EVOLUTIONARY GENETIC ANALYSIS OF PACIFIC SALMON AND TROUT (*ONCORHYNCHUS*).

by
Sheldon John McKay

B.Sc., University of British Columbia, 1990
M.Sc., University of British Columbia, 1993

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

In the Department
of
Biological Sciences

SIMON FRASER UNIVERSITY

July 1997

APPROVAL

Name: Sheldon J. McKay

Degree: Doctor of Philosophy

Title of Thesis:

EVOLUTIONARY GENETIC ANALYSIS OF PACIFIC SALMON AND TROUT (ONCORHYNCHUS).

Examining Committee:

Chair: Dr. A. P. Farrell, Professor

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

Dr. R. Devlin, Adjunct Professor
Department of Biological Sciences, SFU

Dr. A. Beckenbach, Professor
Department of Biological Sciences, SFU

Dr. F. Breden, Associate Professor
Department of Biological Sciences, SFU

Dr. B. Crespi, Assistant Professor
Department of Biological Sciences, SFU
Public Examiner

Dr. M. Blouw, Professor
Faculty of Natural Resources and Environmental Studies
University of Northern British Columbia
External Examiner

Date Approved: July 30, 1997

Abstract

This thesis addresses the topic of molecular evolution at the genus, species, and gene levels. DNA sequence analysis was used to resolve taxonomic and systematic problems in the salmonid genus *Oncorhynchus* and to examine the evolution of duplicated genes. The evolution of Pacific salmon and trout has been intensively studied using a variety of methods, but the early evolutionary history of the genus and the relationships among sockeye, pink and chum salmon remained controversial. In this study, phylogenetic analyses of mitochondrial and nuclear genes provided strong evidence that pink and chum salmon are sister species, but the conflict regarding deeper phylogeny was still unresolved. The new phylogenetic data were combined with previously generated character sets to yield a tree that suggests the ancestor of the Asian *O. masou* species complex was the first lineage to diverge from the proto-*Oncorhynchus* line, which then rapidly radiated to form the other Pacific salmon and trout lineages.

The Asian salmon masu and amago were previously considered to be distinct species. Here, DNA sequences from their mitochondrial genomes were found to be almost identical, but considerable variation was detected in intron sequences of the growth hormone type-2 (GH2) gene. Markedly different allele frequencies suggest that that masu and amago are genetically distinct. The DNA evidence was found to be consistent with a classification scheme placing masu and amago as *O. masou* subspecies.

The genome of the ancestral salmonid is believed to have been doubled in size sometime after it diverged from the related smelt family Osmeridae, producing two

copies of each gene. The evolutionary history of the duplicated, non-allelic salmonid growth hormone genes was examined using DNA sequences. GH1 and GH2 isoforms have been identified in all salmonine (salmon, trout, char) species, but the GH genes of whitefish (subfamily Coregoninae) could not be assigned to either category. Evidence is presented that the two gene pairs diverged independently. The most likely explanation is that disomic inheritance of these genes had not yet been re-established when the salmonine and coregonine lineages diverged.

Acknowledgements

I acknowledge the contributions and influence of my supervisors R. H. Devlin and M. J. Smith. I also thank the other members of my supervisory committee, A. T. Beckenbach and F. Breden. For helpful discussions, moral support and technical assistance, I thank members of the Devlin, Smith, Beckenbach and Breden Labs. In particular I thank Duane Smailus and Karen Beckenbach. I also acknowledge the contribution of the molecular evolution discussion group.

I thank the following agencies for financial support: Fisheries and Oceans Canada, National Science and Engineering Research Council of Canada, Science Council of British Columbia, Simon Fraser University and Sea Spring Salmon Farm.

Most of all I thank my wife Barbara McKay, to whom this work is dedicated.

Preface

Inclusion of co-authored articles and copyrighted materials

This thesis contains material reprinted from published, co-authored articles in the *Canadian Journal of Fisheries and Aquatic Sciences* and *Molecular Marine Biology and Biotechnology*. In both cases, the contribution of co-authors was in an advisory and supervisory capacity. I was the primary contributor in experimental design, data interpretation and preparation of manuscripts.

Chapter 2 is based in part on the article below. The original article was substantially modified to accommodate data acquired after the publication date. ***This article is reprinted with the permission of the National Research Council Press.***

McKay, S.J., Devlin, R.H., and M.J. Smith. 1996. Phylogeny of Pacific salmon and trout based on mitochondrial NADH Dehydrogenase Subunit 3 (ND3) and nuclear Growth Hormone Type-2 (GH2) DNA sequences. Can. J. Fish. Aquat. Sci. 53: 1165-1176.

Appendix 3 is presented as published except for changes in layout to accommodate thesis requirements. A modification in Figure A.3.1 was made to reflect the revised phylogenetic hypothesis forwarded in chapter 2. ***This article is reprinted with the permission of Blackwell Science Inc.***

McKay, S.J., Smith, M.J. and R.H. Devlin. 1997 Polymerase chain reaction-based species identification of salmon and coastal trout in British Columbia. Mol. Mar. Biol. Biotechnol. (In press)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

General Introduction

The family Salmonidae contains is made up of three subfamilies: Coregoninae (whitefish, ciscos), Thymallinae (graylings) and Salmoninae (salmon, trout, char, huchen). Each of these subfamilies is widely distributed in the northern hemisphere (Norden,1961; Vladykov, 1963). Salmonidae appears to have originated from a freshwater ancestor that acquired the ability to descend into the ocean, although a marine origin has also been proposed for this family (See Tchernavin, 1939 for a history of this disputed subject). The presence of several exclusively freshwater genera and complete absence of any entirely marine forms seem more consistent with a freshwater ancestral species that subsequently adapted to a marine environment.

Although the exact relationships among the genera are not always clear, it is widely accepted that Coregoninae was the first to diverge from this lineage, and is a sister taxon to the thymalline and salmonine groups (Stearley and Smith, 1993). The better known salmonine genera include *Hucho* (huchen), *Salvelinus* (char), *Salmo* (Atlantic salmon, brown trout) and *Oncorhychus* (Pacific salmon and trout). Within *Oncorhynchus*, there are eight extant species or species complexes and two fossil species *O. (Rhabdofario) lacustris*, a trout-like fish, and *O. (Smilodonichthys) rastrotus*, the "saber-toothed salmon" (Cope, 1870; Cavender and Miller, 1972; Stearley and Smith, 1993).

Investigation of evolutionary relationships in *Oncorhynchus*

Evolutionary relationships among the Pacific salmon and trout have been the subject of considerable debate. Rainbow and cutthroat trout were originally grouped

together with Atlantic salmon and brown trout in the genus *Salmo*. However, it has long been recognized that these trout were similar to the onchorhynchid species (Regan, 1914; Vladykov, 1963). More recent work has led to the conclusion that Salmo was a paraphyletic assemblage and that rainbow, cutthroat and allied trout are actually part of the monophyletic Pacific salmon and trout clade, *Oncorhynchus* (Smith and Stearley, 1989; Stearley and Smith, 1993). *Oncorhynchus* is believed to have arisen from a single ancestral species derived from the *Salmo* evolutionary line. Neave (1958) proposed that the common ancestor of rainbow and cutthroat trout was the first to diverge from the proto-*Oncorhynchus* evolutionary line about one million years ago, which went on to found the present-day Pacific salmon. However, the age of modern Pacific salmon species has been estimated at least 6 million years based on fossil evidence (Smith, 1992), which indicates that Neave's (1958) time scale for the radiation of *Oncorhynchus* is a substantial underestimate.

Oncorhynchus phylogenies have been reconstructed using a variety of methods (Utter et al., 1973 and references therein; Berg and Ferris, 1984; Thomas et al., 1986; Thomas and Beckenbach, 1989; Grewe et al., 1990; McVeigh and Davidson, 1991; Phillips and Pleyte, 1991; Shedlock et al., 1992, Devlin 1993; Murata et al. 1993,1996; Takasaki et al. 1994; Oohara et al., 1997). A reasonable consensus has been achieved for most species groups, such as (chinook, coho) and (rainbow, cutthroat) (Utter and Allendorf, 1994). However, the earlier evolutionary history of the genus and the relationships among sockeye, pink and chum salmon remained controversial.

In Chapter 2, DNA sequences of the nuclear growth hormone type 2 (GH2) and mitochondrial NADH Dehydrogenase Subunit 3 (ND3) genes were examined in an effort to clarify the basal evolutionary branching order and resolve the relationships

among sockeye, pink and chum. As with previous phylogenetic analyses, the GH2 and ND3 gene trees agreed in terminal species groupings, but differed in their inference of deeper phylogeny. To address this recurring problem in analysis of single genes or small character sets, a combined and comparative analysis of all available data and phylogenetic information was used to infer a tree that better addressed outstanding controversies in the systematics of *Oncorhynchus*.

Resolving systematic problems within the *O. masou* species complex

The taxonomy of *Oncorhynchus* was also not fully resolved at the species-level. Five types of salmon (sockeye, pink, chum, chinook and coho) occur on both sides of the northern Pacific Ocean. Each of these salmon exhibits morphological and ecological differences that have made it possible to assign unambiguous species status. However, the status of the three types of salmon that occur only in Asia (masu, amago and biwa salmon) is less clear. Two competing schemes are in current use for the classification of this complex: one assigns species status to masu (*O. masou*) and groups amago and biwa together as *O. rhodurus* (Kato, 1985; 1991), while the other groups masu (*O. masou masou*), amago (*O. masou ishikawae*) and biwa (*O. masou spp.*) as conspecific races (Kimura, 1990). The geographic ranges and some morphological characters distinguish each of the kinds of salmon, but overall similarity in most morphological and meristic characters along with vague descriptions of the original type specimens (Jordan and McGregor, 1925) have resulted in considerable confusion in their taxonomy and nomenclature (summarized in Table 3.1). In Chapter 3, I examine mitochondrial DNA sequence from the ND3 gene and the control (D-loop) region, where both interspecific (Thomas and Beckenbach, 1989; Shedlock et al,

1992) and intraspecific (Beckenbach et al., 1990; Park et al., 1993) variation in *Oncorhynchus* have previously been observed. Very little DNA sequence variation was detected among mitochondrial sequences of masu and amago, providing no evidence for genetic differentiation between the two. However, analysis of intronic sequences of the nuclear growth hormone type-2 (GH2) gene revealed considerable variation within and between types, providing evidence that masu and amago are genetically distinct, possibly at the subspecific level.

Evolution of duplicated salmonid growth hormone genes

In addition to evolution at the species and generic levels, this thesis also examines the evolution of particular genes. The proto-salmonid lineage that gave rise to subfamilies Coregoninae (*Coregonus, Prosopium, Stenodus*) and Salmoninae (*Salvelinus, Salmo, Oncorhynchus, Hucho, Brachymystax, Salmothymus, Platysalmo, Acantholingua*) is believed to have undergone a genome-doubling event some 25-100 Million years ago (Ohno, 1970; Allendorf and Thorgaard, 1984). Based on comparisons of genome size and chromosome numbers with related families (Hinegardner, 1976; Simon, 1963; Hartley, 1987), the tetraploidization of the salmonid genome must have occurred after Salmonidae diverged from other salmoniform lineages. After a genome is doubled, eventual re-establishment of disomic inheritance can lead to divergence of duplicated genes, many of which are lost. This process is well documented in salmonids, which have lost duplicated copies of approximately 50% of their genes (Allendorf, 1978). In a newly formed tetraploid genome, many multivalent pairing arrangements would be expected at meiosis (Ohno, 1970). These structures are formed by the pairing of multiple sets of homeologous (duplicated and

diverged sets of homologous) chromosomes. The fact that a few multivalent structures are still observed in present-day salmonids indicates that the process of diploidization is not yet complete.

Many duplicated gene pairs still exist as functional, non-allelic isoforms. For example, two isoforms of insulin (Kavsan et al., 1993), insulin-like growth factor (Wallis and Devlin, 1993) and MyoD (Rescan and Gauvry, 1996) have been identified. Among salmonine species, the growth hormone (GH) gene is also represented by non-allelic isoforms: GH1 and GH2 (Agellon et al., 1988a, 1988b; Agellon and Chen, 1986; Johanson et al., 1989; Male et al., 1992, Devlin, 1993; Du et al., 1993; Forbes et al., 1994, McKay et al., 1996). Although selective constraints have caused this gene pair to remain very similar in protein-coding regions, divergence of intronic and flanking DNA sequences indicates that the genes have been separate for a long time (Devlin, 1993). The accumulation of differences between GH1 and GH2 argues that the chromosomes or chromosomal regions on which they reside have completed the process of diploidization.

In Chapter 4, sequence analysis of GH intron D is used to examine the evolutionary history of these duplicated genes in salmonid genera. Analysis of a microsatellite locus nested within this intron (Chapter 3) revealed variation within and among species in the GH2 gene of *Oncorhynchus*, but not in any *Oncorhynchus* GH1 gene or in the GH genes of other salmonid genera. Further, new DNA sequences from intron D of the GH genes in brown trout (*Salmo trutta*), mountain whitefish (*Prosopium williamsonii*) and lake whitefish (*Coregonus clupeaformis*) were used to examine the evolutionary history and patterns of change of GH genes at the generic level. The two GH genes identified in the whitefish species could not be assigned to

the categories represented by the salmonine GH1 and GH2 isoforms, which suggests that the ancestral coregonine separated from the proto-salmonine lineage before the divergence of the GH1 and GH2 genes.

Chapter 2

Toward the Resolution of Pacific Salmon and Trout (*Oncorhynchus*) Phylogeny

Abstract:

The phylogeny of the genus *Oncorhynchus* has been studied previously using a variety of morphological and genetic characters. However, two unresolved systematic problems remain: the position of the masu salmon lineage (*O. masou*) and the relationships within the related group of species that contains sockeye (*O. nerka*), pink (*O. gorbuscha*) and chum (*O. keta*) salmon. Relationships among eight *Oncorhynchus* species and Atlantic salmon (*Salmo salar*) were examined using the nuclear growth hormone type-2 (GH2) and mitochondrial NADH dehydrogenase subunit 3 (ND3) DNA sequences. Phylogenies inferred using cladistic, distance and maximum likelihood approaches were concordant except where the branch leading to the Atlantic salmon outgroup joined the tree. The sequence data generated in this study were also combined with eight other morphological, allozyme and DNA character sets to perform a "total evidence" maximum parsimony analysis. In addition, all available DNA sequence data were combined in a maximum likelihood analysis. The same tree was inferred by both approaches. Strong support is provided that pink and chum salmon are sister species, and that the masu salmon lineage is distinct from those of the other Pacific salmon and trout, forming a sister taxon to the monophyletic North American Pacific salmon and trout lineage.

Introduction:

Historically, the presumed relationships among the Pacific salmon and trout (species designation listed in Table 2.1) have been the subject of considerable debate. Rainbow and cutthroat trout were originally grouped together with Atlantic salmon and brown trout in the genus *Salmo*. More recent work has led to the reclassification of rainbow and cutthroat trout as *Oncorhynchus* species (Smith and Stearley, 1989). The genus *Oncorhynchus* contains all Pacific salmon species, including masu and amago salmon, which are found only in Asia. *Oncorhynchus* is believed to have arisen from a single ancestral species derived from the *Salmo* evolutionary line. Neave (1958) proposed that the common ancestor of rainbow and cutthroat trout was the first to diverge from the proto-*Oncorhynchus* evolutionary line, which then radiated to form the seven extant Pacific salmon species.

Oncorhynchus phylogenies have been reconstructed from morphology, physiology, ontogeny, DNA-DNA hybridization, protein electrophoretic mobility variation, karyology, and DNA analysis (Utter et al., 1973 and references therein; Berg and Ferris, 1984; Thomas et al., 1986; Thomas and Beckenbach, 1989; Grewe et al., 1990; McVeigh and Davidson, 1991; Phillips and Pleyte, 1991; Shedlock et al., 1992, Devlin 1993; Murata et al. 1993,1996; Takasaki et al. 1994; Domanico and Phillips, 1995; Oohara et al., 1997). However, ambiguities still exist regarding the origins of masu salmon and, more generally, the branching order for the more basal lineages such as the common ancestors of the (rainbow, cutthroat) and (chinook, coho) clades. The relationships among sockeye, pink and chum salmon are also controversial.

DNA sequences of the nuclear growth hormone type 2 (GH2) and mitochondrial NADH Dehydrogenase Subunit 3 (ND3) genes have been examined previously in salmonid species (Table 2.1). In this study, a portion of the GH2 locus and the complete ND3 gene were sequenced from species where they had not been characterized, making it possible to examine the relationships among anadromous Pacific trout and all extant salmon species. The phylogenetic schemes inferred here were related to those of other studies to address recurring problems in the systematics of *Oncorhynchus*.

Materials and Methods:

Sample collections, DNA extraction and gene amplification

Species used in this study are listed in Table 2.1. DNA extracted from chum, amago, masu and Atlantic salmon liver samples was used to obtain sequence from the ND3 locus. GH2 sequences were amplified from cutthroat trout, chinook, coho, pink, masu and amago salmon. DNA was extracted from liver tissue according to the method of Devlin (1991). The concentration of DNA samples was determined with a Hoeffer DNA flourometer. The PCR and sequencing primers used (based on consensus sequences of salmonid species) are listed in Table 2.2 and their map positions are shown in Figure 2.1. PCR amplifications were carried out in 25-100 μL volumes containing 1X PCR buffer, (based on 'Medium' Buffer [Idaho Technologies] but with 1.5% w/v Ficoll), 6 ng/μL template DNA, 0.025 units/μL Taq polymerase (Bethesda Research Laboratories), 200 uM each deoxynucleotide-triphosphate (dNTPs), and approximately 1 pmol/μL of each amplification primer. Amplifications

Table 2.1. Species used in this study.

Species	Common Name	Origin[a]	Locus	Accession #
O. clarki	cutthroat trout	Coastal Cutthroat, Vancouver Island,	GH2[a] ND3[b]	Genbank U28156 NS
O. mykiss	rainbow trout		GH2[c] ND3[b]	Genbank J03797 NS
O. tshawytscha	chinook	Chilliwack Hatchery, B.C.	GH2[a] ND3[b]	Genbank U28157 NS
O. kisutch	coho	Chilliwack Hatchery, B.C.	GH2[a] ND3[b]	Genbank U28359 NS
O. nerka	sockeye		GH2[d] ND3[b]	Genbank U14535 NS
O. gorbuscha	pink	Weaver Creek Hatchery, B.C.	GH2[a] ND3[b]	Genbank U28360 NS
O. keta	chum	Weaver Creek Hatchery, B.C.	GH2[e] ND3[a]	NS Genbank U28365
O. masou masou	masu	Hokadate, Japan	GH2[a] ND3[a]	Genbank U28361 Genbank U28364
O. masou ishikawae	amago	Tamaki, Japan	GH2[a] ND3[a]	Genbank U28362 Genbank U28363
Salmo salar	Atlantic	Cultured, Sea Spring Salmon Farm, Chemainus, B.C.	GH2[f] ND3[a]	Genbank M21573 Genbank U28366

Note: NS, taken from reference and not located in database search.
[a]This study
[b]Thomas and Beckenbach (1989)
[c]Agellon et al. (1988)
[d]Devlin (1993)
[e]X. Shen, Y. Wang, M. Wett, D.Liu, and F.C. Leung, unpublished data
[f]Johansen et al. (1989)

were carried out primarily in a Perkin Elmer 9600 thermal cycler. Some amplifications were also carried out on Biometra and Idaho Technologies thermal cyclers.

PCR amplifications were performed with 30 cycles. Denaturation, annealing and extension times were varied according to the thermal cycler used and the size of the expected amplification product.

Primers (GH45 and GH47), designed to specifically amplify the GH2 gene, were based on conserved sequences from the promoter and terminator regions identified by the alignment of all available GH sequence data from several salmonid species. Other GH sequencing and PCR primers (Figure 2.1;Table 2.2) were designed based on intron D and flanking sequences of sockeye salmon GH1 and GH2 and, in the case of GH48-53, based on the alignment of all previously published GH2 intron D sequences.

Multiple amplification products were often observed when using GH primers with a genomic DNA template. To isolate GH2 sequences, a portion of the complete GH2 PCR product (from GH45 and GH47) was reamplified using internal primers GH7 and GH30, or GH7 and GH36. These reamplification products were compared to the amplification products from a genomic DNA template using agarose gel electrophoresis. In each case, the GH30 or 36 and GH7 product amplified from GH2 had the same electrophoretic mobility as one of the genomic DNA amplification products. Wherever possible, the genomic (GH36/GH7 or GH30/GH7) DNA amplification product corresponding to GH2 was isolated for cloning. In the case of chinook salmon, where the GH2-specific product could not be unambiguously distinguished from that of GH1 using agarose gel electrophoresis, the GH7/30 product reamplified from the GH2 PCR product was cloned.

A mitochondrial DNA fragment containing ND3 was amplified using primers (ARG and GLY) based on conserved regions of the genes for $tRNA^{ARG}$ and $tRNA^{GLY}$, which flank ND3 in vertebrate mitochondrial genomes. To facilitate the sequencing of ND3 from Atlantic salmon, for which the ARG primer worked poorly, the internal primers ND3A and ND3B, based on the alignment of all *Oncorhynchus* ND3 sequences, were subsequently designed (Table 2.2; Figure 2.1).

Table 2.2. PCR and sequencing primers used in this study

Primer	Sequence (5'->3')
GH7	CTTATGCATGTCCTTCTTGAA
GH8	TGTGGCCTTCAAGTGAATTC
GH9	TATACAGAATCTGACTGCAG
GH16	TTGTTAATCTTTGTGAAAA
GH30	TTTCTCTACGTCTACATTCT
GH36	GTCCTGAAGCTGCTCCG
GH45[a]	GTACGCGGCCGCC(C/G)GAACTCATGGAAAAATTC
	NotI
GH47[a]	GTACGCGGCCGCATGTACTAATCTAAAATGTC
	NotI
GH48	CAAT(G/T)ACCATTTGTGGT
GH49	CA(C/T)GCTCTACTACAGGTA
GH50	AC(A/G)CCTCAAAATA(A/C)GG(C/T)C
GH51	GTCAAGCTGATACAACTC
GH52	AGTGAAATACAACTATGC
GH53	ACAGAGAGAGATCGATGG
ARG[a]	ATGCGGATCCT(T/C)TTGAGCCGAAATCA
	BamHI
GLY[a]	ACGTGAATTCGTA(T/G)(A/G)(A/C)GTG(A/G)CTTCCAA
	EcoRI
ND3A	CAAAT(C/T)TC(C/T)CC(A/C)GACGCA
ND3B	CATTCTAAGCCTCCTTGGG

[a]The four nt at the 3' end plus the *NotI*, *BamHI* and *EcoRI* restriction sites (underlined) are not present in the template sequence

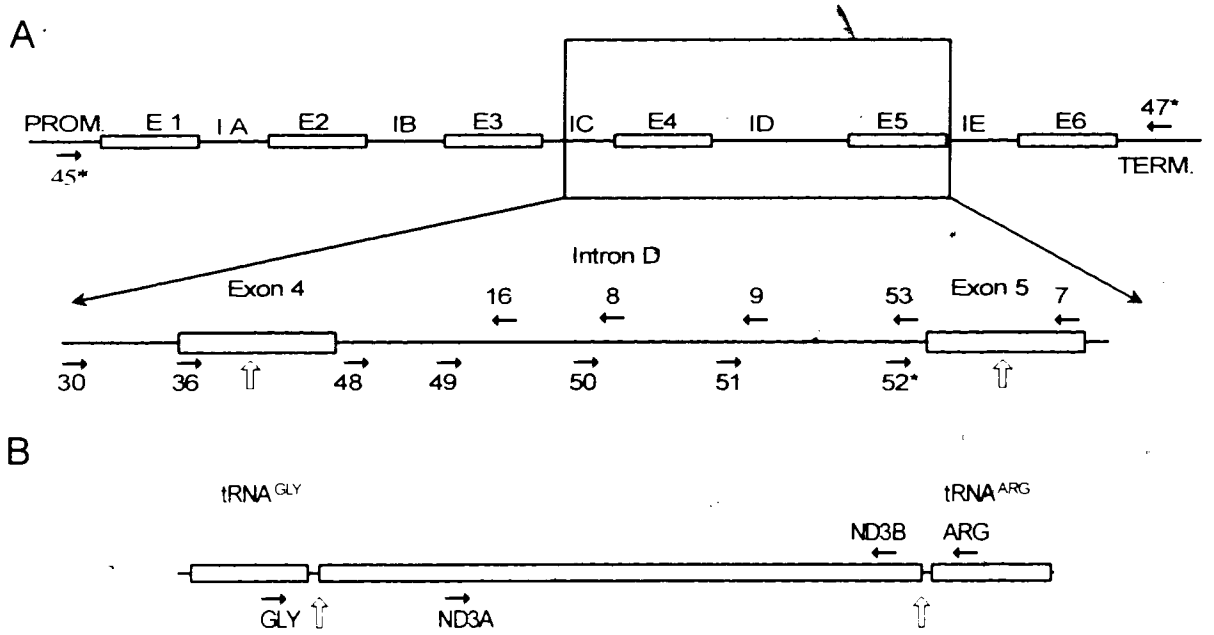


Figure 2.1. Map of the locations of GH2 and mitochondrial ND3 gene amplification and sequencing primers. Horizontal arrows represent the position of each primer. Open, vertical arrows delimit sequenced regions. A) Growth hormone loci. E1-5 are exons and IA-E are introns. Primers were designed from aligned GH1 and GH2 sequences, except for those marked with (*), which are GH2 specific. B) Mitochondrial ND3 sequence primers.

DNA cloning and sequencing

PCR amplification products to be cloned were purified by electrophoresis in low melting point agarose (Nusieve-GTG, FMC Biochemical), followed by isolation of DNA from excised bands using the Magic or Wizard PCRprep kits (Promega). The ND3 and GH2 amplification products were blunt-end cloned into pCRscript, a pBluescript derivative, using the pCRscript cloning kit (Stratagene). Sequencing of the clones was performed on both strands using the single- and double-stranded methods described in the Sequenase 2.0 sequencing kit (United States Biochemical Corp.). Various combinations of the primers described in Figure 2.1 and Table 2.2 were used in sequencing reactions. To compensate for the inherent error rate of Taq polymerase (Saiki et al., 1988; Tindall and Kunkel, 1988; Keohavang and Thilly, 1989) and possible differences due to allelism in heterozygous individuals, a minimum of two clones were sequenced for each species. Sequence differences between clones (usually single nucleotide differences) were encountered at a rate of about one per 520 bases. Ambiguities were resolved by direct sequencing of PCR products or by sequencing the region in question from a third clone and accepting the consensus between two of the three sequences. Raw sequence data were processed and assembled using PC\Gene (Intelligenetics, Mountain View, CA). The final DNA sequences have been submitted to Genbank (Accession numbers are listed in Table 2.1.)

Sequence and phylogenetic analysis of GH2 and ND3

In addition to the sequences determined in this study, published sequence data from other species (Table 2.1) were incorporated into the GH2 and ND3 data sets. Sequences were manually aligned using the Eyeball Sequence Editor (ESEE v1.09d;

Cabot and Beckenbach 1989). The sequenced GH2 fragment contained intron D plus 100 nt of 5' and 3' flanking exon sequence. The complete ND3 coding sequence was determined.

Cladistic, distance and maximum likelihood approaches to phylogeny reconstruction were used in this study to evaluate consistency among methods. Maximum parsimony analysis was performed using the DNAPARS program of the PHYLIP v3.5 package (Felsenstein 1993). Bootstrap analyses (2000 replicates) were performed with the taxon-input order randomized once for each replicate. Neighbor-joining bootstrap trees (Saitou and Nei, 1987) were constructed from Kimura 2-parameter (Kimura, 1980) corrected distance matrices using the NEIGHBOR program in PHYLIP v3.5. Maximum likelihood analysis was performed with DNAML in the PHYLIP package. To search for the best tree, the global rearrangement option was selected and the taxon-input order was randomized 10 times. To compare the likelihood values of alternative tree topologies, the user defined tree option was selected. With this option, DNAML performs a statistical analysis to determine whether the likelihood values of alternative trees are significantly worse than that of the best, or maximum likelihood tree (Kishino and Hasegawa, 1989).

Gaps introduced to maximize alignment of the GH2 intron sequence alignment were reduced to one site. Normally, gap sites can be scored as a character state in parsimony analysis but are ignored when calculating distance measures. In order to ensure that exactly the same data were considered with all methods of phylogeny reconstruction, each of the reduced gap sites was weighted equivalent to one transitional (G<-> A, or T<->C) change. The 100 nt of flanking 5' and 3' sequence determined in this study was retained in the GH2 data set.

Other DNA data sets

Recently, much of the mitochondrial genome has been sequenced from the nine species used in the present study. The complete mitochondrial control (D-loop) region (Shedlock et al., 1992) and complete or partial sequences of the ATPase 6, COIII, ND4L, $tRNA^{ARG}$ and $tRNA^{GLY}$ genes have been published (Thomas and Beckenbach, 1989; Domanico and Phillips, 1995; Oohara et al., 1997). All but the tRNA genes, which were not sufficiently variable for phylogenetic analysis at this taxonomic level, were reanalyzed in the present study. Analyzing each sequence as described above for ND3 and GH2 ensured consistency of methods.

In order to evaluate the performance of individual-gene data sets, sequences were used as reported except that distance-based and initial parsimony analyses were performed on each gene or region individually, rather than treating the entire contiguous region together as reported by Oohara et al. (1997). Sequence alignment of the protein-coding mitochondrial genes was unambiguous. A few sites involved in discrepancies discussed by Oohara et al. (1997) were removed. The D-loop sequence reported by Shedlock et al., (1992) had many small gaps introduced to maximize the alignment and was ambiguous in some regions. To avoid such ambiguities and comparison of non-homologous sites, all positions involved in insertions or deletions were removed from the data set.

Total evidence and maximum likelihood analysis of combined data sets

The criteria for inclusion of each character set in combined analyses were 1) availability of published data, 2) completeness (only character sets which included at least six taxa were used), 3) relevance to the branching order of the (sockeye, pink,

chum) clade (only data sets with all three taxa represented were used). Total evidence (Kluge 1989) analysis was performed on a pooled data set containing all informative sites from the ND3 and GH2 sequences identified in this study, as well as from morphological data (Stearley and Smith, 1993), protein variations (Utter et al., 1973; Tsuyuki and Roberts, 1963), DNA restriction site (Phillips et al., 1992) and sequence data (Shedlock et al., 1992; Thomas and Beckenbach, 1989; Oohara et al., 1997). All data were converted to the same notation by encoding character states from morphological data as 0=G, 1=A, 2=T; presence/absence restriction site and protein electrophoretic mobility variant data as "+" =G and "–" =A. For the single gene (and D-loop) data sets, the DNA-based phylogenetic analysis (described above) included only sites represented in all nine taxa. The DNA sequence of the full-length GH2 genes of chinook and masu salmon were also determined (Appendix 1). The sequence of the entire gene is also known for Atlantic, chum and sockeye salmon and rainbow trout. (Table 2.1). The new GH2 data was added to the partial GH2 sequences for the remaining three species (with gap sites reduced as described above). The expanded data set was used only in combined analyses. The total evidence phylogeny was inferred using parsimony analysis (DNAPARS) as described above. DNAML was used to infer the maximum likelihood tree of the combined data set (all DNA sequence data pooled) and to compare the likelihood values of different trees using single gene data sets and various combined data sets.

Results:

Masu and amago are virtually identical at the DNA sequence level

Masu and amago salmon have been considered either distinct species (Kato, 1991) or conspecific races (Kimura, 1990). The surprising finding that their ND3 genes are identical at the DNA sequence level (Figure 2.2), and that their GH intron D (GH2ID) sequences (Figure 2.3) are almost so, is not compatible with a long separation of these two types of salmon. The relationship between masu and amago salmon is discussed in chapter 3. For the purpose of the phylogenetic analyses presented here, the masu salmon DNA sequences were used to represent the (masu, amago) lineage.

Insertion/deletion patterns in GH intron D

The total aligned length of the GH2 sequence fragments used in this study was 1406 nt. Individual sequences ranged from 635 to 1376 nt in length due to numerous insertion or deletion sites (Figure 2.3). GH1 and GH2 are duplicated, paralogous genes, presumably resulting from the tetraploidization of the ancestral salmonid genome (Ohno, 1970; Allendorf and Thorgaard, 1984). The GH1 and GH2 lineages are clearly distinct and the two genes display little evidence of recent intergenic recombination after their divergence (Devlin 1993). This is consistent with the fact that several deletion sites are common to all GH2 intron sequences of *Oncorhynchus* species examined here, but absent in the GH1 introns from chinook, Atlantic and sockeye salmon (Figure 2.4).

Gaps revealed by sequence alignment of the intron show that such events are common in the evolution of these sequences (Devlin, 1993).

Figure 2.2. Aligned DNA sequence of the mitochondrial ND3 gene for nine salmonid taxa. Codon triplets are separated by spaces. Dots (.) indicate nucleotide identity with the initial sequence, Atlantic salmon. Species designations are listed in Table 3.1.

```
Atlantic  ATG AAC TTA ATT ACA ATA ATT ATT GCT ATT ACC ATT ACA CTA TCG GCA GTA CTA GCC ACT ATT TCC TTC TGA CTA CCA CAA ATA ACG CCC  90
sockeye   ... ... ... G.. ... .C. ..C ..C A.. ... ... ..C ... ..G ..C ... ... ... ... ... ... ..T ... ... ... ... ... ..T T.C ...
chum      ... ... ... ... ... .C. ..C ... A.. ..C ... ..C ... T.G ..C ... ... ... ... ... ..C ..T ... ... T.. ... ... ..C T.C ...
pink      ... ... C.. ... ... .CG ..C ... A.. ... ... ..C ... ..G ..C ... ... ... ... ... ..C ..T ... ... T.G ... ... ..T T.C ...
chinook   ... ... ... ... ... .C. ..C ..C A.. ... ... ..C ... T.G ..C ... ... ... ... ... ... ..T ... ... T.. ..T ... ..C T.C ..A
coho      ... ... ... ... ... .C. ..C ..C A.. ..C ... ... ... ..G ..T ... ... ... ... ... G.. ..T ... ... ... ... ... ..C T.C ...
rainbow   ... ..T ... ..C ... .C. ..C ..C A.. ... ... ..C ... T.. ..C ... ... ... ... ... ... ..T ... ... T.. ... ... ..C T.C ..A
cutthroat ... ... ... ... ... .C. ..C ..C A.. ..C ... ..C ... T.G ..C ... ... ... ... ... ... ..T ... ... T.. ... ... ..C T.C ..A
masu      ... ... ... ... ... .C. ... ..C A.. ... ... ..C ... ... ..T ... ... ... ... ... ... ..T ... ... T.. ... ... ..T T.T ..A

Atlantic  GAC GCA GAA AAA CTA TCC CCC TAC GAA TGT GGC TTC GAT CCC CTA GGA TCC GCC CGC CTA CCC TTC TCC CTG CGC TTC TTT CTA ATT GCC  180
sockeye   ... ... ... ..G T.. ... ... ... ... ... ..A ..T ..C ... T.G ... ... ... ..T ..C ... ... ... T.A ..A ..T ... ... ..C ...
chum      ... ... ... ..G T.G ..T ... ..T ..G ..C ..A ..T ..C ..A ... ..G ... ... ... ..C ... ... ..T T.A ... ... ... T.. ... ...
pink      ... ... ... ..G T.. ... ... ... ... ... ..A ... ..C ... T.G ..G ..T ... ..T ... ... ... ..T T.A ... ... ... ... ..C ...
chinook   ... ... ..G ..G T.. ... ... ... ... ... ..A ..T ... ..A T.. ..G ... ... ... ... ... ... ... T.A ... ... ... ... ..C ...
coho      ... ... ..G ... ..G ... ... ... ... ... ..A ..T ... ..A T.. ..G ... ... ... ..G ... ... ... T.A ... ... ... ... ..C ...
rainbow   ... ... ..G ..G T.. ... ... ... ... ... ..A ..T ..C ... T.. ..G ... ... ... ..G ... ... ... T.A ... ... ... ... ..C ...
cutthroat ... ... ..G ..G T.. ... ... ... ... ... ..A ..T ..C ... ... ... ... ... ... ..G ..T ... ... T.A ... ... ... ... ..C ..T
masu      ... ... ..G ... T.G ... ... ... ... ... ..A ..T ..C ..T T.. ... ... ... ..T ... ... ... ... T.A ... ... ..C ... ..C ...

Atlantic  ATT CTA TTT CTC CTA TTT GAT CTA GAA ATC GCC CTC CTA CTC CCC CTT CCC TGA GGA GAT CAA CTC ACC ACC CCA GCC CTC ACA CTT GCC  270
sockeye   ..C ..G ..C ... ..G ..C ..C ... ... ... ... ..T ..T ..G ... ..G ..T ..G ..G ... ..G ... .AT G.. ..G A.. ..A ... ..C CT.
chum      ..C ..C ... ... ... ... ... ... ... ..T ... ... ..T ... ... T.A ..T ..G ..G ... ... ... .GT ... ... A.. ..G ... ... ATT
pink      ..C ..C ..C ... ... ..C ... ... ... ..T ... ... ..C ... ... ..A ... ... ..G ... ..G ... .AT ... ... A.T ..A ... ... ATT
chinook   ..C ... ..C ... ... ... ... ... ... ..T ... ... ..T ..A ..T ..A ..A ... ..G ... ..G ... GA. ... ... A.. T.A ... ... ATT
coho      ..C ... ..C ..A ... ... ... ... ... ..T ... ... ..T ..A ..T ..G ..A ... ... ... ..G ... .A. ... ..T A.. T.G ... ..C .TT
rainbow   ..C T.. ..C ... ... ... ... ... ... ... ... ... ..T T.G ... ..A ..T ... ..G ... ... ... CA. ... ..G A.. ..G ... ..C AT.
cutthroat ..C ... ..C ... ... ... ..C ... ... ..T ... ... ..T T.A ... ..G ..T ... ..G ... ... ... .A. ... ... A.. ..A ... ... AT.
masu      ..C ..G ..C ..T ... ... ..C T.G ... ... ... ... ..T ..A ... ..G ..T ... ... ... ... ... .A. ... ..C ... ..A ... ..C .T.

Atlantic  TGA TCC GCT GCC GTA CTT GCT CTC CTC ACT CTT GGC CTA ATC TAT GAA TGG ACC CAG GGA GGC CTT GAA TGA GCC GAA TAG  351
sockeye   ... ... A.. ... ... ..C ..C ... ..T ... ... ... T.. ..T ... ..G ..A ... ..A ... ... T.A ... ... ... ... ..A
chum      ... ... A.. ... ... ..C ..C ... ..T ... ... ... T.. ..T ... ..G ..A ... ..A ... ... T.G ... ... ... ... ...
pink      ... ... A.. ... ... ..C ..C ... ..T ... ... ... T.. ..T ... ..G ..A ... ..A ... ... T.A ... ... ... ... ..A
chinook   ... ... A.. ... ... ..C ..C ... ..T ... ... ... T.. ..T ... ... ..A ... ..A ... ... T.A ... ... ... ... ..A
coho      ..G ... A.. ... ... ..C ..C ... ..T ... ... ... T.. ..T ... ... ..A ... ..A ... ... T.A ... ... ... ... ..A
rainbow   ... ... A.. ... ..T ..A ..C ..T ..T ... ... ... T.. ... ... ... ..A ... ..A ... ... T.A ... ... ... ..G ..A
cutthroat ... ... A.. ... ..T ..A ..C ..T ..T ... ... ... T.. ... ... ... ..A ... ..A ... ... T.A ... ..G ... ... ..A
masu      ... ... A.. ..T ... ... ..C ... ..T ... ..A ... T.. ... ... ... ..A ... ..A ... ... T.A ... ... ... ... ...
```

Figure 2.3. Aligned nucleotide sequence of a portion of the GH2 locus from nine salmonid taxa, comprised of intron D and portions of flanking exons. Dots (.) indicate nucleotide identity with the initial sequence, Atlantic salmon. Dashes (-) represent gaps introduced to produce optimal sequence alignment. 100 nt each at the 3' and 5' ends are coding sequence from exons four and five, respectively.

```
Atlantic  CCAGACCCTG GCCATCTCCA ACAGCCTAAT GGTCAGAAAC TCCAACCAGA TCTCTGAGAA GCTCAGCGAC CTCAAAGTGG GCATCAATCT GCTCATCAAG  100
sockeye   .......... A......... .......... .......... .......... .......... .......... .......... .......C.. .......G..
chum      .......... A......... .......... .......... .......... .......... .......... .......... .......C.. .......G..
pink      G......... A......... .......... .......... .......... .......... .......... .......... .......C.. .......G..
chinook   .......... A......... .......... .......... .A........ ....C..... .......... .......... .......C.. ..........
coho      .......... A......... .......... .......... .......... ....C..... .......... .......... .......C.. ..........
rainbow   .......... A......... .......... .......... .......... .......... .......... .......... .......C.. ..........
cutthroat .......... A......... .......... .......... .......... .......... .......... ........T. .......C.. ..........
masu      .......... A......... .......... .......... .......... .......... A......... .......... .......C.. ..........

Atlantic  GTAAAG--AA AGGAGGGAGA ACAATGACCA TTTGTGGTGC CGCACTTTGT GCACTGTAAA CCACAAGGCA -TTTTTAACT CAAATACTTC TAGTAAGTTG  200
sockeye   ....T.GT-- ---------- -....T.... .......... .......... ...------- ---------- .......... .......... ..........
chum      ....T.GT-- ---------- -....T.... .......... .......... ...------- ---------- .......... .......... ..........
pink      ....T.GT-- ---------- -....T.... .......... .......... ...------- ---------- .......... .......... ..........
chinook   ....T.GT-- ---------- -....T.... .......... .......... ...------- ---------- .......... .......... ..........
coho      ....T.GT-- ---------- -....T.... .......... .......... ..G------- ---------- .......... .......... ..........
rainbow   ....T.GT-- ---------- -......... .........A .......... CG.------- ---------- .......... .......... ..........
cutthroat ....T.GT-- ---------- -....T.... .C.......A .......... ...------- ---------- .......... .......... ..........
masu      ....T.GT-- ---------- -....T.... .......... .......... ...------- ---------- .......... .......... ..........

Atlantic  AACTCAGTCA ATGAAAAGTC ATTATTACTT AAAATGTCTA TGTGGTACTG GCTCAAATCT AAATGAGTCA CATTAATGCA ATTTTTTTAA GTTATAACAA  300
sockeye   ..G....... .......... .......... C......... .......... .......... .......... ...C...... .......A.. ..........
chum      ..G....... .......... .......... C......... .......... .......... .......... ...C...... .......G.. ..........
pink      ..G....... .......... .......... .......... .......... .......A.. .......... ...C...... .......G.. ..........
chinook   ..G....... .......... .......... C......... .......... .......... .......... ...C...... .......A.. ..........
coho      ..G....... .......... .......... CG........ .........A .......... .......G.. ...C....AT .......A.. ..........
rainbow   ..G....... .......... .......... C......... .......... .......... .......... ...C...... .......A.. .......G..
cutthroat ..G....... .......... .......... C......... .......... .......... .......... ...C...... .......A.. ......G...
masu      ..G....... .......... ...G...... C......... .......... .......... .......... ...C...... .G.....A.. ..........

Atlantic  ATTAACTTTT TACCCAGCAT GCTCTACTAC AGGTATATTT TTTGGAATTG TTTTTAATAT CTGTGTTTTT GCATGTA--- ----CTATTT GCATTTTTGC  400
sockeye   .......... .........C .......... .....G.... .......... .......... .......... .......CAG TACA------ ----------
chum      .......... .......... .....G.... .....G.... ..-....... ....-..... .......... .......CAG TACA------ ----------
pink      .......... .......... .......... ..T..G.... ..-....... .......... .......... ...C...CAG TACA------ ----------
chinook   ..G....... .......... .......... .....G.... .......... .......... .......... .......CAG TACA------ ----------
coho      .......... A......... .......... .....G.... .......... .......... .......... ...C...CAG TACA------ ----------
rainbow   ..G....... .......... .......... .....G.... .......... .......... .......... .......CAG TACA------ ----------
cutthroat ..G....... .......... .......... .....G.... .......... .......... .......... .......CAG TACA------ ----------
masu      ..G....... .......... .......... .....G.... .......... ....G..... .......... .......CAG TACA------ ----------

Atlantic  ATTGAGTGAT TGATTGATT- -------AAT TTTATGCTTC ACACAGATAT ATAACATACA TTTTTCTACG TTTTCACAAA GATAAATAAC A-------TA  500
sockeye   -......... .....C...C ATT----C.. C.......A. .......... .....G.... .........A .......... .......... .TACAAGG..
chum      -......... .........G ATTGATTC.. C.......A. .......... .....G.... ..G......A .......... .......... CTACAAGG..
pink      -......... .........G ATT----C.. C.......A. .......... .....G.... ..G......A .......... .......... CTACAAGG..
chinook   -......... .....----. -------G.. C.......A. .......... .....G.... .........A .......... .......... .TACAAGG..
coho      -......... .....----. -------C.. C.......A. .......... .....G.... .........A .......... .......... .TACAAGGA.
rainbow   -......... .........G ATT----C.. C.......A. .......... ..G..G.... .........A .......... .G........ .TACAAGG..
cutthroat -......... ........GG ATT----C.. C.......A. .......... .....G.... .........A .......... .......... .TACAAGG..
masu      -......... .........- -------C.. C.......A. .......... .....G.... .........A .......... .......... .TACAAGG..
```

Figure 2.3. (Cont'd)

```
Atlantic  CCGGAATTTT GCAAACCTAC TTGCAGGCCT GATGTGGCCT GTAAA-CCAT GAGTTTCAAG GCCACTGTAT TAGGGTAAAG CTACACCTCA AAATAAGGCC  600
sockeye   .T........ .......... .......... .......... ......... ........-. .......... .T........ .......... ..........
chum      .T........ .......... .......... .......... ......... ........-. .......... .T........ .......... ..........
pink      .T........ .......... .......... .......... ......... A.A------- -...-..AG. .T........ .......... T.........
chinook   .......... .......... .......... .......... ......... ........-. .......... .T........ .......... ...G......
coho      .......... .......... .......... .......... .....A.... ........-. .......... .T........ .......... .........A
rainbow   ...A...... .......... .......... .......... ......... ........-. .......... .T...A.... .......... ..........
cutthroat .......... .......... .......... .......... ......... ........-. .......... .T...A.... .......... ..........
masu      .......... .......... .......T.- ---------- -----.---- ---------- ---------- ---------- ---------- ----------

Atlantic  TT-ATGAGAT ATGTAATATA TTGTTATAAA GAGTTTAACT ATAATGATAA TATTTGCCTA GAAAATCACT TGATGGCCAC AGGACTGAAA ATGAATGACA  700
sockeye   .....A.... .......... .......... .......... .......... .......... .......... ...A...... .......... ..T.......
chum      ..T..A.... .......... .......... .T........ .......G.. .......... .......... ...A.TT... .......... ..T.C.....
pink      .....A.... .......... .......... .......... .......... .......... .......... ...A.TT... ....A..... ..T.......
chinook   .....A.... .......... .......... .......... .......... .........G .......... ...A...... .......... ..T.......
coho      .....A.... .......... .......... .......... .......... .......... ....-..... ...A.....A .......... ...T......
rainbow   .....A.... .......... .......... .......... .......... .......... .......... ...A...... .......... ..T.......
cutthroat .....A.... .......... .......... .......... .......... .......... .......... ...A...... .......... ..T.......
masu      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

Atlantic  ACAAACATGT CTCTGTCGCT AACAAATACA GTCATGGGTG ATAACTCGAC AATTCACTCA AAAGGCAAGG CACACTTGGA AATTATATTG GAGACATGGC  800
lake      .......... .......A.. .......... .......... .......T.. ....G..C.. .......... .......... .......... ..........
sockeye   .........- ---------- ---------- ---------- .......T.. .......... .......... .......... .......... ..........
chum      ...G.....- ---------- ---------- ---------- .......T.. .......... .......... .......... .......... ..........
pink      .......--- ---------- ---------- --------.. .......T.. .......... .......... .......... .......... ..........
chinook   .........- ---------- ---------- ---------- .......T.. .......... .......... .......... .......... ..........
coho      .........- ---------- ---------- ---------- .......T.. .......... .......... .......... .......... ....T.....
rainbow   .........- ---------- ---------- ---------- ....T..T.. .......... .......... .......... ..G....... ........T.
cutthroat .......A.- ---------- ---------- ---------- .......T.. .......... ...------- ---------- ---------- ----------
masu      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

Atlantic  ATTAGTGCGG GCATTACTAA TAAATGTCAA GCTGATACCA CTCAAATCTC AACC-TCTAC AGGGTGACTC TATAGGTTTG AGTAATGACT ATAAAAATCA  900
sockeye   -......G.. .......... A......... .......... .......... ....C..... ....C..... .......... .......... ..........
chum      -......G.. .......... A......... .......... .......... ....C..... ....CA.... .......... .......... ..........
pink      -......G.. .......... A......... .......... .......... ....C..... ....C..... ...G...... .......... ..........
chinook   -......G.. .......... A......... .......... .......... ....C..... ....C..... .......... .......... ..........
coho      -......G.. .T........ A......... .......... ---------- -...C..... ....C..... .......... .......... .......C..
rainbow   -......G.. .......... A......... .......... .......... ....C..... ....C..... .......... .......... ...-......
cutthroat ---------- ---------- ---------- ---------- ---------- ----.----- ---------- ---------- ---------- ----------
masu      ---------- ---------- ---------- ---------- ---------- ----.----- ---------- ---------- ---------- ----------

Atlantic  CTTTAAGTAA CTGTAGTCAG ATTCTGTTTA TTAAGTGCAA CGGTTTCCTC AAAAGTTTTG AGTAATGACA GCACATTGGG GTTTACAGTG TGGTTATTAT 1000
sockeye   ........G. .......... .......A.. .......... .......... .......... .......... ...T...... .......A.. ..........
chum      ........G. .......... .......A.. .......... .......... .......... .......... .A.C..C... ...T...... .......A.. ..........
pink      .....C..G. .......... .......A.. .......... .......... .......... .A........ ...T...... .......A.. ..........
chinook   ......A.G. .......... .......A.. .......... .......... .......... .......... .......... .......A.. .........A
coho      ........G. .......... .......A.. .......... ..T-...... .......... .......... .......... .......A.. ..........
rainbow   .......CG. .......... .......A.. .......... .......... .......... .......... .......... .......A.. .....G....
cutthroat ---------- ---------- ---------- ---------- ---------- ---------- ---------- -......... .......A.. ..........
masu      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
```

Figure 2.3. (Cont'd)

```
Atlantic  CTTCCACTGA CATGAAAGTG AAATACAACT ATGCTTTCCT AGTTAGAAAG CATAGTGTAG GACTACGTAC GAGGTCTTCT CAGCAGATCT TTCAGTGCTT  1100
sockeye   .......... .......... .......... .........- ---------- ---------- ---------- ---------- ----------
chum      .......... .--------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
pink      .......... .--------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
chinook   .......... .......... .......... .......... .......... .......... ...C.....T .-CC...... .......... .....C....
coho      .......... .......... .......... .......... .......... .......... ...C.....T .-CT...... .......... .....C....
rainbow   .......... .......... .......... .......... .......... .......... ...C.....T .-TC...... .......... .....C....
cutthroat .......... .......... .......... .......... .......... .......... ...C.....T .-TC...... .......... .....C....
masu      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

Atlantic  TACATTGTGA TGTGGTAACT CACCTCATAT ATATAGTCAC TAATAGTGAC TATATCAGTA ACACCCCATT CAATGACTGA ATATTGTCCC ATTCAAGGAC  1200
sockeye   ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
chum      ---------- ----------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
pink      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
chinook   .......... ..G....... ........TG ..--CA.... .......... .......... .......... .......... ......G... .....T...
coho      .....C.... C.G.T..... ........CG ..--CA.... .......... .......... .......... .......... ......G... ..........
rainbow   .C........ C.G....... ........CG ..--CA.... .......... .......... .......... .......... ......G... ..........
cutthroat .......... C.G....... ........CG ..--CA.... .......... .......... .......... .......... ......G... ..........
masu      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

Atlantic  ATCTATGCAT --GTCTTTTG CTATATGTGC TTTCTGAATG GCCCAATAAA CAAATATTGA TATGCACGCA TCCACCCCAC CATGCATCTC TCTCTGTCTC  1300
sockeye   ----...... GT.......T .......... ..GTA..... ...A...... ...G...... .......A.. .......... .......... ..........
chum      ---------- ..-------- ---------- ---------- ---------- ---------- ---------- --....A... .......... ..........
pink      ---------- ..-------- ---------- ---------- ---------- ---------- ---------- --....A... .......... ..........
chinook   ..T....... GT........ .......... ..GTA..... ...A...... .G.G...... .......A.. .......... .......... ..........
coho      ..T....... GT........ .......... ..GTA..... ...A...... ...G...... .......A.. .......... .......... ..........
rainbow   ..T....... GT........ .C........ ..GTA..... ...A...... ...G.....T .......A.C .......... .......... ..........
cutthroat ..T....... GT........ .C........ ..GTA..... ...A...... GT.G.....C .......A.C .......... .......... ..........
masu      ---------- ---------- ---------- ---------- ----...... ...G...... ...C...A.. .......... .....G.... ..........

Atlantic  CCACAGGGGA GCCAGGATGG CGTACTGAGC CTGGATGACA ATGACTCTCA GCATCTGCCT CCCTACGGGA ACTACTACCA GAACCTGGGG GGCGATGGCA  1400
sockeye   .......... .......A.. .A........ .......... .......... .........C .......... .......... .......... .....C....
chum      .......... .......A.. G......... .......... .......... .........C .......... .......... .......... .....C....
pink      .......... .......A.. .......... .......... .......... .........C .......... .......... .......... .....C....
chinook   .......... .......... .......... .......... .......... .........C .......... .......... .......... .....C....
coho      .-........ .......... .......... .......... .......... .........C .......... ......T... .......... .....C..
rainbow   .......... .......... .......... .......... .......... .........C .......... .......... .......... .....C..
cutthroat .-........ .......... .......... .......... .......... .........C .......... .......... .......... .....C..
masu      T......... .......... ..A....... .......... .......... .........C .......... .......... .......... .....C..

Atlantic  ACATCA 1406
sockeye   ..G...
chum      ..G...
pink      ..G...
chinook   ..G...
coho      ..G...
rainbow   ..G...
cutthroat ..G...
masu      ..G...
```

Table 2.3. Pair-wise Kimura 2-parameter distance comparisons (in percent) based on sequence data. ND3 distances are above the diagonal, GH2 below. GH2 distances were calculated from sequence used in phylogenetic analysis: all gaps were reduced to one site and weighted equivalent to one transition.

	sock.	chum	pink	chin.	coho	rain.	cutt.	masu	amago	Atla.
sockeye		13.5	9.2	10.9	12.3	11.3	10.9	11.6	11.6	20.5
chum	1.88		10.0	10.2	13.1	13.0	11.6	16.8	16.8	19.3
pink	2.23	1.36		8.6	13.1	12.7	12.0	15.0	15.0	19.0
chinook	1.71	2.93	3.28		6.3	7.6	6.9	11.7	11.7	18.0
coho	3.10	4.35	4.34	2.40		11.2	10.6	11.3	11.3	17.2
rainbow	2.40	3.63	3.98	2.05	3.81		5.7	11.3	11.3	18.8
cutthroat	1.88	3.10	3.45	1.53	2.92	1.86		10.6	10.6	18.1
masu	2.58	3.63	4.16	2.23	3.99	2.90	2.40		0	16.1
amago	2.23	3.28	3.81	1.88	3.63	2.55	2.05	0.68		16.1
Atlantic	4.90	5.99	5.98	4.72	6.36	5.25	5.07	5.81	5.44	

```
Atlantic    GH2 ...AGTTGAAGTCA--GTCAATGAAA...//...TCTAAATGAG---TCACATTAAT...
sockeye     GH2 ...AGTTGAAGTCA--GTCAATGAAA...//...TCTAAATGAG---TCACATCAAT
chum        GH2 ...AGTTGAAGTCA--GTCAATGAAA...//...TCTAAATGAG---TCACATCAAT
pink        GH2 ...AGTTGAAGTCA--GTCAATGAAA...//...ACTAAATGAG---TCACATCAAT
chinook     GH2 ...AGTTGAAGTCA--GTCAATGAAA...//...TCTAAATGAG---TCACATCAAT
coho        GH2 ...AGTTGAAGTCA--GTCAATGAAA...//...TCTAAATGAG---TCACATCAAT
rainbow     GH2 ...AGTTGAAGTCA--GTCAATGAAA...//...TCTAAATGAG---TCACATCAAT
cutthroat   GH2 ...AGTTGAAGTCA--GTCAATGAAA...//...TCTAAATGAG---TCACATCAAT
masu        GH2 ...AGTTGAAGTCA--GTCAATGAAA...//...TCTAAATGAG---TCACATCAAT
chinook     GH1 ...AGTTGAAGTCAAGGTCAATGAAA...//...ACTAAATGAGAAGTCACATCAAT
Atlantic    GH1 ...AGTTGAAGTCAAAGTCAATGAAA...//...ACTAAATGAGAAGTGACATCAAC
sockeye     GH1 ...AGTTGAAGTCAAAGTCAATGAAA...//...ACTAAATGAGAAGTCACATCAAT...
```

Figure 2.4. Insertion or deletion sites in GH1 and GH2 intron D sequences. Dashes (-) represent gaps introduced to produce optimal sequence alignment. The presence of gaps specific to the GH1 or GH2 isoforms reveals that the two loci have been separate since before the divergence of Pacific and Atlantic salmon.

Shared, derived (synapomorphic) deletions of identical length and position involving two or more but less than (n-2) taxa can be used as phylogenetically informative character states. For example, pink and chum salmon share gaps not present in other taxa (nt positions 343, 1011-1272), supporting a close relationship between the two species.

Phylogenetic inference using mitochondrial and nuclear sequences

To evaluate consistency among methods and between data sets, three approaches to phylogeny reconstruction (maximum parsimony, maximum likelihood and neighbor-joining distance analyses) were used. For each data set except ND4L, all three methods produced the same trees (Figure 2.5). With the exception of the placement of the outgroup, there was good agreement between gene trees except for the D-loop and ND4L. Bootstrap testing was performed with 2000 replicates for both the neighbor-joining and parsimony methods. The bootstrap confidence levels (BCLs), shown at the nodes in phylogenetic trees, represent the percentage of replicates in which that particular node or branch-point occurred. The BCL values tended to be higher at terminal nodes, providing support for the species pairs (chinook, coho), (masu, amago), (cutthroat, rainbow) and the group (sockeye, (pink, chum)). The consistent monophyly observed with (rainbow, cutthroat) and (chinook, coho) clades is also well supported by previous phylogenetic analysis (Table 2.4). The previously controversial grouping (sockeye, (pink, chum)) (Table 2.4; Stearley and Smith, 1993) is well supported by most inferred trees.

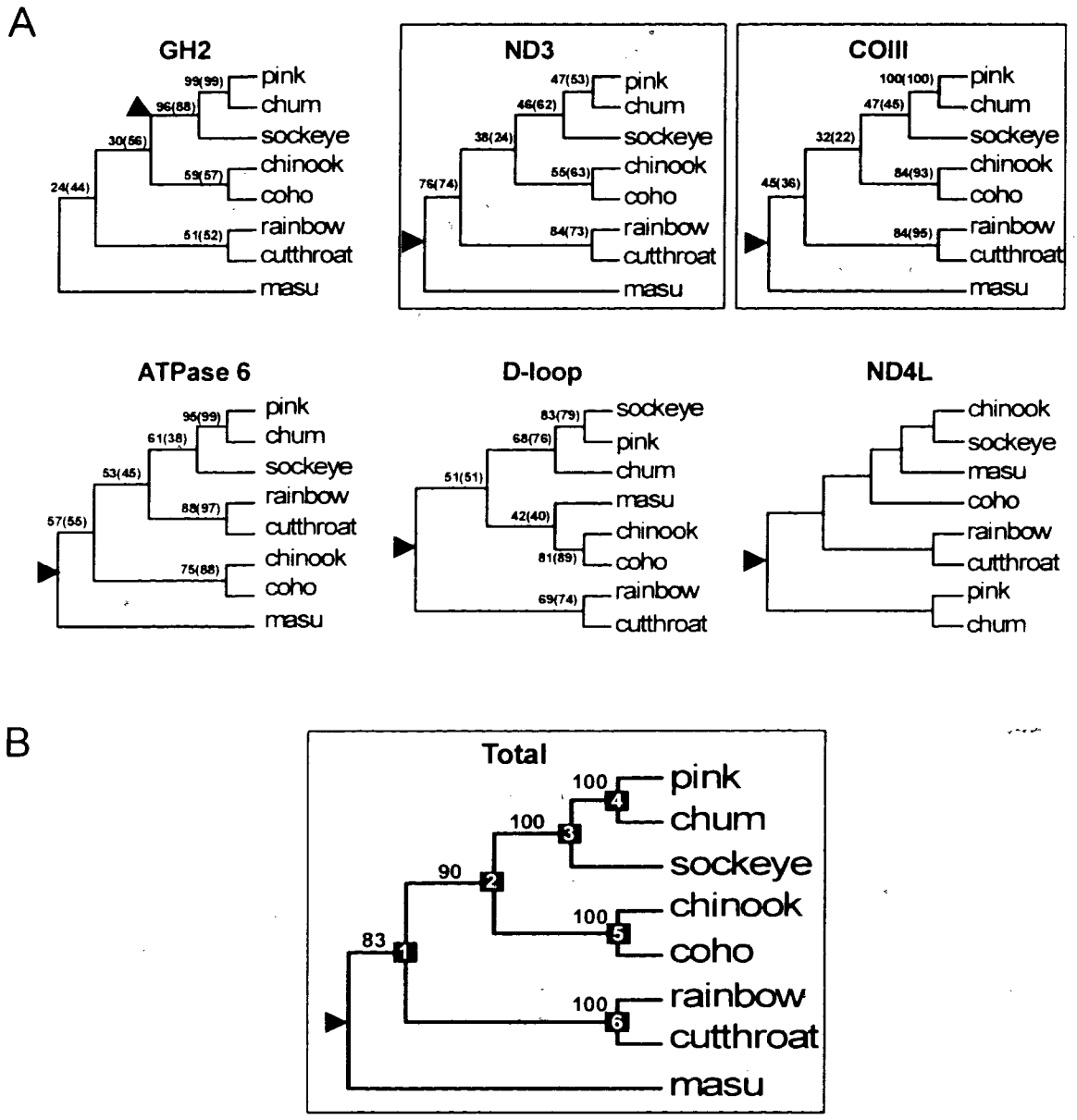


Figure 2.5. Congruent *Oncorhynchus* trees from three methods of phylogenetic inference. Arrowheads indicate the position where the branch leading to the outgroup joins the tree. The outgroup was Atlantic salmon. Parsimony and neighbor-joining (in parentheses) bootstrap confidence levels (BCLs) are given at the relevant nodes. A) Individual genes. Except for ND4L, each data set produced identical neighbor-joining, maximum parsimony and maximum likelihood trees. The ND4L tree is the majority-rule consensus of the three methods. B) The total evidence tree with BCL values. The tree was produced by maximum parsimony analysis of 10 pooled character sets, including the DNA sequences used to generate the other trees in this figure. Maximum likelihood analysis of all DNA sequences in a pooled data set produced an identical tree. Note that the same tree (boxed) was recovered for ND3, COIII and the total evidence analysis. The nodes are numbered to facilitate discussion.

Table 2.4. Phylogenetic studies of *Oncorhynchus*. Nodes refer to the total evidence tree (Figure 2.5)

Node	Supporting	Conflicting
1	1,19	2,3,4,5,6,7
2	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16	19
3	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,19	
4	1,2,8,10,11,16,19,20	3,5,6,7,18,12,13,14,18
5	1,2,3,6,7,8,10,11,12,13,14,15,16	4
6	1,2,3,5,6,8,12,13,15	4

1, this study (ND3, COIII); 2, this study (GH2); 3, Smith and Stearley (1989); 4, Stearley and Smith (1993); 5, Shedlock et al. (1992); 6, Phillips and Pleyte (1991); 7, Hikita (1963); 8, Grewe et al. (1990); 9, Tsuyuki and Roberts (1963); 10, Murata et al. (1993); 11, Thomas et al. (1986); 12, Thomas and Beckenbach (1989); 13, Utter et al. (1973); 14, Tsuyuki and Roberts (1966); 15, Gorshkov and Gorshkova (1981); 16, Murata et al., (1996); 17, McVeigh and Davidson (1991); 18, Simon (1963); 19, Oohara et al. (1997); 20, Domanico and Phillips (1995)

Table 2.5. The contribution of each character set to the phylogenetic analysis. Bootstrap confidence levels (BCLs) are shown for each node in the total evidence tree (Figure 2.5). The effect of removing each character set from the combined parsimony analysis can be seen by the change in the BCLs.

	TOTAL	ND4L	ATPase	COIII	ND3	D-LOOP	GH2	OTHER[a]
#Sites[b]	420	19	96	84	52	51	52	66
Node 1	83	85	56	64	59	93	83	99
Node 2	90	90	97	90	87	93	95	38[c]
Node 3	100	100	100	100	100	100	100	100
Node 4	100	100	93	83	100	100	99	100
Node 5	100	100	100	100	100	100	100	100
Node 6	100	100	99	100	100	100	100	100

[a]This character set was assembled from morphological (Stearley and Smith, 1993), allozyme (Utter et al., 1973; Tsuyuki and Roberts, 1963), Ribosomal DNA restriction site (Phillips et al., 1992), and SINE repeat element insertion site data (Murata et al., 1993, 1996)

[b]Refers to the number of phylogenetically informative (synapomorphic) characters used by parsimony analysis

[c]The indicated BCL refers to a node not included in the bootstrap consensus tree for this partial data set. The tree recovered was identical to the ATPase 6 gene tree (Figure 2.5).

To resolve the rooting of the *Oncorhynchus* phylogenetic tree and address ambiguities in the systematics of the (sockeye, pink, chum) group, data from other studies were used in combination with the GH2 and ND3 data sets to construct a total evidence estimate of the species phylogeny. The total evidence approach introduced by Kluge (1989) uses all available informative characters pooled into a single data set for maximum parsimony analysis. The total evidence character set was assembled from the data generated in this and nine previously published studies (Tsuyuki and Roberts 1963; Utter et al. 1973; Thomas et al. 1986; Shedlock et al. 1992; Phillips et al., 1992; Murata et al., 1993, 1996; Stearley and Smith, 1993; Oohara et al., 1997). Except for the placement of the outgroup root, the total evidence tree had the same topology as most others shown in Figure 2.5. Similarly, maximum likelihood analysis (Felsenstein, 1981) was performed on a combined data set assembled from all available DNA sequence data. The maximum likelihood tree inferred by this approach was identical to the total evidence tree.

Contribution of each data set to total evidence analysis

The total evidence tree recovered by analyzing all available data was identical to the ND3 and COIII trees, except that the BCLs of most nodes were improved. To assess the impact of different character sets on the combined analysis, each was removed in turn and the change in bootstrap confidence levels at each node was observed (Table 2.5). The small, non-DNA-sequence character sets, composed of morphological, biochemical, restriction site and SINE (short, interspersed, repetitive element) insertion site data, were combined into a single set, referred to as "other"

(Table 2.5). Overall, the impact of removing individual character sets was minor in terms of tree topology, with all but one of the subset trees recovered being identical. Different genes had different influences on the BCLs, with those of the deeper, more controversial nodes being the most affected. Unlike other subsets, the DNA-only ("other" characters removed) character set recovered a tree like that of the ATPase-6 gene. This combined with the effect on BCLs caused by the removal of ATPase 6 from the subset data suggests that the phylogenetic signal from this gene dominates at that node in the absence of non-DNA characters. This is not surprising, as the ATPase 6 data set contributed more informative characters than any of the others. The D-loop and ND4L sequences produced trees quite different than the others, but did not appear to have a substantial confounding influence of the total evidence tree.

Maximum likelihood evaluation of inferred phylogenetic trees

The tree inferred by total evidence and combined maximum likelihood analyses and the individual gene trees were evaluated by comparing their likelihood (L) values calculated from individual and pooled data sets (Figure 2.6; Table 2.6; Ln L values listed in Appendix 2). To test alternative positions for sockeye salmon, alternative branching orders for the (sockeye, (pink, chum)) clade were also tested for all trees except ND4L, which did not have this clade (Figure 2.5). Statistical analysis of differences in Ln L can be used to reject trees (hypotheses) whose L values are significantly lower than of the best (highest L) tree (Kishino and Hasegawa, 1989). Among the single gene data sets, only the COIII and ATPase 6 data provided statistical arguments for rejecting most of the alternative trees (Table 2.6). The pooled data set of all available DNA sequence

data had sufficient resolving power to reject all alternatives to the maximum likelihood tree except that of ATPase 6.

The contribution of each data set to the likelihood comparisons was evaluated by removing each in turn from the combined data set. As with the total evidence analysis, the removal of the GH2, ATPase 6, COIII and ND3 sequences had some effect on the resolving power of the data.

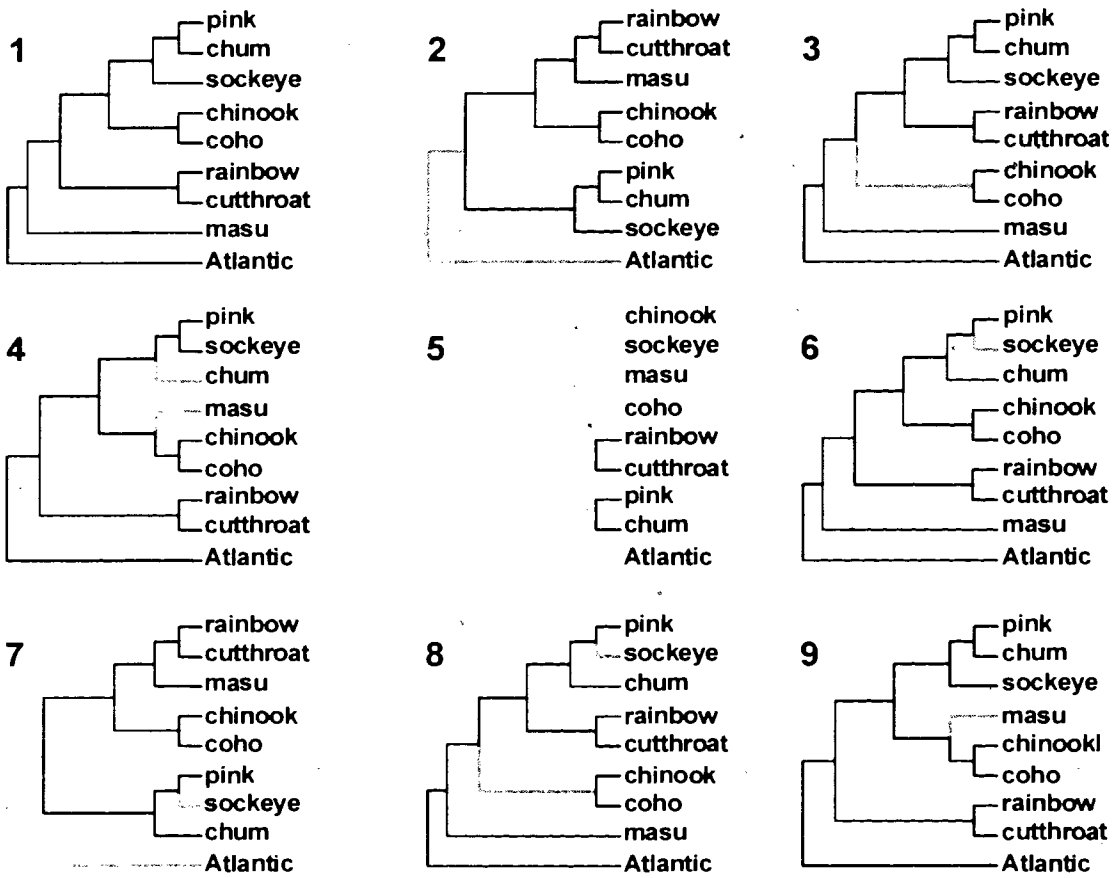


Figure 2.6. Trees used to evaluate maximum likelihood differences. Branches whose placement differs from tree 1 are shaded. 1 is the total evidence tree, which is identical to the COIII and ND3 trees (Figure 2.5). 2, 3, 4 and 5 are the GH2, ATPase 6, D-loop and ND4L consensus trees, respectively. Trees 6-9 are the same as Trees 1-4, except for the position of sockeye salmon.

Table 2.6. Summary of comparisons of the likelihood values of nine trees (Figure 2.6) with single gene and combined data sets. Likelihood estimates values were calculated with the program DNAML v3.57c (Felsenstein, 1993), which uses a model for sequence evolution described by Felsenstein (1981) and updated as described in the program documentation. DNAML performs statistical evaluation of differences in Ln likelihood values after the method of Kishino and Hasegawa (1989). "All" is the pooled set of all available sequence data. "- gene name" data sets include all available sequence except for that gene. "Best" indicates the tree with the highest likelihood value. (+) indicates trees with significantly worse likelihood values. (-) indicates trees with likelihood values that were not significantly different from that of the best tree.

	GH2*	ATPase	COIII	ND3	ND4L	D-loop	All	- GH2	- ATPase	- COIII	- ND3	- ND4L	- D-loop
# Sites	2435	689	780	351	228	910	5353	2918	4704	4573	5002	5125	4443
Tree 1	Best	-	Best	Best	-	-	Best	Best	Best	-	-	Best	Best
Tree 2	-	-	-	-	-	-	+	+	+	+	-	+	+
Tree 3	+	Best	-	-	-	-	-	-	-	Best	Best	-	-
Tree 4	+	+	+	-	-	Best	+	+	+	+	+	+	+
Tree 5	-	+	+	+	Best	+	+	+	+	+	+	+	+
Tree 6	-	+	+	-	-	-	+	+	+	+	+	+	+
Tree 7	-	+	+	-	-	-	+	+	+	+	+	+	+
Tree 8	-	+	+	-	-	-	+	+	+	+	+	+	+
Tree 9	-	-	-	+	-	-	+	-	-	-	-	+	+

*The GH2 data set used in the combined maximum likelihood analyses contained the full-length sequences of sockeye, chum, chinook, masu and Atlantic salmon and rainbow trout in addition to the intron D (plus flanking regions) sequence used to infer the tree in Figure 2.5. Note that tree 1 (the total evidence and combined maximum likelihood tree) has a higher likelihood value than the GH2 intron D tree (tree 2).

Removal of each of these data sets increased the relative L value of the modified D-loop tree (tree 9, Figure 2.6), but still made it possible to reject most other trees (Table 2.6). The contribution of the D-loop and ND4L data sets was less substantial, as removal of each of these data sets did not change the outcome relative to the complete data set (Table 2.6)

Discussion:

In this study I have examined patterns of change in the DNA sequences of the GH2 and ND3 genes, and used them in an effort to resolve systematic problems among *Oncorhynchus* species. In order to address the recurring problem of conflicting gene trees, DNA sequence data and other character types from this and previous studies were used in combined analysis with parsimony and maximum likelihood approaches. The resulting tree resolves outstanding conflicts in the phylogenetic analysis of this genus.

Resolving the relationships among *Oncorhynchus* species

The phylogenetic relationships among members of the genus *Oncorhynchus* have been the source of debate for a considerable period. Originally, the genus *Salmo* encompassed salmonid species from both Pacific and Atlantic drainages. Due to similarities between Pacific trout and Atlantic salmon in characters such as the number of anal fin rays and life histories, rainbow and cutthroat trout were retained in *Salmo* when the Pacific salmon were classified as *Oncorhynchus*. However, increasing resolution of systematic analysis brought about by additional morphological and

biochemical characters (reviewed by Smith and Stearley 1989) suggested a closer relationship to other Pacific salmonids, leading to the eventual placement of rainbow and cutthroat trout in *Oncorhynchus*.

I have examined the phylogeny of the genus *Oncorhynchus* by comparing the genealogies of nuclear and mitochondrial loci. The rationale for examining a variety of DNA sequences was to perform independent phylogenetic analyses to determine whether the conclusions were complementary. Biases introduced by the examination of sequence data from a single locus may cause inferred genealogies to differ among loci (Friedlander et al., 1994). In fact, trees based on genes or contiguous blocks of DNA sequence sampled from the mitochondrial genome often recover different trees (Cummings et al., 1995). Confounding influences, such as 1) differing rates of change of separate loci, lineages or genomes, 2) introgression due to interspecific hybridization and 3) homoplasy due to multiple substitutions at the same site, may play larger or smaller roles based on the dynamics of local evolution of a particular locus. Another consideration is that the examination of only one representative from each species may introduce a bias if there is considerable intraspecific variation or if the genotype of the sampled individual was a result of past introgressive hybridization events. In this case, the recovery of several different trees from six different DNA sequences provides a strong empirical justification for conservative interpretation of individual gene trees.

The use of different approaches to phylogenetic reconstruction reduces the impact that biases inherent to particular methods can have upon the inferred phylogeny. Although self-consistency within a data set will often support the same

conclusions based on different approaches to phylogenetic analysis (as was observed in this study), applying different methods of analysis to the same data may not necessarily satisfy the condition of independence. However, concordance between proposed trees based upon a variety of systems and genetic loci using both cladistic and distance approaches can be taken as an intuitive measure of confidence in a tree topology.

Despite elements of similarity, individual genes trees often disagreed on the deeper phylogeny. To resolve such conflicts, total evidence and maximum likelihood analyses of combined data sets were performed. Both methods recovered the same tree, which is identical to the ND3 and COIII trees (Figure 2.5). Under this hypothesis, the masu lineage is distinct from that of all other Pacific salmon and trout. This conflicts with the previous consensus of *Oncorhynchus* phylogeny (Stearley and Smith, 1993; McKay et al., 1996), which placed the pacific trout basal to the masu lineage. However, the total evidence and combined maximum likelihood analyses presented here are based on much larger character sets.

This work has been preceded by a number of other molecular phylogenetic studies of salmonid phylogeny based on mitochondrial DNA sequences (Thomas and Beckenbach 1989; Shedlock et al. 1992; Domanico and Phillips,1995; Oohara and Okazaki, 1997), growth hormone sequences (Devlin 1993), mitochondrial and nuclear restriction site differences (Thomas et al. 1986; Grewe et al. 1990; Phillips and Pleyte 1991; Phillips et al., 1992), protein variations (Utter et al. 1973; Tsuyuki and Roberts 1963;1966) and insertion patterns of short interspersed repetitive elements (SINEs; Takasaki et al. 1994; Murata et al. 1993,1996). The groupings of species produced by

terminal (more recent) and penultimate nodes in the consensus tree are all well supported by such analyses (Table 2.4): (pink, chum, sockeye), (chinook, coho), and (rainbow, cutthroat) are all robust clades both in terms of BCLs and concordance with trees inferred from other molecular data. Except for the basal branching order, the phylogenies reconstructed in this study were concordant not only between alternative methods of phylogenetic inference, but also between different genes.

The ATPase 6 (Figure 2.5) has a reversed arrangement for the (rainbow, cutthroat) and (chinook, coho) lineages. Although both the total evidence and combined maximum likelihood evidence tree places the (rainbow, cutthroat) clade more basally, the Kishino and Hasegawa (1989) test detects no significant difference in the likelihood of either branching order (Table 2.6; Appendix 2). The monophyly of all North American Pacific salmon with respect to masu and Pacific trout has not previously been a source of disagreement between different phylogenetic hypotheses (Table 2.4). The node in the total evidence tree that supports their monophyly is moderately well supported by its BCL. However, BCL values are generally more informative about the self-consistency of a data set than as a test of a phylogenetic hypothesis. This does not mean that the branching order of these two lineages is irresolvable. Classical taxonomy is based on well-defined, presumably irreversible cladistic characters that are common to members of the clade they define. The presence or absence of inserted repeat elements at orthologous loci in the nuclear genome represents such a character. SINE repeats Hpa-341 (Murata et al., 1993) and Hpa 391 (Murata et al., 1996) are inserted at orthologous loci in all North American Pacific salmon but *not* rainbow or cutthroat trout, which argues that these salmon are

part of a monophyletic clade distinct from the (rainbow, cutthroat) lineage.

Uncertainty in the relationships among sockeye, pink and chum salmon.

Most phylogenetic trees inferred from DNA sequence data agree on the pairing of pink and chum salmon as sister species. This is consistent with their similar life histories. Previously, the systematic consensus has been to group sockeye and pink as sister species. This association is borne out by morphology (Smith 1992; Stearley and Smith 1993), karyology (Simon 1963; Gorshkov and Gorshkova 1981), and other character types (Table 2.4). Smith (1992) asserted that the conflicting evidence observed by Thomas et al. (1986) with restriction analysis of mitochondrial DNA, and similarities in the life histories of pink and chum salmon can be explained by introgression due to hybridization. However, this assertion was made based primarily on only four morphological characters and in the absence of most currently available DNA sequence data. The phylogenetic trees observed in this study strongly support the branching order (sockeye, (pink, chum)).

Further synapomorphic cladistic characters as described above are represented by deletions in the GH2 intron D. Two deletions were present in chum and pink but not sockeye salmon, providing unambiguous evidence that the GH2 loci in these species are more closely related than either is to sockeye GH2. A closer relationship between these species has also been inferred by Murata et al. (1993; 1996) based on amplification of SINE repeat elements (However, see Takasaki et al. (1997) for an alternative interpretation). Further evidence was provided by likelihood analyses; the Ln L values calculated by DNAML for trees placing pink and sockeye as sister species

were all significantly worse than that of the maximum likelihood tree (Table 2.6; Appendix 2). Thus, the consensus of all available DNA evidence places pink and chum as sister-species.

Phylogenetic signal of individual character sets

The ATPase 6 and COIII data sets appeared to have a strong phylogenetic signal, as reflected by their provision of statistical arguments to reject most alternative trees (Table 2.6). The GH2 and ND3 data sets were able to reject fewer alternative trees, while almost all trees were equally supported by the D-loop and ND4L data. The overall contribution of the data sets to the combined maximum likelihood and total evidence analyses was measured by removing each from the combined character sets. The removal of the ATPase and COIII genes had the strongest impact on the BCL values of nodes in the total evidence tree (Table 2.5). The effect of removing the GH2, ND3 and D-loop data was less substantial, while the ND4L data made almost no contribution. Although the number of informative sites contributed by each data set is also a factor in the total evidence analysis, the stronger contributions of ATPase 6 and COIII to the final outcome parallel their relatively higher phylogenetic signal (Table 2.6). For the combined maximum likelihood analysis, the ND4L and D-loop sequences had almost no effect on the final outcome, which is consistent with the lack of phylogenetic signal inferred by likelihood analysis of the individual data sets.

The ND4L and D-loop data sets each produced trees that differed substantially from the consensus of other analyses. The large body of work on *Oncorhynchus* phylogeny and the availability of several independent character sets makes it possible

to evaluate the outcomes of phylogenetic analyses of individual genes. Although the deeper phylogeny and relationships among sockeye, pink and chum salmon are controversial, the consistent monophyly observed among the groups (sockeye, (pink, chum)), (chinook, coho), (rainbow, cutthroat) and ((chinook, coho), (sockeye, (pink, chum))) likely reflect the actual evolutionary history of the genus.

The goal of a comparative approach is not to reject data sets based solely on their non-conformance to the hypothesis being tested, rather it is to evaluate the reliability of particular genes or character sets for phylogenetic analysis. Such information would make it possible to avoid the use of unreliable genes or regions in other groups of species where extensive data are not available. Used in relative isolation, such data could result in a seriously flawed inference of phylogeny.

The ND4L data set provides very few informative sites (Table 2.5), and infers very different trees with parsimony, neighbor-joining and maximum likelihood analyses (not shown). Few of the well-supported clades in *Oncorhynchus* phylogeny (Table 2.4) appear in the consensus of the ND4L trees. The impact of this gene on the total evidence and combined maximum likelihood analyses was minimal (Tables 2.5; 2.6) This is likely due in part to the small number of characters relative to the pooled data set (228 aligned nucleotide positions). The weak or conflicting phylogenetic signal evident from the lack of consistency between different methods of phylogenetic inference for ND4L may be due to very different rates of sequence substitution in different lineages. The rates of each lineage are compared by measuring their divergence from the undisputed Atlantic salmon outgroup (see below for a discussion of relative rate tests). The ND4L genes of *Oncorhynchus* species differed from the

Atlantic salmon gene from between 6.9% and 11.9%. In contrast, most of the other genes examined had divergence values that were much more similar to one another, which is consistent with greater uniformity in the rate at which particular lineages accumulate mutations.

The departure of the D-loop tree from the phylogenetic consensus was less substantial. This data set was more self-consistent, as reflected by the recovery of the same tree with all three methods of phylogenetic inference. Since the other DNA data sets were from protein-coding regions, the alignment of most sequences was not ambiguous. However, in the case of the D-loop, the aligned sequences reported by Shedlock et al. (1992) contained many small alignment gaps interspersed in the sequence to maximize sequence identity. Such an approach may lead to ambiguities that allow the comparison of non-homologous nucleotide positions. A more conservative approach would be to realign the sequences allowing fewer gaps and more nucleotide substitutions, or to remove all regions where unambiguous alignment is not easily accomplished.

Dating divergence events in *Oncorhynchus* evolution

Based on the analysis of fossil specimens found in Idaho (Smith, 1992), pink, chum and sockeye salmon have been separate and distinct species for at least six million years. Using salmon growth hormone sequences, Devlin (1993) has estimated that the establishment of disomy in Salmonidae occurred at least 27.2 million years ago, which is consistent with dating of a proto-salmonid fossil (*Eosalmo driftwoodensis*) to the middle Eocene (Wilson, 1977), and that Pacific and Atlantic salmonids diverged

a minimum of 19.9 million years ago. Examination of the level of DNA sequence divergence observed in this study makes it possible to estimate the rate of divergence among *Oncorhynchus* species. Assuming a constant molecular clock within *Oncorhynchus*, the accumulation rate of substitutions for ND3 was estimated as (10/6)/2, or 0.83%/MY (percent per million years), based on 10% divergence between pink and chum salmon and an approximate date of six million years ago (MYA) for the node defining the (pink, chum) clade (Smith, 1992). The mitochondrial genomes of poikilotherms have been shown to evolve at a lower rate than their homiothermic counterparts (Martin and Palumbi 1993). A lower clock rate for salmon mitochondrial DNA is consistent with similar observations from Perciformes spp. (Cantatore et al. 1994), and turtles (Avise et al. 1992). Moreover, lower rates observed in warm-blooded vertebrates such as cetaceans (Hoelzel et al. 1991) cast doubt on the concept of a universal molecular clock rate for higher vertebrates. The pair-wise distance between pink and chum using the GH2 sequence data is 1.4%, corresponding to a divergence rate of 0.11 %/MY, approximately seven-fold lower than the ND3 rate.

All rate estimates must be accepted with the caveat that they are vulnerable to violations of the assumption of a constant molecular clock. The validity of this assumption can be tested with a relative rate test (Sarich and Wilson 1973; Li et al. 1987). *Oncorhynchus* species are monophyletic with respect to Atlantic salmon. If the clock rate is constant between lineages, all taxa should be approximately the same distance from this outgroup. Since the level of DNA sequence divergence between pink and chum was used to calibrate the molecular clock, it is important to determine whether the average mutation rate in this lineage is equal to those of the other

Oncorhynchus species. For the ND3 sequence data (Table 2.3), the average pair-wise distance between the (pink, chum) clade and Atlantic salmon is 19.2%, indicating that these species have accumulated sequence differences 4.0% faster than the genus average. This was calculated using the formula 100%-([average (pink, chum) species rate/average rate]*100%). Similarly, the relative rate of this clade was +0.8% for ATPase 6, +6.7% for COIII, +8.1% for GH2, +12.0% for the D-loop and -22.2% for ND4L. Applying an arbitrary cut-off value of ±10%, and bearing in mind concerns expressed above regarding their phylogenetic information content, the D-loop and ND4L data sets were not used in the calculation of estimated times for evolutionary branch points (discussed below).

In protein-coding sequences and functional non-coding sequences, selective constraints lead to unequal rates of variation at some positions. For example, most variation in coding sequence is at the degenerate first and third positions of codons. Because of the high rate of change in mitochondrial DNA, such variable sites can undergo undetected multiple substitutions, leading to an underestimation of the actual distance between related sequences. To minimize this effect for time estimates based on mitochondrial DNA, only variable nucleotide positions were used to calculate distance measures. Under these conditions, the Kimura 2-parameter correction (Kimura, 1980) for unobserved multiple substitutions produced higher (presumably more realistic) estimates of the degree of saturation. The time estimations based on the GH2 sequence were uniformly higher than the mitochondrial DNA estimates. The recalculation of distances using only variable sites in the mitochondrial DNA substantially reduced the disparity between the nuclear and mitochondrial gene-based

time estimates.

Applying the molecular clock estimates discussed above, a crude time scale was applied to the divergence or speciation events in *Oncorhynchus* phylogeny (Figure 2.7). Time estimates were calculated with the formula d/2k, where d= the pair-wise distance between taxa (or average distance between clades) and k= molecular clock rate for that locus. The time estimates based on each of the sequences varied considerably, as is reflected in the large standard error of the mean values (Figure 2.7). The time estimates in this study are consistently higher than those observed by Shedlock et al. (1992) with the D-loop sequence. However, it should be noted that rather than calibrating their molecular clock with dated fossil evidence, they based their time estimates on the mutation rate of the mammalian D-loop. Generally, the wide range of time estimates for each node, particularly the divergence of *Oncorhynchus* and *Salmo*, provides a compelling argument for cautious interpretation of time estimates extrapolated using single-gene DNA sequence divergence.

Based on the mean of the divergence times calculated with four DNA sequence data sets, I estimate that the minimum age of *Oncorhynchus*, or the time since it diverged from the ancestor it shares with *Salmo*, is approximately 18-24 MY (Figure 2.7). Some nine million years later, the first in a rapid series of speciation or divergence events occurred, leading to the radiation of four main lineages, which in turn gave rise to the eight Pacific salmon and trout species or species complexes. The distance between the first, second and third internal nodes in the phylogenetic tree was essentially zero (slightly exaggerated in Figure 2.7 to show inferred branching order), indicating that the radiation leading to the four main groups was extremely rapid on this

time scale. The rapidity with which the first three divergence events occurred in the tree is most likely the source of conflicting phylogenetic hypotheses. Despite the large amount of attention paid to this group of species, poor agreement has been achieved with regard to the deeper phylogeny of *Oncorhynchus*.

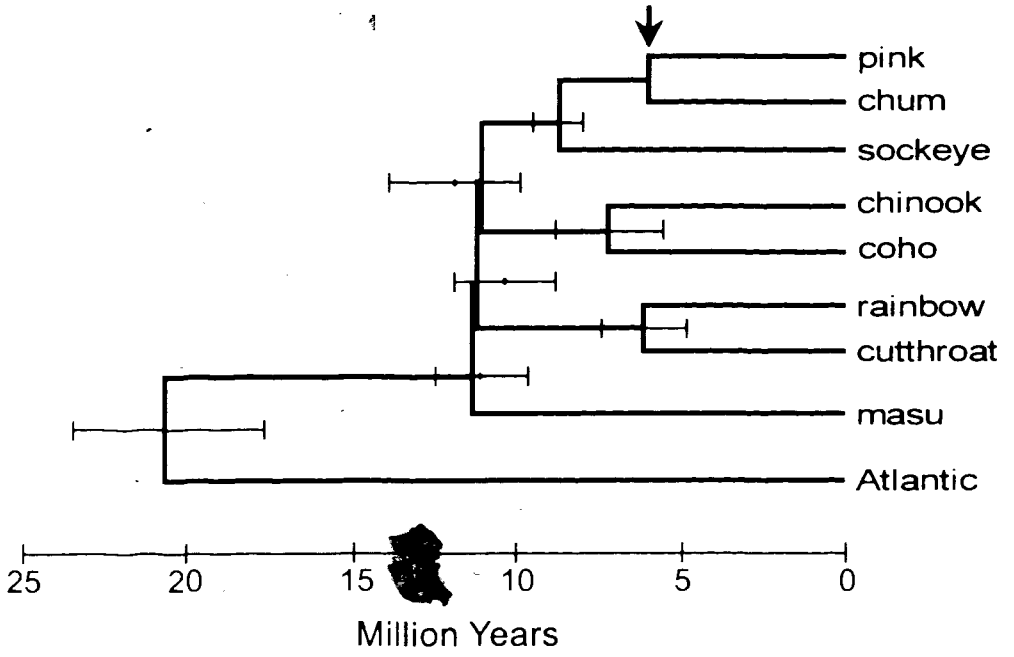


Figure 2.7. The evolution of Oncorhynchus based on the inferred total evidence phylogeny. The time of each branching point was extrapolated from the pink/chum split (arrow), which has been dated through fossil evidence to at least 6 million years ago (Smith, 1992). Horizontal bars represent the mean (+/- standard error) of time estimates from the GH2, ND3, ATPase 6 and COIII genes. The first three internal branching points occurred at approximately the same time. These nodes are shifted from their respective mean time estimates to prevent negative branch lengths.

It seems likely that the abundance of conflicting phylogenetic hypotheses can be attributed to the nature of the evolutionary processes being studied. In terms of more basal phylogeny, short internodal intervals would have allowed only minimal

accumulation of phylogenetically informative changes between lineages, which have had approximately ten million years to become swamped by uninformative, apomorphic changes. Based on the above time estimates, and accepting their limitations, the masu lineage diverged from the proto-*Oncorhynchus* line 9-12 MYA. Subsequent divergence events in *Oncorhynchus* must have occurred over a very short time. Evidence of a similar radiation of species has not been observed in the closely related genus *Salmo*, which occupies a similar range in the Atlantic basin. This suggests that geologic or climatic conditions unique to the North Pacific basin opened up a new series of ecological niches, leading to the episodic bursts of speciation observed in the inferred *Oncorhynchus* phylogeny.

Smith (1981) observed that the fossil record of the late Cenozoic fishes west of the North American continental divide contains only about one quarter of the diversity of contemporaneous species as that of more eastern regions. The lower diversity is attributed to a much higher rate of extinction, which is consistent with geologic and climactic instability in Pacific drainages. Other evidence of a distinction between the Pacific and Atlantic basins comes from the Ocean Drilling Program (ODP), which has revealed a paleoceanographic phenomenon termed the biogenic bloom. The biogenic bloom hypothesis deals with a several-fold increase in surficial productivity, which is believed to be related to phytoplankton abundance, measured from ODP holes in the Indian and Equatorial and North Pacific Oceans (e.g. Dickens et al., 1996). Although a link between *Oncorhynchus* evolution and these general observations would be conjectural at best, the lack of a parallel radiation in the Atlantic genes *Salmo* could be tied to the relative stasis of late Cenozoic Atlantic drainages.

Chapter 3

Clarification of the genetic relationship between masu and amago salmon of Japan through mitochondrial and nuclear DNA sequence analysis.

Abstract:

Historically, the taxonomy and nomenclature of Japanese salmon have been in a state of confusion. Masu, amago and biwa salmon have been variously classified as distinct species, subspecies, or often conflicting or overlapping combinations of the two. In particular, the taxonomy of masu and amago salmon is obscured by their similarity in ecological and morphological traits. Here, DNA sequence analysis of the nuclear and mitochondrial loci is applied to clarify the genetic relationship between masu and amago salmon. No type-specific variation was detected in the mitochondrial ND3 gene or control (D-loop) region. However, considerable variation was detected in intronic sequences of the nuclear GH2 gene. Although no fixed differences were observed between masu and amago, the frequency of single nucleotide substitution alleles in intron C and size variants at a microsatellite locus nested within intron D differed markedly, providing genetic evidence to support a taxonomic distinction between the two types. The genetic data were related to previous mitochondrial DNA sequence analyses and alternative classification schemes for masu and amago salmon. The best-supported scheme arranges masu and amago as subspecies *Oncorhynchus masou masou* Brevoort (masu) and *Oncorhynchus masou ishikawae* Jordan and McGregor (amago).

Introduction:

The genus *Oncorhynchus* contains eight types of Pacific salmon and the recently re-classified rainbow, cutthroat and allied trout species (Smith and Stearley, 1989). Five types of salmon, sockeye, pink, chum, chinook and coho, occur on both sides of the northern Pacific Ocean. Each of these types exhibits marked morphological and ecological differences that have made it possible to assign unambiguous species status. This group of salmon is believed to have descended from a single common ancestor that diverged from other Pacific salmon and trout lineages at least 10 million years ago (Chapter 2). Three types of salmon that occur only in Asia represent the masu lineage: masu (sakuramasu), amago (satsukimasu) and biwa (biwamasu) salmon. Two classification schemes are in current use for this group of salmon. One assigns specific status to masu (*O. masou*) and groups amago and biwa together as *O. rhodurus* (Kato, 1985; 1991), while the other groups masu (*O. masou masou*), amago (*O. masou ishikawae*) and biwa (*O. masou spp.*) as conspecific races (Kimura, 1990).

Table 3.1. Outline of the *Oncorhynchus masou* species complex.

Type	Life History	Red spots[a] Juvenile	Adult	Synonyms[b]
sakuramasu	anadromous	absent	absent	*Salmo masou, O. masou, O. perryi, O. yessoensis,*
yamame	fluvial	absent	absent	*S. macrostoma, S. perryi, S. masou, O. kisutch, O. macrostomus, O. ishikawae, O. m. masou*
satsukimasu	anadromous	present	present	*S. masou, O. masou, S. perryi, O. ishikawae,*
amago	fluvial	present	present	*O. macrostomus, O. milktschitsch, O. rhodurus, O. r. macrostomus, O. m. rhodurus, S.(O.) m. iwame, O. ishikawai, O. m. ishikawae, S.(O.) m. macrostomus*
biwamasu	lacustrine	present	absent	*S. perryi, S. masou, O. masou, O. rhodurus, S. (0.) m. macrostomus, O. m. rhodurus, O. m. spp. O. r. rhodurus,*

[a]Red spots are a diagnostic character, generally used to distinguish between the different types.
[b]A detailed examination of holotypes and chronology of nomenclature are presented in Kimura (1990).

The root of their names in the Japanese vernacular is "masu", which means trout. Unlike the North American Pacific salmon, this group has retained more primitive, trout-like life history traits: sea-run forms, particularly satsukimasu, do not venture as far into the Ocean, and land-locked forms do not always die after spawning. The trout-like character of these fish is consistent with their basal position in inferred phylogenetic trees for *Oncorhynchus* (Chapter 2; Stearley and Smith, 1993, Oohara et al., 1997).

The geographic range of masu (Table 3.1; hereafter collectively referring to sakuramasu and yamame) salmon stretches northward as far as the Kamchatka Peninsula. Yamame, the land-locked form, occurs as far south as Taiwan and Formosa. The distribution of amago (Table 3.1; collectively referring to the land-locked form, amago and the anadromous form, satsukimasu) and biwa salmon are more restricted, with amago occurring primarily on the Pacific side of Southern Japan, and the biwa salmon native only to lake Biwa and associated drainages. The range of biwa salmon is completely within that of amago, but masu does not currently occur sympatrically with either of the other types (Oshima, 1957; Kimura, 1989). Historically, marked similarity in morphological and meristic characters and vague descriptions of original type specimens (Jordan and McGregor, 1925) have led to confusion in their taxonomy and nomenclature (Table 3.1). Differences in scale morphology and the presence of red spots above and below the lateral line of juvenile and adult fish are diagnostic characters for distinguishing between the three types. DNA sequence analysis of the mitochondrial genome demonstrated that the lacustrine biwa salmon is probably the oldest lineage of the *O. masou* species complex (Oohara and Okazaki, 1996). However, molecular differences between the masu and amago types are less

pronounced; much of their mitochondrial genomes are nearly identical in sequence (Oohara and Okazaki, 1996; McKay et al., 1996).

In this study, we examined additional mitochondrial DNA sequence from the ND3 gene and the control (D-loop) region, where both interspecific (Thomas and Beckenbach, 1989; Shedlock et al, 1992) and intraspecific (Beckènbach et al., 1990; Park et al., 1993) variation in *Oncorhynchus* have previously been observed. Very little DNA sequence variation was detected among mitochondrial sequences of masu and amago. However, analysis of intronic sequences of the nuclear growth hormone type-2 (GH2) gene revealed considerable variation within and between types, providing evidence that masu and amago are genetically distinct.

Materials and methods:

DNA extraction, gene amplification and sequence analysis

Strains and sample origins are listed in Table 3.2. Samples of liver or fin tissue from fish specimens were stored in 70% ethanol at ambient temperature until use. DNA was isolated from tissue samples using Proteinase K digestion followed by extraction with organic solvents as described previously (Devlin et al., 1991). Polymerase chain reaction (PCR; Saiki et al., 1988) amplification was performed on 200-500 ng of genomic DNA template with either Ultratherm (Bio/Can Scientific) or Taq (Bethesda Research Laboratories-BRL) DNA Polymerase using the reagents and instructions provided by the manufacturer. Typically, the thermal profile of a PCR consisted of 2-4 min. incubation at 94° C, followed by 30 cycles of 30 s at 94°, 30 s at 55°, 60 s at 72°, followed by a 4 min. incubation at 72°. PCR amplification products

were prepared for sequencing by purification with Wizard PCR-Prep or DNA Clean-Up kits (Promega). Where necessary, multiple amplification products were separated by electrophoresis in low-melting-point agarose using standard methods (Sambrook et. al., 1989). Amplification products were sequenced directly using either the Sequenase v2.0 or Thermosequenase sequencing kits (Amersham-United States Biochemicals). Sequencing, electrophoresis and autoradiography were performed according to the manufacturer's instructions.

PCR and sequencing primers

A portion of the mitochondrial control region was amplified using the F+ (5'-TTC CTG TCA AAC CCC TAA ACC AGG-3') and F- (5' CCA TCT TAA CAG CTT CAG-3') primer pair described in Shedlock et al. (1992). 185 nt of DNA sequence corresponding to the 3' end of the aligned sequence reported by Shedlock et al. (1992), was obtained.

Two portions of the GH2 gene were amplified (Figure 3.1). Primers GH 41 (5'-ATG GAA AAC CAA CGG CT-3') and GH28 (5'-GTC TGG CTA GGG TAC TCC CA-3') were used to amplify a segment containing introns B, C and flanking regions. This primer combination produced two amplification products corresponding to GH1 and GH2. The GH2 product was identified by comparison with sequences from sockeye salmon GH1 and GH2 genes (Devlin, 1993), from which primers GH41 and GH28 were designed. The entire 451 nt intron C sequence was determined using primer GH28 and the opposing primer GH27 (5'- ATA TTC CTG CTG GAC TTC TG-3').

The second portion of the gene was obtained with primers GH57 (5'-GCT CAT CAA GGT AAT GGT CA-3') and GH7 (5'-CTT ATG CAT GTC CTT CTT GAA-3'), which

specifically amplify a segment of GH2 containing intron D and exon 5 (McKay et al., 1997). The same segment plus the extreme 3' end of exon 4 was also amplified from both the GH1 and GH2 genes using primers GH7 and GH56 (5'-AAG CTC AGC GAC CTC AAA GT-3').

Table 3.2. The names and geographic origins of strains used in this study.

Type	Strain	n	Origin	Island
Amago	AS	3	Hida-gawa, Gifu Prefecture[a]	Honshu
Amago	AP	3	Hida-gawa, Gifu Prefecture[a]	Honshu
Amago	AY	3	Fuji-gawa, Yamanshi Prefecture	Honshu
Amago	AE	3	Ehime Prefecture	Shikoku
Amago	AM	3	Miya-gawa, Mie Prefecture	Honshu
Amago	AU	26	Unknown[b]	
Amago	AMI	2	Maze, Mie Prefecture	Honshu
Amago	AT	2	Misugi, Gifu Prefecture	Honshu
Masu	MK	10	Shokanbetsu-gawa	Hokkaido
Masu	MS	10	Shiribetsu-gawa	Hokkaido
Masu	MKA	4	Kawauchi-gawa, Aomori Prefecture	Honshu
Masu	MO	3	Oohata-gawa, Aomori Prefecture	Honshu
Masu	MU	26	Unknown[c]	
Masu	MP	1	Unknown[d]	

[a]7th generation cultured strain of known parentage
[b]farmed or hatchery-reared strains, natal rivers unknown, National Research Institute for Aquaculture, Gifu Prefecture, Honshu, Japan
[c]farmed or hatchery-reared strains, natal rivers unknown, Mori hatchery, Hokkaido
[d]farmed or hatchery-reared strain, natal river unknown, Kunsan, Korea

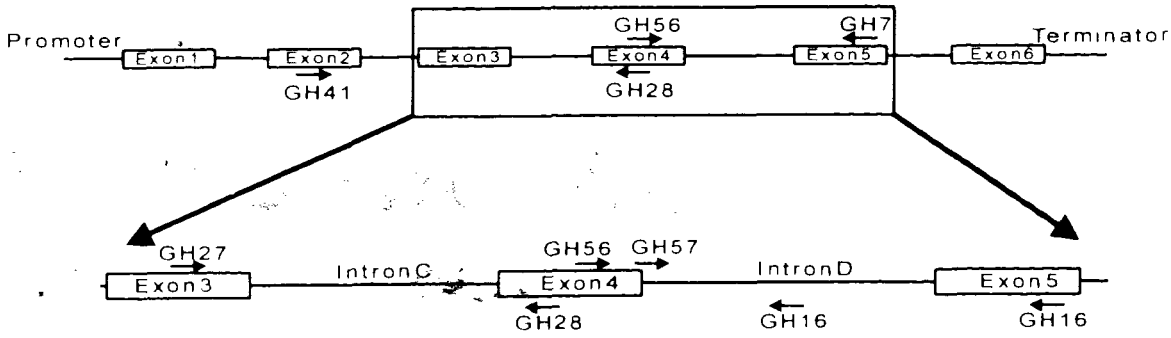


Figure 3.1. Map of Oncorhynchus growth hormone genes. The position and orientation of PCR and sequencing primers are indicated by small arrows. Protein coding sequence (exons) are represented as open boxes.

Sequence from the 3' end was obtained using primer GH57 (GH2) or GH56 (GH1 and GH2). In some cases, the opposite strand was read using primer GH7 or GH16 (5'-TTG TTA ATC TTT GTG AAA A-3').

Direct PCR sequencing of PCR products from heterozygous individuals

Direct sequencing of amplification products from individuals heterozygous at variable positions in the GH2 gene produced sequence ambiguities (Figure 3.2A). Two bands of equal intensity occurring at the same position in the sequence were interpreted as having resulted from amplification of two alleles differing at that position. Such ambiguities never involved more than two nucleotides at one position. The possibility that the two-fold ambiguities were amplification artifacts resulting from misincorporation of nucleotides by Taq DNA Polymerase was ruled out for two reasons: 1) the site and type of virtually all observed sequence ambiguities was the same in several individuals, each of which represented independent DNA extractions, PCR amplifications, and sequencing experiments, and 2) in the case of intron D, two independent PCR amplifications with different primer pairs (GH56/7 vs. GH57/7) from six fish produced identical sequences, including the position and nature of each ambiguity.

A second type of heterozygote was observed in GH2 intron D (Figure 3.2B). A four nt microsatellite repeat varied between three and five iterations (discussed below). Direct PCR sequencing from heterozygous individuals produced clean sequence upstream of the repeat region. The region immediately downstream of the heterozygous repeat produced two superimposed sequences, one being shifted out of

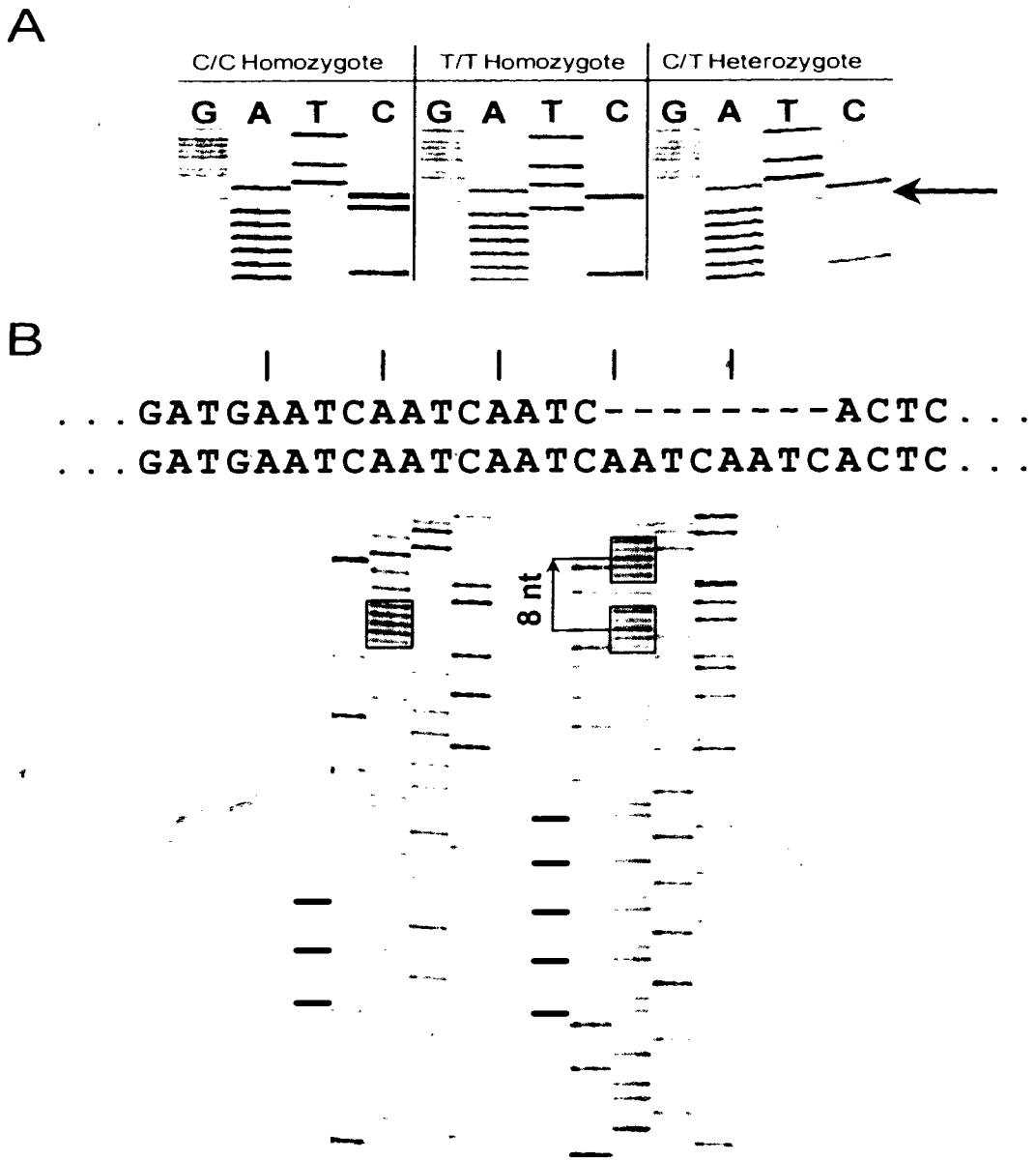


Figure 3.2. Direct PCR sequencing of heterozygous individuals. A) Single nucleotide substitutions. The sequence shown corresponds to the complement of positions 466-484 in the aligned GH2 sequences presented in Chapter 2. B) Variation in number of repeat units at a microsatellite locus nested within GH2 intron D. The sequence shown corresponds to the complement of positions 329-381 of the same alignment as in panel A. Left-$(GATT)_3$ homozygote, Right-$(GATT)_3/(GATT)_5$ heterozygote. Note that the run of five A's (boxed) is shifted out of register by 8 nt (two repeat units) in the heterozygote.

register by either four or eight nucleotides (one or two iterations of the repeat unit). Alleles were scored by counting the number of iterations of the repeat unit, then observing the displacement of easily identified sequence motifs downstream of the repeat, such as the run of five A's shown in Figure 3.2B. The reliability of this scoring method was confirmed by reproducing the results in some cases by sequencing both strands, and in others by independent PCR reactions as described above. In addition, the genotypes scored by sequence analysis were confirmed in 24 individuals by denaturing polyacrylamide gel electrophoresis of full-length alpha-^{32}P dATP-labelled PCR products (not shown).

Results:

Mitochondrial DNA sequence analysis

Overall, the ND3 gene has a relatively high substitution rate in salmonid fishes (Thomas and Beckenbach, 1989; McKay et al., 1996). However, the complete sequence of the ND3 gene (351 nt) was found to be identical between a masu sampled in Hokkaido and an amago from southern Honshu. With the exception of a silent substitution in one masu individual (Figure 3.3A), complete sequence identity in the ND3 gene was also observed among an additional three masu and three amago sampled from the same locations. Silent substitutions are changes in protein-coding DNA sequence that do not affect the translated amino acid sequence. Two additional haplotypes, reported by Oohara and Okazaki (1996), that differ by single silent substitutions were not observed among the individuals sampled in this study (Figure 3.3A).

Similar results were obtained with the mitochondrial control region. The 3' end of this region is highly variable among salmonid fishes (Shedlock et al., 1992), but very little variation was detected masu and amago individuals. A 185 nt region was sequenced from 14 amago and 6 masu individuals (Figure 3.3B). Two haplotypes, differing by a single, transitional substitution, were observed. The most common haplotype was present in all but one fish. The haplotypes observed in this study differ from the masu sequence reported by Shedlock et al., (1992) by a single nucleotide substitution, as well as several single-nucleotide gaps. As was observed with the ND3 genes, the most commonly observed haplotypes were found in both masu and amago salmon, providing no evidence for a genetic distinction between the two types.

Variation in intronic sequences of the GH2 gene

The complete DNA sequence of intron C from masu and amago individuals was determined. A total of 16 fish were sampled, with two representatives from each of four geographically isolated populations (Table 3.2) represented in each sample group. To avoid confusion about geographic origin, only wild strains from known sampling locations were analyzed. Considerable variation was observed in intron C, both within and between the two types (Table 3.3). Seven nucleotide positions varied among individuals. Comparison of variation within types revealed that the amago sample group was more genetically heterogeneous, as reflected by the higher degree of heterozygosity with respect to the masu sample group. Although no fixed differences were observed between masu and amago, particular nucleotides at variable positions tended to be more common within one type than another. For example, an "A"

occurred at position 269 with a frequency of 0.875 (14/16 haploid genomes) in masu, but only 0.375 in the amago sample group. In addition, polymorphism at positions 140 and 182 were confined to masu and variation at position 425 was specific to amago. These observations suggest that masu and amago are genetically distinct.

The sequence GH2 intron D from masu and amago salmon has been reported previously (McKay et al., 1996). In this study, analysis of the 5' end of intron D from 44 amago and 52 masu salmon revealed a variable microsatellite locus nested within the intron. A direct, tandem repeat of a (GATT) sequence motif was found to vary between three and five iterations. Genetic heterogeneity at this locus was high, with greater than half of the individuals tested being heterozygous. Sequence of the same region of the GH1 gene was also obtained from three masu and three amago. Similar variation was not detected within this gene: the (GATT) core repeat sequence was present in only two iterations in each of the six individuals tested.

In addition to variation in the number of (GATT) repeat units in GH2 intron D, two (G<->A) transitional substitutions at positions 206 and 224 of the aligned intron sequence reported by McKay et al. (1996) were found to vary within and among the masu and amago sample groups. A "G" was observed rarely at position 206 (G^{206}), with an overall frequency of 0.08 (14/188 haploid genomes). G^{206} is likely physically linked on the same chromosome as a $(GATT)_4$ allele; 14/14 individuals with a G^{206} allele also had at least one copy of the $(GATT)_4$ variant, which was either homozygous, or heterozygous with $(GATT)_3$ or $(GATT)_5$. A "G" occurred more commonly at position 224, with a frequency of 0.28. G^{224} is almost certainly linked to the $(GATT)_3$ variant.

Figure 3.3. Mitochondrial DNA haplotypes. A) The ND3 gene. 1 is the sequence of the haplotype from individuals MU1, MU2, MUB, AMI1, AMIA, AT1, ATA; 2 is the haplotype of individual MUA, 3 and 4 are the haplotypes reported in Oohara and Okazaki (1996). B) The 3' end of the mitochondrial control region (D-loop). 1 is the sequence reported in Shedlock et al. (1992), 2 is the haplotype of individual AE1, 3 is the haplotype observed in all other individuals examined.

A

```
1       ATG AAC TTA ATT ACA ACA ATT ATC ACT ATT ACC ATC ACA CTA TCT GCA GTA  51
2       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
3       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
4       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

1       CTA GCC ACT ATT TCT TTC TGA TTA CCA CAA ATT TCT CCA GAC GCA GAG AAA  102
2       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
3       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
4       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

1       TTG TCC CCC TAC GAA TGT GGA TTT GAC CCT TTA GGA TCC GCC CGT CTA CCC  153
2       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
3       ... ... ... ... ... ... ... ..C ... ... ... ... ... ... ... ... ...
4       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

1       TTC TCC TTA CGC TTC TTC CTA ATC GCC ATC CTG TTC CTT CTA TTT GAC TTG  204
2       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..A
3       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
4       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

1       GAA ATC GCC CTC CTT CTA CCC CTG CCT TGA GGA GAT CAA CTC AAC ACC CCC  255
2       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
3       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
4       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

1       GCC CTA ACA CTC GTC TGA TCC ACT GCT GTA CTT GCC CTC CTT ACT CTA GGC  306
2       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
3       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
4       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..C ...

1       TTA ATC TAT GAA TGA ACC CAA GGA GGC TTA GAA TGA GCC GAA TAG 351
2       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
3       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
4       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
```

B

```
1       TATATACATT AATGAACTTT TGATGTACTT TATTGCATTT GGCACCGACA GCGCTGT-AT 60
2       .......... ...A...... .......... .......... .......... .......G..
3       .......... ...A...... .......... .......... .......... .......A..

1       ACGTACACTT TCATAATTAA AGTATACATT AATAAACTTT TCGATCCATT TAACAGCACC 120
2       .......... .......... .......... .......... .......... ..........
3       .......... .......... .......... .......... .......... ..........

1       TGGCACCGAC AACGCTATCA TAAATGCCAT TTCCCGGCAC AACCCGCTG- CTGGCGTAGC 180
2       .......... .......... .......... ....-.-... .........G ..........
3       .......... .......... .......... ....-.-... .........G ..........

1       TTAAC 185
2       .....
3       .....
```

In 43/43 individuals with G^{224}, the $(GATT)_3$ variant was also present. In addition, homozygous G^{224}/ G^{224} individuals were always homozygous $(GATT)_3/(GATT)_3$. A causal relationship between these nucleotide substitutions and the number of repeat iterations is unlikely, as the same locus (nested within GH2) in other *Oncorhynchus* species varies from two to five iterations of the (GATT) repeat while having an "A" at positions 206 and 224 (Figure 2.3).

Table 3.3. Variable positions within GH2 intron C of wild masu and amago salmon. Strain designations are defined in Table 3.2.

	25	140	166	182	269	367	425
Amago							
AM1	C	A	T	G	G	C	G
AY2	C	A	T	G	G	C/T	G
AS2	C	A	T/C	G	G/A	C/T	G
AS1	C	A	T/C	G	G/A	T	G
AY1	C/A	A	T/C	G	G/A	C/T	G/A
AE1	C/A	A	T/C	G	G/A	C/T	G
AE2	C/A	A	T/C	G	G/A	C/T	G
AM2	C/A	A	T/C	G	G/A	C/T	G
Masu							
MS1	A	A	C	G	A	T	G
MOA	A	A	C	G	A	T	G
MOB	A	A	C	G	A	T	G
MKA2	A	A	C	G/A	A	T	G
MKA	A	A	C	G/A	A	T	G
MK1	C/A	A	C	G/A	A	T	G
MKA1	C/A	A	T/C	G	G/A	C/T	G
MSA	C	A/G	T/C	G	G/A	C/T	G

Microsatellite allele frequencies differ between masu and amago salmon

As was observed with the single nucleotide substitutions in GH2 intron C, the distribution of the $(GATT)_n$ alleles of the microsatellite locus within intron D are not equal between masu and amago (Figure 3.4A). Taken overall, the $(GATT)_3$ allele is more common in masu, while the frequency of the $(GATT)_5$ allele is more than two-fold higher in amago (Figure 3.4B). The observed differences in total allele frequencies were found to be statistically significant using chi-squared analysis ($p=0.015$). Because salmon in Japan have a history of being transplanted, and many of the sampled individuals were of uncertain parentage, the sample populations were divided into two categories. Wild fish (or their descendents), taken from known geographic locations, were analyzed separately from cultured or hatchery-reared fish of unknown geographic origin, hereafter referred to collectively as "cultured". By treating the two categories separately, it was revealed that the allele frequencies differ markedly between wild and cultured fish (Figure 3.4). Among wild fish, the $(GATT)_5$ allele is clearly the most common in amago (n=19), and

the $(GATT)_3$ allele was observed only in a single heterozygous individual. In wild masu (n=26), the three allele frequencies are more similar, with $(GATT)_4$ slightly more common than the others. The overall difference in allele frequencies between wild masu and amago was significant ($p<0.005$).

In contrast, $(GATT)_4$ was the least common variant among the remaining masu (n=24) samples (Figure 3.4). The $(GATT)_3$ allele was the most common among both cultured masu and cultured amago (n=25). Unlike the wild fish, the three alleles were more equally represented among cultured amago. The overall differences in allele

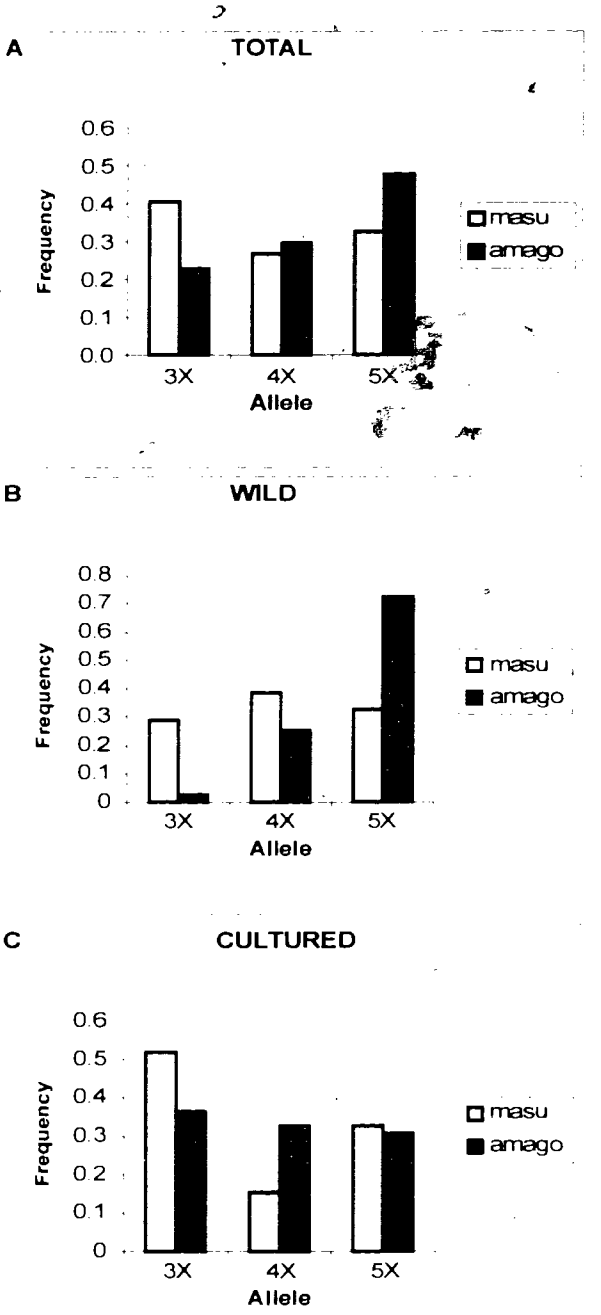


Figure 3.4. Allele frequencies of the (GATT) microsatellite locus within GH2 intron D. 3X, 4X and 5X refer to $(GATT)_3$, $(GATT)_4$ and $(GATT)_5$. A) Overall allele frequencies. B) Allele frequencies in wild fish only. C) Allele frequencies in cultured or hatchery-reared fish of unknown origin.

frequencies between the two types of cultured fish were not statistically significant. Masu and amago are known to hybridize readily under hatchery conditions (Oshima, 1955), and produce viable offspring. The markedly higher incidence of the $(GATT)_3$ allele among cultured vs. wild amago salmon suggests that that introgression of this allele from masu to amago may have occurred among captive populations.

Discussion:

Because it tends to evolve relatively rapidly (Brown et al., 1979), analysis of the mitochondrial genome is commonly used to study relationships among conspecific populations. Unfortunately, the initial confusion surrounding the taxonomy and nomenclature of masu and amago salmon was not alleviated by analysis of mitochondrial DNA sequence. It has been argued that these two types are simply morphs of the same species (Imanishi, 1951). A pronounced difference between biwa salmon and the other types was supported by mitochondrial sequence data (Oohara and Okazaki, 1996), but there is no convincing evidence from the mitochondrial genome supporting a genetic distinction between masu and amago.

Variation in the GH2 gene supports a genetic distinction

Substantial variation was observed within intronic sequences from the GH2 gene among and between masu and amago salmon. None of the observed differences were fixed between types, but masu and amago clearly differed in patterns of single nucleotide substitutions in intron C (Table 3.3) and in allele frequencies at the (GATT) microsatellite locus nested within intron D (Figure 3.4). While the overall

frequency of the three observed microsatellite alleles were similar, the $(GATT)_5$ allele was much more common in amago salmon, while the $(GATT)_3$ form was extremely rare in this type. These allele frequency differences provide evidence that the two types are genetically distinct, and that recent interbreeding between wild masu and amago salmon has likely not occurred.

Microsatellite allele frequencies differ between cultured and wild fish

Allele frequencies at the microsatellite locus were markedly different between cultured and wild sample groups (Figure 3.4). Unlike the wild fish, the differences in allele frequencies between cultured sample groups of each type were not found to be statistically significant. The $(GATT)_3$ allele was very rare among wild amago but was the most common among the cultured fish. The distribution of mitochondrial haplotypes was also found to vary between wild and cultured sample groups (Oohara and Okazaki, 1996), with cultured amago and masu more similar than wild amago and masu. The differing frequencies could be the result of a founder effect in the population or populations used to establish the cultured strains. This scenario is unlikely, however, as cultured strains of both types have probably been established from a number of wild populations. Our observations and those of Oohara and Okazaki (1996) are consistent with recent introgressive hybridization between captive masu and amago, but lack of information on the geographic origin and history of cultured strains precludes resolution of this question.

Recent history of the GH2 microsatellite locus

Although the three microsatellite alleles (Figures 3.3, 3.4) were scored only by the number of iterations of the (GATT) repeat, there are at least five alleles if one considers the (G<->A) substitutions at positions 206 and 224 of intron D. G^{206} was always associated with the $(GATT)_4$ variant, but a $(GATT)_4$ allele was also observed with an "A" at that position, which indicates there are at least two $(GATT)_4$ alleles. Similarly, G^{224} was always associated with the $(GATT)_3$ variant, but $(GATT)_3$ was also observed with an "A" at that position. The more conservative scoring of alleles based strictly on the number of repeat unit iterations was used for two reasons: 1) the sequence of upstream region was not determined in all individuals for which the microsatellite alleles were scored, and 2) with the exception of the two sequences reported by McKay et al. (1996), direct, physical linkage between the (G<->A) substitutions and particular $(GATT)_n$ variants was not demonstrated by sequencing of cloned alleles.

Nevertheless, a strong association was observed between G^{206} and $(GATT)_4$, and between G^{224} and $(GATT)_3$. This information can be used to infer patterns of evolutionary change at the microsatellite locus. For example, G^{224}-$(GATT)_3$ is the more common $(GATT)_3$ allele. Since an association between a "G" at position 224 and $(GATT)_4$ was not observed in the sampled population (n=94), recent expansion of G^{224}-$(GATT)_3$ to G^{224}-$(GATT)_{4\ or\ 5}$ has probably not occurred. Likewise, evidence of expansion of G^{206}-$(GATT)_4$ to G^{206}-$(GATT)_5$ was not observed. The rare A^{224}-$(GATT)_3$ allele could be the result of contraction of A^{224}-$(GATT)_{4\ or\ 5}$ but could also have resulted from inter-allelic recombination between the two variable positions. Overall, it was

possible to infer that the microsatellite alleles have probably not undergone recent expansion from $(GATT)_3$ to $(GATT)_{4 \text{ or } 5}$ or from $(GATT)_4$ to $(GATT)_5$, but recent contraction of alleles by loss of one or more repeat iterations could not be ruled out.

Evaluation of alternative classification schemes

In the classification scheme reviewed by Kato (1991), masu and amago are the distinct species *O. masou* and *O. rhodurus*, respectively. This is consistent with the fact that masu and amago have essentially non-overlapping geographic distributions, consistent differences in coloration, and differing scale morphology. However, the strong similarity in mitochondrial DNA sequences between masu and amago is unlike observed differences between other closely-related pairs in *Oncorhynchus*. For example, the smallest distance observed in the ND3 gene between species pairs was that of rainbow and cutthroat trout, which differ by 5.7% (Chapter 2). These two species also differ by 6.2% in the portion of the mitochondrial control region analyzed in this study (Shedlock et al., 1992). Since other related pairs of species in *Oncorhynchus* have accumulated measurable differences in the DNA of their mitochondrial genomes, it would be reasonable to expect at least some differences between the mitochondrial genomes of masu and amago if they were distinct species. The observation of no type-specific sequence divergence in the ND3 gene and mitochondrial control region argues that these two types diverged from each other much later than any of the other salmon and trout that have undisputed species status.

On first inspection, failure to detect type-specific differences in the mitochondrial genome, while considerable genetic heterogeneity was observed in a nuclear gene,

appears contradictory. Similar results were also obtained when comparing Atlantic salmon (*Salmo salar*) populations in North Wales (O'Connell et al., 1996). However, these seemingly contradictory findings are both consistent with Kimura's (1990) proposal that masu and amago salmon be recognized as distinct subspecies within an *O. masou* complex. This arrangement allows for a close relationship between types (as supported by the mitochondrial data), while having a clear, sub-specific distinction of the types, which is consistent with the differing GH2 allele frequencies, morphological characters, and geographic ranges.

In contrast to that of the mitochondrion, the nuclear genome is inherited in a diploid and bi-parental manner, which allows more potential for polymorphism between closely related individuals. Because the mitochondrial genome is hemizygous and inherited only from the maternal parent, its effective population size is only ¼ that of alleles of nuclear genes. This means that particular mitochondrial haplotypes have a higher probability of drifting to fixation. If this process occurred in the common ancestor of masu and amago, it is likely that most of the nuclear gene polymorphism observed in this study predates the separation of masu and amago populations. I propose that masu and amago salmon have diverged very recently on a macro-evolutionary time scale. A relatively recent divergence of the two types would have allowed insufficient time for a substantial number of differences to accumulate between the mitochondrial genomes. However, enough time has elapsed for genetic drift between the reproductively isolated populations to have produced the dissimilar allele frequencies observed for the GH2 gene and its nested microsatellite locus.

Chapter 4

Evolutionary behavior of duplicated growth hormone genes in salmonid fishes

Abstract:

The proto-salmonid lineage is believed to have undergone a genome-doubling event. In the process of re-diploidization of a genome, mutation of duplicated genes results in their divergence, of which the most extreme form is complete loss of one copy of the gene. Present day salmonids have lost one copy of approximately 50% of their duplicated genes, indicating that re-establishment of disomic inheritance is well underway. Among salmonine species (salmon, trout, char), the growth hormone (GH) gene is represented by two functional, non-allelic isoforms: GH1 and GH2, which argues that each gene has re-established disomic inheritance. In this study, DNA sequence analysis was used to examine the evolutionary history of GH genes in salmonids. A microsatellite locus nested within the fourth intron of all GH genes was invariant in most genera. However, this locus was found to vary both within and among species in the GH2 of *Oncorhynchus*, suggesting it has undergone an evolutionary process unique to this lineage. The overall history of GH genes in Salmonidae was examined by comparing these genes between representative species of the subfamilies Coregoninae (whitefish, ciscos) and Salmoninae. The two GH genes identified in the whitefish species could not be assigned to the salmonine GH1 and GH2 categories, suggesting that the ancestral coregonine and salmonine lineages diverged before the duplicated GH genes had established disomic inheritance.

Introduction:

The proto-salmonid lineage that gave rise to subfamilies Coregoninae (*Coregonus, Prosopium, Stenodus*) and Salmoninae (*Salvelinus, Salmo, Oncorhynchus, Brachymystax, Hucho, Salmothymus, Acantholingua*) is believed to have undergone a genome-doubling event some 25-100 Million years ago (Ohno, 1970; Allendorf and Thorgaard, 1984). Based on comparisons of genome size and chromosome numbers with the related but non-tetraploid smelt family Osmeridae (Hinegardner, 1976; Simon, 1963; Hartley, 1987), the tetraploidization of the salmonid lineage must have occurred after Salmonidae and Osmeridae diverged. Autotetraploidization of a genome (doubling of endogenous chromosomes) produces two identical copies of each gene. In the process of diploidization, or the re-establishment of disomic inheritance, mutation of duplicated genes results in functional or structural divergence. Because newly duplicated genes are functionally redundant, a relaxation in selective constraints can allow the complete loss of one copy of the gene, most likely as a result of nonsense mutations. Present day salmonids have lost duplicated copies of approximately 50% of their genes (Allendorf, 1978), indicating that the process of diploidization of the ancestral tetraploid genome is well underway. In the newly-formed tetraploid genome, many multivalent pairing arrangements would be expected at meiosis (Ohno, 1970). These structures are formed by the pairing of multiple sets of homeologous (duplicated sets of homologous) chromosomes. The fact that a few multivalent structures are still observed in present-day salmonids indicates that the process of diploidization is not yet complete.

The chromosomes of the ancestral salmonid are believed to have been primarily

acrocentric (referring to the subterminal position of the centromere), but the process of Robertsonian fusion (Robertson, 1916) has created many metacentric chromosomes (Ohno, 1970), which are a common feature in present-day salmonid karyotypes. The high frequency of this type of rearrangement is reflected by the degree of variation in chromosome number among and within species. For example, closely related species such as pink and chum salmon have very different chromosome numbers and acrocentric/metacentric ratios (Simon, 1963). Chromosome fusions and other rearrangements likely contribute to the process of genome diploidization by reducing the pairing affinity of homeologous chromosomes. In the absence of meiotic pairing between homeologues, the duplicated, paralogous genes are no longer homogenized by intergenic recombination or gene conversion. This means that the duplicated genes are free to diverge by accumulating mutations.

In salmonids, duplicated isozyme loci are a well documented phenomena (e.g. Lim and Bailey, 1977; Allendorf, 1978). In addition, duplicated, non-allelic forms of a number of genes, such as insulin (Kavsan et al., 1993), insulin-like growth factor (Wallis and Devlin, 1993) and MyoD (Rescan and Gauvry, 1996) have been identified. Among salmonine species (salmon, trout, char), the growth hormone (GH) gene is also represented by two functional, non-allelic isoforms: GH1 and GH2 (Agellon et al., 1988a, 1988b; Agellon and Chen, 1986; Johanson et al., 1989; Male et al., 1992, Devlin, 1993; Du et al., 1993; Forbes et al., 1994, McKay et al., 1996). Although selective constraints have caused this gene pair to remain very similar in protein coding regions, divergence of intronic and flanking DNA sequences indicates that the genes have been separate for a considerable period. The accumulated differences

between these genes argue that the chromosomes or chromosomal regions on which they reside have completed the process of diploidization. The fourth intron (intron D), the largest in salmonid GH genes, in particular has accumulated many species-specific (McKay et al., 1997) and isoform-specific changes that shed some light on its evolutionary history within and among salmonine species (Devlin, 1993; McKay et al, 1996).

In this study, sequence analysis of GH intron D is used to examine the evolutionary history of these duplicated genes in salmonid genera. Analysis of a microsatellite locus nested within this intron (Chapter 3) revealed variation within and among species in the GH2 gene of *Oncorhynchus*, but not in any *Oncorhynchus* GH1 gene or in the GH genes of other salmonid genera. Further, new DNA sequence from intron D of the GH genes in brown trout (*Salmo trutta*), mountain whitefish (*Prosopium williamsonii*) and lake whitefish (*Coregonus clupeaformis*) was used to examine the evolutionary history and patterns of change of GH genes at the generic level. The two GH genes identified in the whitefish species could not be assigned to the categories represented by the salmonine GH1 and GH2 isoforms, suggesting that the ancestral coregonine separated from the proto-salmonine lineage before the divergence of its GH1 and GH2 genes.

Materials and methods:

Species used in this study

The Pacific salmon species masu, chinook, coho, sockeye, pink and chum, as well as rainbow and coastal cutthroat trout were included in this study (species

designations are listed in Table 2.1). The GH intron D sequences of many of these species has been reported previously (McKay et al., 1996 and references therein; Blackhall, 1994). New sequence data were generated from the 3' end of this intron for at least four individuals from each of the Pacific salmon and trout species to assess intraspecific variation. The sequence of the entire intron was also obtained for GH genes of brown trout, mountain whitefish and lake whitefish.

DNA sequence analysis of GH intron D

The region of GH genes that contained intron D was amplified using the polymerase chain reaction (Saiki et al., 1988) with primers GH56 and GH7 as described in Chapter 3. An ancient GH2-like pseudogene is present on the Y-chromosome of most *Oncorhynchus* species (Du et al., 1993; R.H. Devlin, unpublished results). Amplification of the male-specific pseudogene was avoided by using female fish. In all *Oncorhynchus* species and brown trout, the two amplification products corresponding to GH1 and GH2 differed in size. Amplification products were isolated by electrophoresis in low-melting-temperature agarose using standard methods (Sambrook et al., 1989), followed by purification using the Wizard PCR-Prep kit (Promega). Gel-purified amplification products were sequenced directly with the Thermosequenase cycle-sequencing kit (Amersham-United States Biochemicals).

In both lake and mountain whitefish, the GH56 and GH7 primers produced amplification products that migrated as a single band using standard agarose gel electrophoresis. The presence of two, co-migrating amplification products was demonstrated by direct sequencing with the primer GH50 (Table 2.2). Two related

sequences, differing by single nucleotide substitutions and an insertion or deletion (beyond which the sequence was not readable) were superimposed on the autoradiogram. The single band observed by electrophoresis was purified as above, and the amplification products were subsequently cloned using the pCRScript cloning kit (Stratagene).

Restriction endonuclease digestion with the enzyme combination *PstI, SstI* and *SstII* (BRL Life Technologies) identified two classes of clones for the whitefish amplification products. Sequence analysis of the 5' end of the insert from representative clones with primer GH56 revealed that the two classes were different forms of GH intron D. To compensate for the error rate of Taq DNA polymerase (Saiki et al., 1988; Tindall and Kunkel, 1988; Keohavang and Thilly, 1989), two clones of each class were pooled in a 1:1 ratio for sequence analysis. For mountain whitefish, a conflict at one nucleotide position was resolved by sequencing a third clone derived from a different PCR experiment. In the case of lake whitefish, only two clones were recovered, both corresponding to the same GH isoform. A single conflict corresponding to a (T<->C) transition remained unresolved because a third clone was not available. This nucleotide position was treated as missing data.

Sequencing of clones or purified PCR products from all species was performed using a strategy similar to that described in Chapter 2. This analysis differed in that new primers, GH62 (5'-CATTATGCTTTCTAACTA-3'), GH63 (5'-TATAATTTCCCAGTGTGC-3') and GH64 (5'-TTTACCCTAATACAGTGG-3') were used. These primers were designed using an alignment of all known salmonid GH intron D sequences at positions roughly corresponding to those of GH9, GH8 and

GH16, respectively. The latter three primers, which were based on sockeye salmon GH sequences (Devlin, 1993) did not work well for brown trout or whitefish. The complete nucleotide sequence of intron D was determined for intron D of GH1 and GH2 of brown trout, two GH genes (identity discussed below) of mountain whitefish, and one GH gene from lake whitefish. To screen for intraspecific variation in allele size of the nested microsatellite locus, partial sequences from the 5' end of the intron were generated for GH1 and GH2 of all *Oncorhynchus* species listed above using primer GH56 or GH57. For each species, three or four fish were examined simultaneously by pooling separate amplification products in equimolar ratios.

Results and Discussion:

A conserved microsatellite locus is nested within GH intron D

The salmonine growth hormone genes GH1 and GH2 are distinct, non-recombining, paralogous loci (Devlin, 1993) and a tandem duplication of a $(GATT)_n$ tetra-nucleotide is present in the fourth intron of both genes (Figure 4.1). In each, the repeat tract is flanked by related tetra-nucleotide motifs that almost always match the core repeat in three of four positions. Overall, the nested microsatellite loci are in a similar sequence context; the average sequence identity between the GH1 and GH2 introns is 91.2 ± 0.2%. Despite their similar DNA sequence, the paralogous microsatellite loci have met different evolutionary fates. While the GH1 form has a constant GATT repeat number of two (Figure 4.1), the GH2 form varies both within and among species. $(GATT)_3$ is common to the GH2 genes of Atlantic salmon, brown trout, lake trout (*Salvelinus namaycush*), the two whitefish GH loci and an ancient, GH2-derived pseudogene (Du et al., 1993; McKay et al., 1996). It is likely that this form represents the ancestral state for salmonids. The repeat region, sequenced from 3-6

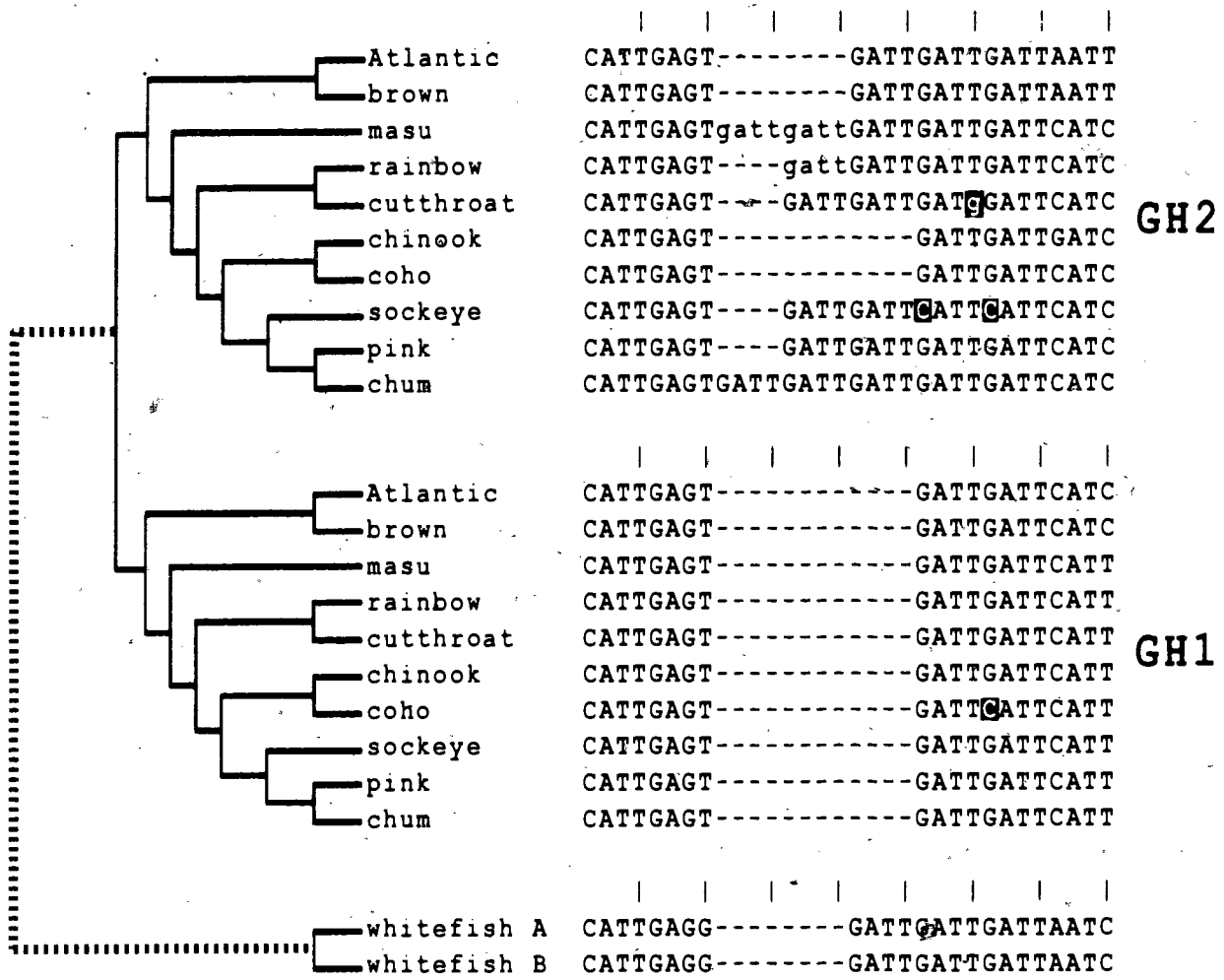


Figure 4.1. The structure of a $(GATT)_n$ microsatellite locus nested within growth hormone intron D. Evolutionary branching of GH genealogies are based on hypothesized evolutionary relationships discussed in Chapter 2. Shaded characters represent nucleotide substitutions within core repeat units. Lower case letters refer to sites that vary intraspecifically.

$(GATT)_4$ that are linked to these alleles had occurred, one would also expect to observe the A->G substitution associated with more than $(GATT)_4$ size category. Except for chinook and coho, the number of repeats differs even between more closely related species pairs. Intraspecific polymorphism was observed in only two species, but lower levels of variation could have gone undetected due to the small number of individuals examined. However, the lack of evidence for extensive variation within most species suggests that the interspecific variation is due either to ancestral polymorphism, or sporadic mutation of this locus occurring episodically over an evolutionary time scale. Such an interpretation is consistent with the observation that repeats have been disrupted by nucleotide substitutions in some species.

If a replication slippage model for microsatellite mutation applies to these loci, it is likely that the observed nucleotide substitutions occurred after the repeat numbers became fixed. The $(GATT)_2(CATT)_2$ form observed only in sockeye GH2 could be the result of a G -> C mutation followed by an expansion of the repeat unit, but this is not consistent with simple replication slippage. Comparison of orthologous cow and goat microsatellite loci revealed that disruption of formerly perfect repeats by nucleotide substitutions greatly reduced the amount of observed variability (Pépin et al., 1994), suggesting that mismatches within short repeats inhibit replication slippage. An independent G->C transversion was observed in an analogous position in coho GH1, indicating that such a substitution is not unlikely on this time scale. A G at that position in two adjacent repeat units of the sockeye GH2 locus could be coincidental rather than the result of amplification. Further evidence that mutations in repeat number predate the observed nucleotide substitutions is provided by the cutthroat trout sequences. Although the coastal (McKay et al., 1996), westslope and Yellowstone (Blackhall, 1994) varieties of cutthroat trout all have four copies of the repeat unit, a (T->G) substitution occurred in one of the repeats in the coastal form after they diverged (Figure 4.1).

The sequence alignment in Figure 4.1 shows the (GATT) repeat in the same 5'-3' orientation as the host gene. First inspection of the observed sequences provides no immediate indication whether mutation is acting on the (GATT) repeat or its complement, (AATC).

Given the similar sequence context and structure of the microsatellite loci, it is surprising that only the *Oncorhynchus* GH2 locus was observed to vary within and among species. Messier et al. (1996) observed that a minimum of two tetra-nucleotide repeats in the primate n-globin pseudogene are required for their expansion. Unlike other closely related pairs of species, coho and chinook have the same GH2 form, $(GATT)_2$. Assuming that $(GATT)_3$ is the ancestral form for both GH1 and GH2, a similar loss of one repeat unit occurred in the antecedent of salmonine GH1 genes (Figure 4.1). The lack of observed variation in species with the $(GATT)_2$ form suggests that at least three repeat units represent the critical threshold for variation at these loci. A simple replication slippage model can be invoked to account for amplification of at least two repeats, but it is not clear why a minimum of three units were associated with size variation in the salmonine GH genes.

Further, lake trout GH2 (McKay et al., 1996), Atlantic salmon and brown trout GH1 and GH2, whitefish GHA and GHB all have three repeat units, despite having been separate longer than the GH2 genes within *Oncorhynchus*. Although a minimum of three repeat units may confer the potential for size variation, it seems possible that factors other than the number of repeat units in the *Oncorhynchus* GH2 locus may also be involved. The fact that variation was only observed in *Oncorhynchus* species could be due to an extrinsic factor specific to this genus, such as a mutation in one the

components of a DNA repair mechanism or replication complex in the proto-*Oncorhynchus* line. If such an extrinsic factor were involved, a more generalized effect on the variability other microsatellite loci in this genus would be predicted. To account for the lack of variability in GH1 under this model, it would be necessary to stipulate a three repeat-unit minimum for replication slippage to occur. However, such a model is not supported by analysis of two other microsatellite loci, where variation in *Oncorhynchus* was no more extensive than in *Salmo* or *Salvelinus* species (Morris et al., 1996).

It has recently been demonstrated in the yeast *Saccharomyces cerivisiae* that mutation of a tri-nucleotide microsatellite repeat is greatly influenced by its orientation with respect to the direction of DNA replication (Freudenreich et al., 1997). When this repeat is replicated in the direction of the lagging strand, the mutation rate is greatly increased, presumably due to the formation of hairpin loops in the Okazaki fragments of the lagging template strand. This model is not directly applicable to a GATT core repeat, as its poor self-complementarity makes such secondary structures unlikely. However, in the *Oncorhynchus* GH2 loci, a related tetra-nucleotide (Figure 4.1) at the 3' end of the GATT repeat tract forms an interrupted, inverted repeat (GATtcATC) with the adjacent GATT. Such an inverted repeat has the potential to form a hairpin loop. However, similar structures are also possible in the Atlantic salmon GH1, brown trout GH1 and both whitefish GH isoforms, where no interspecific or intergenic size differences were observed.

The lack of variability observed in GH1 and all other GH loci except *Oncorhynchus* GH2 suggests that little or no variation in repeat number is the norm for

these loci. If this is the case, the loss of one repeat unit from the antecedent of all GH1 loci may not be responsible for its lack of variability. A trait that distinguishes the *Oncorhynchus* GH2 locus from all other GH types is that it was involved in a chromosomal rearrangement that resulted in the duplication of GH2. Phylogenetic analysis of a Y-linked GH2 pseudogene indicates that it diverged from GH2 after the separation of *Oncorhynchus* and *Salmo* but before *Oncorhynchus* radiated to form the contemporary species (Du et al., 1993). To account for the unusual behavior of the *Oncorhynchus* GH2 microsatellite locus, I propose a model that incorporates the hairpin-mediated strand slippage of Freudenreich et al. (1997). The following assumptions are required: 1) the GH2 gene in *Oncorhynchus* was involved in a complex chromosome rearrangement and is inverted with respect to the other GH2 genes and GH paralogues, 2) the orientation of the inverted GH2 is such that the 5'-$(GATT)_nCATC$-3' is the lagging strand template for DNA replication, 3) the hairpin-mediated mutation process is directional, resulting primarily in expansion of the locus and, conversely, that deletion of repeat units occurs by a different mechanism. Under this model, the Okazaki fragment would occasionally become dissociated from the lagging strand template, allowing the formation of a hairpin loop that would result in slippage by one repeat unit when the fragment reassociates with the template and primes DNA synthesis (Figure 4.2).

This would result in the addition of one repeat unit to the 3' end of complementary strand, such that the polarity of the expansion is opposite to the orientation of the gene. If a hairpin loop were to form in the template strand rather than the Okazaki fragment, it would result in the deletion of one repeat unit and the adjacent

CATC tetra-nucleotide, which was not observed in any of the GH loci. The loss of one repeat unit in the (chinook, coho) lineage could account for the lack of variability between these species. Messier et al.'s (1996) postulated two repeat-unit minimum for variation to occur may not be applicable under this model, as the proposed secondary structure could be too unstable if only four nucleotides at the 3' end of the loop were available to anneal to the template and reprime DNA synthesis (Figure 4.2).

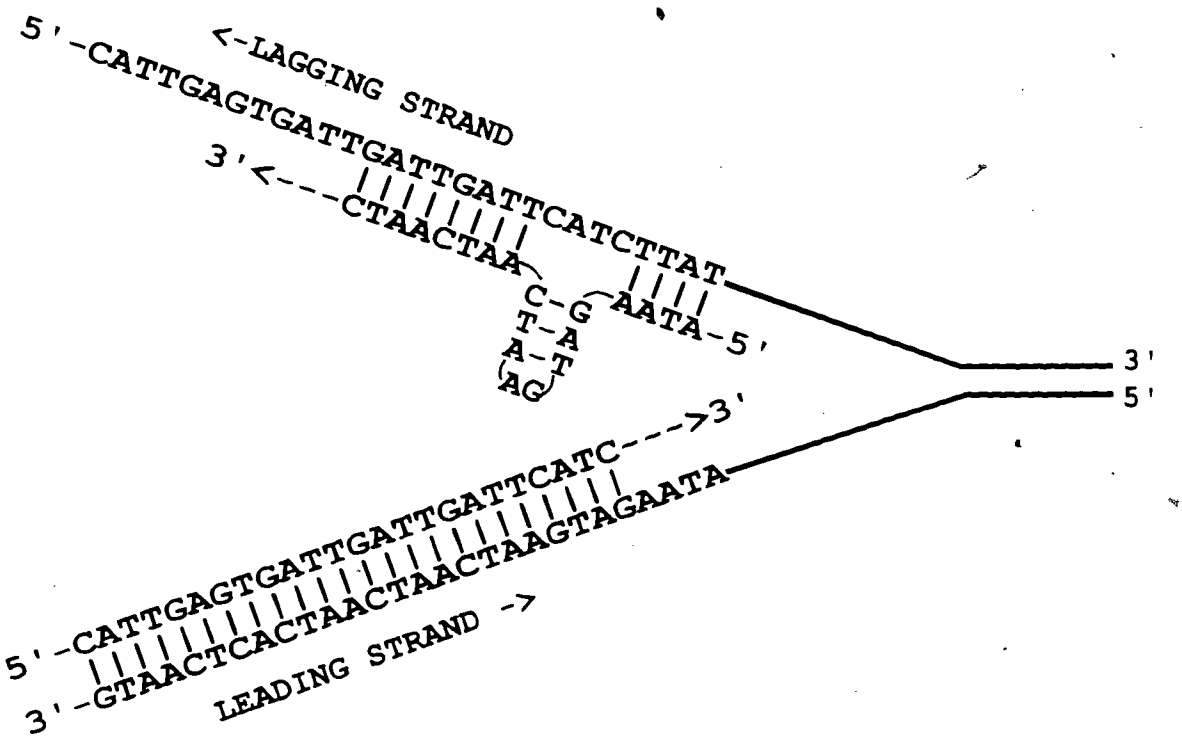


Figure 4.2. A model for expansion of the microsatellite locus by hairpin loop-mediated replication slippage. The model is modified from that of Freudenreich et al. (1997). It is based on the assumptions that the (5'-GATgaATC-3') inverted repeat can form a hairpin loop on the Okazaki fragment, and that the slippage process is more likely to occur in lagging strand replication. Evidence for deletion caused by the formation of a hairpin-loop in the lagging strand template sequence was not observed.

The GH1 and GH2 isoforms are not present in all salmonids

DNA sequence analysis of intron D from whitefish and representative *Salmo* and *Oncorhynchus* species suggests that the GH1 and GH2 isoforms common to the salmonine lineages are not represented in Coregoninae. Full length nucleotide sequences were obtained for intron D of two GH isoforms of brown trout and mountain whitefish (Figure 4.3). Sequence from an additional GH gene in lake whitefish was also obtained. With the exception of a Y-linked, GH2-derived pseudogene in *Oncorhynchus* (Du et al., 1993), only two isoforms have been identified in the salmonine genera (Agellon et al., 1988a, 1988b; Agellon and Chen, 1986; Johanson et al., 1989; Male et al., 1992, Devlin, 1993; Du et al., 1993; Forbes et al., 1994, McKay et al., 1996; Baxter et al., 1996). PCR amplification products produced with primers GH56 and GH7, which anneal to the conserved coding regions that flank intron D, produced two products for each salmonine species tested, which is consistent with there being two growth hormone genes. Similarly, only two amplification products were identified in lake and mountain whitefish. The sequence of a full-length growth hormone gene was obtained for the German lake whitefish *Coregonus lavaretus* (J. Trautner, personal communication), but its relationship to the GH1 and GH2 genes is not clear. Similarly, it was not possible to unambiguously assign either of the other coregonine GH genes to the GH1 or GH2 categories.

For the purpose of this discussion, the mountain whitefish GH genes and their corresponding orthologues from the lake whitefish species were named GHA and GHB Sequence analysis of GH intron sequences has previously revealed that certain deletions or insertions are characteristic of a particular isoform, and can be used as

Figure 4.3. The complete nucleotide sequence of GH intron D from representative salmonid species. Sequence identity is indicated by (.). Alignment gaps are represented by (-). The sequences were aligned manually. Species names are as follows: Ss-*Salmo salar*, St-*Salmo trutta*, Ot-*Oncorhynchus tshawytscha*, On-*Oncorhynchus nerka*, Pw-*Prosopium williamsonii*, Cl-*Coregonus lavaretus*, Cc-*Coregonus clupeaformis*. Cl-GHA is from J. Trautner (Personal communication). The St, Pw and Cc sequences were generated in this study. Sources for other sequences are listed in Chapter 2.

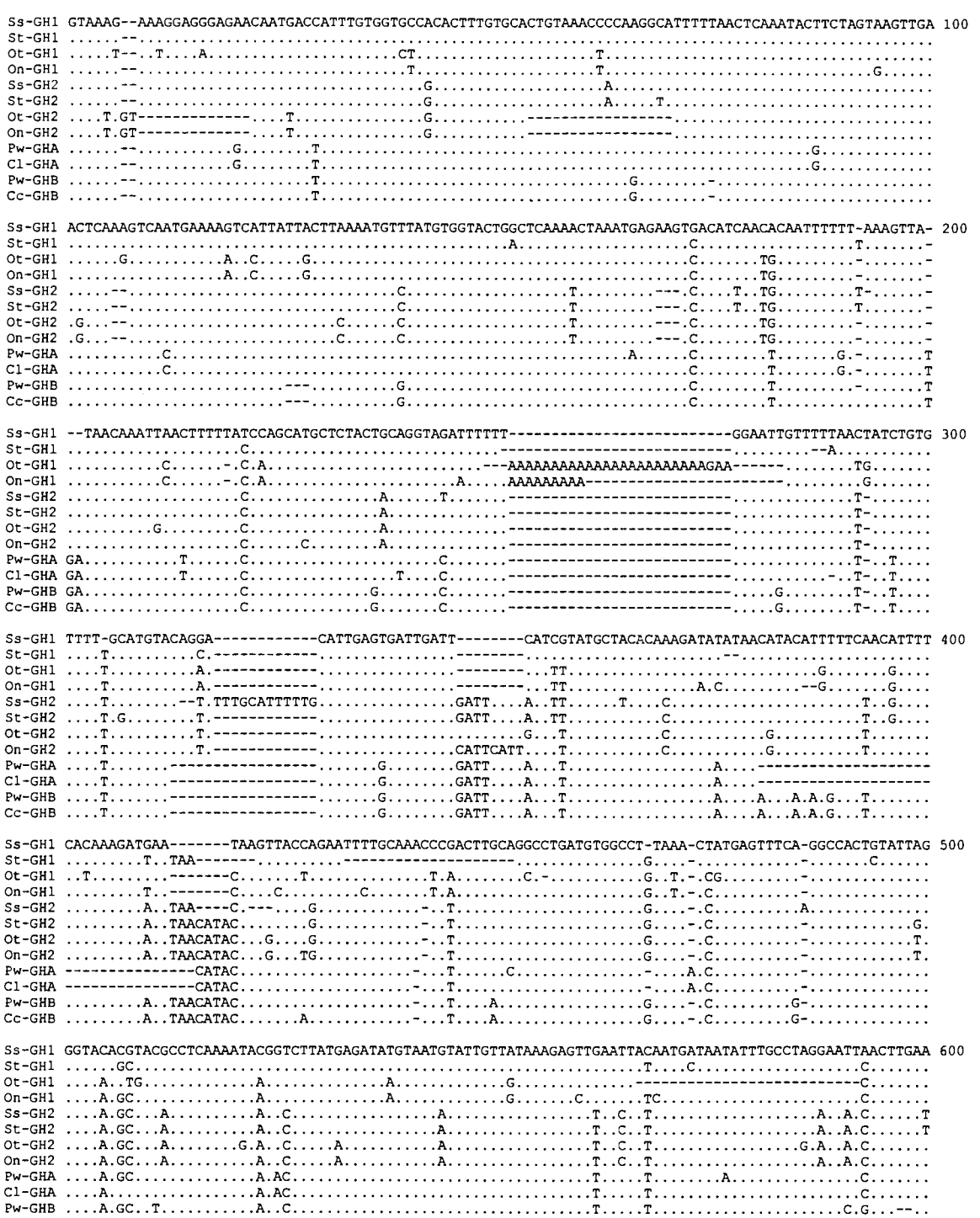
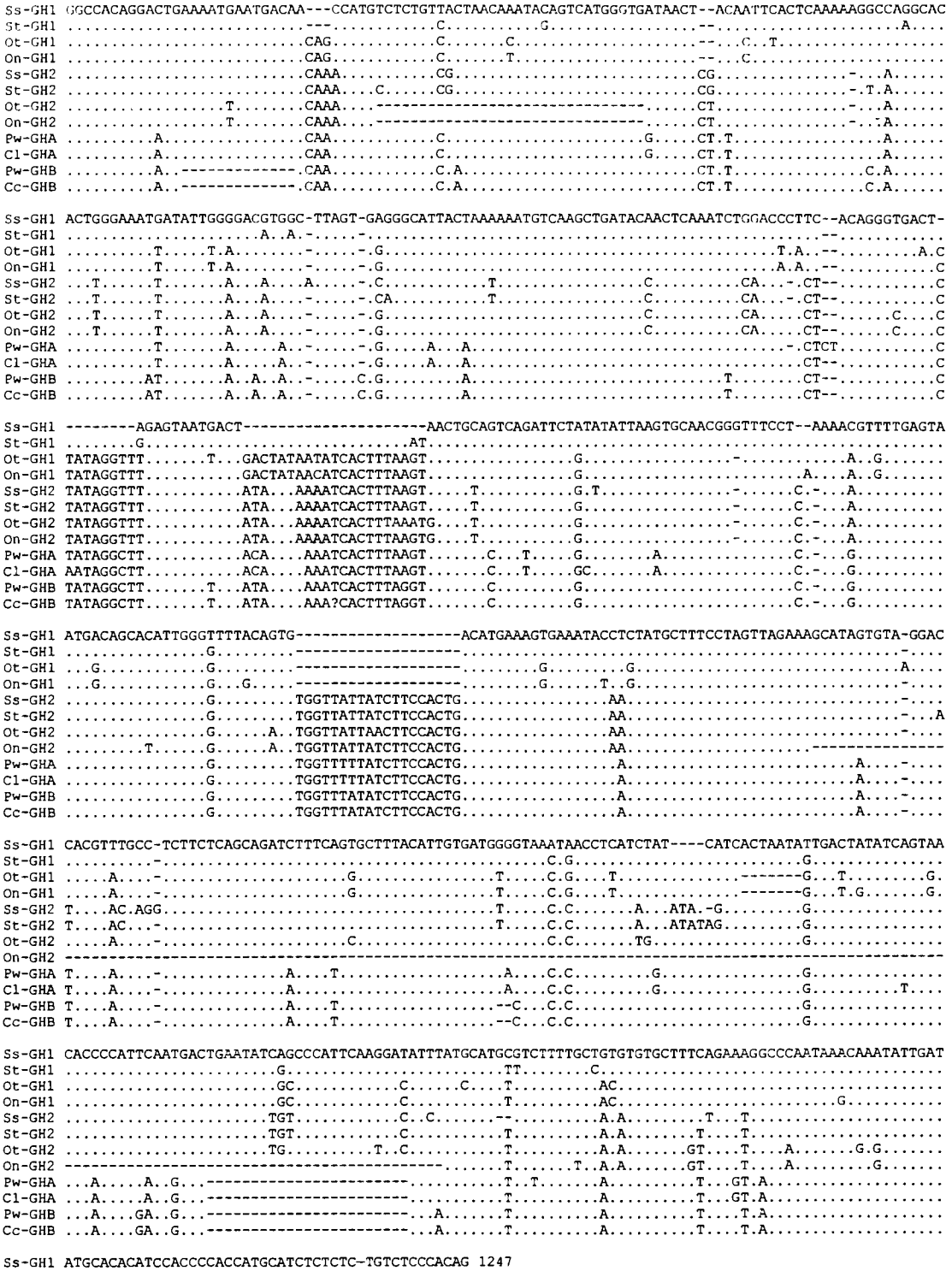


Figure 4.3. (Cont'd)



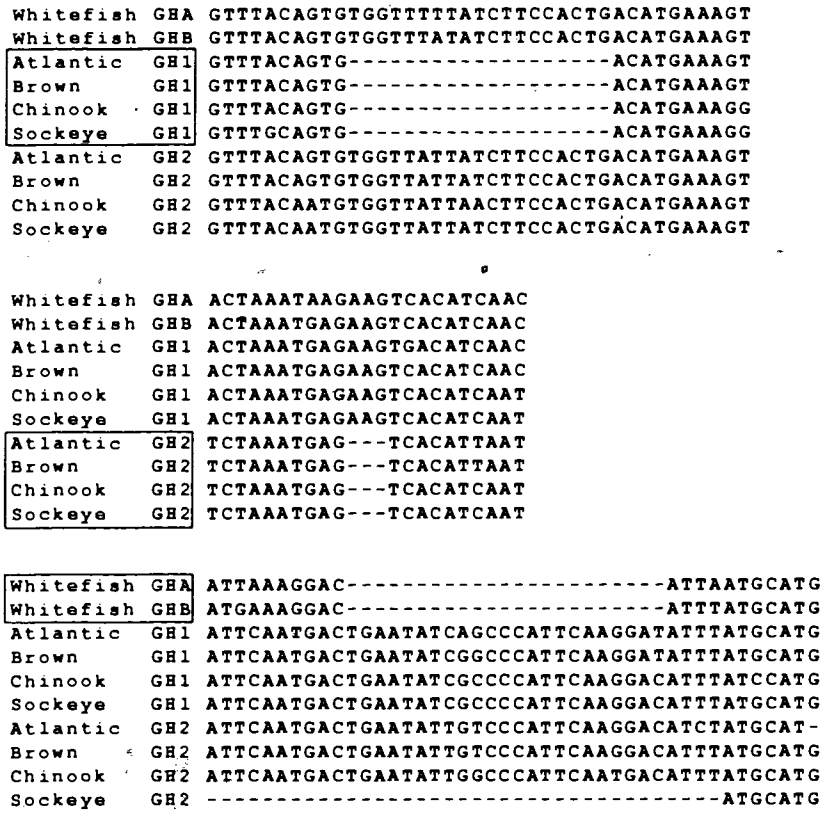


Figure 4.4. Characteristic insertions or deletions in duplicated growth hormone genes. The boxed sequence names refer to genes for which each feature is characteristic.

diagnostic features (Devlin, 1993; McKay et al., 1996). Aligned salmonine and coregonine sequences indicate that such features can not be used to assign the more distantly-related whitefish GH genes as GH1 or GH2 (Figure 4.4), as the whitefish introns possessed features of both. Pair-wise distance measures do nothing to clarify the relationship (Table 4.1): the average whitefish GHA-GH1distance (8.1 ±0.5%) is the same as the GHA-GH2 distance (7.9 ±0.3%). Similar distances were obtained for whitefish GHB (7.9 ±0.5% vs. 7.6 ±0.3%). A surprising finding was that the GHA and GHB introns differ by only 3.1%, which is three-fold less than the average GH1-GH2 distance of 8.8 ±0.2%. This difference is not consistent with each of the paralogous gene pairs having diverged at the same time, as would be expected if diploid inheritance of these genes had been established before the coregonine and salmonine

Table 4.1 Pair-wise Kimura 2-parameter distance comparisons (in percent) based on growth hormone intron D sequence data. To consider only sequences common to all genes, sites containing alignment gaps were deleted.

	STGH1	OTGH1	ONGH1	SSGH2	STGH2	OTGH2	ONGH2	PWGHA	PWGHB
SSGH1	2.13	6.49	6.33	7.66	7.65	8.70	8.68	7.07	6.78
STGH1		7.36	6.76	7.80	7.78	8.84	8.82	7.49	7.36
OTGH1			3.08	9.04	9.02	10.25	9.94	9.17	8.73
ONGH1				8.70	8.69	9.91	9.89	8.69	8.56
SSGH2					1.32	3.64	3.78	7.48	7.20
STGH2						3.22	3.36	7.33	6.90
OTGH2							1.05	8.37	8.10
ONGH2								8.52	8.24
PWGHA									3.08

lineages diverged. Assuming that divergence between isoforms occurred only after homologous or homeologous exchange due to recombination or gene conversion ceased, the greater degree of similarity between GHA and GHB may indicate that recombination or genes conversion between these isoforms stopped occurring more

recently than it did between GH1 and GH2.

Phylogenetic analyses using the neighbor-joining, maximum parsimony, and maximum likelihood methods all produced the same tree (Figure 4.5). The inferred relationships between the genes indicates that GHA and GHB have a much stronger

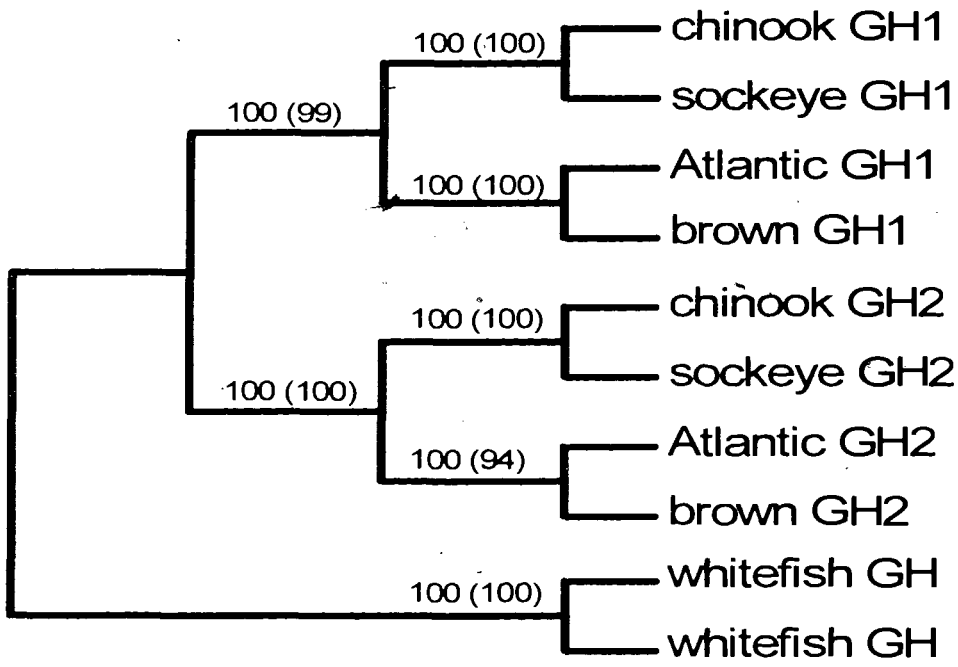
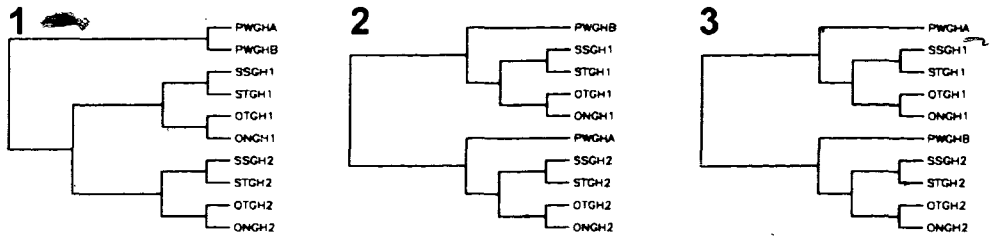


Figure 4.5. Inferred genealogical tree for duplicated growth hormone genes. The tree represents a consensus phenogram generated by maximum parsimony, neighbor-joining and maximum likelihood analyses. Numbers at nodes represent parsimony and neighbor-joining (in parentheses) bootstrap confidence levels for 2000 replicates. All gap sites in the sequence alignment were removed for phylogenetic analysis.

phylogenetic affinity for each other than for GH1 or GH2, which form monophyletic clades distinct from the GHA and GHB. This finding was very unambiguous, as indicated by universally high bootstrap confidence levels, and by statistical analysis of maximum likelihood ratios (Table 4.2; Kishino and Hasegawa, 1989).

Table 4.2. Statistical evaluation of branching order in growth hormone genealogies. The Ln Likelihood (L) values of trees 1-3 were compared using the Kishino and Hasegawa (1989) test in the program DNAML v3.57c (Felsenstein, 1993). Tree 1 places the whitefish GH genes in a separate lineage distinct from those of GH1 or GH2; trees 2 and 3 test alternative arrangements that separate whitefish GHA and GHB in the GH1 or GH2 clades.



Tree	Ln L	Diff. Ln L	St. Dev.	Significantly worse?
1	-2112.78208	<------ best		
2	-2152.48488	-39.70280	13.7797	Yes
3	-2152.48715	-39.70508	13.7729	Yes

The most parsimonious explanation for these findings is that diploid inheritance of the duplicated GH gene had not been established when the Coregoninae diverged from the salmonid evolutionary line. There are several alternative models to explain these observations. The implicit assumption of each is that disomic inheritance of the GH paralogues had been established before the radiation of Salmonidae, and that there are only two ancestral GH genes for all extant salmonids. Alternative explanations can be discounted as follows. In the first model, there was a slowdown in the rate of fixation of mutations in the coregonine lineage (GHA vs. GHB), resulting in a greater similarity to the ancestral GH sequence and less divergence between these paralogues. Because a non-salmonid GH outgroup was not available, the relative

rates of GHA/B vs. GH1/2 could not be established. However, the assumption that disomic inheritance predates the coregonine/salmonine split means some divergence must already have occurred between the GH paralogues. This argument can not account for the complete lack of phylogenetic affinity of GHA and GHB for either GH1 or GH2. A rate slowdown in the coregonine lineage would have resulted in less homoplasy, thus less noise to swamp out whatever phylogenetic signal was present before it diverged from the salmonine line. This conflicts with the observed tree. A second explanation is that one of the original coregonine GH genes was lost, and the remaining one was subsequently duplicated. This three-step scenario is less parsimonious. Since it is assumed that there were only two ancestral salmonid GH genes, the more recently duplicated gene pair would have to be closer to either GH1 or GH2, which conflicts with the observed distances. Under a third model, there are more than two GH genes in coregonine species and GH1 and GH2 have not yet been identified. There is ample precedence from salmonine species that only two functional GH genes are present (Agellon et al., 1988a, 1988b; Agellon and Chen, 1986; Johanson et al., 1989; Male et al., 1992, Devlin, 1993; Du et al., 1993; Forbes et al., 1994; Baxter et al., 1996). Moreover, GHA and GHB were detected using PCR primers designed based on the conserved coding regions of GH1 and GH2 genes. It seems more likely that GH1 and GH2 would be detected more easily than less closely related, non-orthologous genes. However, it should be noted that preliminary results indicate that this conserved primer pair produces at least three amplification products in arctic grayling (subfamily Thymallinae, *Thymallus arcticus*), which provides some impetus for a more rigorous examination of this question.

Based on the high degree of variability in chromosome number and structure, and the fact that multivalent figures can still be observed at meiosis, it can be argued that the process of diploidization of the salmonid genome is ongoing. It seems likely that many loci were inherited tetrasomically when the subfamily Coregoninae split from the salmonid lineage. From the evidence presented here, it is likely that the duplicated growth hormone gene had not completely established disomic inheritance at this point. The inferred growth hormone genealogy and DNA sequence divergence data suggest that the chromosomes containing these genes only became fully diploidized after the two lineages diverged.

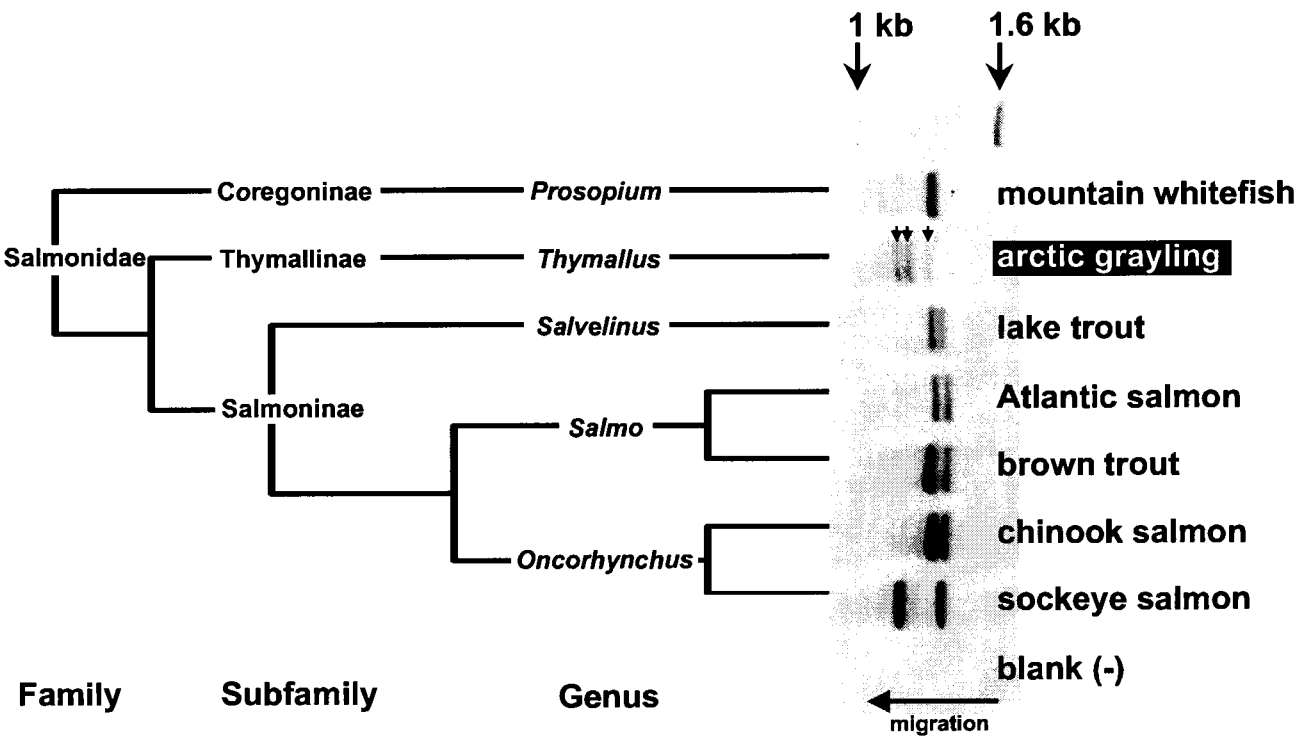


Figure 4.7. PCR amplification of GH intron D from representative salmonid species. Intron D plus flanking exon sequences was amplified using primers GH 56 and GH7 (Chapter 3). Two amplification products were identified for all species except arctic grayling (*Thymallus arcticus*), which produced three (small arrows).

Chapter 5

General Conclusion

The objective of this thesis was to resolve evolutionary relationships among Pacific salmon and trout of the genus *Oncorhynchus*. Through the course of genus-level phylogenetic analysis (Chapter 2), the unexpected findings that masu and amago salmon are probably not distinct species, and that the duplicated growth hormone genes have behaved differently in other salmonid genera, gave rise the investigations described in Chapters 3 and 4. The end product is a downward progression from genus to species to individual genes. The underlying theme is the use of DNA sequence analysis to uncover patterns of variation in present-day species, and to infer evolutionary relationships therefrom.

Gene trees vs. phylogeny

Although some species in *Oncorhynchus* had been analyzed at the DNA sequence level, not all were represented in molecular phylogenetic analysis. I used the sequence of a nuclear gene (GH2) and a mitochondrial gene (ND3) to study the phylogeny of all Pacific salmon and representative trout species. The use of one gene from each genome was intended to assess the degree to which the two data sets agreed, as agreement between independent analyses lends some intuitive measure of confidence to particular conclusions. In this case, the history of disagreement between independent studies persisted; each of the genes inferred different trees. Moreover, as additional data sets were assembled from other mitochondrial genes, it became

apparent that even genes from the same region of the mitochondrial genome could not agree on the deeper evolutionary branching order.

As was shown for the ND4L and D-loop data sets, some DNA sequences are less reliable as indicators of species phylogeny due to large rate variations among lineages or alignment ambiguities. However, the failure of COIII and ND3 to agree with the tree inferred by ATPase 6 is troublesome. These three genes are all from the same contiguous stretch of DNA in the mitochondrial genome (Thomas and Beckenbach, 1989; Oohara et al., 1997). It is very unlikely that their evolutionary histories differ. Although the phylogeny inferred using the DNA sequence of a single gene may coincide with the species phylogeny, deciding which is the true tree is problematic.

Is a star phylogeny resolvable?

From the estimated divergence times (Figure 2.7), it can be seen that the first three branches in the inferred *Oncorhynchus* phylogenetic tree occurred over a very short interval. This could account for the poor resolution of the exact branching order. The most conservative approach would be to interpret the tree as a star phylogeny, with the controversial nodes collapsed to a basal polytomy. Rather than reflecting an evolutionary reality, however, this would likely attest to the poor resolving power of phylogenetic inference based only on extant species. This is not a general limitation, rather it refers specifically to the case where a weak phylogenetic signal is built-up during a rapid succession of speciation events.

Over the relatively long interval between the ancient radiation of lineages and

the sampling of extant species, a weak signal would be obscured by the accumulation of uninformative changes. In this case, the ideal gene for phylogenetic analysis would have to have been rapidly evolving when the species radiated but, paradoxically, would have to be evolving slowly enough that few uninformative changes could have accumulated since that time. Although slowdowns in the rate at which mutations are accumulated are possible, it seems improbable that they could occur independently in each of the new lineages created by a burst of speciation.

In this study, I tried to resolve the phylogeny of *Oncorhynchus* using a combined approach that involved pooling all available data into one large character set. The rationale for such an approach is as follows. I assume that each data set contains some signal from the true phylogeny. In many cases the stochastic accumulation of noise such as homoplasy due to multiple substitutions or convergent evolution of other characters may obscure this signal. In cases where the tree inferred from a particular data set does not represent the actual phylogeny, the signal has been swamped out. Because many analyses have produced discordant trees that usually disagree in more basal branching order, the underlying signal for controversial nodes must be generally weak relative to the accumulated noise. Assuming minimal confounding factors such as introgressive hybridization or non-venereal (horizontal) gene transfer, the phylogenetic signal should carry the same information for each data set, whereas the accumulation of noise is arguably random.

It follows that in pooled data sets, the signal accumulates additively, while the random background noise would tend not to be reinforced in the same way. The advantage of such an approach is that it can take a weak signal into account even in

data sets that do not recover the correct tree when treated in isolation. This effect is quantifiable. The BCL values at basal nodes in the total evidence tree exceeded those of any individual data set (Figure 2.5). Although BCLs are a better indicator of self-consistency than as a test of confidence in a phylogenetic hypothesis, it is clear that a signal that does not dominate in all individual character sets is reinforced when all data are considered together. Whether the reinforced signal is that of the true phylogeny is debatable. However, it is interesting that maximum likelihood analysis of the pooled sequence data converged on precisely the same tree. These two approaches are at least partially independent, as maximum likelihood estimation uses all nucleotide positions (Table 2.6), while parsimony considers only synapomorphic characters (Table 2.5).

The taxonomic status of masu and amago salmon

Two competing classification schemes are in current usage for masu and amago salmon. They are either considered separate species (Kato, 1991), or races (Kimura, 1990). The initial finding that their ND3 genes and a portion of the D-loop regions were virtually identical, combined with a more extensive analysis of a large portion of the mitochondrial DNA (Oohara and Okazaki, 1996), implies that they can not be distinguished based on fixed differences in this genome. The seemingly paradoxical finding that their nuclear GH2 gene is more variable also fails to provide a clear distinction of the type associated with separate species. For example, no other pairs of related species in *Oncorhynchus* have such similar ND3 or GH2 genes. The allelic variation of GH2 appears to predate the separation of masu and amago, as

almost all alleles are present in both. The fact that the allele frequencies differ substantially between the types does provide a genetic basis for a distinction, but the overall morphological, meristic and mitochondrial DNA similarities argue against a classification scheme that assigns species status to these salmon. Because the rate at which mutations are accumulated can vary even among closely related groups, a species definition based on DNA sequence divergence is difficult to apply. Nevertheless, the most reasonable explanation for virtually identical mitochondrial genomes is that masu and amago share a very recent common ancestor. Whether this is due to recent divergence or to coalescence of two lineages by introgressive hybridization is not clear. The fact that the two types hybridize readily when brought together (Oshima, 1955) is consistent with the genetic homogenization observed between cultured populations of both varieties (Figure 3.4).

It is possible that the larger degree of variation observed in the nuclear genome indicates that masu and amago were once distinct lineages, and that the mitochondrial genome of one was introgressed into the other. Because of the broad geographic range of sampling sites, such an exchange would have to have predated the spread of masu throughout Japan. An alternative explanation is that the two lineages have only recently diverged. The higher degree of variability in the GH2 is not necessarily inconsistent with this idea. With an effective population size ¼ that of the nuclear genome, it is possible that the lack of type-specific variation in the mitochondrial genome is the result of fixation by random drift in a recent common ancestor of both types. The fact that the $(GATT)_5$ allele appears to drifting toward fixation in amago but not in masu argues that substantial gene-flow between contemporary wild populations

of the two types has not occurred (Figure 3.4). Regardless of their recent evolutionary history, the observed overall similarity between masu and amago is more consistent with the classification scheme (reviewed by Kimura, 1990), which treats masu and amago as conspecific races.

Evolution of duplicated growth hormone genes

In Chapter 4, evidence is presented that coregonine fishes have growth hormone genes that do not fall into the categories defined by GH1 and GH2, the two functional growth hormone genes of salmonine fishes. This has led to a re-evaluation of the idea that the two GH genes in the ancestral salmonid had established disomic inheritance and started to diverge before the radiation of Salmonidae (Devlin, 1993). The two GH genes isolated from whitefish are more similar to one another than GH1 is to GH2, and are equally dissimilar from both GH1 and GH2. This implies that the evolutionary history of coregonine GH genes differs from those of Salmoninae. The most parsimonious explanation is that GHA and GHB in whitefish share a more recent common ancestor than GH2. This implies that GH1 and GH2 lost the ability to be homogenized by homologous or homeologous pairing and recombination earlier in their history.

Another possible explanation is that there are (or were) more than two GH genes in the coregonine lineage. Under this scenario, one of the GH genes was duplicated, resulting in GHA and GHB. The failure to detect another coregonine GH gene with conserved PCR primers implies that it has been lost or has diverged in its protein coding sequences. There is a sizeable body of evidence that salmonines have

only two functional GH genes (Agellon et al., 1988a, 1988b; Agellon and Chen, 1986; Johanson et al., 1989; Male et al., 1992, Devlin, 1993; Du et al., 1993; Forbes et al., 1994; Baxter et al., 1996). Further, evidence from a genomic Southern blot using a probe from the conserved GH coding region indicates that there are only two GH genes in *Coregonus lavaretus*, a German relative of lake and mountain whitefish (J. Trautner, personal communication). Assuming that the ancestral GH genes diverged before the subfamilies Coregoninae and Salmoninae and that the whitefish GH genes are the result of a more recent duplication, the whitefish GHA and GHB should both resemble one of the salmonine isoforms more closely. That fact that they do not suggests that they are not the result of a recent duplication. It is possible that a more ancient duplication occurred when the ancestral GH paralogues were still very similar to one another, but the passage of time since that event would have allowed substantial accumulation of differences between GHA and GHB. This conflicts with the relatively high degree of sequence identity observed in the intron sequences of these genes.

Although the DNA sequence analysis reveals patterns that argue against GHA and GHB having resulted form an independent duplication event, there is insufficient evidence to entirely discount the possible existence of more than two GH genes in some salmonid lineages. For example, GH2 is known to have been duplicated early in the history of *Oncorhynchus* (Du et al., 1993). Conserved PCR primers from the fourth and fifth exons of GH genes, designed to amplify across intron D (Figure 3.1), recover two amplification products from whitefish, salmon, char and trout (Figure 4.7). However, arctic grayling, which represents the salmonid subfamily

Thymallinae, produced three amplification products. This suggests that there may be at least three conserved GH genes in this lineage. The assertion that there are only two GH genes in all salmonid lineages probably requires further investigation.

Toward a model for microsatellite evolution

A model to explain the evolution of the $(GATT)_n$ microsatellite locus in *Oncorhynchus* GH2 is proposed in Chapter 4. This model seeks to explain the finding that this locus is variable only within Oncorhynchus GH2. Although some sequence differences exist between the GH1 and GH2 loci, no unique sequence element of the locus or flanking regions can explain why it has been amplified in GH2 of *Oncorhynchus*, but not *Salmo* (Atlantic salmon and brown trout), *Salvelinus* (char), or in GHA or B of *Coregonus* and *Prosopium* (whitefish). Although the paralogous GH1 microsatellite locus in *Oncorhynchus* has contracted by one repeat unit, no variation was observed within or among species. A simple replication slippage model with a three-iteration minimum for variation can not satisfactorily explain why no variation was observed in similar sequences from the GH genes of four other genera, all of which also have three repeat units. The single known feature which distinguished *Oncorhynchus* GH2 from all others is that it was involved in a chromosome rearrangement early in its history (Du et al., 1993).

It has recently been demonstrated that a yeast microsatellite locus capable of forming hairpin-loops is much more variable in a particular orientation with respect to the direction of DNA replication (Freudenreich et al., 1997). I have proposed a similar a model for evolution of the GH2 locus that is based on several assumptions: 1) GH2

in *Oncorhynchus* has been inverted with respect to the direction of DNA replication, and to the orientation of all other salmonid GH genes, 2) a hairpin loop formed at the 3' end of the GATT repeat is sufficiently stable to occasionally mediate replication slippage, resulting in the addition of one repeat unit, 3) a minimum of three repeats is required for this to occur and 4) contraction of the locus occurs by a more general replication slippage mechanism, and reduction to two repeat iterations precludes further variation.

Although assumptions 3 and 4 are consistent with the observed sequence variation, assumptions 1 and 2 are untested. Barring the ability to test these assumptions, the model must remain conjectural. Knowledge of salmonid karyology is not sufficiently detailed to evaluate the orientation of the GH2 locus, so direct verification of assumption 1 is not currently possible. However, it is conceivable that both assumptions could be tested in vitro. The region in question could be placed in alternative orientations in a genetic construct, such as a yeast artificial chromosome, and be tested for variability after passage through many generations in cultured yeast. If an orientation-dependent, hairpin-mediated slippage mechanism does apply, the short life-span and concomitant high frequency of DNA replication could result in variation in one orientation but not the other.

Application of DNA sequence data to fisheries research

Understanding the evolutionary relationships among salmonid species has direct and indirect implications for conservation and fisheries genetics. A secondary motive existed for generating new DNA sequence data for the nuclear GH2 gene. By

obtaining sequence information for all salmon and trout species that occur in British Columbia, it was possible to design a simple, PCR-based method of species identification. Appendix 3 describes a series of experiments directed toward this goal. Although a descriptive report of this nature does not fall within the parameters defined by the theme of this thesis, Appendix 3 is included to demonstrate the practical application of information used in a more theoretical approach.

The work described in Appendix 3 also served to address a theoretical consideration raised in chapter two, namely the effect of intraspecific variation on phylogenies inferred from individual species representatives. Deletions are important source of variation in GH introns (Devlin, 1993). Intron D, the subject of much of this thesis, is particularly variable in this regard (McKay et al., 1996). With the possible exception of chum salmon, the evaluation of representatives of several populations for each *Oncorhynchus* species described in Appendix 3 demonstrated that no detectable changes in intron size or restriction sites are present in the GH2 gene of any of the species used in this thesis.

Oncorhynchus Phylogeny: Where to go from here?

The economic and recreational importance of salmon and trout species makes them a much loved, and consequently much studied group of fish. Apart from the intellectual appeal of solving long-standing problems regarding their taxonomy and nomenclature, there are modern issues that render an understanding of evolutionary relationships among these fish more than just an academic question. Although the focus of the phylogenetic analysis described in this thesis is more on the genus and species levels, it bears at least indirectly upon challenges facing the increasingly

managed salmon and trout populations. Knowledge about the nature and relationships of species lays the foundation for the emerging field of conservation genetics; a field whose robust growth is inversely correlated to the health of endangered stocks.

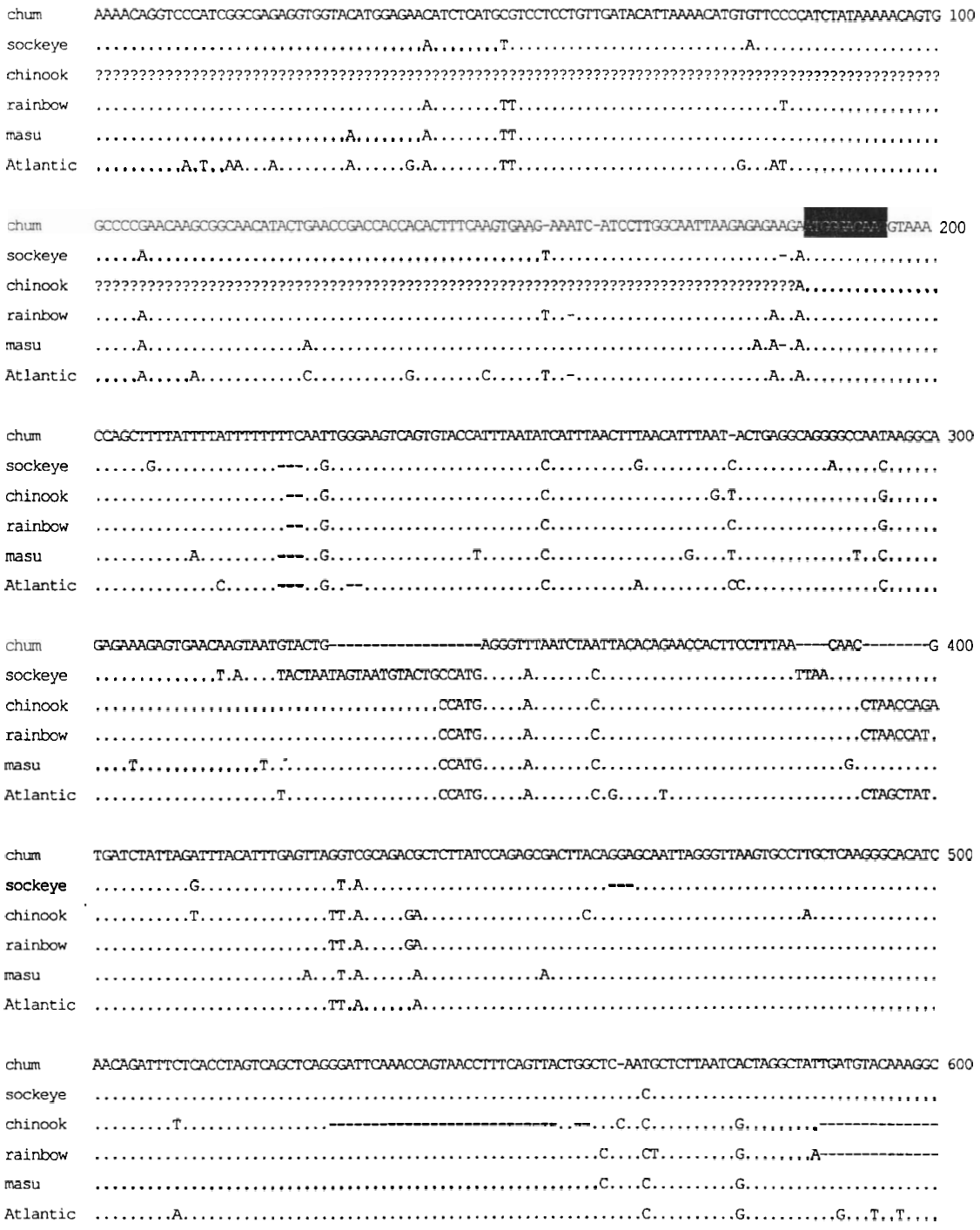
The synthetic treatment of salmon phylogeny described in Chapter 2 provides good evidence that pink and chum salmon clade is monophyletic, which has been the source of disagreement in the past. Other elements of the total evidence tree, such as the monophyly of all North American pacific salmon group and the (rainbow, cutthroat) clade are convincing given the phylogenetic consensus in these areas. A certain measure of caution must be used in accepting the relative branching order of the Asiatic salmon and Pacific trout groups (nodes 1 and 2 in the total evidence tree). A previous total evidence analysis using less mitochondrial DNA sequence (McKay et al., 1996) found the positions of nodes 1 and two to be reversed, which agreed with the phylogenetic consensus at that time.

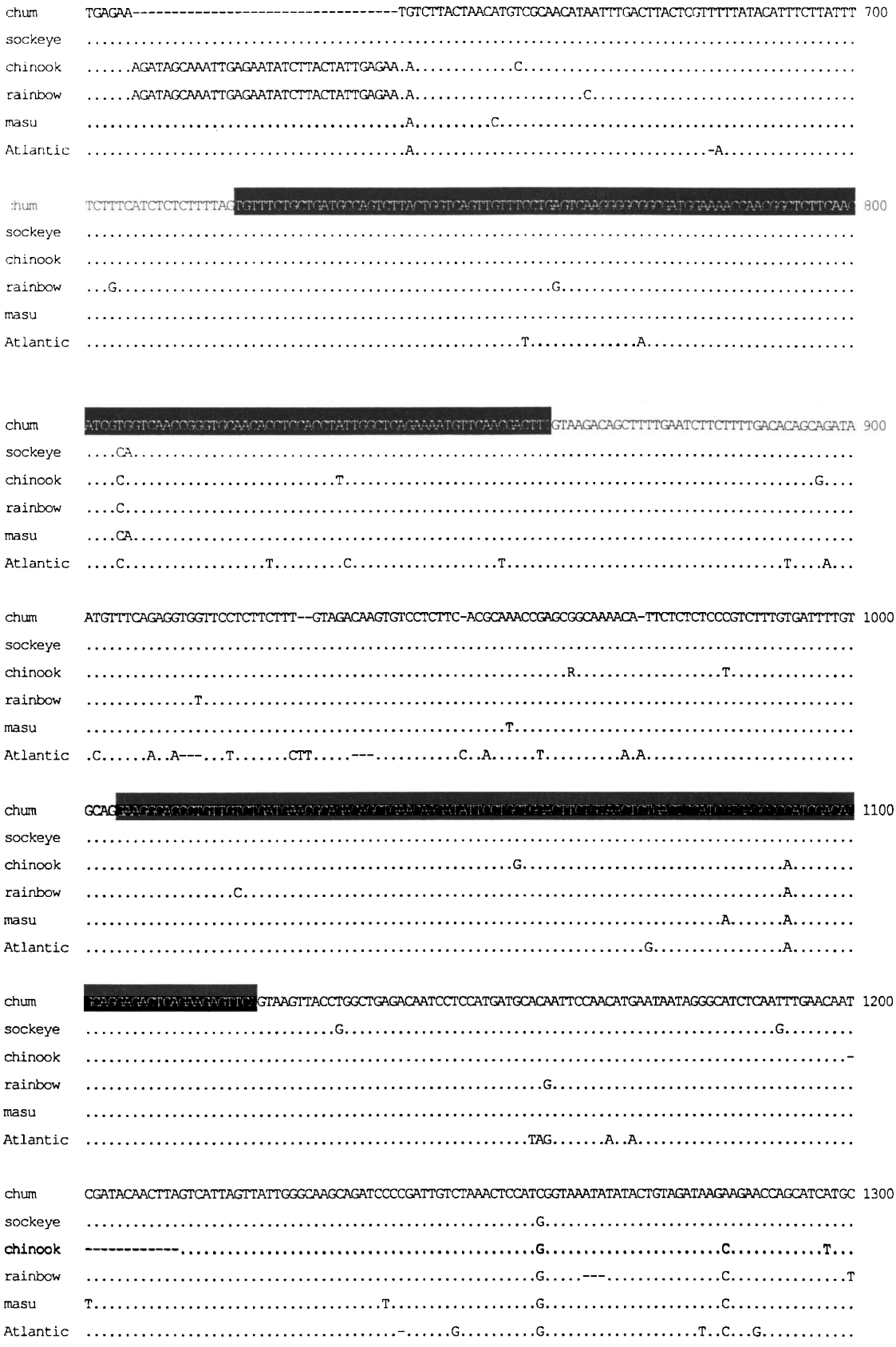
Since the speciation events that created these nodes were estimated to have occurred at or about the same time (Figure 2.7), the resolution of their exact order may require further analysis. Considering only sequences represented in all nine taxa, the majority of the data are from the mitochondrial genome. If there were a bias imposed by the preponderance of one data type, then the statistical support provided by the analysis presented here could be a reflection of the mitochondrial genome tree, which could differ from the species tree. If the true mitochondrial and nuclear trees agree with each other and the actual phylogeny of *Oncorhynchus*, then inclusion of more nuclear DNA sequence data in a combined analysis would only serve to increase the confidence in the conclusions regarding the order of the first two nodes. If the true

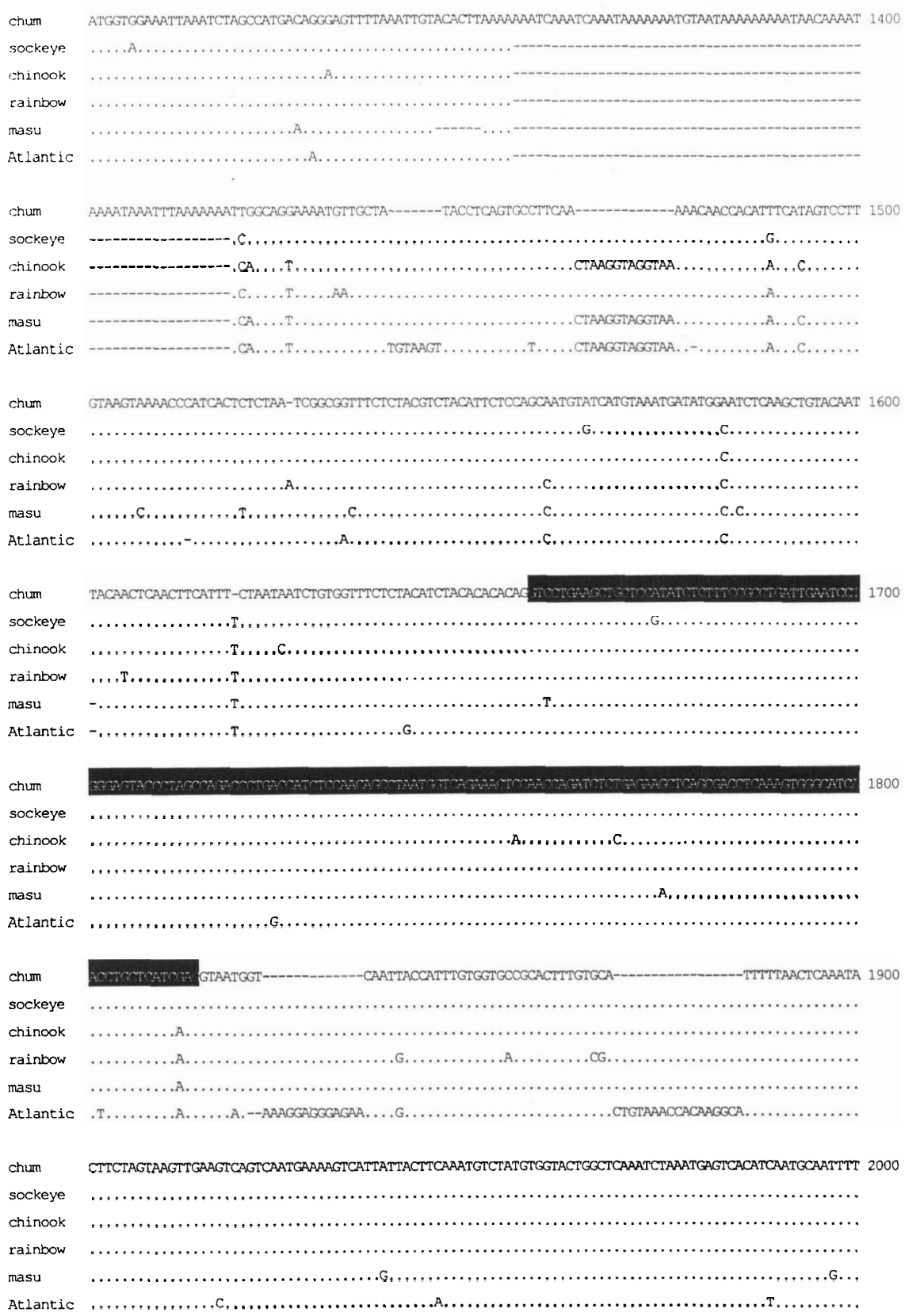
nuclear and mitochondrial trees were to disagree, an expanded nuclear component of the combined data set would cause a reduction in support for the basal branching order inferred in this study, in which case a basal tritomy in the tree would likely better reflect evolutionary reality.

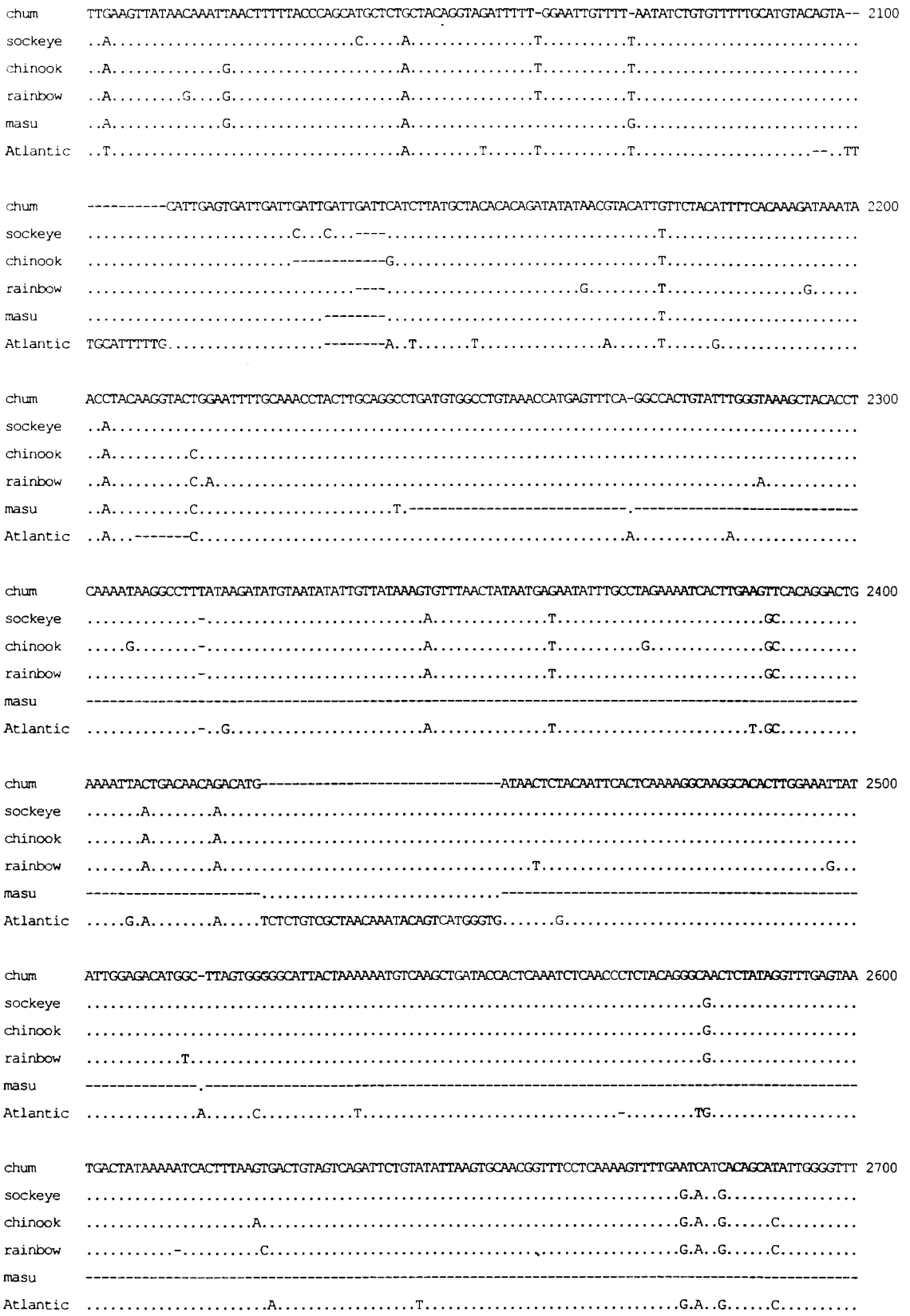
Appendix 1

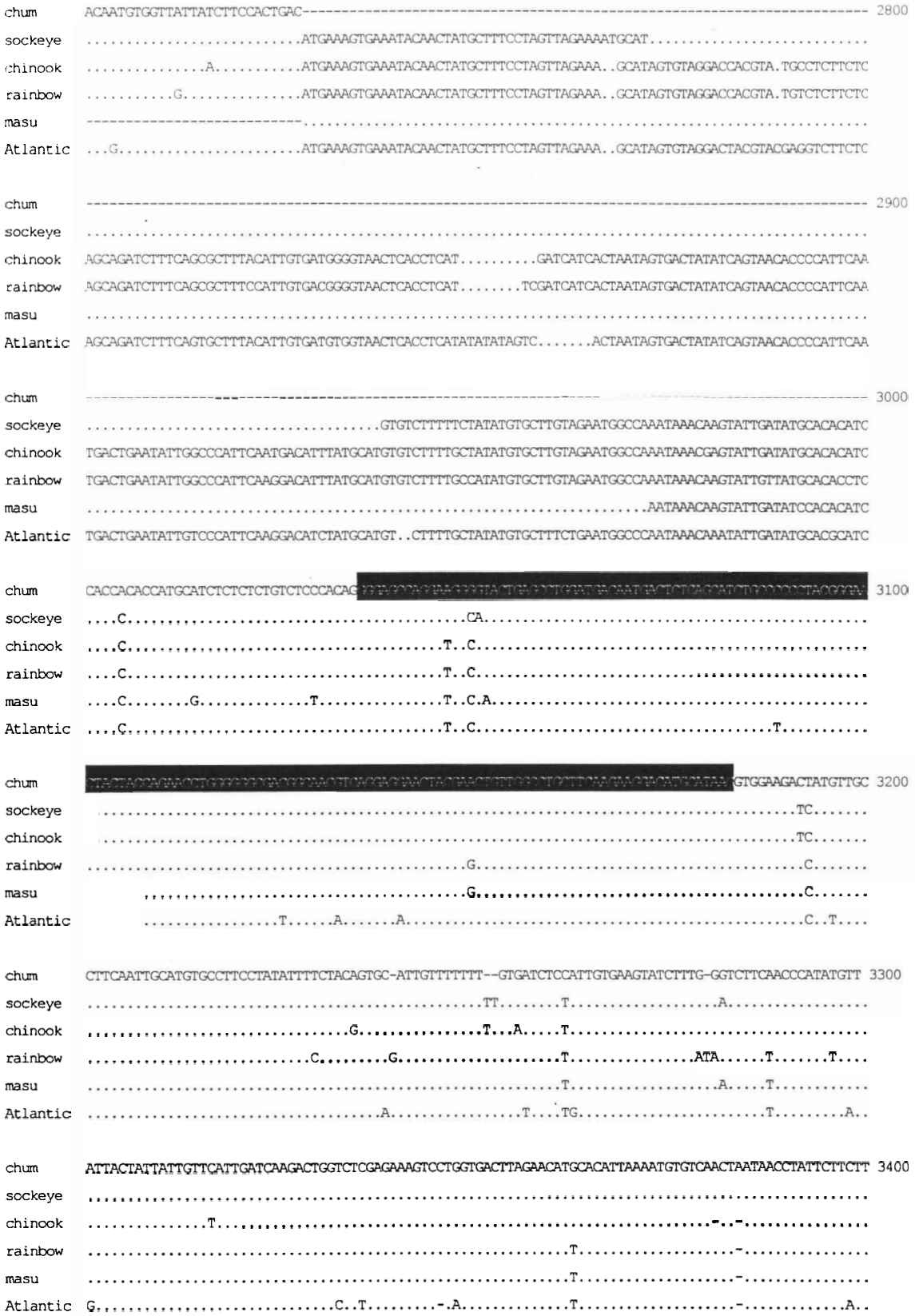
Aligned DNA sequence of complete the GH2 genes used in phylogenetic analysis (Chapter 2). The chinook and masu sequences were generated in this study. Dots (.) indicate identical sequence. Alignment gaps are indicated by (-). Exon sequences are shaded.

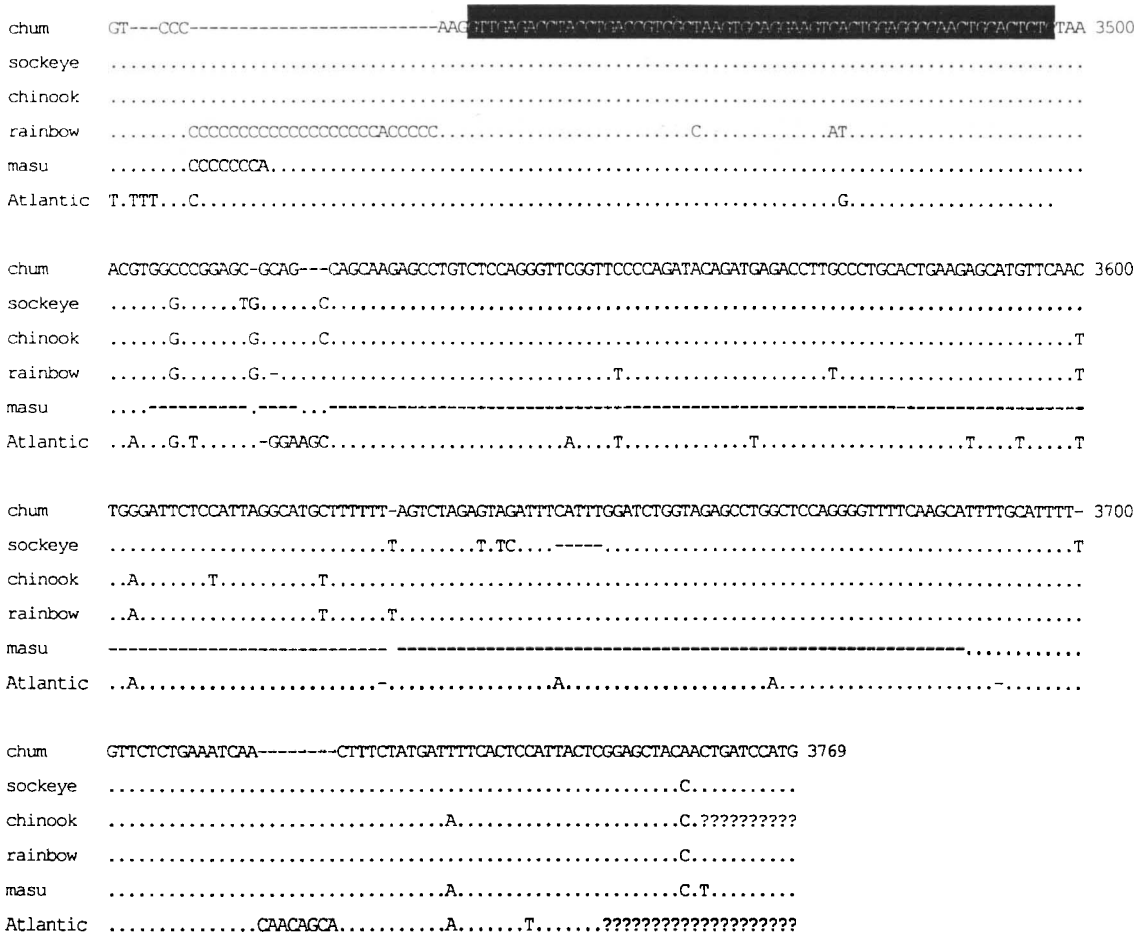



chum TGAGAA----------TGTCTTACTAACATGTCGCAACATAATTTGACTTACTCGTTTTTATACATTTCTTATTT 700
sockeye .....
chinook .....AGATAGCAAATTGAGAATATCTTACTATTGAGAA.A.....C.....
rainbow .....AGATAGCAAATTGAGAATATCTTACTATTGAGAA.A.....C.....
masu .....A.....C.....
Atlantic .....A.....-A.....
chum TCTTTCATCTCTCTTTTAG 800
sockeye .....
chinook .....
rainbow ...G.....G.....
masu .....
Atlantic .....T.....A.....
chum GTAAGACAGCTTTTGAATCTTCTTTTGACACAGCAGATA 900
sockeye ....CA.....
chinook ....C.....T.....G....
rainbow ....C.....
masu ....CA.....
Atlantic ....C.....T.....C.....T.....T....A...
chum ATGTTTCAGAGGTGGTTCCTCTTCTTT--GTAGACAAGTGTCCTCTTC-ACGCAAACCGAGCGGCAAAACA-TTCTCTCTCCCGTCTTTGTGATTTTGT 1000
sockeye .....
chinook .....R.....T.....
rainbow .....T.....
masu .....T.....
Atlantic .C.....A..A---...T.....CTT.....---.....C..A.....T.....A.A.....
chum GCAG 1100
sockeye .....
chinook .....G.....A.....
rainbow .....C.....A.....
masu .....A.....A.....
Atlantic .....G.....A.....
chum GTAAGTTACCTGGCTGAGACAATCCTCCATGATGCACAATTCCAACATGAATAATAGGGCATCTCAATTTGAACAAT 1200
sockeye .....G.....G.....
chinook .....-
rainbow .....G.....
masu .....
Atlantic .....TAG.....A..A.....
chum CGATACAACTTAGTCATTAGTTATTGGGCAAGCAGATCCCCGATTGTCTAAACTCCATCGGTAAATATATACTGTAGATAAGAAGAACCAGCATCATGC 1300
sockeye .....G.....
chinook ----------.....G.....C.....T...
rainbow .....G.....---.....C.....T
masu T.....T.....G.....C.....
Atlantic .....-.....G.....G.....T..C...G.....


chum ATGGTGGAAATTAAATCTAGCCATGACAGGGAGTTTTAAATTGTACACTTAAAAAAATCAAATCAAATAAAAAAATGTAATAAAAAAAAATAACAAAAT 1400
sockeye .....A.....
chinook .....A.....
rainbow .....
masu .....A.....
Atlantic .....A.....
chum AAAATAAATTTAAAAAAATTGGCAGGAAAATGTTGCTA-------TACCTCAGTGCCTTCAA----------AAACAACCACATTTCATAGTCCTT 1500
sockeye ----------.C.....G.....
chinook ----------.CA....T.....CTAAGGTAGGTAA.....A...C.....
rainbow ----------.C.....T.....AA.....A.....
masu ----------.CA....T.....CTAAGGTAGGTAA.....A...C.....
Atlantic ----------.CA....T.....TGTAAGT.....T.....CTAAGGTAGGTAA..-.....A...C.....
chum GTAAGTAAAACCCATCACTCTCTAA-TCGGCGGTTTCTCTACGTCTACATTCTCCAGCAATGTATCATGTAAATGATATGGAATCTCAAGCTGTACAAT 1600
sockeye .....G.....C.....
chinook .....C.....
rainbow .....A.....C.....C.....
masu .....C.....T.....C.....C.....C.C.....
Atlantic .....-.....A.....C.....C.....
chum TACAACTCAACTTCATTT-CTAATAATCTGTGGTTTCTCTACATCTACACACACAG[redacted]1700
sockeye .....T.....G.....
chinook .....T.....C.....
rainbow ....T.....T.....
masu -.....T.....T.....
Atlantic -.....T.....G.....
chum [redacted]1800
sockeye .....
chinook .....A.....C.....
rainbow .....
masu .....A.....
Atlantic .....G.....
chum [redacted]GTAATGGT----------CAATTACCATTTGTGGTGCCGCACTTTGTGCA----------TTTTTAACTCAAATA 1900
sockeye .....
chinook .....A.....
rainbow .....A.....G.....A.....CG.....
masu .....A.....
Atlantic .T.....A.....A.--AAAGGAGGGAGAA....G.....CTGTAAACCACAAGGCA.....
chum CTTCTAGTAAGTTGAAGTCAGTCAATGAAAAGTCATTATTACTTCAAATGTCTATGTGGTACTGGCTCAAATCTAAATGAGTCACATCAATGCAATTTT 2000
sockeye .....
chinook .....
rainbow .....
masu .....G.....G..
Atlantic .....C.....A.....T.....


chum TTGAAGTTATAACAAATTAACTTTTTACCCAGCATGCTCTGCTACAGGTAGATTTTT-GGAATTGTTTT-AATATCTGTGTTTTTGCATGTACAGTA-- 2100
sockeye ..A.....C.....A.....T.....T.....
chinook ..A.....G.....A.....T.....T.....
rainbow ..A.....G....G.....A.....T.....T.....
masu ..A.....G.....A.....G.....
Atlantic ..T.....A.....T.....T.....T.....--..TT
chum ----------CATTGAGTGATTGATTGATTGATTCATCTTATGCTACACACAGATATATAACGTACATTGTTCTACATTTTCACAAAGATAAATA 2200
sockeye .....C...C..----.....T.....
chinook .....----------G.....T.....
rainbow .....----.....G.....T.....G.....
masu .....-------.....T.....
Atlantic TGCATTTTTG.....-------A..T.....T.....A.....T.....G.....
chum ACCTACAAGGTACTGGAATTTTGCAAACCTACTTGCAGGCCTGATGTGGCCTGTAAACCATGAGTTTCA-GGCCACTGTATTTGGGTAAAGCTACACCT 2300
sockeye ..A.....
chinook ..A.....C.....
rainbow ..A.....C.A.....A.....
masu ..A.....C.....T.----------.----------
Atlantic ..A...-------C.....A.....A.....
chum CAAAATAAGGCCTTTATAAGATATGTAATATATTGTTATAAAGTGTTTAACTATAATGAGAATATTTGCCTAGAAAATCACTTGAAGTTCACAGGACTG 2400
sockeye .....-.....A.....T.....GC.....
chinook .....G.....-.....A.....T.....G.....GC.....
rainbow .....-.....A.....T.....GC.....
masu ----------
Atlantic .....-..G.....A.....T.....T.GC.....
chum AAAATTACTGACAACAGACATG----------ATAACTCTACAATTCACTCAAAAGGCAAGGCACACTTGGAAATTAT 2500
sockeye .....A.....A.....
chinook .....A.....A.....
rainbow .....A.....A.....T.....G...
masu ----------.----------
Atlantic .....G.A.....A.....TCTCTGTCGCTAACAAATACAGTCATGGGTG.....G.....
chum ATTGGAGACATGGC-TTAGTGGGGGCATTACTAAAAAATGTCAAGCTGATACCACTCAAATCTCAACCCTCTACAGGGCAACTCTATAGGTTTGAGTAA 2600
sockeye .....G.....
chinook .....G.....
rainbow .....T.....G.....
masu ----------.----------
Atlantic .....A.....C.....T.....-.....TG.....
chum TGACTATAAAAATCACTTTAAGTGACTGTAGTCAGATTCTGTATATTAAGTGCAACGGTTTCCTCAAAAGTTTTGAATCATCACAGCATATTGGGGTTT 2700
sockeye .....G.A..G.....
chinook .....A.....G.A..G.....C.....
rainbow .....-.....C.....G.A..G.....C.....
masu ----------
Atlantic .....A.....T.....G.A..G.....C.....


chum ACAATGTGGTTATTATCTTCCACTGAC---------- 2800
sockeye .....ATGAAAGTGAAATACAACTATGCTTTCCTAGTTAGAAAATGCAT.....
chinook .....A.....ATGAAAGTGAAATACAACTATGCTTTCCTAGTTAGAAA..GCATAGTGTAGGACCACGTA.TGCCTCTTCTC
rainbow .....G.....ATGAAAGTGAAATACAACTATGCTTTCCTAGTTAGAAA..GCATAGTGTAGGACCACGTA.TGTCTCTTCTC
masu ----------.....
Atlantic ...G.....ATGAAAGTGAAATACAACTATGCTTTCCTAGTTAGAAA..GCATAGTGTAGGACTACGTACGAGGTCTTCTC
chum ---------- 2900
sockeye .....
chinook AGCAGATCTTTCAGCGCTTTACATTGTGATGGGGTAACTCACCTCAT.....GATCATCACTAATAGTGACTATATCAGTAACACCCCATTCAA
rainbow AGCAGATCTTTCAGCGCTTTCCATTGTGACGGGGTAACTCACCTCAT.....TCGATCATCACTAATAGTGACTATATCAGTAACACCCCATTCAA
masu .....
Atlantic AGCAGATCTTTCAGTGCTTTACATTGTGATGTGGTAACTCACCTCATATATATAGTC.....ACTAATAGTGACTATATCAGTAACACCCCATTCAA
chum ---------- 3000
sockeye .....GTGTCTTTTTCTATATGTGCTTGTAGAATGGCCAAATAAACAAGTATTGATATGCACACATC
chinook TGACTGAATATTGGCCCATTCAATGACATTTATGCATGTGTCTTTTGCTATATGTGCTTGTAGAATGGCCAAATAAACGAGTATTGATATGCACACATC
rainbow TGACTGAATATTGGCCCATTCAAGGACATTTATGCATGTGTCTTTTGCCATATGTGCTTGTAGAATGGCCAAATAAACAAGTATTGTTATGCACACCTC
masu .....AATAAACAAGTATTGATATCCACACATC
Atlantic TGACTGAATATTGTCCCATTCAAGGACATCTATGCATGT..CTTTTGCTATATGTGCTTTCTGAATGGCCCAATAAACAAATATTGATATGCACGCATC
chum 3100
sockeye ....C.....CA.....
chinook ....C.....T..C.....
rainbow ....C.....T..C.....
masu ....C.....G.....T.....T..C.A.....
Atlantic ....C.....T..C.....T.....
chum GTGGAAGACTATGTTGC 3200
sockeye .....TC.....
chinook .....TC.....
rainbow .....G.....C.....
masu .....G.....C.....
Atlantic .....T.....A.....A.....C..T....
chum CTTCAATTGCATGTGCCTTCCTATATTTTCTACAGTGC-ATTGTTTTTTT--GTGATCTCCATTGTGAAGTATCTTTG-GGTCTTCAACCCATATGTT 3300
sockeye .....TT.....T.....A.....
chinook .....G.....T...A.....T.....
rainbow .....C.....G.....T.....ATA.....T.....T....
masu .....T.....A.....T.....
Atlantic .....A.....T....TG.....T.....A..
chum ATTACTATTATTGTTCATTGATCAAGACTGGTCTCGAGAAAGTCCTGGTGACTTAGAACATGCACATTAAAATGTGTCAACTAATAACCTATTCTTCTT 3400
sockeye .....
chinook .....T.....-..-.....
rainbow .....T.....-.....
masu .....T.....-.....
Atlantic G.....C..T.....-.A.....T.....-.....A..


chum GT---CCC----------AAG[redacted]TAA 3500
sockeye
chinook
rainbow .....CCCCCCCCCCCCCCCCCCACCCCC.....C.....AT.....
masu .....CCCCCCCA.....
Atlantic T.TTT...C.....G.....
chum ACGTGGCCCGGAGC-GCAG---CAGCAAGAGCCTGTCTCCAGGGTTCGGTTCCCCAGATACAGATGAGACCTTGCCCTGCACTGAAGAGCATGTTCAAC 3600
sockeye .....G.....TG.....C.....
chinook .....G.....G.....C.....T
rainbow .....G.....G.-.....T.....T.....T
masu ....----------.----...----------
Atlantic ..A...G.T.....-GGAAGC.....A....T.....T.....T....T.....T
chum TGGGATTCTCCATTAGGCATGCTTTTTT-AGTCTAGAGTAGATTTCATTTGGATCTGGTAGAGCCTGGCTCCAGGGGTTTTCAAGCATTTTGCATTTT- 3700
sockeye .....T.....T.TC....-----.....T
chinook ..A.....T.....T.....
rainbow ..A.....T.....T.....
masu ---------- ----------.....
Atlantic ..A.....-.....A.....A.....-.....
chum GTTCTCTGAAATCAA--------CTTTCTATGATTTTCACTCCATTACTCGGAGCTACAACTGATCCATG 3769
sockeye .....C.....
chinook .....A.....C.??????????
rainbow .....C.....
masu .....A.....C.T.....
Atlantic .....CAACAGCA.....A.....T.....??????????????????

Appendix 2

The likelihood values of nine alternative trees (Figure 2.6) were compared using DNAML in the PHYLIP package (Felsenstein, 1993). The model for DNA sequence evolution used by this program is outlined in Felsenstein (1991) and updated as described in the program documentation. DNAML calculates the likelihood of recovering the observed sequence data given a particular tree under the above model. Statistical significance of differences in observed Ln likelihood values were determined using the method of Kishino and Hasegawa (1989), which is included in the DNAML program. The values in tables A.2.1 to A.2.6 were calculated using single gene sequence data sets. The values in tables A.2.7 to A.2.15 were calculated using all available sequence data (5353 aligned nucleotide positions) minus the single-gene data set shown in the table caption.

Table A.2.1. GH2

Tree	Ln L	Diff. Ln L	St. Dev.	Significantly worse?
1	-4970.00240	<------ best		
2	-4975.32670	-5.32431	8.0300	No
3	-4971.08712	-1.08472	10.0621	No
4	-4997.83946	-27.83706	12.6139	Yes
5	-5043.37331	-73.37091	19.9136	Yes
6	-4988.28324	-18.28084	10.2720	No
7	-4993.46700	-23.46460	12.9485	No
8	-4989.31460	-19.31221	14.2748	No
9	-4979.85456	-9.85217	7.6029	No

Table A.2.2. ATPase 6

Tree	Ln L	Diff. Ln L	St. Dev.	Significantly worse?
1	-2366.30976	-4.81858	4.2000	No
2	-2372.25154	-10.76036	10.9861	No
3	-2361.49118	<------ best		
4	-2419.03369	-57.54251	16.0198	Yes
5	-2403.22411	-41.73293	15.2012	Yes
6	-2409.06213	-47.57094	14.3446	Yes
7	-2414.07721	-52.58603	17.3783	Yes
8	-2404.65541	-43.16423	13.9448	Yes
9	-2376.20302	-14.71184	8.4625	No

Table A.2.3. COIII

Tree	Ln L	Diff. Ln L	St. Dev.	Significantly worse?
1	-2426.16055	<------ best		
2	-2434.53419	-8.37363	8.7410	No
3	-2430.36009	-4.19953	7.5817	No
4	-2484.29664	-58.13608	16.6239	Yes
5	-2469.79229	-43.63173	18.9724	Yes
6	-2474.02125	-47.86070	13.9713	Yes
7	-2479.58406	-53.42350	15.8286	Yes
8	-2480.53054	-54.36998	15.6351	Yes
9	-2437.80383	-11.64327	9.3843	No

Table A.2.4. ND3

Tree	Ln L	Diff. Ln L	St. Dev.	Significantly worse?
1	-1322.36086	<------ best		
2	-1333.89755	-11.53669	7.4901	No
3	-1325.77232	-3.41146	3.4611	No
4	-1340.04917	-17.68831	9.2069	No
5	-1358.67170	-36.31084	12.5743	Yes
6	-1323.39763	-1.03677	6.2652	No
7	-1334.58901	-12.22815	9.7421	No
8	-1327.37471	-5.01385	7.3258	No
9	-1337.09831	-14.73745	6.9954	Yes

Table A.2.5. ND4L

Tree	Ln L	Diff. Ln L	St. Dev.	Significantly worse?
1	-682.43950	-5.03112	12.3819	No
2	-684.09193	-6.68355	12.3260	No
3	-682.54642	-5.13804	12.3871	No
4	-686.19697	-8.78859	10.2559	No
5	-677.40838	<------ best		
6	-686.29482	-8.88644	13.3006	No
7	-687.01318	-9.60480	13.8776	No
8	-686.60760	-9.19922	13.2406	No
9	-682.08091	-4.67253	9.4370	No

Table A.2.6. D-LOOP

Tree	Ln L	Diff. Ln L	St. Dev.	Significantly worse?
1	-2443.52970	-21.85309	11.9651	No
2	-2443.91730	-22.24069	12.1941	No
3	-2441.32531	-19.64870	12.2422	No
4	-2421.67661	<------ best		
5	-2474.80448	-53.12787	19.2472	Yes
6	-2433.34927	-11.67266	7.8293	No
7	-2433.78125	-12.10464	8.1971	No
8	-2430.93818	-9.26157	8.1419	No
9	-2430.59999	-8.92338	8.8844	No

Table A.2.7. All sequence data

Tree	Ln L	Diff. Ln L	St. Dev.	Significantly worse?
1	-14479.80932	<------ best		
2	-14519.31149	-39.50218	17.5540	Yes
3	-14482.16984	-2.36052	12.6675	No
4	-14623.71178	-143.90247	30.8342	Yes
5	-14699.15303	-219.34371	37.4617	Yes
6	-14587.05359	-107.24428	24.9784	Yes
7	-14617.10367	-137.29435	30.1926	Yes
8	-14590.95513	-111.14581	28.0379	Yes
9	-14517.48240	-37.67308	18.2557	Yes

Table A.2.8. NO GH2

Tree	Ln L	Diff. Ln L	St. Dev.	Significantly worse?
1	-9350.11367	<------ best		
2	-9382.28412	-32.17044	14.8654	Yes
3	-9350.85508	-0.74141	8.7843	No
4	-9467.70778	-117.59411	27.5403	Yes
5	-9497.04755	-146.93387	31.6405	Yes
6	-9440.48838	-90.37471	22.5928	Yes
7	-9462.45583	-112.34216	26.7657	Yes
8	-9442.72818	-92.61451	24.3597	Yes
9	-9378.14807	-28.03440	15.7210	No

Table A.2.9. NO ATPase 6

Tree	Ln L	Diff. Ln L	St. Dev.	Significantly worse?
1	-12035.79024	<------ best		
2	-12066.43085	-30.64061	15.2897	Yes
3	-12047.71217	-11.92192	12.6072	No
4	-12127.52866	-91.73842	26.0264	Yes
5	-12220.64862	-184.85838	35.5799	Yes
6	-12097.70315	-61.91291	20.2069	Yes
7	-12122.34051	-86.55027	25.0855	Yes
8	-12110.23154	-74.44130	23.7444	Yes
9	-12064.65105	-28.86081	16.4705	No

Table A.2.10. No COIII

Tree	Ln L	Diff. Ln L	St. Dev.	Significantly worse?
1	-12027.86452	-2.72755	10.3794	No
2	-12058.16079	-33.02382	16.3993	Yes
3	-12025.13697	<------ best		
4	-12112.02514	-86.88817	26.2210	Yes
5	-12202.94958	-177.81261	33.6467	Yes
6	-12084.61744	-59.48047	23.4200	Yes
7	-12109.79220	-84.65523	26.5662	Yes
8	-12083.02954	-57.89258	21.2838	Yes
9	-12054.14558	-29.00861	15.9431	No

Table A.2.11. No ND3

Tree	Ln L	Diff. Ln L	St. Dev.	Significantly worse?
1	-13105.54687	-4.43103	12.4789	No
2	-13131.40725	-30.29141	17.1265	No
3	-13101.11584	<------ best		
4	-13225.07463	-123.95879	29.8194	Yes
5	-13283.95383	-182.83799	36.1895	Yes
6	-13211.41198	-110.29613	27.3782	Yes
7	-13228.25562	-127.13978	29.6801	Yes
8	-13208.16819	-107.05235	25.0054	Yes
9	-13122.68126	-21.56541	16.8588	No

Table A.2.12. No ND4L

Tree	Ln L	Diff. Ln L	St. Dev.	Significantly worse?
1	-13787.18067	<------ best		
2	-13823.65629	-36.47563	17.2414	Yes
3	-13789.62273	-2.44207	12.7688	No
4	-13925.49110	-138.31043	30.0640	Yes
5	-14010.70990	-223.52923	35.7123	Yes
6	-13888.76655	-101.58588	24.5630	Yes
7	-13917.43026	-130.24959	29.6462	Yes
8	-13892.01972	-104.83905	27.6948	Yes
9	-13825.40702	-38.22636	17.4673	Yes

Table A.2.13. No D-loop

Tree	Ln L	Diff. Ln L	St. Dev.	Significantly worse?
1	-12009.31094	<------ best		
2	-12048.51372	-39.20278	17.6374	Yes
3	-12013.60570	-4.29476	12.0598	No
4	-12176.87493	-167.56399	29.1518	Yes
5	-12198.03337	-188.72243	34.3201	Yes
6	-12040.74636	-31.43542	12.4009	Yes
7	-12129.04076	-119.72982	23.8460	Yes
8	-12160.20491	-150.89397	29.3679	Yes
9	-12136.06227	-126.75133	26.7448	Yes
10	-12057.89885	-48.58791	17.1282	Yes

Table A.2.14. No D-loop or ND4L

Tree	Ln L	Diff. Ln L	St. Dev.	Significantly worse?
1	-11318.53461	<------ best		
2	-11354.55376	-36.01915	17.3106	Yes
3	-11322.83141	-4.29680	12.1643	No
4	-11480.44767	-161.91306	28.3251	Yes
5	-11510.38903	-191.85442	32.6518	Yes
6	-11432.77311	-114.23850	23.3644	Yes
7	-11462.42439	-143.88978	28.7722	Yes
8	-11439.03040	-120.49579	26.3384	Yes
9	-11367.20424	-48.66963	16.2930	Yes

Appendix 3

Polymerase chain reaction-based species identification of salmon and coastal trout in British Columbia.

Abstract:

The west coast of North America has seven native species of anadromous salmon and trout (*Oncorhynchus spp.*), introduced brown trout (*Salmo trutta*), and low numbers of Atlantic salmon (*S. salar*) that have presumably escaped from fish farms. Species identification based on morphology of intact juvenile or adult specimens is not usually difficult, but in cases where only anonymous tissue samples, larvae, or suspected hybrids are examined, molecular methods of identification are often required. Current molecular species identification techniques involve electrophoresis of proteins, and restriction mapping or sequence analysis of mitochondrial or genomic DNA. Here, the development of a new, DNA-based species identification method using the polymerase chain reaction to amplify a portion of the growth hormone type-2 gene is described. No intraspecific variation was detected when this species identification method for *Oncorhynchus* and *Salmo* species was tested on representatives of 31 different populations collected from 19 locations on the west coast of North America. The test was also applied to anonymous samples, interspecific hybrids, suspected feral Atlantic salmon larvae, and to commercially prepared fresh, previously frozen, and smoked fish samples.

Introduction:

The genus *Oncorhynchus* is believed to have arisen from a common ancestor that diverged from the Atlantic salmon lineage 18-20 million years ago (Devlin, 1993; McKay et al., 1996; Figure A.3.1). On the Pacific coast of North America, *Oncorhynchus* is represented by native populations of chinook (*O. tshawytscha*), coho (*O. kisutch*), sockeye/kokanee (*O. nerka*), pink (*O. gorbuscha*) and chum (*O. keta*) salmon, and has recently been expanded to include steelhead/rainbow (*O. mykiss*) and cutthroat trout (*O. clarki)* (Smith and Stearley, 1989; Stearley and Smith, 1993). Masu (*O. masou*) and amago (*O. rhodurus*) salmon do not occur outside of Asia. The more distantly related *Salmo* species are not native to Pacific drainages.

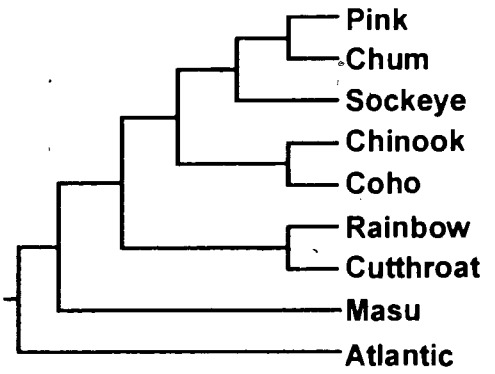


Figure A.3.1 Evolutionary relationships among Pacific salmon and trout. The phylogenetic tree was inferred using total evidence cladistic analysis of a number of morphological and molecular character sets (Chapter 2).

Introduced brown trout (*Salmo trutta*) have become established in some British Columbia water systems, but early attempts to establish Atlantic salmon (*Salmo salar*) in local rivers failed (McKinnel et al., 1996). Thousands of Atlantic salmon escape

yearly from damaged sea pens but have yet to establish feral populations, suggesting that the species is not well suited to life in the Northern Pacific basin. Although there is ample evidence that domesticated salmonids tend to be less successful when in direct competition with their wild counterparts (Bams, 1976; Reisenbirchler and MacIntyre, 1977; Fraser, 1981; MacLean et al., 1981; Chilcote et al., 1986; Skaala et al. 1990; 1991), the perceived threat of the establishment of Atlantic salmon in local rivers persists (for a thorough discussion of this subject, see McKinnel et al., 1996).

Each of the anadromous salmon and trout species has clear morphological, meristic, and behavioral characters that normally make species identification of intact adult or juvenile specimens relatively straightforward (Carl et al., 1977; Scott and Crossman, 1973; McPhail and Carveth, 1993). However, circumstances sometimes arise where a clear identification is not always possible: larvae, anonymous or processed tissue samples and exceptional individuals, such as interspecific hybrids, are less amenable to easy identification (Wilkins et al., 1994). In such cases, tests based upon molecular rather than macroscopic characters can be employed.

In fish, molecular species identification has been carried out by detecting protein variation with starch gel electrophoresis, peptide mapping of the myosin heavy chain (Rehbein, 1992), liquid chromatography or high performance liquid chromatography (Osman, et al., 1987; Armstrong et al., 1992) and isoelectric focusing of water-soluble sarcoplasmic proteins (Lundstrom, 1979;1983; Durand and Landrein, 1982; Neti and Rehbein, 1988; Rehbein, 1990; Rehbein et al., 1995). DNA-based analyses, such as PCR (Polymerase Chain Reaction)-single stranded conformational polymorphism (Hara et al., 1994), random amplified polymorphic DNA (RAPD) analysis (Bardakci and

Skibinski, 1994), and DNA sequence or restriction enzyme site analysis of mitochondrial (Bartlett and Davidson, 1991; Woodley et al., 1994) and nuclear loci (Silberman and Walsh, 1992) also have been used. In Pacific salmon and trout, DNA-based species identification has been accomplished by Southern or PCR analysis of nuclear growth hormone and insulin-like growth factor genes (R.H. Devlin, unpublished; Wallis and Devlin, 1993) and Atlantic and brown trout and their hybrids have been studied by DNA sequence analysis of mitochondrial loci (McGowan and Davidson, 1992; Youngson et al., 1992, Pendas et al., 1995). For most *Oncorhynchus* species and Atlantic salmon, DNA sequence or restriction site data are available for the mitochondrial D-loop (Shedlock et al., 1992), mitochondrial NADH Dehydrogenase Subunit 3 (ND3) and nuclear growth hormone type-2 (GH2) genes (McKay et al., 1996), SINE repeat elements (Murata et al., 1993; Takasaki et al., 1994), and nuclear ribosomal DNA (Phillips et al., 1992). However, the applicability of these methods for species diagnosis has not been tested for most salmonid species.

This paper describes the development and application of a molecular species identification method designed to distinguish all native and exotic anadromous salmon and trout species from the west coast of North America. The test, based on PCR technology (Saiki et al., 1988), is designed to improve the ease of species identification, and to expand the range of species and sample types that can be analyzed.

Material and Methods:

Sample collection

Liver or fin tissue was collected from wild or native, hatchery-reared fish from 19

different locations in coastal north America (Figure A.3.2; Table A.3.1), with the exception of the New Zealand domestic chinook salmon, which are derived from a Sacramento River strain transported there in 1905. Upon collection in the field, tissue samples were placed on dry ice or in 70% ethanol. In addition, commercial fish products (origin of fish unknown) representing all species examined in this study except cutthroat trout were purchased from retailers in the Vancouver, Canada area. Fresh, previously frozen and smoked fish samples purchased at retail seafood outlets were transported to the laboratory at ambient temperature in the original packaging. For long-term storage, all tissue samples were either stored in 70% ethanol at ambient temperature or frozen at -80°C.

DNA preparation and PCR amplification of DNA samples

DNA extraction was performed by Proteinase K digestion and organic extraction as described (Devlin et al., 1991). DNA quantity was estimated using a Hoeffer DNA Flourometer and/or Agarose-gel electrophoresis with Ethidium Bromide staining. PCR primers were designed to amplify a portion of the type-2 salmon growth hormone gene (GH2) containing all (for Pacific salmon/trout) or a portion (for Atlantic salmon) of the fourth intron and fifth exon (Figure A.3.xA). The primers GH57 (5'-TGCTCATCAAGGTAATGGTCA-3') and GH58 (5'-TGTTTTTGCATGTACTATTTG-3') were designed based on the aligned DNA sequence of GH2 from Atlantic salmon and all anadromous Pacific salmon and trout occurring in British Columbia (McKay et al., 1996 and references therein). GH7 (5'-CTTATGCATGTCCTTCTTGAA-3') was

Table A.3.1. Populations tested in this study. Except where indicated, two individuals were sampled from each location and place names refer to rivers.

Species	Sampling location
sockeye/kokanee (O. nerka)	Henderson Lake, Weaver Creek, Williston Lake
chum (O. keta)	Big Qualicum, Chilliwack (1), Inch Creek (1), Nitinat lake(1), Snootli, Weaver Creek
pink (O. gorbuscha)	Puntledge, Weaver Creek, Henderson Lake
chinook (O. tshawytscha)	Big Qualicum, Chilliwack, Chehalis, Coquitlam, Nimpkish, Puntledge, Quinsam, Sacremento+
coho (O. kisutch)	Big Qualicum, Capilano, Chilliwack, Inch Creek (1), Skeena, Alsea* (1)
rainbow/steelhead (O. mykiss)	Abbotsford Trout Hatchery, Chilliwack, Pennask lake
coastal cutthroat (O. clarki)	Chehalis, Fraser, Taylor, Upper Quinsam
Atlantic (S. salar)	Domestic (McConnel strain)

+California, via transplanted New Zealand stock, *Oregon

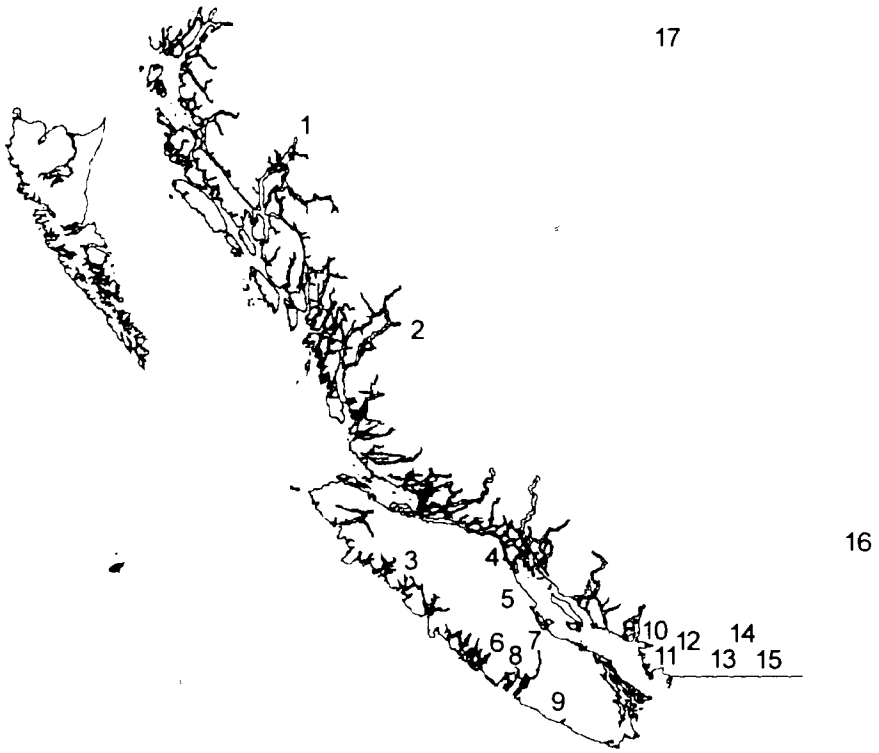


Figure A.3.2. Canada's West Coast. Numbers represent sampling locations: 1) Skeena River, 2) Snootli River, 3) Nimpkish River, 4) Quinsam River, 5) Puntledge River, 6) Taylor River, 7) Big Qualicum River, 8) Henderson Lake, 9) Nitinat Lake, 10) Capilano River, 11) Fraser River, 12) Inch Creek, 13) Chehalis River, 14) Weaver Creek, 15) Chilliwack River, 16) Pennask lake, 17) Williston Lake. Samples were also taken from the Alsea (Oregon) and Sacramento (California) rivers (not shown on map).

designed based on the aligned sequences of the sockeye salmon GH1 and GH2 genes (Devlin, 1993). GH 57 spans the 5' boundary of the fourth intron, GH58 is near the 5' end of the same intron, and GH7 anneals to a site within the fifth exon, immediately downstream of the fourth intron. The combinations of GH57/7 and GH58/7 specifically amplify a GH2 fragment from Pacific salmon/trout and Atlantic salmon, respectively, and produce no amplification product for brown trout. For samples of unknown identity, all three primers were used together. Typically, PCR reactions were performed in 50-100 μl volumes, with 6 ng/μl template DNA, 1X PCR Buffer (Bethesda Research Laboratories-Life Technologies), 0.2 mM of each of the four deoxynucleotide-tri-phosphates, 1.5 mM $MgCl_2$, 0.5 pmol/μl of each primer, 0.025 U/μl of Taq DNA Polymerase (BRL-Life Technologies). Reactions were carried out in thin walled 200 μl tubes (ABI-Perkin Elmer or Fisher Scientific) for 5 cycles of 30s at 95°C, 30s at 58°C, and 60s at 72°C, then 25 cycles of 30s at 95°C, 30s at 55°C, and 60s at 72°C in an MJ-Research "DNA Engine" Twin-Block thermal cycler using a heated lid with no mineral oil overlay. The initial five cycles with a higher annealing temperature were used to eliminate competing amplification products occasionally observed when the reactions were carried out at lower stringency. Occasionally, PCR reactions were performed in thick-walled 600 μl tubes (Eppendorf) with a mineral oil overlay using a Perkin-Elmer-Cetus 480 thermal cycler with the above incubations times doubled. The ND3 gene was PCR amplified and sequenced as described in McKay et al. (1996).

Restriction endonuclease digestion of PCR amplification products

The expected length and restriction maps of PCR products (Table A.3.2) were predicted from GH2 sequences for each species (McKay et al., 1996 and references therein) using the program PC\GENE (Intelligenetics, Mountainview, CA). PCR products were digested with the restriction endonucleases *AluI* and *HpaII* (BRL-Life Technologies). In cases where pink and chum salmon samples were analyzed, a separate aliquot was also digested with *HinFI*. PCR products were digested by diluting a 5-20 μl aliquot 4-fold in 1X REact 1 or REact 2 Buffer (BRL) with 1-5 U of each restriction enzyme, and incubating at least two hours at 37°C. Digestion products were electrophoresed using 1XTBE (89 mM Tris, 89 mM Boric Acid, 1 mM EDTA) running buffer and 2.5% (*AluI/HpaII*) or 4% (*HinfI*) MetaPhor Agarose (FMC Biochemicals).

Results and Discussion:

A molecular test for species identification:

For molecular species identification, the need for relatively large amounts of high-quality starting material or certain types of tissue can be a limitation in situations where appropriate collection or storage is not possible. To avoid these problems, PCR was used to amplify minute quantities of DNA extracted from a variety of tissue types. The use of PCR analysis coupled with agarose gel electrophoresis is a relatively simple approach that can be carried out with a minimum of equipment. A nuclear, rather than a mitochondrial, gene was chosen for amplification due to the higher degree of variation observed in mitochondrial genomes (Brown et al., 1979) and

Table A.3.2. GH 57/58 and 7 PCR-amplification products and predicted fragments resulting from restriction endonuclease digestion*.

Species	PCR product (nt)	Restriction sites HpaII	AluI	Digestion Products (nt)
sockeye/kokanee	1122	-	2	467, 437, 218
chum	1019	-	2	439, 361, 219
pink	1007	-	2	428, 361, 218
chinook	1266	1	2	619, 352, 218, 77
coho	1243	1	2	813, 221, 131, 78
rainbow/steelhead	1273	-	2	618, 437, 218
coastal cutthroat	1066	1	1	629, 360, 77
Atlantic salmon	1064	1	2	619, 250, 117, 78

*Based on the GH2 DNA sequences reported by McKay et al. (1996)

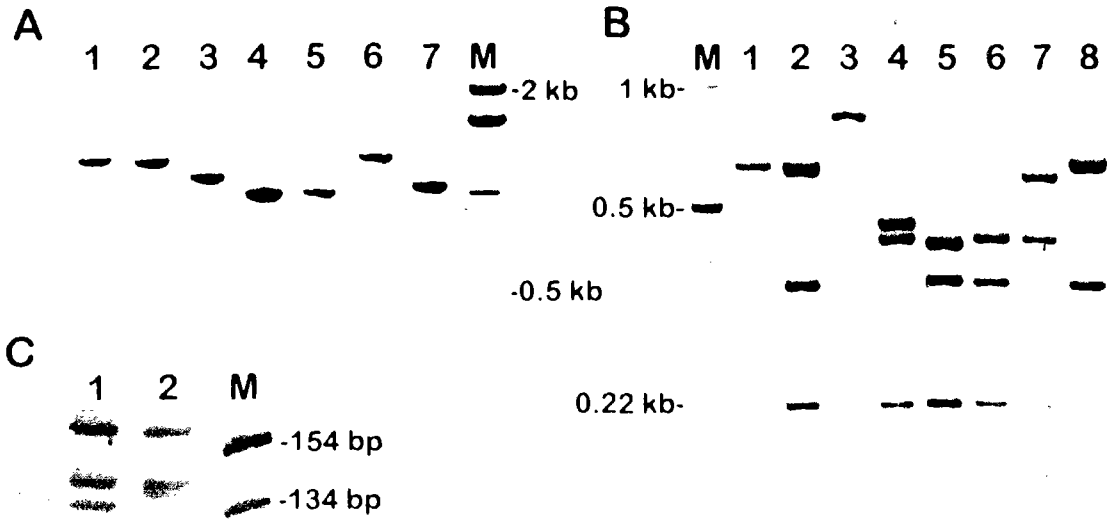


Figure A.3.3. Species-specific profiles generated by amplification of a portion of the GH2 locus. A) Unrestricted GH57/7 PCR products. Lane 1) chinook, 2) coho, 3) sockeye, 4) pink, 5) chum, 6) rainbow, 7) cutthroat, M BRL 1 kb molecular size marker B) AluI/HpaII-digested GH57/7 and 58/7 PCR products. Lane M) BRL 1 kb molecular size marker, 1) Atlantic, 2) chinook, 3) coho, 4) sockeye, 5) pink, 6) chum, 7) rainbow, 8) cutthroat C) diagnostic bands: HinfI-digested GH57/7 PCR products. Lane 1) pink, 2) chum, M) BRL 1 kb molecular size marker.

because nuclear loci can be used to identify hybrids. The higher variability in mitochondrial DNA makes it more useful for the study of allopatric or sympatric populations at an infraspecific level (e.g. Birt et al., 1991; Cronin et al., 1993; Park et al., 1993; Ward et al., 1994; Bickham et al., 1995). In this study, primers based on the GH2 sequences of 11 salmon and trout species (McKay et al., 1996 and references therein) were used to amplify of a portion of the GH2 gene from Pacific salmon and trout as well as Atlantic salmon DNA. DNA sequence conservation in this gene is high enough to allow amplification with the same primer set for all species within *Oncorhynchus*, but variation in the sequence is sufficient to produce easily detectable differences among individual species.

A major trend in the evolution of the GH2 locus is a reduction in overall size by deletion events in the non-coding intron sequences (Devlin, 1993). Certain species, particularly masu, pink and chum salmon, have lost much of the fourth intron (McKay et al., 1996). Differing patterns of insertion or deletion events result in three size categories of PCR amplification products containing this intron (Table A.3.2, Figure A.3.3A). The combination of primers GH57, 58 and 7 produces an amplification product from the DNA of all anadromous salmonids of the west coast except brown trout. Amplification products of the primer combination GH57/7 vary from 1007-1273 nt in length, with chinook, coho and steelhead/rainbow all having larger products of similar size, sockeye and coastal cutthroat intermediate, and pink and chum forming the smallest size category. GH57 does not produce an amplification product from Atlantic salmon but the GH58/7 produces an amplification product specific to this species. Neither GH57 nor GH58 produced an amplification product from the DNA of

individuals sampled from the British Columbia strain of introduced brown trout tested in this study; the presence of amplifiable DNA in the sample was verified with the primer pair GH30 (McKay et al., 1996) and GH7, which amplifies the same gene fragment plus the entire fourth exon from both GH loci in all *Oncorhynchus*, *Salmo*, and *Salvelinus* species tested (not shown). By high resolution electrophoresis of the PCR products, it is possible to distinguish between products of similar sizes, except for coastal cutthroat and Atlantic salmon. The small size differences between some species renders direct comparison of product mobility on the same agarose gel a necessity, and this becomes quite laborious for large numbers of samples or in cases where no *a priori* information on the species of the sample (expected size of amplification product) is available. In order to reduce ambiguity and further meet the criterion of improved ease of species identification, an additional step was added to the analysis.

There are two *AluI* and one *HpaII* restriction endonuclease sites in the consensus sequence of the predicted amplification products. When the products are digested with both of these enzymes, the resulting fragments vary considerably in size and number due to deletions and nucleotide substitutions affecting the restriction sites (Table A.3.2; Figure A.3.3B). For example, the *HpaII* site (CCGG) is destroyed by single transitional substitutions in steelhead/rainbow trout (G->A, position 4) and in sockeye, pink, and chum salmon (C->T, position 2). One of the *AluI* sites is completely deleted in coastal cutthroat trout and coho salmon. Thus, the number and size of bands observed using agarose gel electrophoresis of restriction enzyme-digested GH57/58/7 amplification products form profiles unique to each species. The

profiles are sufficiently different to allow unambiguous identification in most cases without direct, side-by-side comparison to reference standards run on the same gel.

The pink and chum GH2 DNA sequences are very closely related (McKay et al., 1996), and their profiles only differ by 11 nt and 1 nt in the top and bottom bands, respectively (Table A.3.2; Figure A.3.3B). Although the size difference in the top band is resolvable with agarose gel electrophoresis, direct comparison among samples run on the same gel is necessary to unambiguously distinguish between these two species when *AluI* and *HpaII* are used (see hybrid analysis below and Figure A.3.4). This can be accomplished by re-running suspected pink or chum samples adjacent to known standards, or, alternatively, by digesting the suspected pink or chum amplification products with *HinfI*. This enzyme produces a clear distinction between pink and chum samples by producing diagnostic bands in the 130-140 nucleotide size range that are easily distinguished with a high-resolution agarose gel (Figure A.3.3C). The species identification method requires only minimal quantities of starting material and can be performed rapidly, usually in a single long day or over two days without difficulty.

Intraspecific variation in GH2 sequences

Intraspecific size or sequence differences in the portion of GH2 used in this study could potentially pose a problem by affecting diagnostic species profiles. Sequence changes could destroy or create restriction sites, or change the size of restriction fragments. The GH2 intron examined in this study does not appear to be evolving very quickly in terms of sequence composition, but does tend to accumulate insertions and deletion, detected by aligning the sequences of different species. There

is some evidence that a large deletion observed in the fourth intron of coastal cutthroat GH2 is recently acquired (McKay et al., 1996). Sequences obtained from the same locus in two inland races of *O. clarki*, westslope and Yellowstone cutthroat, lack this deletion and have an overall structure similar to the rainbow trout GH2 (Blackhall, 1994). Although, their banding pattern would differ from that of coastal cutthroat, the *HpaII* site absent from rainbow trout is present in the inland cutthroat races, which would make it possible to resolve them. However, these stocks are not anadromous and do not occur in coastal river systems.

Although the chum salmon sampled in this study had identical profiles, and the predicted digestion products for *AluI/HpaII* digestion did not differ in size from those observed, the chum GH2 (Shen et al., 1993 unpublished Genbank submission) sequence used in McKay et al. (1996) did not predict the actual *HinfI* restriction pattern for this species. Based on the observed fragment sizes, a site lost by a G->A transition in the chum sampled by Shen et al. (1993) is intact in all of the chum salmon individuals tested in this study (Table A.3.2).

Two lines of evidence suggest that there is insufficient intraspecific variation in the GH2 gene to confound this method among anadromous salmonid stocks in western Canada: 1) A total of 31 populations were sampled from 19 locations ranging from Northern British Columbia as far south as Northern California (Figure A.3.2; Table A.3.1). No intraspecific variation was observed between sampled individuals of any of the seven indigenous Pacific salmon and trout species tested. Between three and eight populations were sampled for each species (Table A.3.1). Coho, sockeye and kokanee (landlocked sockeye) samples were taken from either side of a North/South

phylogeographic break, presumably resulting from recolonization from the northern Beringia and southern Columbia refugia at the end of the most recent glaciation (Lindsey and McPhail, 1986; McPhail and Lindsey, 1970;1986), observed in populations of many Northern Pacific marine fauna (Bickham et al., 1995; Gharret et al., 1987; Wilson et al., 1987; Cronin et al., 1993; Burg, pers. comm.; Arndt, pers. comm.). 2) Allopatric subspecies of both *O. nerka* and *O. mykiss* show no detectable variation in diagnostic species profiles. Anadromous sockeye and steelhead populations were compared to kokanee and rainbow populations located as far inland in Williston and Pennask lakes, respectively (Figure A.3.2). These populations have likely been reproductively isolated for a considerable period. Assuming that these lakes were reinvaded by land-locked forms shortly after the end of the last glaciation, they may have been isolated as long as 10,000 years.

Application of the species identification strategy to analysis of interspecific hybrids and unknown wild fish.

The species identification test has also been applied to address local fisheries management issues. Repeated sightings of escaped Atlantic salmon in coastal waters and river systems have caused some concerns that local salmon stocks could be endangered (McKinnel et al., 1996). Suspected feral Atlantic salmon larvae recovered from the location of one such sighting were found to be chinook salmon when tested with our method.

In cases where entire fish specimens are available, species-identification based on gross morphology and coloration is usually possible. Hybridization between

sympatric salmonid species may render identification less clear-cut, as morphological and meristic characters may be characteristic of one or the other parent, or may be intermediate between the two (Wilkins et al., 1994). Numerous cases of interspecific and intergeneric hybridization have been observed between members of the subfamily Salmoninae. For an extensive bibliography of hybrid studies, see Dangel et al., 1973; Chevassus, 1979; 1983. We have examined some experimental hybrids produced under hatchery conditions to determine whether the test described above could resolve the identity of both parent species. Chinook/coho, chinook/sockeye, and pink/chum hybrids were all tested, and in each case showed diagnostic bands from both parent species (Figure A.3.4), demonstrating that this test is suitable for hybrid identification.

In order to attempt to identify species in cases where the sample origin was not immediately obvious, the method was tested on commercially processed samples. Seven fresh, two previously frozen and six smoked fish samples were purchased at various locations in the Vancouver, Canada area. In most cases, the species was clearly identified on the packaging. The species of all the fresh and previously frozen, as well as three of the smoked samples were successfully identified. Some of the smoked salmon samples yielded degraded DNA. In our hands, it was not possible to amplify the GH2 fragment used in this study from these samples. Of the successfully tested samples, all but one were confirmed as the species indicated on the label or at the vendors' establishment. One of the smoked salmon samples labeled as sockeye was clearly identified as chum salmon by our test.

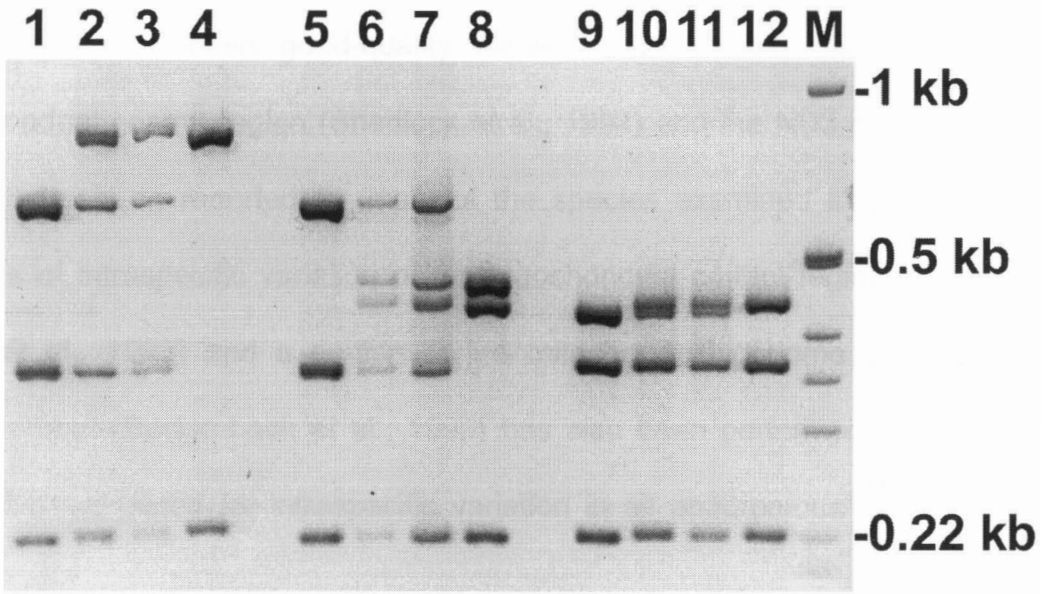


Figure A.3.4. Species analysis of experimentally produced hybrids. Lane 1) chinook, 2,3) chinook/coho hybrids, 4) coho, 5) chinook, 6,7) chinook/sockeye hybrids, 8) sockeye, 9) pink, 10,11) pink/chum hybrids, 12) chum, M) BRL 1 kb molecular size marker. Note the tight doublet band present in the pink/chum hybrids (lanes 10, 11).

Confirmation of species identification using mitochondrial DNA sequence analysis

Sequencing of salmonid mitochondrial DNA fragments has been demonstrated as an alternative approach to species identification (Bartlett and Davidson, 1991). This approach has not heretofore been widely applied to all the species represented in this study, but could be used as an alternative means of species identification (when it is not possible to obtain good-quality nuclear DNA). DNA sequences from the mitochondrial control region (Shedlock et al., 1992) and the ND3 gene (McKay et al., 1996) have been reported for most of the species examined in this study. Some analysis of intraspecific variation of the mitochondrial control region in chum salmon (Park et al., 1993) and a portion of the mitochondrial genome containing ND3 in rainbow trout (Beckenbach et al., 1990) has also been performed, but neither locus has yet been tested for intraspecific variation in all anadromous salmonids. In this study,

we used sequence from the ND3 gene to resolve a conflict regarding the identification of trout samples. Each of the anadromous salmonid species examined in this study has numerous unique nucleotide substitutions in the ND3 gene (Figure A.3.5). Because the degree of intraspecific variation has not been fully characterized for all species, it is not known whether the changes are diagnostic for a particular species, but a close match with all or most of the variable sites of one species makes it possible to identify the species of an unknown sample with reasonable certainty. For example, four liver samples received on the same date were indicated to be taken from coastal

cutthroat. However, one of the samples produced a rainbow trout profile using the PCR method, with no cutthroat trout bands present. To determine whether this individual was actually a rainbow trout, the ND3 gene (351 nt) was PCR amplified and sequenced. The ND3 sequence obtained from the aberrant fish matched the rainbow trout sequence at 19 of the 20 nucleotides that differ between rainbow and coastal cutthroat trout (not shown) including all of the unique sites shown in Figure A.3.5. The only difference observed was a silent change in the third position of the stop codon.

In conjunction with the PCR species identification test results, such a close match indicates that the fish in question was a rainbow trout and that the hatchery population from which these fish were sampled contained both rainbow and cutthroat trout.

The DNA sequence of the ND3 gene of brown trout used in the comparison described above can also be used to confirm identification of this species when no amplification is observed with primers GH 57 or 58 and GH7. In contentious cases, where samples are misidentified or an ambiguous result is observed, a combination of the two approaches would be appropriate.

The development of a new, nuclear DNA-based species identification test has increased the ease with which an unambiguous species identification of anonymous tissue samples can be performed, and the range of species that it is possible to identify. In addition to testing the samples listed in Table A.3.1, the use of this method for identification of anonymous samples was evaluated in a blind test. Eleven randomly selected DNA samples were provided by an individual not involved in this study. These samples were processed as described in the materials and methods by

one of the authors. The results were analyzed by two individuals with no prior knowledge of the sample origins. Each individual was able to identify all samples correctly.

```
                                  1111111111111122222233333333
                    1111246778901124566788899112579022333344
                   6702787565873847383582034256960370170346 28
sockeye            CTGTCATCTAACCCCAACCATCCCCCGTTCGATTAAACTAAA
chum               ..A.........TTG...CCT....AC.TA........G..
pink               .CA..G...G.........CC.....ACC.A..........
chinook            ..A.......T........CC.....AC..A..........
coho               ..A....T...........CC....AAC..AG.........
rainbow            T.AC..............CC...T.AC..A.........G
cutthroat          ..A..............TCC.T...AC..A.........G.
atlantic           ..A.T..GC..G....C..CC..T..ACA.A.C.GG..CT..
brown              ..A...C....AT....T.CC.....ACG.A..C..GTC...
```

Figure A.3.5. Nucleotide positions in the ND3 gene that show apomorphic (unique) substitutions in the eight anadromous salmonid species examined. Numbers refer to nucleotide positions (1-351). Dots represent identity with the sockeye sequence.

The applicability of the test to widely separated North American stocks indicates that the test shows promise for more global application. However, confirmation of these results with particular populations of interest that do not fall within geographical areas covered in this study is recommended before large-scale application. The use of this method in the analysis of hybrids, commercial samples, and randomly selected unknown samples has demonstrated the reliability of the test in a variety of contexts. Potential applications for this test include forensics and fisheries enforcement, further analysis of hybrids, and identification of embryos, alevins and fry.

References:

Avise, J.C., Bowen, B.W., Lamb, T., Meylan, A.B., and E. Bermingham 1992. Mitochondrial DNA evolution at a turtle's pace: evidence for a low genetic variability and reduced microevolutionary rate in the Testudines. Mol. Biol. Evol. 9: 457-473.

Agellon, L.B., Davies, S.L. Chen, T.T., and D.A. Powers. 1988. Structure of a fish (rainbow trout) growth hormone gene and its evolutionary implications. Proc. Natl. Acad. Sci. USA 85: 5136-5140.

Agellon, L.B., Davies, S.L. Lim, C.N., Chen, T.T., and D.A. Powers. 1988. Rainbow trout has two genes for growth hormone. Mol. Reprod. Dev. 1: 11-17

Agellon, L.B., and T.T. Chen. 1986. Rainbow trout growth hormone: molecular cloning of cDNA and expression in *Eschericia coli*. DNA (N.Y.) 5: 463-471

Allendorf, F.W. 1978. Protein polymorphism and the rate of loss of duplicate gene expression. Nature (Lond.) 271: 76-79.

Allendorf, F.W., and G.H. Thorgaard. 1984. Tetraploidy and the evolution of salmonid fishes, p. 1-53. In B.J. Turner [ed.] The evolutionary genetics of fishes. Plenum Press, New York, N.Y.

Armstrong, S.G., Leach, D.N., and S. G. Wyllie. 1992. The use of HPLC protein profiles in fish species identification. Food. Chem. 44: 147-155.

Bams, R.A. 1976. Survival and propensity for homing as affected by the presence or absence of locally adapted paternal genes in two transplanted populations of pink salmon (*Oncorhynchus gorbuscha*). J. Fish. Res. Board Can. 33: 2716-2725.

Bardakci, F., and D.O.F. Skibinski. 1994. Application of the RAPD technique in tilapia fish: Species and subspecies identification. Heredity 1994 73: 117-123.

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

Berg, W.J., and S.D. Ferris. 1984. Restriction endonuclease analysis of salmonid mitochondrial DNA. Can. J. Fish Aquat. Sci. 41: 1041-1047.

Bickham, J.W., Wood, C.C., and J.C. Patton. 1995. Biogeographic implications of cytochrome b sequences and allozymes in sockeye (*Oncorhynchus nerka*). J. Hered. 86: 140-144.

Birt, T.P., Green, J.M., and W.S. Davidson. 1991 Mitochondrial DNA variation reveals genetically distinct sympatric populations of anadromous and nonanadromous Atlantic salmon, *Salmo salar*. Can. J. Fish. Aquat. Sci. 48: 577-583

Blackhall, W.J. 1994. A molecular study of introgressive hybridization between westslope cutthroat trout and rainbow trout. M.Sc. Thesis. University of Alberta, Edmonton.

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

Cabot., E.L., and A.T. Beckenbach. 1989. Simultaneous editing of multiple nucleic acid and protein sequences with ESEE. Comput. Appl. Biosci. 5: 233-234.

Cantatore, P., M. Roberti, G. Pesole, A. Ludovico, F. Milella, M.N. Gedaleta, and C. Saccone 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.

Carl, G.C., Clemens, W.A., and C.C. Lindsey. 1977. *The freshwater fishes of British Columbia*. British Columbia Provincial Museum, Handbook No. 5.

Cavender, T.M., and R.R. Miller. 1972. *Smilodonichthys rastrotus*, a new pliocene salmonid fish from western United States. University of Oregon, Museum of Natural History Bulletin 18, Eugene Oregon.

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

Chevassus, B. 1983. Hybridization in fish. Aquaculture 33: 245-262.

Chilcote, M.W., Leider, S.A., and J.J. Loch. 1986. Differential reproductive viability of hatchery and wild summer-run steelhead under natural conditions. Trans. Am. Fish. Soc. 115: 726-735.

Cope, E.D. 1870. On the fishes of a freshwater tertiary [lake] in Idaho, discovered by Capt. Clarence King. Proc. Amer. Philos. Soc. 11: 538-546.

Cronin, M.A., Spearman, W.J., Wilmot, R.L., Patton, J.C., 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.

Cummings, M.P., Otto, S.P.,and J. Wakeley. 1995. Sampling properties of DNA sequence data in phylogenetic analysis. Mol. Biol. Evol. 12: 814-822.

Dangel, J.R., Macy, P.T., and F.C. Withler. 1973. Annotated bibliography of interspecific hybridization of fishes of the subfamily Salmoninae. NOAA Tech. Memor. NMFS NWFC 1: 1-48.

Devlin, R.H., McNeil, B.K. Groves, T.D.D., and E.M. Donaldson. 1991.Isolation of a Y-chromosomal DNA probe capable of determining genetic sex in chinook salmon *Oncorhynchus* tshawytscha). Can. J. Fish. Aquat. Sci. 48: 1606-1612.

Devlin, R. H. 1993. Sequence of sockeye salmon type 1 and 2 growth hormone genes and the relationship of rainbow trout with Atlantic and Pacific salmon. Can. J. Fish. Aquat. Sci. 50: 1738-1748.

Dickens, G.M., and R.M. Owen. 1996. Sediment geochemical evidence for an early middle gilbert (early pliocene) productivity peak in the North Pacific red clay formation province. Marine Micropaleontology 27:107-120.

Domanico, M.J. and R. B. Phillips. 1995. Phylogenetic analysis of Pacific salmon (genus *Oncorhynchus*) based on mitochondrial DNA sequence data. Mol. Phylogenet. Evol. 4:366-71.

Du, S.J., Devlin, R.H., and C.L. Hew. 1993. Genomic structure of growth hormone genes in chinook salmon (*Oncorhynchus* tshawytscha): presence of two functional genes, GH-I, GH-II and a male-specific pseudogene, GH-Ψ. DNA Cell Biol. 12: 739-751.

Durand, P., and A. Landrein. 1982. Electrofocalisation en gel d'agarose, application a la determination d'especes de poissons et coquillages Rev. Trav. Inst. Peches. Marit, Nantes 46: 299-305.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17:368-376.

Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle WA.

Forbes, S.H., Knudson, K.L., North, T.W., and F.W. Allendorf. 1994. One of two growth hormone genes in coho salmon is sex linked. Proc. Natl. Acad. Sci. USA 91: 1628-1631.

Fraser, J.M. 1981. Comparative survival and growth of planted, wild, hybrid and domestic strains of brook trout (*Salvelinus fontinalis*) in Ontario lakes. Can. J. Fish. Aquat. Sci. 38:.1672-1684.

Freudenreich, C.H., Stavenhagen, J.B., and V. Zakian. 1997. Stability of a CTG/CAG trinucleotide repeat in yeast is dependent on its orientation in the genome. Mol. Cell. Biol. 17: 2090-2098.

Friedlander, T.P., Regier, J.G., and C. Mitter. 1994. Phylogenetic information content of five nuclear gene sequences in animals: initial assessment of character sets from concordance and divergence studies. Syst. Biol. 43: 511-525.

Gharret, A.J., Shirley, S.M., and G.R. Tromble. 1987. Genetic relationships among populations of Alaskan chinook salmon (*Oncorhynchus tshawytscha*). Can. J. Fish. Aquat. Sci. 44: 765-774.

Gorshkov, S.A., and G.V. Gorshkova 1981. Analysis of relations between species of Pacific salmons from the genera *Oncorhynchus* and *Salmo*. Zool. Zh. 60: 84-96 (English summary).

Grewe, P.M., Billington, N., and P.D.N. Hebert 1990. Phylogenetic relationships among members of Salvelinus inferred from mitochondrial DNA divergence. Can. J. Fish. Aquat. Sci. 47: 984-991.

Hara, M., Noguchi, M., Naito, E., Dewa, K., and H. Yamanouchi. 1994. PCR-SSCP-ho

o mochiita gyoshu hanbetsu. (Ribosomal RNA gene typing of fish genome using PCR-SSCP method.) Bull. Japan Sea Natl. Res. Inst. Nissuiken Hokoku 44: 131-138.

Hartley, S. 1987. The chromosomes of salmonid fishes. Biol. Rev. 62: 197-414.

Hikita, T. 1963. Ecological and morphological studies of the genus *Oncorhynchus* (Salmonidae) with particular consideration on phylogeny. Sci. Rep. Hokkaido Salmon Hatchery 17: 1-97.

Hinegardner, R. 1976. Evolution of geome size, p. 179-199 In F.J. Ayala [ed.] Molecular Evolution. Sinauer Associates, Sunderlan, Mass.

Hoelzel, A.R., Hancock, J.M., and G.A. Dover. 1991. Evolution of the Cetacean mitochondrial D-loop region. Mol. Biol. Evol. 8:475-493.

Imanishi, K. 1951. Iwana and yamame. Tech. Assoc. Jpn. For. Guidebk. For. Ser. 35:36p. (In Japanese).

Johansen, B., Johnson, O.C., and S. Valla. 1989. The complete nucleotide sequence of the growth-hormone gene from Atlantic salmon *(Salmo salar)*. Gene (Amst.) 77: 317-324.

Jordan, D.S., and E.A. McGregor. 1925. Family Salmonidae. In *Record of fishes obtained by David Starr Jordan, 1922*. (Jordan, D.S., and C.L. Hubbs, eds.) Mem. Carneg. Mus. 10: 93-347

Kato, F. 1985. A note on the scientific name and phylogeny of the amago salmon. Bull. Fukui Municipal Mus. Nat. Hist. 32: 47-54. (English summary)

Kato, F. 1991. Life histories of masu and amago salmon. In *Pacific salmon Life Histories*. (Groot., C. and L. Margolis, eds.) UBC Press, Vancouver, Canada.

Kavsan, V, Koval, A., Petrenko, O, Roberts, C.T. Jr., and D LeRoith. 1993. Two insulin genes are present in the salmon genome. Biochem. Biophys. Res. Commun. 191: 1373-1378.

Keohavang, P., and W.G. Tilly. 1989. Fidelity of DNA polymerase in DNA amplification. Proc. Natl. Acad. Sci. USA 86: 9253-9257.

Kimura, S. 1990. On the type specimens of *Salmo macrostoma, Oncorhynchus ishikawae* and *O. rhodurus*. Bull. Inst. Zool., Academia Sinica 29: 1-16.

Kimura, S. 1989. The yamame, land-locked salmon of Kyushu Island, Japan. Physiol. Ecol. Japan, Spec. Vol. 1: 77-92.

Kimura, M. 1980. A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. J. Mol. Evol. 6: 111-120.

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

Kluge, A. 1989. A concern for evidence and a phylogenetic hypothesis of relationships among Epicrates (Boidae, Serpentes). Syst. Zool. 38: 7-25.

Li, W.H., Tanimura, M., and P.M. Sharp. 1987. An evaluation of the molecular clock hypothesis using mammalian DNA sequences. J. Mol. Evol. 25:330-342.

Lindsey, C.C., and McPhail, J.D. 1986. Zoogeography of fishes of the Yukon and Mackenzie basins. In: *Zoogeography of North American freshwater fishes* (Hocutt, C.H. and Wiley, E.O., eds) New York: Wiley, 639-674.

Lorens, J. B., Nerland, J.A.H., Male, R., Lossius, I., Telle, W., and G. Totland. 1989. The nucleotide sequence of Atlantic salmon growth hormone cDNA. Nucleic Acids Res. 17: 2352.

Lundstrom, R.C. 1979. Fish species identification by thin layer isoelectric focusing. J. Assoc. Off. Anal. Chem. 62:624-629.

Lundstrom, R.C. 1983. Fish species identification by agarose gel isoelectric focusing: Collaborative study. J. Assoc. Off. Anal. Chem. 66: 123-127.

Male, R., J.A.H. Nerland, J.B. Lorens, W. Telle, I. Lossius, and G.K. Totland. 1992. The complete nucleotide sequence of the Atlantic salmon growth hormone I gene. Biochem. Biophys. Acta 1130: 345-348.

MacGowan, C., and W.S. Davidson. 1992. Unidirectional natural hybridization between brown trout (*Salmo trutta*) and Atlantic salmon (*S. salar*) in Newfoundland. Can. J. Fish. Aquat. Sci. 49: 1953-1958.

MacLean, J.A., Evans, D.O., Martin, N.V., and R.L. Desjardine. 1981. Survival, growth, spawning distribution and movements of introduced and native lake trout (*Salvelinus namaycush*) in two inland Ontario lakes. Can. J. Fish. Aquat. Sci. 38: 1685-1700.

Martin, A.P., and S.R. Palumbi. 1993. Body size, metabolic rate, generation time, and the molecular clock. Proc. Natl. Acad. Sci. USA 90: 4087-4091.

McKay, S.J., Devlin, R.H., and M.J. Smith. 1996. Phylogeny of Pacific salmon and trout based on mitochondrial NADH Dehydrogenase Subunit 3 (ND3) and nuclear Growth Hormone Type-2 (GH2) DNA sequences. Can. J. Fish. Aquat Sci. 53: 1165-1176.

McKay, S.J., Smith, M.J. and R.H. Devlin. 1997 Polymerase chain reaction-based species identification of salmon and coastal trout in British Columbia. Mol. Mar. Biol. Biotechnol. (In Press).

McKinnel, S., Thomson, A.J., Black, E.A., Wing, B.L., Guthrie III, J.F., Koerner, J.F., and J.H. Helle. 1996. Atlantic Salmon in the North Pacific. *Aq. Res.* 28: 145-157.

McPhail, J.D., and R. Carveth. 1993. *Field key to the freshwater fishes of British Columbia.* Fish Museum, Department of Zoology, University of British Columbia.

McPhail, J.D., and C.C. Lindsey. 1970. Freshwater fishes of northwestern Canada and Alaska. Fish. Res. Board. Can. Bull. 173: 1-381.

McPhail, J.D., and C.C. Lindsey. 1986. Zoogeography of freshwater fishes of Cascadia (the Columbia system and rivers north to the Stikine) In: *Zoogeography of North American freshwater fishes* (Hocutt, C.H. and Wiley, E.O., eds) New York: Wiley, 615-637.

McVeigh, H.P., and W.S. Davidson 1991. A salmonid phylogeny inferred from mitochondrial cytochrome b gene sequences. J. Fish. Biol. 39 (Supplement A): 277-282.

Messier, W, Li, S.H., and C.B. Stewart. 1996. The birth of microsatellites. (letter)

Nature 381: 483.

Morris, D.B., Richard, K.R., and J.M. Wright. 1996. Microsatellites from rainbow trout (*Oncorhynchus mykiss*) and their use for genetic study of salmonids. Can. J. Fish. Aquat. Sci. 53: 120-126.

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

Murata, S., Takasaki, N., Saitoh, M., and N. Okada. 1993. Determination of the phylogenetic relationships among Pacific salmonids by using short interspersed elements (SINEs) as temporal landmarks of evolution. Proc. Natl. Acad. Sci. USA 90: 6995-6999.

Murata, S., Takasaki, N., Saitoh, N., Tachida, H. and N. Okada. 1996. Details of retropositional genome dynamics that provide a rationale for a generic division: the distinct branching of all the Pacific salmon and trout (*Oncorhynchus*) from the Atlantic salmon and trout (*Salmo*). Genetics 142: 915-126.

Neave, F. 1958. The origin and speciation of *Oncorhynchus*. Trans. R. Soc. Can. 52: 25-39.

Neti, G., and H. Rehbein. 1988. Fish species identification by fish muscle dry powder reference material. J. Sci. Food. Agric. 46: 81-91.

Norden, C.R. 1961. Comparative osteology of representative salmonid fishes, with particular reference to the grayling (*Thymallus arcticus*) and its phylogeny. J. Fish. Res. Board Can. 18:679-671.

O'Connell, M, Skibinski, D.O.F., and J.A. Beardmore. 1996. Allozyme and mtDNA divergence between Atlantic salmon populations in North Wales. J. Fish Biol. 48: 1023-1026.

Ohno, S. 1970. Evolution by gene duplication. Springer-Verlag, New York, N.Y.

Oohara, I., and T. Okazaki. 1996. Genetic relationship among three subspecies of *Oncorhynchus masou* determined by mitochondrial DNA sequence analysis. Zool. Sci. 13: 189-198.

Oohara, I., Sawano, K, and T. Okazaki. 1997. Mitochondrial DNA sequence analysis of the masu salmon-phylogeny in the genus *Oncorhynchus*. Mol. Phylogenet. Evol. 7: 71-78

Oshima, M. 1955. *Masu salmon (Oncorhynchus masu* Brevoort), *and biwa salmon, Oncorhynchus rhodurus* (Jordan and McGregor). Nireshobo, Tokyo Japan, 79pp. (In Japanese)

Oshima, M. 1957. *Studies on the dimorphic salmons, Oncorhynchus masou* (Brevoort) *and Oncorhynchus rhodurus* Jordan and McGregor, *found in Japan and adjacent territories.* Nireshobou, Sapporo, 79pp. (In Japanese)

Osman, M.A., Ashoor, S.H., and P.C. Marsh. 1987. Liquid chromatographic identification of common fish species. J. Assoc. Off. Anal. Chem. 70: 618-625.

Park. L.K., Brainard, M.A., Dightman, D.A., and D.A. Winans. 1993. Low levels of intraspecific variation in the mitochondrial DNA of chum salmon (*Oncorhynchus keta*). Mol. Mar. Biol. Biotechnol. 2: 362-370.

Pendas, A.M., Moran, P., Martinez, J.L., and E. Garcia-Vazquez. 1995. Applications of 5S rDNA in Atlantic salmon, brown trout, and in Atlantic salmon x brown trout hybrid identification. Mol. Ecol. 4: 275-276.

Pépin, L., Amigues, Y., Lépingle, A., Berthier, J-L, Bensaid, A., and D. Vaiman. 1994. Sequence conservation of microsatellites between *Bos taurus* (cattle), *Capra hircus* (goat) and related species. Examples of use in parentage testing and phylogeny analysis. Heredity 74: 53-61

Phillips, R.B., Pleyte, K.A., and R. Brown. 1992. Salmonid phylogeny inferred from ribosomal DNA restriction maps. Can. J. Fish. Aquat. Sci. 49: 2345-2353.

Phillips, R.B., and K.A. Pleyte. 1991. Nuclear DNA and salmonid phylogenetics. J. Fish Biol. 39: 259-275.

Regan, C.T. 1914. The systematic arrangement of the fishes of the family Salmonidae. Annals and Magazine of Natural History 13 (series 8): 405-408

Rehbein, H. 1990. Electrophoretic techniques for species identification of fishery products. Lebensmittel Unters. Forsch. 191: 1-10.

Rehbein, H. 1992. Fish species identification by peptide mapping of the myosin heavy chain. Electrophoresis 13: 805-806.

Rehbein, H., Etienne, M., Jerome, M., Hattula, T., Knudsen, L.B., Jessen, F., Luten, J.B., Bouquet, W., Mackie, I.M., Ritchie, A.H., Martin, R., and R. Mendes. 1995. Influence of variation in methodology on the reliability of the isoelectric focusing method of fish species identification. Food Chem. 52: 193-197.

Reisenbirchler, R.R., and J.D. MacIntyre. 1977. Genetic differences in growth and survival of juvenile hatchery and wild steelhead trout. J. Fish. Res. Board Can. 34: 123-128.

Rescan, P-Y, and L. Gauvry. 1996. Genome of the rainbow trout (*Oncorhynchus mykiss*) encodes two distinct muscle regulatory factors with homology to MyoD. Comp. Biochem. Physiol. 113B: 711-715.

Robertson, W.R.B. 1916. Chromosome studies I: Taxonomic relationships shown in the chromosomes of tettigidae and acrididae: V-shaped chromosomes and their significance in acrididae, locustidae, and gryllidae: Chromosomes and variation. J. Morphol. 27: 179-331.

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

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

Sambrook, J., Fritsch, E.F. and T. Maniatis. 1989. Molecular cloning: a laboratory manual (2nd Ed.) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA.Sarich, V.M., and A.C. Wilson 1973. Generation time and genomic evolution in primates. Science 179: 1144-1147

Sarich, V.M. and A.C. Wilson. 1973. Generation time and genomic evolution in primates. Science (Washington, DC) 179: 1144-1147

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

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

Silberman, J.D., and P.J. Walsh. 1992. Species identification of spiny lobster phyllosome larvae via ribosomal DNA analysis. Mol. Mar. Biol. Biotechnol. 1: 195-205.

Simon, R.C. 1963. Chromosome morphology and species evolution in the five North American species of Pacific salmon. J. Morphol. 112: 77-97.

Skaala, O., Dahle, D., Dorstad, K.E., and G. Naevdal. 1990. Interactions between natural and farmed fish populations: information from genetic markers. J. Fish Biol. 36: 449-460.

Skaala, O. Jorstad, K.E., and R. Borgstrom. 1991. Genetic impact on wild populations from fish farming: experiments with genetically tagged brown trout (*Salmo trutta*). In: *Abstracts from the international symposium on biological interactions of enhanced and wild salmonids*. June 17-20, 1991. Nanaimo British Columbia, Canada p. 22.

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

Smith, G.R, and R.F. Stearley. 1989. The classification and scientific names of rainbow and cutthroat trouts. Fisheries 14: 4-10.

Smith, G.R. 1992. Introgression in fishes: significance for paleontology, cladistics, and evolutionary rates. Syst. Biol. 1: 41-57.

Stearley, R., and G. R. Smith 1993. Phylogeny of the Pacific trouts and salmons (*Oncorhynchus*) and genera of the family Salmonidae. Trans. Am. Fish. Soc. 122: 1-33.

Takasaki, N., Murata, S., Saitoh, M., Kobayashi, T.. Park, L., and N. Okada. 1994. Species-specific amplification of tRNA-derived short interspersed repetitive elements (SINEs) by retroposition: a process of parasitization of entire genomes during the evolution of salmonids. Proc Natl. Acad. Sci. USA 91: 10153-10157.

Takasaki, N., Yamaki, T., Mamada, M., Park, L., and N. Okada. 1997. The salmon SmaI family of short interspersed repetitive elements (SINEs): interspecific and intraspecific variation of the insertion if SINEs in the genomes of chum and pink salmon. Genetics 146: 369-380

Tchernavin, V. 1939. The origin of salmon: Is its ancestry marine or freshwater? Salmon and trout magazine 95: 120-140.

Thomas, W.K., Withler, R.E., and A.T Beckenbach. 1986. Mitochondrial DNA analysis of Pacific salmonid evolution. Can J. Zool. 64: 1058-1064.

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.

Tindall, K.R., and T. Kunkel. 1988. Fidelity of DNA synthesis by *Thermus Aquaticus* DNA polymerase. Biochemistry 27: 6008-6013.

Tsuyuki, H., and E. Roberts. 1966. Inter-species relationships within the genus *Oncorhynchus* based on biochemical systematics. J. Fish. Res. Board. Can. 23: 101-107.

Tsuyuki, H., and E. Roberts. 1963. Species differences of some members of Salmonidae based on their muscle myogen patterns. J. Fish. Res. Board. Can. 20: 101-104.

Utter, F.M., Allendorf, F.W., and H.O. Hodgins. 1973. Genetic variability and relationships in Pacific salmon and related trout based on protein variations. Syst. Zool. 22: 257-270.

Utter, F.M. and F.W. Allendorf. 1994. Phylogenetic relationships among species of *Oncorhynchus*: a consensus view. Conserv. Biol. 8: 864-867.

Valdes, A.M., Slatkin, M., and N. B. Freimer. 1993. Allele Frequencies at Microsatellite loci: the stepwise mutation model revisited. Genetics 133: 737-749.

Vladykov, V.D. 1963. A review of salmonid genera and their broad geographical distribution. Trans. R. Soc. Can. I(IV), Sec. III: 459-504.

Wallis, A.E. and R.H. Devlin. 1993. Duplicate insulin-like growth factor-I genes in salmon display alternative splicing pathways. Mol. Endocrinol. 7: 409-422.

Wilkins, N.P., Courtney, H.P., Gosling, E., Linnane, A., Jordan, C. and A. Curatolo. 1994 Morphometric and meristic characters in salmon, *Salmo salar* L., trout, Salmon trutta L, and their hybrids. Aquacult. Fish. Manage. 25:505-518.

Wilson, M.V.H. 1977. Middle Eocene freshwater fishes from British Columbia. Life Sci. Contrib. R. Ont. Mus. 113: 1-61.

Woodley, C.M., Chapman, R.W., Webster, L.F. and D.F Carter. 1994. The 12s-16s rRNA of mitochondrial DNA provides unambiguous identification of shark species. 3rd International Marine Biotechnology Conference: Program, Abstrcats and List of Participants., Tromsoe University, Tromsoe, Norway. p134.

Youngson, A.F., Knox, D., and R. Johnstone. 1992. Wild adult hybrids of *Salmo salar* and *Salmo trutta*. J. Fish Biol. 40: 817-820.