic salmon populations, it was possible to confirm these previously characterized cut-sites (Figure 6). When comparing the restriction maps of European and North American Atlantic salmon, it can be seen that the mtDNA from North American Atlantic salmon failed to cut at three locations with three enzymes as compared to those of European Atlantic salmon. They failed to digest (1) by *Bgl*II in the 12S rRNA gene at bp #1436, (2) *BstE*II in the *COX2* gene at bp #8495 and (3) *Xba*I in the *ND5* gene at bp #13274. The rest of the predicted restriction fragment sizes were comparable to those published (Bermingham *et al.*, 1991).

Furthermore, when examining the multiple sequence alignment, several sites were observed to have substitutions only in either North American or European samples. One of these sites (base pair #1441) coincides with the *Bgl*II recognition site, which mutated the cut site; therefore, the samples were no longer cleaved by *Bgl*II. By having the sequence data, the locations of the markers that were previously used to discriminate population structures could be deduced from the sequence data.

Figure 5 **Confirmation of previously characterized restriction sites that discriminate between European and North American Atlantic salmon populations.**

a) Restriction map of "European" Atlantic salmon

b) Restriction map of "North American" Atlantic salmon

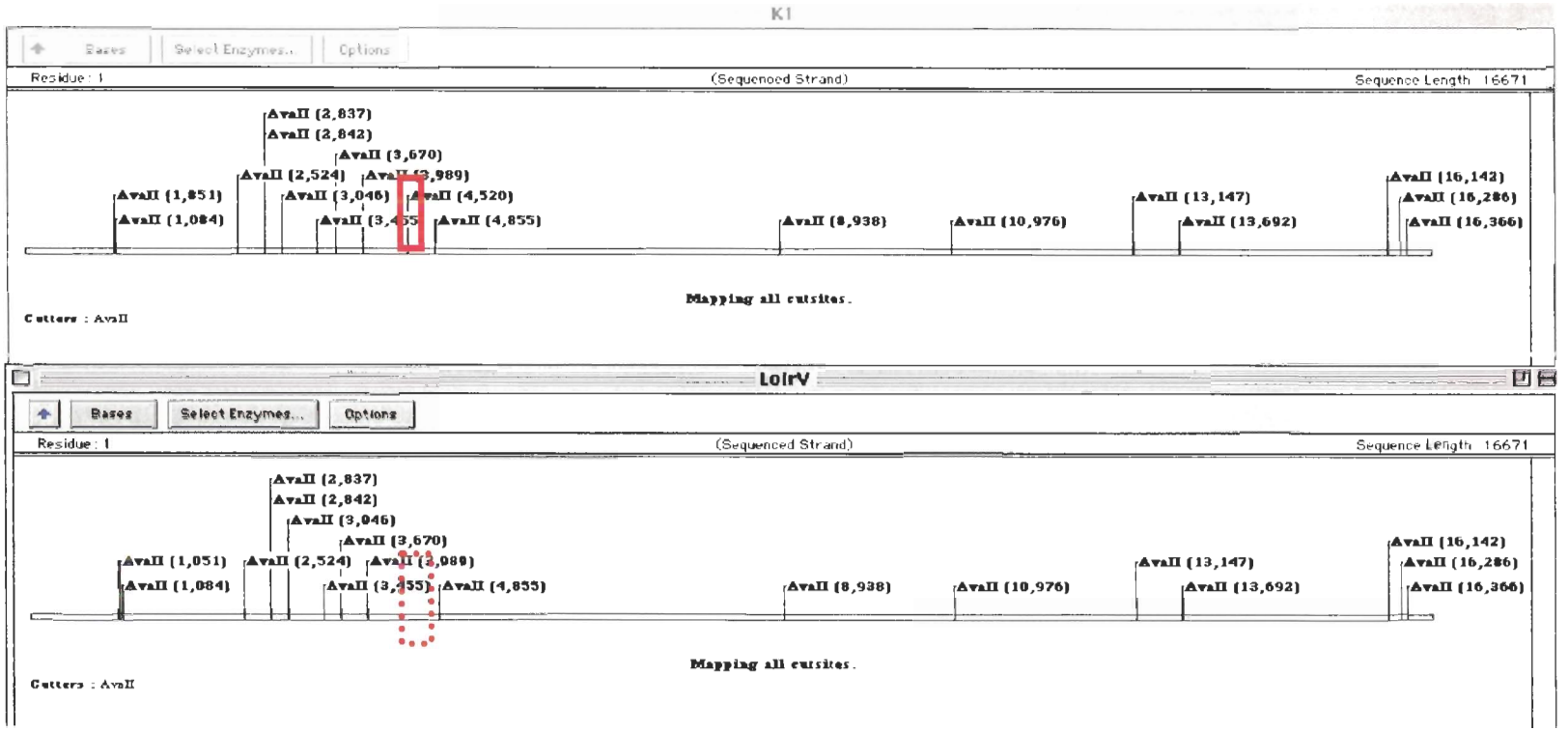
c) Restriction fragment profiles of "North American" and "European" Atlantic salmon mitochondrial genome digested with *Xba*I, *Bst*EII and *Bgl*II.

Enzyme	*"North American"*	*"European"*	*Enzyme*	*"North American"*	*"European"*	*Enzyme*	*"North American"*	*"European"*
*Xba*I	5168	5030	*Bst*EII	16674	-----	*Bgl*II	16674	-----
	3699	3699		-----	10262		-----	15855
	3270	3270		-----	6412		-----	819
	2279	2279						
	2258	2258						
	-----	138						

Restriction digests can be used to further divide the classification into subgroups corresponding to rivers and broader geographical areas. *In silico* restriction analysis of Atlantic salmon mtDNA with a five bases cutter (*Ava*II) was performed. This enzyme has been shown to produce digestion patterns with a 950 bp fragment in Atlantic salmon from the River Atran, which was not observed in any of the salmon from River Lule (Palva *et al.*, 1989). These studies suggested that a discrete mtDNA marker could distinguish between Baltic and other European Atlantic salmon. However, when creating restriction maps of sequences in the current study with *Ava*II, the distribution of *Ava*II cut sites does not conform to the geographical designations in this case (Figure 7). *Ava*II has been previously shown to differentiate two mtDNA genotypes from European fish by only a single *Ava*II restriction site (Birt *et al.*, 1991) where samples from the Baltic region should be cut 18 times, one more than that of the Southern European genotype. However, this enzyme is not successful in this differentiation because not all Baltic samples had 18 restriction sites, suggesting that the sample sizes from the previous study might not be large enough. Two enzymes with four base pair recognition sites (*Hae*III and *Hinf*I) each produced two restriction digest patterns that unambiguously distinguished the 23 non-anadromous Atlantic salmon from the nine anadromous salmon derived from the Neva River (Davidson *et al.*, 1989a). These data proved that restriction digests could be very useful for Atlantic salmon population genetic studies, but it was difficult to obtain informative data with the limited technology at the time.

Figure 6 The comparison of the restriction maps of two European Atlantic salmon mitochondrial genomes cleaved by Ava**II**.

The box marks the site that was previously reported to have geographic discrimination based on the absence and presence of that particular cut site.



4.4.3 Partial sequencing of cytochrome *b* revealed two main genotypes

The most accurate method for measuring genetic variation in the mitochondrial genome within a species is to compare actual sequences (Kocher *et al.*, 1989) The ability to obtain sequence data was previously limited by the ability to isolate highly purified mtDNA, clone specific restriction fragments and sequence them. This was too time-consuming and expensive a procedure to be used for routine large-scale population genetic surveys. Kocher *et al.* (1989) described the amplification and sequencing of mtDNA with conserved primers, and opened the door for a new tool for population biology. This technology was quickly employed to estimate the extent of intra-specific variation and population sub-structuring in several fish species such as Atlantic cod (Carr and Marshall, 1991), tuna species (Bartlett and Davidson, 1991), capelin (Birt *et al.*, 1995), turbot (Vis *et al.*, 1997) and salmonid species (Bernatchez *et al.*, 1992; Davidson *et al.*, 1989b; Domanico and Phillips, 1995; Mcveigh and Davidson, 1991; Palsson and Arnason, 1994).

Cytochrome *b* sequences were used to study the population dynamics of Atlantic salmon from North American and European river systems (McVeigh *et al.*, 1991). From sampling 295 bp of the cytochrome *b* gene of 60 individuals, only three cytochrome *b* genotypes (I, II, and III) could be identified. The distribution of the genotypes was not uniform. All European samples gave genotype I. Samples from North American had a mixture of genotype I, II, and III. The sample size used in the experiment was too small for statistical significance. Palsson and Arnason (1994) used the same region of the gene to investigate the genotypes of fifteen salmon from Iceland. More than half of the samples were classified as genotype I, but none of the fish had genotype II or III.

Instead, three additional genotypes (IV, V and VI) were identified. These genotypes differed from genotype I by only a single nucleotide. There is a connection between the cytochrome *b* genotypes defined by partial cytochrome *b* sequencing and the European and North American haplotypes defined by restriction enzyme analysis (Birt *et al.*, 1991; Mcveigh and Davidson, 1991). In both studies, the European genotype defined by restriction digests with *Bgl*II and *BstE*II corresponded to cytochrome *b* genotype I, whereas fishes classified with the North American restriction site genotype possessed genotype II or III. The same correlation was observed in the fourteen samples from a wide range of populations in the current study, and seven more genotypes were identified.

The six genotypes identified previously, genotypes I-III from McVeigh *et al.* (1991) and genotypes IV-VI from Palsson and Arnason (1994) were based on only two to five mutational sites in cytochrome *b*. Furthermore, the results from Palsson and Arnason (1994) were based on only 15 samples from Iceland. Using samples from a wide range of populations on both sides of the Atlantic Ocean, and all 1140 nucleotides from the cytochrome *b* gene, a total of ten genotypes was identified (Table 11). The 295 bp segment from previous studies corresponds to positions 96-390 of the whole 1140 bp cytochrome *b* gene. Genotypes I-III were identified. The locations of the two variable sites that were previously used for genotyping, are numbers 108 and 165. Genotypes IV-VI were not identified in the samples that were subject to study. One explanation could be that the current study did not sample Atlantic salmon from Iceland, which has been predicted as being a refugium during the last the Ice Age (Palsson and Arnason, 1994). The analysis of the full gene identified seven other genotypes, all with mutations outside

of the 295 bp fragment. Genotype IV is closely related to genotype II of North American origin and Genotype V-X differ from genotype I (European genotype) by one or two mutations. These results illustrate the benefit of using longer sequences (1140 bp vs. 295 bp) to detect differences between individuals and potentially populations.

Table 11 **Variable sites in the full cytochrome *b* gene found by comparing fourteen Atlantic salmon.**

108* and 165* mark the two mutational site used to define genotype I, II, III in *McVeigh et al.* (1991). Genotype I – European; II and III – North American, same as McVeigh *et al.* (1991). Genotype IV from Conne River, Nfld, Canada. Genotype V – X are one or two nucleotide differences diverged from main European genotype I. {Origin: Genotype V = West Greenland; VI = Loire Valley, France; VII = Almond River, Scotland; VIII and IX = Baltic Sea; X = Sweden}

	Variable Sites															
Genotype	27	102	108*	132	165*	441	558	783	795	897	911	945	978	1038	1047	
I	G	C	C	A	C	A	A	C	A	G	T	G	G	A	C	European
II	•	•	A	•	•	•	•	•	•	•	•	•	A	C	•	North American
III	•	•	A	•	T	•	G	A	•	C	•	•	A	G	T	North American
IV	•	•	A	G	•	•	•	A	•	C	•	•	A	C	•	North American
V	A	•	•	•	•	G	•	•	•	•	•	•	•	•	•	European – West Greenland
VI	•	T	•	•	•	•	•	•	•	•	•	•	•	•	•	European – Loire Valley
VII	•	•	•	•	•	•	•	•	G	•	•	•	•	•	•	European – Almon River
VIII	•	•	•	•	•	•	•	•	G	•	•	•	A	•	•	European – Baltic Sea
IX	•	•	•	•	•	•	•	•	•	•	G	A	•	•	•	European – Baltic Sea
X	•	•	•	•	•	•	•	•	•	•	•	•	A	•	•	European – Sweden

4.4.4 PCR-RFLP of NADH dehydrogenase genes

Another method that commonly utilizes mtDNA for large-scale population studies is the combination of PCR and restriction fragment length polymorphism (PCR-RFLP). The advantage of this method is the ability to amplify a fragment, and perform a quick survey to find any mutation sites that might coincide with restriction sites (Avise, 2004). This method enables screening of large population sizes with less time and less quantity of sample.

Several studies used universal primers for the amplification step. However, if these primers were designed from a variety of species (Cronin *et al.*, 1993), the success of amplification could be reduced. Nielsen *et al.* (1998) modified the three sets of conserved primers from Cronin *et al.* (1993) to be more specific for salmonid fishes by using the complete nucleotide sequence of the mtDNA from rainbow trout (Zardoya *et al.*, 1995) as a template. The three pairs of primers were used to amplify a larger part of the mitochondrial genome compared to partial cytochrome *b* sequencing. Thirty Atlantic salmon samples from four different European salmon populations covering a wide range of the continent were amplified. A total of eleven genotypes were found in the 30 samples; however, there was no evidence of any phylogeographical grouping of the genotypes.

The three pairs of primers (Nielsen *et al.*, 1998) were also able to successfully amplify the fragments in three other salmonid species (brown trout, whitefish and grayling). However, it should be noted that two of the three pairs of primer sequences were found to differ from the sequence of Atlantic salmon mtDNA by single nucleotide substitutions. The correct primer sequences for Atlantic salmon should be:

ND1-F (5'-GCCTCGCCTGTTTACCAAAAACAT-3') and ND1-R (5'-GGTATGGGCCCGAAAGCTTA-3') for NADH dehydrogenase-1;

ND3/4-F (5'-TTAA**C**ACGTATAAGTGACTTCCAA-3') and ND3/4-R (5'-TTTTGGTTCCTAAGACCAATGGAT-3') for NADH dehydrogenase-3/4 fragment; and ND5/6-F (5'-AA**C**AGCTCATCCATTGGTCTTAGG-3') and ND5/6-R (5'-TAACAACGG**C**GGTTTTTCAAGTCA-3') for NADH dehydrogenase-5/6 fragment. (Note: the underline marks sites which differ from Nielsen's primers by a single transition substitution)

The three mtDNA fragments were subsequently digested with restriction endonucleases. The 2.0 Kb *ND1* fragments gave polymorphic banding patterns when cut with four enzymes; the 2.3 Kb *ND3/4* and 2.4 Kb *ND5/6* fragments gave polymorphic banding patterns with three enzymes (Nielsen *et al.*, 1998). A total of eleven haplotypes could be identified in the 30 individuals. Only two variants were observed per restriction enzyme used. These primers were applied to the newly obtained whole mitochondrial genome sequences. An *in silico* method was used to examine the sequences amplified from primers. With *Alu*I (5'-AGCT-3'), nine recognition sites could be observed in the *ND5/6* fragment creating ten restriction fragments, but only 3 or 4 of the ten fragments (depends on the polymorphism) resembled the size of the 4 or 5 fragments found by Nielsen *et al.* (1998). The other seven fragments were all less than 120 bp in size. Because of the low resolution of agarose gels, these fragments were probably not seen. The shortcoming of this method lies in the difficulty of extracting data from electrophoretic gels, and sharing these data between laboratories. PCR and direct sequencing analysis can overcome the problem by yielding molecular sequence data.

Nonetheless, PCR-RFLP is still very useful in large-scale studies for detecting known sequence variations. This simple DNA-based analytical method is now being used for authentication of ingredients in seafood products (Bossier, 1999). For example: PCR-RFLP analysis of a 558 bp of the cytochrome *b* gene was able to successfully identify three closely related gadoid species originating from Alaska and Iceland (Aranishi *et al.*, 2005). This method is reliable and inexpensive, and suitable to be routinely applied for diagnostic studies of commercial seafood products.

4.4.5 Advantages of mtDNA compared to microsatellites for population studies

The advantage of using actual sequences as population data is widely accepted today (Ferguson *et al.*, 1995). The number of repeating units of a DNA segment can be used as a genetic marker as well. Microsatellites are one to six base pairs of DNA in a unit that could repeat up to 100 times at some loci. The difference in the number of repeats at a variety of loci can be used as a population marker. King *et al.* (2001) used the variation at twelve microsatellite loci to investigate the genetic variation of Atlantic salmon from 29 locations covering North America and in Europe. The group identified 266 alleles, with some hypervariable loci having up to 37 alleles. The level of genetic variation is summarized by gene diversity and the average frequency of unique alleles. Genetic distance also measures the differences between each pair of populations. Similar to previous results, a genetic discontinuity between North American and European populations was observed using the microsatellite data (King *et al.*, 2001). These results support the two distinct genotypes previously observed by allozyme and restriction enzyme analysis (Bermingham *et al.*, 1991; Gyllensten and Wilson, 1987; Stahl, 1987).

Using microsatellite data, problems can be seen in the misclassification of fish to population, at a rate of 25% with 10 microsatellite loci (Cornuet *et al.*, 1999). This could be owing to the fact that microsatellite profiles are unique from individual to individual. Therefore, it would be hard to classify each member of a particular population. Furthermore, not all microsatellites are highly variable (Ferguson *et al.*, 1995). Therefore, microsatellite loci must be chosen with care. Since Atlantic salmon have gone through whole genome tetraploidization and undergoing re-diploidization, duplicated copies of potentially all genes and microsatellites will add to the complexity in the analysis.

On the other hand, mitochondrial genomes are haploid and maternally inherited. This allows the differentiation between male and female-mediated gene flow (Li, 1997), and its sequence carries the history of the lineage without having undergone recombination (Wilson *et al.*, 1985). Also, mtDNA is more subject to drift, so bottlenecks can be detected easier (Birky *et al.*, 1983). These characteristics make mtDNA sequence an excellent type of molecule for population studies. The ease of obtaining mtDNA sequence data with the assistance of the established fish mtDNA primers for PCR, strengthens the usefulness of this molecule in large-scale population studies. In the current experiment, the genotyping using the whole cytochrome *b* sequence (1140 bp) was able to reveal seven more genotypes than that identified with 295 bp partial gene in the previous studies.

4.4.6 Future studies

The use of mtDNA markers assists researchers to evaluate population genetics on a large scale. Oligonucleotide DNA microarray hybridization is a method developed for the rapid detection of genotypes (Lander, 1999), and has since been tested on the D-loop of chum salmon populations (Moriya *et al.*, 2004). This technique involves the immobilization of 17 to 20-mer oligonucleotides, representing all variation previously identified at polymorphic sites, on a glass slide. To detect mitochondrial genome variation, variable regions of the mtDNA are amplified via PCR and hybridized on to the glass slide. Since all allelic variations are represented on the DNA microarray, the genotype based on the mitochondrial DNA is indicated by the strength of hybridization to a specific allelic sequence. An intense signal denotes a perfect match, representing the presence of that particular nucleotide sequence variant. This method enables the screening of known nucleotide variable sites in under four hours. Examining the fourteen Atlantic salmon whole mitochondrial genome sequences from various populations, 198 variable sites could be identified. Of which, 87 sites are able to differentiate between European and North American samples; and 111 sites are population specific. The combination of these sites and other previously identified, could be used to make a microarray for this species. The technological advance enables the generation of a high-throughput set of data that surveys a large sample of world-wide populations in a short amount of time.

4.5 Conclusions

The sampling of a wide range of whole Atlantic salmon mitochondrial genomes gave insight into the mode and evolution of the mitochondrial genome. In the genus *Salmo*, the mtDNA has a bias against guanine in the base composition on the H-strand. Closely related individual Atlantic salmon samples show more transition substitutions than transversions. These translate into a lower occurrence of non-synonymous amino acid changes. By examining the region with the highest number of substitutions, the fastest evolving gene is *ND3*, consistent with that found in other salmonids. More detailed information about using mtDNA as a genetic marker allowed the confirmation of previously described genotypes, and will facilitate the identification of new genotypes. Furthermore, a list of possible markers for population studies was developed, and this may prove useful for the production of a mtDNA microarray.

REFERENCE LIST

Alves, M. J., H. Coelho, M. J. Collares-Pereira and M. M. Coelho, 2001. Mitochondrial DNA variation in the highly endangered cyprinid fish *Anaecypris hispanica*: Importance for conservation. *Heredity* **87:** 463-473.

Anderson, S., M. H. L. Debruijn, A. R. Coulson, I. C. Eperon and F. Sanger *et al*, 1982. Complete sequence of bovine mitochondrial-DNA - conserved features of the mammalian mitochondrial genome. *J. Mol. Biol.* **156:** 683-717.

Anderson, S., A. T. Bankier, B. G. Barrell, M. H. L. Debruijn and A. R. Coulson *et al*, 1981. Sequence and organization of the human mitochondrial genome. *Nature* **290:** 457-465.

Apostolidis, A. P., C. Triantaphyllidis, A. Kouvatsi and P. S. Economidis, 1997. Mitochondrial DNA sequence variation and phylogeography among *Salmo trutta* L. (greek brown trout) populations. *Mol. Ecol.* **6:** 531-542.

Aquadro, C. F., and B. D. Greenberg, 1983. Human mitochondrial-DNA variation and evolution - analysis of nucleotide-sequences from 7 individuals. *Genetics* **103:** 287-312.

Aranishi, F., T. Okimoto and S. Izumi, 2005. Identification of gadoid species (pisces, gadidae) by PCR-RFLP analysis. *J. Appl. Genet.* **46:** 69-73.

Avise, J. C., J. Arnold, R. M. Ball, E. Bermingham and T. Lamb *et al*, 1987. Intraspecific phylogeography - the mitochondrial-DNA bridge between population-genetics and systematics. *Annual Review of Ecology and Systematics* **18:** 489-522.

Avise, J. C., 2004. *Molecular Markers, Natural History, and Evolution*. Sinauer Associates, Sunderland, Mass. 55-114.

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.

Bermingham, E., S. H. Forbes, K. Friedland and C. Pla, 1991. Discrimination between atlantic salmon (*Salmo-salar*) of North-American and European origin using restriction

analyses of mitochondrial-DNA. *Canadian Journal of Fisheries and Aquatic Sciences* **48:** 884-893.

Bernatchez, L., and R. G. Danzmann, 1993. Congruence in control-region sequence and restriction-site variation in mitochondrial-DNA of brook charr (*Salvelinus-fontinalis-mitchill*). *Mol. Biol. Evol.* **10:** 1002-1014.

Bernatchez, L., R. Guyomard and F. Bonhomme, 1992. DNA sequence variation of the mitochondrial control region among geographically and morphologically remote European brown trout *Salmo trutta* populations. *Mol. Ecol.* **1:** 161-173.

Bibb, M. J., R. A. Vanetten, C. T. Wright, M. W. Walberg and D. A. Clayton, 1981. Sequence and gene organization of mouse mitochondrial-DNA. *Cell* **26:** 167-180.

Bielawski, J. P., and J. R. Gold, 2002. Mutation patterns of mitochondrial H- and L,strand DNA in closely related cyprinid fishes. *Genetics* **161:** 1589-1597.

Billington, N., and P. D. N. Hebert, 1988. Mitochondrial-DNA variation in great-lakes walleye (*Stizostedion-vitreum*) populations. *Can. J. Fish. Aquat. Sci.* **45:** 643-654.

Billington, N., P. D. N. Hebert and R. D. Ward, 1988. Evidence of introgressive hybridization in the genus *Stizostedion* - interspecific transfer of mitochondrial-DNA between sauger and walleye. *Can. J. Fish. Aquat. Sci.* **45:** 2035-2041.

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

Birt, T. P., J. M. Green and W. S. Davidson, 1991. Mitochondrial-DNA variation reveals genetically distinct sympatric populations of anadromous and nonanadromous atlantic salmon, *Salmo salar*. *Canadian Journal of Fisheries and Aquatic Sciences* **48:** 577-582.

Birt, T. P., J. M. Green and W. S. Davidson, 1986. Analysis of mitochondrial-DNA in allopatric anadromous and nonanadromous Atlantic salmon, *Salmo-salar*. *Canadian Journal of Zoology-Revue Canadienne De Zoologie* **64:** 118-120.

Birt, T. P., V. L. Friesen, R. D. Birt, J. M. Green and W. S. Davidson, 1995. Mitochondrial DNA variation in Atlantic capelin, *Mallotus villosus*: A comparison of restriction and sequence analyses. *Mol. Ecol.* **4:** 771-776.

Blouin, M. S., C. A. Yowell, C. H. Courtney and J. B. Dame, 1998. Substitution bias, rapid saturation, and the use of mtDNA for nematode systematics. *Mol. Biol. Evol.* **15:** 1719-1727.

Boore, J. L., and W. M. Brown, 1998. Big trees from little genomes: Mitochondrial gene order as a phylogenetic tool. *Current Opinion in Genetics & Development* **8:** 668-674.

Boore, J. L., J. R. Macey and M. Medina, 2005. Sequencing and comparing whole mitochondrial genomes of animals. *Methods Enzymol.* **395:** 311-348.

Bossier, P., 1999. Authentication of seafood products by DNA patterns. *J. Food Sci.* **64:** 189-193.

Broughton, R. E., J. E. Milam and B. A. Roe, 2001. The complete sequence of the zebrafish (*Danio rerio*) mitochondrial genome and evolutionary patterns in vertebrate mitochondrial DNA. *Genome Res.* **11:** 1958-1967.

Brown, K. H., and G. H. Thorgaard, 2002. Mitochondrial and nuclear inheritance in an androgenetic line of rainbow trout, *Oncorhynchus mykiss*. *Aquaculture* **204:** 323-335.

Brown, W. M., and J. Vinograd, 1974. Restriction endonuclease cleavage maps of animal mitochondrial dnas. *Proc. Natl. Acad. Sci. U. S. A.* **71:** 4617-4621.

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

Brown, W. M., E. M. Prager, A. Wang and A. C. Wilson, 1982. Mitochondrial-DNA sequences of primates - tempo and mode of evolution. *J. Mol. Evol.* **18:** 225-239.

Brunner, P. C., M. R. Douglas, A. Osinov, C. C. Wilson and L. Bernatchez, 2001. Holarctic phylogeography of Arctic charr (*Salvelinus alpinus L.*) inferred from mitochondrial DNA sequences. *Evolution Int. J. Org. Evolution* **55:** 573-586.

Cann, R. L., W. M. Brown and A. C. Wilson, 1984. Polymorphic sites and the mechanism of evolution in human mitochondrial-DNA. *Genetics* **106:** 479-499.

Cantatore, P., M. Roberti, G. Pesole, A. Ludovico and F. Milella *et al*, 1994. Evolutionary analysis of cytochrome b sequences in some perciformes: Evidence for a slower rate of evolution than in mammals. *J. Mol. Evol.* **39:** 589-597.

Carr, S. M., and H. D. Marshall, 1991. A direct approach to the measurement of genetic-variation in fish populations - applications of the polymerase chain-reaction to studies of atlantic cod, *Gadus-morhua* L. *J. Fish Biol.* **39:** 101-107.

Carr, S. M., D. S. Kivlichan, P. Pepin and D. C. Crutcher, 1999. Molecular systematics of gadid fishes: Implications for the biogeographic origins of pacific species. *Can. J. Zool. - Rev. Can. Zool.* **77:** 19-26.

Carr, S. M., A. J. Snellen, K. A. Howse and J. S. Wroblewski, 1995. Mitochondrial-DNA sequence variation and genetic stock structure of Atlantic cod (*Gadus-morhua*) from bay and offshore locations on the Newfoundland continental-shelf. *Mol. Ecol.* **4:** 79-88.

Chang, D. D., and D. A. Clayton, 1985. Priming of human mitochondrial-DNA replication occurs at the light-strand promoter. *Proceedings of the National Academy of Sciences of the United States of America* **82:** 351-355.

Chevassus, B., 1979. Hybridization in salmonids - results and perspectives. *Aquaculture* **17:** 113-128.

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

Coates, B. S., D. V. Sumerford, R. L. Hellmich and L. C. Lewis, 2005. Partial mitochondrial genome sequences of *Ostrinia nubilalis* and *Ostrinia furnicalis*. *Int. J. Biol. Sci.* **1:** 13-18.

Cornuet, J. M., S. Piry, G. Luikart, A. Estoup and M. Solignac, 1999. New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* **153:** 1989-2000.

Crespi, B. J., and M. J. Fulton, 2004. Molecular systematics of salmonidae: Combined nuclear data yields a robust phylogeny. *Mol. Phylogenet. Evol.* **31:** 658-679.

Cronin, M. A., W. J. Spearman, R. L. Wilmot, J. C. Patton and J. W. Bickham, 1993. Mitochondrial-DNA variation in chinook (*Oncorhynchus-tshawytscha*) and chum salmon (*O keta*) detected by restriction enzyme analysis of polymerase chain-reaction (PCR) products. *Can. J. Fish. Aquat. Sci.* **50:** 708-715.

Curole, J. P., and T. D. Kocher, 1999. Mitogenomics: Digging deeper with complete mitochondrial genomes. *Trends in Ecology and Evolution* **14:** 394-398.

Dahle, G., 1991. Cod, *Gadus-morhua L*, populations identified by mitochondrial-DNA. *J. Fish Biol.* **38:** 295-303.

Davidson, W. S., T. P. Birt and J. M. Green, 1989a. A review of genetic variation in Atlantic salmon, *Salmo salar L.*, and its importance for stock identification, enhancement programmes and aquaculture. *J. Fish Biol.* **34:** 547-560.

Davidson, W. S., T. P. Birt and J. M. Green, 1989b. Organisation of the mitochondrial genome from Atlantic salmon (*Salmo salar*). *Genome* **32:** 340-342.

Doiron, S., L. Bernatchez and P. U. Blier, 2002. A comparative mitogenomic analysis of the potential adaptive value of arctic charr mtDNA introgression in brook charr populations (*Salvelinus fontinalis mitchill*). *Molecular Biology and Evolution* **19:** 1902-1909.

Domanico, M. J., and R. B. Phillips, 1995. Phylogenetic analysis of pacific salmon (genus *Oncorhynchus*) based on mitochondrial DNA sequence data. *Mol. Phylogenet. Evol.* **4:** 366-371.

Faber, J. E., and C. A. Stepien, 1997. The utility of mitochondrial DNA control region sequences for analyzing phylogenetic relationships among populations, species, genera, and percidae., pp. 129-143 in *Molecular Systematics of Fishes*, edited by T. D. Kocher and C. A. Stepien. Academic Press, San Diego.

Faith, J. J., and D. D. Pollock, 2003. Likelihood analysis of asymmetrical mutation bias gradients in vertebrate mitochondrial genomes. *Genetics* **165:** 735-745.

Ferguson, A., J. B. Taggart, P. A. Prodohl, O. McMeel and C. Thompson *et al*, 1995. The application of molecular markers to the study and conservation of fish populations, with special reference to *Salmo*. *Journal of Fish Biology* **47:** 103-126.

Friesen, V. L., W. A. Montevecchi and W. S. Davidson, 1993. Cytochrome-*b* nucleotide-sequence variation among the atlantic alcidae. *Hereditas* **119:** 245-252.

Froufe, E., I. Knizhin and S. Weiss, 2005. Phylogenetic analysis of the genus *thymallus* (grayling) based on mtDNA control region and ATPase 6 genes, with inferences on control region constraints and broad-scale eurasian phylogeography. *Molecular Phylogenetics and Evolution* **34:** 106-117.

Gibson, A., V. Gowri-Shankar, P. G. Higgs and M. Rattray, 2005. A comprehensive analysis of mammalian mitochondrial genome base composition and improved phylogenetic methods. *Mol. Biol. Evol.* **22:** 251-264.

Gyllensten, U., and A. C. Wilson, 1987. Mitochondrial DNA of salmonids: Inter- and intraspecific variability dected with restriction enzymes, pp. 420 in *Population Genetics & Fishery Management*, edited by N. Ryman and F. Utter. Washington Sea Grant Program : Distributed by University of Washington Press, Seattle.

Hasegawa, M., and T. Hashimoto, 1993. Ribosomal-rna trees misleading. *Nature* **361:** 23-23.

Hovey, S. J., D. P. F. King, D. Thompson and A. Scott, 1989. Mitochondrial-DNA and allozyme analysis of Atlantic salmon, *Salmo-salar L*, in england and wales. *Journal of Fish Biology* **35:** 253-260.

Hurst, C. D., S. E. Bartlett, W. S. Davidson and I. J. Bruce, 1999. The complete mitochondrial DNA sequence of the Atlantic salmon, *Salmo salar*. *Gene* **239:** 237-242.

Iguchi, K., K. Watanabe and M. Nishida, 1999. Reduced mitochondrial DNA variation in hatchery populations of ayu (*Plecoglossus altivelis*) cultured for multiple generations. *Aquaculture* **178:** 235-243.

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.

Ishiguro, N. B., M. Miya and M. Nishida, 2003. Basal euteleostean relationships: A mitogenomic perspective on the phylogenetic reality of the "protacanthopterygii". *Mol. Phylogenet. Evol.* **27:** 476-488.

King, T. L., S. T. Kalinowski, W. B. Schill, A. P. Spidle and B. A. Lubinski, 2001. Population structure of atlantic salmon (*Salmo salar L.*): A range-wide perspective from microsatellite DNA variation. *Mol. Ecol.* **10:** 807-821.

Knox, D., and E. Verspoor, 1991. A mitochondrial-DNA restriction-fragment-length-polymorphism of potential use for discrimination of farmed norwegian and wild Atlantic salmon populations in scotland. *Aquaculture* **98:** 249-257.

Kocher, T. D., W. K. Thomas, A. Meyer, S. V. Edwards and S. Paabo *et al*, 1989. Dynamics of mitochondrial DNA evolution in animals: Amplification and sequencing with conserved primers. *Proc. Natl. Acad. Sci. U. S. A.* **86:** 6196-6200.

Koljonen, M. L., H. Jansson, T. Paaver, O. Vasin and J. Koskiniemi, 1999. Phylogeographic lineages and differentiation pattern of atlantic salmon (*Salmo salar*) in the baltic sea with management implications. *Can. J. Fish. Aquat. Sci.* **56:** 1766-1780.

Krajewski, C., and D. G. King, 1996. Molecular divergence and phylogeny: Rates and patterns of cytochrome b evolution in cranes. *Mol. Biol. Evol.* **13:** 21-30.

Kumar, S., K. Tamura and M. Nei, 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* **5:** 150-163.

Lander, E. S., 1999. Array of hope. *Nat. Genet.* **21:** 3-4.

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 .3. techniques and potential applications. *J. Mol. Evol.* **17:** 214-226.

Lee, W. J., and T. D. Kocher, 1995. Complete sequence of a sea lamprey (*Petromyzon marinus*) mitochondrial genome: Early establishment of the vertebrate genome organization. *Genetics* **139:** 873-887.

Li, W. H., C. I. Wu and C. C. Luo, 1984. Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. *J. Mol. Evol.* **21:** 58-71.

Li, W., 1997. *Molecular Evolution.* Sinauer Associates, Sunderland, Mass. 177-213.

MacIntyre, R. J., 1985. *Molecular Evolutionary Genetics.* Plenum Press, New York. 95-130.

Martin, A. P., and S. R. Palumbi, 1993. Protein evolution in different cellular environments: Cytochrome b in sharks and mammals. *Mol. Biol. Evol.* **10:** 873-891.

Martins, C., Wasko A.P., Oliveira, C. and Foresti, F. 2003. Mitochondrial DNA variation in wild populations of *Leporinus elongatus* from the Parana River basin. *Genet. Mol. Biol.* **26**: 33-38.

Mcveigh, H. P., and W. S. Davidson, 1991. A salmon id phylogeny inferred from mitochondrial cytochrome b gene sequences. *J. Fish Biol.* **39:** 277-282.

McVeigh, H. P., S. E. Bartlett and W. S. Davidson, 1991. Polymerase chain-reaction direct sequence-analysis of the cytochrome B gene in *Salmo-salar*. *Aquaculture* **95:** 225-233.

Meyer, A., 1993. Phylogenetic-relationships and evolutionary processes in east-african cichlid fishes. *Trends in Ecology and Evolution* **8:** 279-284.

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

Miya, M., and M. Nishida, 2000. Use of mitogenomic information in teleostean molecular phylogenetics: A tree-based exploration under the maximum-parsimony optimality criterion. *Molecular Phylogenetics and Evolution* **17:** 437-455.

Miya, M., and M. Nishida, 1999. Organization of the mitochondrial genome of a deep-sea fish, *Gonostoma gracile* (teleostei: Stomiiformes): First example of transfer RNA gene rearrangements in bony fishes. *Mar. Biotechnol.* (NY) **1:** 416-0426.

Moriya, S., S. Urawa, O. Suzuki, A. Urano and S. Abe, 2004. DNA microarray for rapid detection of mitochondrial DNA haplotypes of chum salmon. *Mar. Biotechnol.* **6:** 430-434.

Neave, F. 1958. The origin and speciation of Oncorhynchus. *Transactions of the Royal Society of Canada.* Vol. LII, Series 3, Sec. 5: 25-39.

Nei, M., 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89:** 583-590.

Nelson, J. S., 1994. *Fishes of the World.* J. Wiley, New York. 189-194.

Nielsen, E. E., M. M. Hansen and K. -. D. Mensberg, 1998. Improved primer sequences for the mitochondrial ND1, ND3/4 and ND5/6 segments in salmonid fishes: Application to RFLP analysis of atlantic salmon. *J. Fish Biol.* **53:** 216-220.

Nielsen J.L., C. Gan, J.M. Wright, W.K. Thomas, 1994. Phylogeographic patterns in California steelhead as determined by MtDNA and microsatellite analyses *Calif. Coop. Oceanic Fish. Investig. Rep.* **35**: 90-92.

Nielsen, E. E., M. M. Hansen and V. Loeschcke, 1996. Genetic structure of european populations of *Salmo salar L* (atlantic salmon) inferred from mitochondrial DNA. *Heredity* **77:** 351-358.

Oakley, T. H., and R. B. Phillips, 1999. Phylogeny of salmonine fishes based on growth hormone introns: Atlantic (*Salmo*) and pacific (*Oncorhynchus*) salmon are not sister taxa. *Molecular Phylogenetics and Evolution* **11:** 381-393.

Oconnell, M., D. O. F. Skibinski and J. A. Beardmore, 1996. Allozyme and mtDNA divergence between atlantic salmon populations in north wales. *Journal of Fish Biology* **48:** 1023-1026.

Palsson, S., and E. Arnason, 1994. Sequence variation for cytochrome B genes of 3 salmonid species from iceland. *Aquaculture* **128:** 29-39.

Palva, T. K., H. Lehvaslaiho and E. T. Palva, 1989. Identification of anadromous and non-anadromous salmon stocks in finland by mitochondrial-DNA analysis. *Aquaculture* **81:** 237-244.

Payne, R. H., A. R. Child and A. Forrest, 1971. Geographical variation in atlantic salmon. *Nature* **231:** 250-end.

Phillips, R.B., Pleyte, K.A. and Ihssen, P.E. 1989. Patterns of chromosomal nucleolar organizer region (NOR) variation in fishes of the genus Salvelinus. *Copeia* 1: 47-53.

Reyes, A., C. Gissi, G. Pesole and C. Saccone, 1998. Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. *Mol. Biol. Evol.* **15:** 957-966.

Roe, B. A., D. P. Ma, R. K. Wilson and J. F. H. Wong, 1985. The complete nucleotide-sequence of the *Xenopus-laevis* mitochondrial genome. *J. Biol. Chem.* **260:** 9759-9774.

Rozen, S., and H. Skaletsky, 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* **132:** 365-386.

Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf and R. Higuchi *et al*, 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA-polymerase. *Science* **239:** 487-491.

Shadel, G. S., and D. A. Clayton, 1997. Mitochondrial DNA maintenance in vertebrates. *Annu. Rev. Biochem.* **66:** 409-435.

Shedlock, A. M., J. D. Parker, D. A. Crispin, T. W. Pietsch and G. C. Burmer, 1992. Evolution of the salmonid mitochondrial control region. *Mol. Phylogenet. Evol.* **1:** 179-192.

Sivasundar, A., E. Bermingham and G. Orti, 2001. Population structure and biogeography of migratory freshwater fishes (prochilodus : Characiformes) in major south american rivers. *Mol. Ecol.* **10:** 407-417.

Sorenson, M. D., J. C. Ast, D. E. Dimcheff, T. Yuri and D. P. Mindell, 1999. Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. *Mol. Phylogenet. Evol.* **12:** 105-114.

Stahl, G., 1987. Genetic population structure of atlantic salmon., pp. 121-140 in *Population Genetics & Fishery Management*, edited by N. Ryman and F. Utter. Washington Sea Grant Program : Distributed by University of Washington Press, Seattle.

Stepien, C. A., and T. D. Kocher, 1997. Molecules and morphology in studies of fish evolution, pp. 1-11 in *Molecular Systematics of Fishes*, edited by T. D. Kocher and C. A. Stepien. Academic Press, San Diego.

Tanaka, M., and T. Ozawa, 1994. Strand asymmetry in human mitochondrial-DNA mutations. *Genomics* **22:** 327-335.

Thomas, W. K., and A. C. Wilson, 1991. Mode and tempo of molecular evolution in the nematode *Caenorhabditis* - cytochrome oxidase-ii and calmodulin sequences. *Genetics* **128:** 269-279.

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., R. E. Withler and A. T. Beckenbach, 1986. Mitochondrial-DNA analysis of pacific salmonid evolution. *Canadian Journal of Zoology-Revue Canadienne De Zoologie* **64:** 1058-1064.

Thompson, J. D., D. G. Higgins and T. J. Gibson, 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22:** 4673-4680.

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.

Utter, F., 1987. Allozyme, pp. 420 in *Population Genetics & Fishery Management*, edited by N. Ryman and F. Utter. Washington Sea Grant Program : Distributed by University of Washington Press, Seattle.

Verspoor, E., E. M. McCarthy and D. Knox, 1999. The phylogeography of european atlantic salmon (*Salmo salar L.*) based on RFLP analysis of the ND1/16sRNA region of the mtDNA. *Biol. J. Linnean Soc.* **68:** 129-146.

Verspoor, E., J. A. Beardmore, S. Consuegra, C. Garcia de Leaniz and K. Hindar *et al*, 2005. Population structure in the atlantic salmon: Insights from 40 years of research into genetic protein variation. *J. Fish Biol.* **67:** 3-54.

Vis, M. L., S. M. Carr, W. R. Bowering and W. S. Davidson, 1997. Greenland halibut (*Reinhardtius hippoglossoides*) in the north atlantic are genetically homogeneous. Can. J. *Fish. Aquat. Sci.* **54:** 1813-1821.

Vladykov, V. D., 1963. Review of salmonid genera and their broad geographical distribution. *Transactions of the Royal Society of Canada* **1:** 459-&.

Wilhelm, V., J. Villegas, A. Miquel, E. Engel S. Bernales, Valenzuela P.D. and Burzio L.O. 2003. The complete sequence of the mitochondrial genome of the chinook salmon, *Oncorhynchus tshawytscha*. *Biological Research* **36:** 223-231.

Wilson, A. C., R. L. Cann, S. M. Carr, M. George and U. B. Gyllensten *et al*, 1985. Mitochondrial-DNA and 2 perspectives on evolutionary genetics. *Biol. J. Linnean Soc.* **26:** 375-400.

Xia, X. H., M. S. Hafner and P. D. Sudman, 1996. On transition bias in mitochondrial genes of pocket gophers. *J. Mol. Evol.* **43:** 32-40.

Yamauchi, M. M., M. U. Miya and M. Nishida, 2004. Use of a PCR-based approach for sequencing whole mitochondrial genomes of insects: Two examples (cockroach and dragonfly) based on the method developed for decapod crustaceans. *Insect Mol. Biol.* **13:** 435-442.

Zardoya, R., and A. Meyer, 1996. The complete nucleotide sequence of the mitochondrial genome of the lungfish (*protopterus dolloi*) supports its phylogenetic position as a close relative of land vertebrates. *Genetics* **142:** 1249-1263.

Zardoya, R., A. GarridoPertierra and J. M. Bautista, 1995. The complete nucleotide sequence of the mitochondrial DNA genome of the rainbow trout, *Oncorhynchus mykiss*. *J. Mol. Evol.* **41:** 942-951.

Zhu, D. Q., B. G. M. Jamieson, A. Hugall and C. Moritz, 1994. Sequence evolution and phylogenetic signal in control-region and cytochrome-B sequences of rainbow fishes (melanotaeniidae). *Mol. Biol. Evol.* **11:** 672-683.