

THE EVOLUTION
OF
SALMONID MITOCHONDRIAL DNA.

BY

WILLIAM KELLEY THOMAS

B.Sc. University of Redlands 1981

M.Sc. Simon Fraser University 1984

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
in the Department
of
Biological Sciences

William Kelley Thomas 1988

Simon Fraser University

February 1988

© All rights reserved. This work may not be reproduced in whole or in part, by photocopy or other means, without permission of the author.

APPROVAL

Name: William Kelley Thomas

Degree: Doctor of Philosophy

Title of Thesis:

Evolution of Salmonid Mitochondrial DNA

Examining Committee:

Chairman: Dr. R.W. Mathewes, Professor

Dr. A.T. Beckenbach, Associate
Professor, Senior Supervisor

Dr. D.L. Baillie, Professor

Dr. B.M. Honda, Associate Professor

Dr. M.J. Smith, Professor, Public
Examiner

Dr. R.A. Lansman, Quadra Logic Tech.
Inc., Public Examiner

Dr. W.M. Brown, Professor, Dept. of
Biology, University of Michigan,
Ann Arbor, Michigan, External Examiner

Date Approved

17 Feb 88

PARTIAL COPYRIGHT LICENSE

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

Title of Thesis/Project/Extended Essay

Evolution of Salmonid Mitochondrial DNA

Author:

(signature)

William Kelley Thomas

(name)

17 Feb 1988

(date)

ABSTRACT

The sequence of a 2214 base pair fragment from the mitochondrial genomes of individuals representing six salmonid species has been determined. The levels of pairwise sequence divergence are lower than those estimated from variation at restriction sites but the phylogenetic relationships among species are consistent with the restriction site estimates. The nucleotide composition, codon usage, tRNA structures and junction sequences within this salmon fragment are identified and compared to homologous sequences from other animal mitochondria. Using these comparisons potential constraints on the evolution of the animal mitochondrial genome are identified. In the protein coding regions there exist several potential stem loop structures which may be important in the stabilization or translation of the mitochondrial messages. Although these structures are not completely conserved either among animal groups or within salmonids, they may place important constraints on the mitochondrial genome.

The spectrum of substitutions among the salmonid species shows a high ratio of transitions to transversions as is typical of animal mitochondrial DNA. The large number of substitutions examined in this study allows for an analysis of the ratios of the alternative transversion pathways. The observed transversions do not appear at the same frequency and the spectrum of substitutions in mitochondrial DNA is best explained by a combination of mispairing and unrepaired damage. There are

478 substitutions and a single three base pair deletion among the six species. These differences are distributed among 351 sites in the mitochondrial fragment. Over 30% of the variable sites in the mitochondrial fragment have more than one substitution. Silent positions account for 329 of the variable positions, 12 sites involve non-silent changes and 10 are found in tRNA genes. The distribution of variation in the fragment examined indicates that the level of substitution varies among, as well as within, the protein coding sequences. The tRNA genes also show a complex pattern of variation. The patterns of sequence conservation between salmonid mtDNA and that of other vertebrates and among salmonid species have been used to examine both the potential evolutionary constraints imposed by the function of the mitochondria and the basic mechanisms that lead to changes in mitochondrial genomes.

ACKNOWLEDGEMENTS

I would first like to thank my fellow graduate students and associates for their support and helpful discussions throughout my time at Simon Fraser University. I would especially like to thank Eric Cabot for his invaluable help during my battles with the computers. I also wish to thank the members of my committee for their helpful discussions and comments on this thesis.

I am most indebted to, and gratefully acknowledge the continuous support of my supervisor Andy Beckenbach.

TABLE OF CONTENTS

TITLE PAGE	i
APPROVAL PAGE	ii
ABSTRACT	iii
ACKNOWLEDGEMENTS	v
TABLE OF CONTENTS	vi
LIST OF TABLES	ix
LIST OF FIGURES	x
INTRODUCTION	1
MATERIALS AND METHODS	4
MtdNA Preparation	4
Cloning	4
Directed Deletions	6
Sequencing	7
Oligonucleotides	8
SECTION I: The Trout Mitochondrial Genome	10
Introduction	11
Results and Discussion	15
Comparison of the protein coding sequences	15
Codon usage and nucleotide composition	30
tRNA genes	35
Junction sequences	43
Conclusions	52

TABLE OF CONTENTS (cont.)

SECTION II: Distribution of Variation	54
Introduction	55
Results and Discussion	57
Sequence variation	57
Distribution of variation among genes	74
Intragenic variation	76
Variation and amino acids	81
tRNA genes	84
Non-silent variation	86
Dynamics of sequence evolution	87
Conclusions	89
SECTION III: The Spectrum of Substitutions	90
Introduction	91
Results and Discussion	93
Substitution frequencies	93
Mispairing	93
Polymerase selection	96
DNA damage	96
Deletion	99
Conclusions	99
SECTION IV: Phylogeny of the Salmonid Mitochondrial Genome	101
Introduction	102
Results and Discussion	104

TABLE OF CONTENTS (cont.)

Rates	109
Conclusions	110
CONCLUSIONS	112
LIST OF REFERENCES	114

LIST OF TABLES

Table 1. Salmonids analyzed	5
Table 2. Trout codon usage	31

LIST OF FIGURES

1. Deletion clones	8
2. Fragment sequenced	14
3. Trout sequence	16
4. Protein alignment	19
5. Pairwise comparisons	29
6. Nucleotide composition	33
7. tRNA ^{GLY} structure	37
8. tRNA ^{ARG} structure	38
9. tRNA alignment	39
10. tRNA ^{GLY} splice junction	45
11. tRNA ^{ARG} splice junction	46
12. Secondary structure in ATPase 6	49
13. Secondary structure at ND3 and tRNA ^{ARG} junction	51
14. Salmonid sequence alignments	58
15. Graph of changes per variable site	73
16. Potential and observed variation for each gene	75
17. Distribution of variable sites	78
18. Distribution of variation	79
19. Distribution of potential variation	80
20. Distribution of variation among codons	82
21. Distribution of variation within tRNAs	85
22. Parsimony tree	94
23. Substitution frequencies	95
24. Pairwise nucleotide divergence	105
25. UPGMA tree	108

INTRODUCTION

The high rate of evolution, mode of inheritance and ease of isolation have made the animal mitochondrial genome one of the most useful tools in studies of molecular evolution and population genetics. The animal mitochondrial genome is typically a 16 kb double-stranded DNA circle, coding for a set of 37 genes specifying 22 tRNAs, 13 mRNAs and 2 rRNAs. Gene order changes are limited to highly divergent taxa and most of the observed changes to mitochondria are base substitutions and small length changes.

Factors that affect the evolution of mitochondrial DNA fall into two general categories (Brown et al., 1982; Wilson et al., 1985). The first is the pressure imposed by basic mechanism of mutation in the mitochondrial systems. The high rate of mutation in animal mitochondria is probably due to an ineffective repair system (Brown, George and Wilson, 1979; Brown et al. 1982; Wilson et al., 1985). This high rate of mutation has been critical in the past, allowing the use of mitochondrial DNA as a tool to look at closely related individuals and species. If in fact the high rate is due to a lack of repair mechanisms in mitochondria, the mutational spectrum may reflect basic mechanisms of mutation.

The second group of factors affecting the evolution of mitochondrial DNA are those which affect the fixation of mutations. These include the transmission genetics, inheritance,

and organismal population structures as well as the functional constraints placed on the DNA sequence.

In this thesis, I have investigated the evolution of a 2214 base pair mitochondrial fragment from six closely related salmonid species. This sequence represents almost 15% of the mitochondrial genome. It contains the two prevalent functional gene classes, those for tRNAs and proteins. An understanding of the dynamics of mitochondrial evolution requires the investigation of a range of divergence levels. It is especially important that closely related groups which do not show mutational saturation of potentially variable sites are included. The Pacific salmon represent a closely related species group ideal for such an analysis. Previous analysis with restriction enzymes (Thomas et al., 1986) demonstrated that the levels of sequence divergence among these species range from 2 to 7 percent. This is a range which in primates minimizes the erasure of initial changes by subsequent ones (Brown, 1985).

The thesis is divided into four sections. The first two sections attempt to look at functional constraints in the mitochondrial genome by identifying sequences which are conserved. This follows a basic assumption that sequences which are conserved are functionally important. In section 1, I investigate the constraints in the long term evolution of this fragment by comparing the nucleic acid and amino acid sequences of the trout mitochondrial fragment with the homologous sequence from other animal species. In section II, I extend this

investigation to the distribution of variation in the fragment among the salmonid species. The functions of sequences can have dramatic effects on the types of changes which can occur (Hixson and Brown, 1986). In both sections I attempt to correlate the distribution of variation with known functions of the mitochondrial sequences. The third section investigates the mutational load on the mitochondrial genome and the basic mechanisms of mutation in this system. Following the assumption that mitochondria are essentially without repair, I use the spectrum of mutations which have accumulated in this fragment among these species to test three models of mitochondrial substitution: mispairing, damage and polymerase fidelity. In the final section, I compare the evolutionary relationships determined by direct sequencing of this fragment with those previously estimated with a restriction enzyme method. I also compare the dynamics of the substitutions with those found in primate species and attempt to correlate the molecular relationships of these extant taxa with the fossil record for salmonid fish.

Although I have demonstrated that for some parameters such as the nucleotide composition and codon usage this fragment is representative of the entire genome, it does not represent all functional types of sequence and is not truly representative of the entire mitochondrial genome. A complete understanding of the evolution of the mitochondrial genome requires the complete sequence.

MATERIALS AND METHODS

Adult salmon representing six salmonid species (Table 1) were sampled from populations in British Columbia and California. In all cases, livers were extracted from freshly killed fish, transported on ice and processed within 24 hours of collection. The sockeye salmon was collected from Henderson Lake, Vancouver Island. The chinook and coho salmon were collected from the Harrison River, a major tributary of the lower Fraser River. The domesticated coho salmon was from the Capilano River Hatchery, Vancouver, B.C. The pink salmon was collected from Jones Creek, a Fraser River tributary near Hope, B.C. The rainbow trout sample represents a domesticated strain that originated from the McCloud River, California. The cutthroat trout specimen originated in the lower Fraser River, British Columbia, Canada.

MtDNA Preparation

Mt DNA was extracted from individual adults by the method of Lansman et al. (1981) without the optional sucrose step gradient, except for the use of a Beckman 70.1 Ti rotor in the CsCl ethidium bromide density centrifugation. Approximately 5 to 50 μg of closed circular mtDNA was prepared from 2 to 10 g of liver.

Cloning

Mitochondrial DNA from a single individual representing each species was cloned into the vector pUC 19. 1 μg of each sample was mixed with 0.2 μg of pUC 19 (Pharmacia #27-4951-01) and digested to completion with Hind III in a total volume of 25 μl for 2 hr at 37°C. The restriction enzyme was removed by either

TABLE 1. Pacific North American salmonids analyzed in this study

Genus	Subgenus	Species	Common name
<u>Oncorhynchus</u>		<u>O. kisutch</u>	Coho salmon
		<u>O. tshawytscha</u>	Chinook salmon
		<u>O. nerka</u>	Sockeye salmon
		<u>O. gorbuscha</u>	Pink salmon
<u>Salmo</u>	<u>Parasalmo</u>	<u>S. gairdneri</u>	Rainbow or steel- head trout
		<u>S. clarki</u>	Cutthroat trout

extraction with phenol and subsequent precipitation with ethanol, or by using GENECLEAN, Bio 101. The fragments were resuspended in 10 μ l of H₂O and ligated by adding 6 μ l of ligation mix (20 μ l 10X ligation Buffer (0.66M TRIS pH 7.5, 0.66M MgCl₂); 20 μ l 0.1 M, dithiothreitol (DTT); 20 μ l BSA, 1 mg/ml; 20 μ l 10 mM ATP; and 1 unit of T4 ligase). This mixture was incubated at 15°C overnight. Two μ l of the ligation reaction was mixed with 50 μ l of JM 83 competent cells and left on ice for 1 hr. Five and 45 μ l of the transformation mixture were plated on nutrient plates containing 50 μ g/ml ampicillin, 40 μ g/ml X-gal, and 160 μ g/ml IPTG. Twenty white colonies were collected from each of the species cloned.

Plasmid DNA was prepared from each white colony using a modified boiling method of Holmes and Quigley (1981). The resulting plasmid DNA was digested with Hind III. Clones carrying the 2.3 kb mitochondrial fragment were selected. The homologous 2.3 kb clones from each species were then subjected to large scale amplification and purification by the alkaline lysis method (Birnboim and Doly, 1979) except that all steps prior to precipitation of cell DNA and debris were performed in 500 ml centrifuge bottles. The resulting closed circular plasmid DNA was isolated from CsCl gradients and stored in TE (10 mM Tris, 1 mM EDTA pH. 7.6, Maniatis et al., 1982).

Directed Deletions

Sequential deletions were performed on the rainbow trout (pSg2.3) sockeye salmon (pOn2.3) and pink salmon (pOg2.3) clones

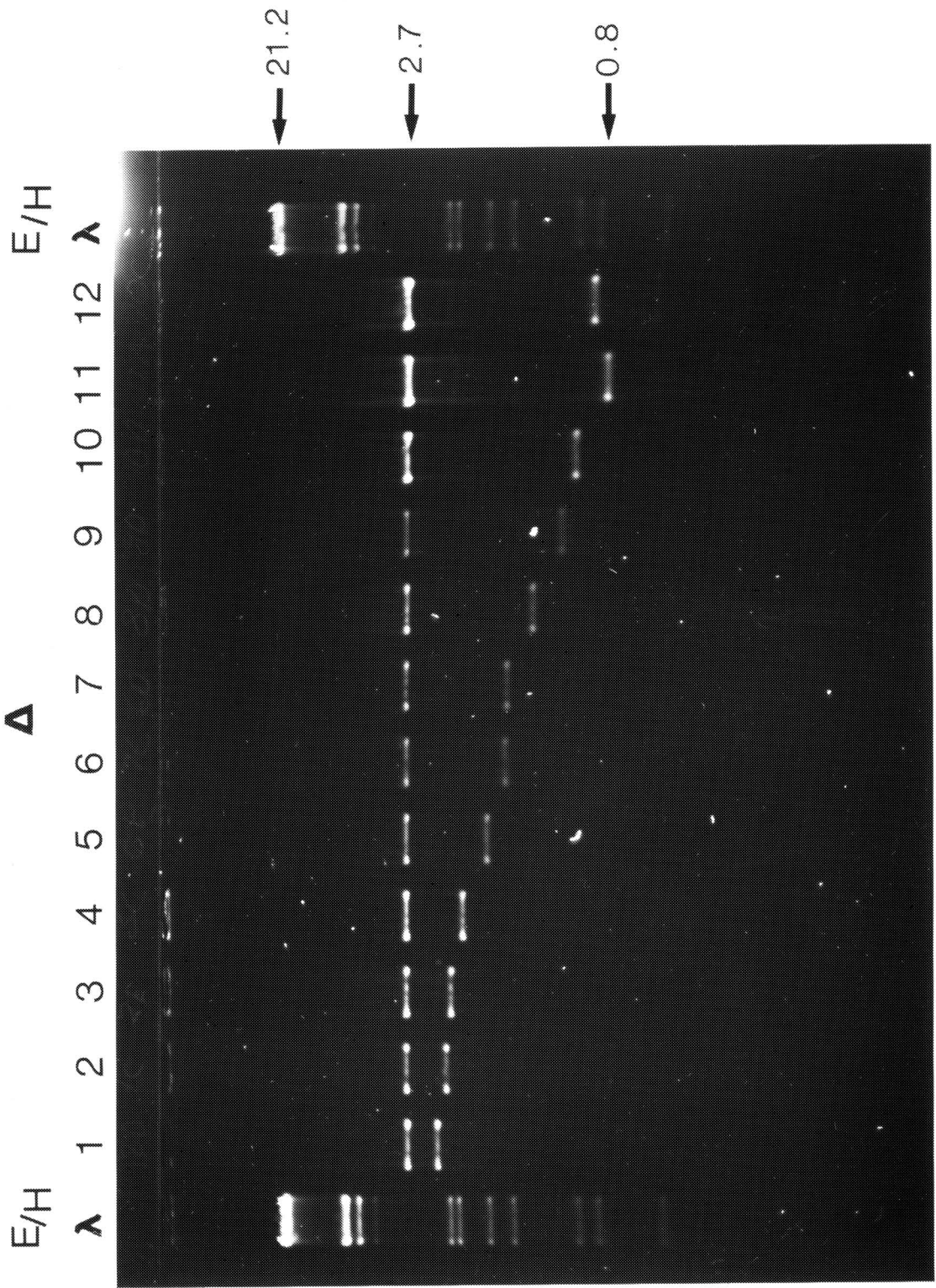
using the method of Henikoff (1984, 1987). Each clone was digested with the restriction enzymes Sal I and Sac I. Each enzyme digested only one site in the polylinker of pUC 19. The 3' protrusion left by the Sac I digestion protects the pUC 19 vector while the Sal I product leaves the mitochondrial fragment susceptible to digestion with exonuclease III. Timed aliquots were taken from an exonuclease digestion. These aliquots were subjected to S1 nuclease and Klenow polymerase prior to ligation and subsequent transformation of JM 83 competent cells. Several successful transformants from each time point were selected and DNA was prepared by the boiling method as described above. The sizes for the resulting clones were determined by digestion with Eco RI and separation on agarose gels (Fig. 1). Clones of appropriate size to produce overlapping sequence were selected. These deletion clones were then sequenced using the universal M13 sequencing primers (Pharmacia).

Sequencing

Template from clones selected for sequencing was prepared from 10 ml overnight cultures by the boiling procedure described above. Supercoiled plasmid DNA was isolated from contaminating RNA, genomic DNA and nicked plasmid by separation on low melting point agarose. The leading (major) DNA band, containing supercoiled plasmid, was cut out of the gel and the DNA was extracted using the phenol method described by Maniatis et al. (1982). The preparation of plasmid template and the sequencing reactions followed the method of Hattori and Sakaki (1986).

FIGURE 1.

Lanes 1 to 12 are Eco RI restriction digested clones resulting from the directed deletion of a complete 2214 bp Hind III clone from rainbow trout. The fragments were separated in a 0.8% agarose gel in 1 X TBE and 0.5 μ g/ml ethidium bromide (Maniatis et al., 1982). The size standard is Eco RI and Hind III double digested lambda DNA (E/H λ).



Oligonucleotides

To facilitate the comparison of all six species and confirm the sequence determined by the deletion method, oligonucleotides were used to prime the sequencing of the opposite strand. The selection of oligonucleotide sequences was based on both an optimal spacing of approximately 200 bp and the need for relatively conserved sequence. Although it was impossible to use completely conserved sequences, no sequences with more than one mismatch in eighteen were allowed, and each was required to have perfect conservation of the 3' ends. The oligonucleotides were prepared in crude form by Tom Atkinson (University of British Columbia) and were purified as in Atkinson and Smith (1985) using the optional sep-pack C18 cartridge (Waters Associates # 31915). Template for sequencing with oligonucleotide primers was prepared by reisolation from CsCl gradients. Once the template was collapsed with NaOH and precipitated by ethanol precipitation it was stored as a dry pellet for up to two months with no noticeable degradation. The sequencing reactions followed as above. Using the S2 sequence apparatus, Bethesda Research Laboratories (BRL) and sharks tooth combs (BRL), the sequencing reactions for all six species could be run side by side.

SECTION I

THE TROUT MITOCHONDRIAL GENOME

INTRODUCTION

An accurate description of how a DNA molecule evolves requires an understanding of the functional constraints on that sequence. Evolution of DNA molecules is the result of mutational pressures, functional constraints, and population dynamics. To estimate the constraints on a sequence one must identify the function of that sequence and the limits imposed by that function. In this section I identify the functions associated with a segment of the trout mitochondrial genome and attempt, through comparisons with other animal mitochondrial sequences, to determine limits imposed by those functions. The trout was selected to represent the salmonids because it was the first species sequenced. Except where noted all comparisons of trout with other groups hold for all salmonids.

The mitochondrial sequence and gene organization has been determined for several animal species (Anderson et al., 1981; Bibb et al., 1981; Anderson et al., 1982; Roe et al., 1985 and Clary and Wolstenholme, 1985). A combination of the sequences and associated transcript mapping (Barrell et al., 1980) demonstrated the extreme economy of the mitochondrial genetic material. The functional maps for all animal mitochondria show that only a very small portion of the DNA does not serve a coding function. The analyses of mitochondrial sequences have also provided the details of codon usage, nucleotide composition and

other functions such as secondary structures involved in mitochondrial RNA splicing. It is clear from comparisons of both divergent groups (mouse-human) and closely related groups (primates) that the function of a sequence has a dramatic effect on its potential for variation (Brown et al., 1985; Hixson and Brown, 1986). It is also clear that other factors including biased codon usage and nucleotide composition affect the evolutionary dynamics of sequences. In the case of Drosophila the biased nucleotide composition is particularly extreme. The nucleotide bias appears to have a dramatic effect on the potential for variation in the Drosophila mitochondrial genome (Wolstenholme et al., 1985; DeSalle et al., 1987). Although other taxonomic groups are less biased, biases do exist in all groups and it is important to understand them.

As pointed out by Brown (1983), among vertebrate species there seems to be a difference in the strand bias between the "cold-" and "warm-blooded" vertebrates. The bias results in the separation into light and heavy strands as a consequence of the G+T content. The sequences of Xenopus and several mammalian mitochondrial genomes has shown that mammals are more asymmetric with respect to G+T content in the L and H strands than Xenopus. Other biases, such as the usage of the TTPu and CTN codons for leucine, are also less asymmetric in Xenopus than in mammals. The trout sequence provides information as to the reduced bias among "cold-blooded" species.

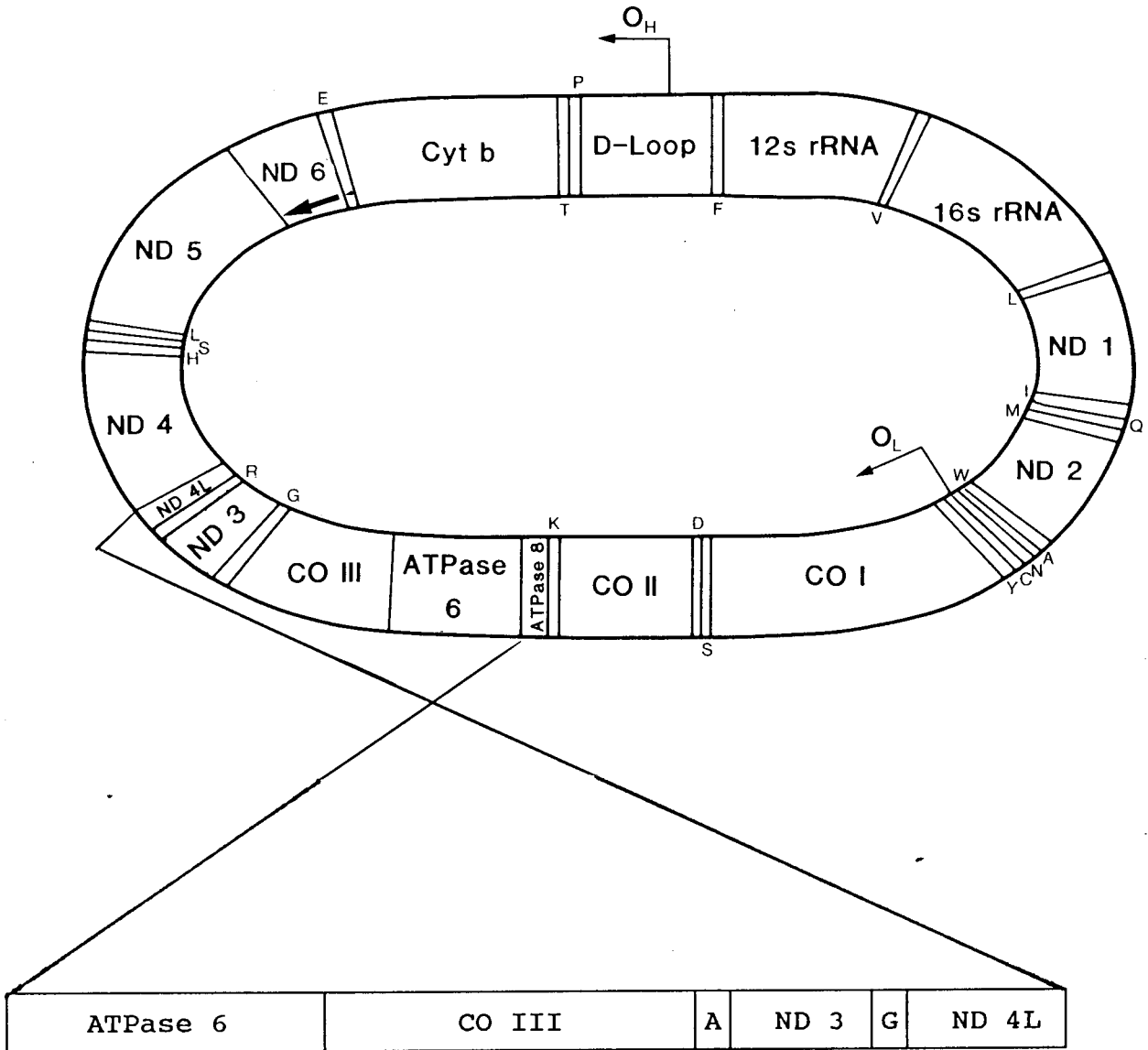
Several studies from our laboratory (Wilson et al., 1985;

Thomas et al., 1986; Wilson et al., 1987) and others (Gyllensten et al., 1985; Gyllensten and Wilson, 1986; Berg and Ferris, 1984) have used the salmon mitochondrial genome to determine the relationships among and within salmon species. In order to interpret the variation found, it is important that we understand the potential for variation in the salmon mitochondrial genome. To this end, I have chosen to sequence a 2214 base pair Hind III fragment containing the coding sequences for ATPase 6, CO III, ND 3, ND 4L, tRNA^{GLY} and tRNA^{ARG} shown in Fig. 2.

By comparing the sequences of the trout mitochondrial genome to those of other vertebrate species, I examine the constraints imposed on the evolution of the mitochondrial genome by its function. Those sequences which are conserved among vertebrate species and in some cases Drosophila are assumed to have important functions. In this analysis I begin by examining the conservation of sequences with known function, the protein and tRNA coding sequences. These comparisons are made through the alignment of the amino acid and nucleic acid sequences of these genes. I also examine the less obvious potential constraints imposed by codon usage and the conservation of signal sequences for processing of the primary transcript of the mitochondrial genome. During the analysis several interesting secondary structures were identified in the DNA sequences. The conservation of these structures among animal species is examined.

FIGURE 2.

The position of the 2214 base pair Hind III fragment relative to the vertebrate mitochondrial functional map. The vertebrate mitochondrial map is represented as two strands, the inner and outer lines represent the heavy and light strands respectively. The template sequences of both rRNA coding genes and all protein coding genes with the exception of ND 6 are on the heavy strand. The tRNA genes are depicted by their single letter amino acid codes next to the strand on which their coding sequences are found. The salmon sequence extends from the ninth nucleotide of ATPase 6 to within 16 nucleotides of the start of ND 4, position 10039-12251 in Xenopus (Roe et al., 1985), 8536-10516 in human (Anderson et al., 1981), 8298-10516 in bovine (Anderson et al., 1982) and 7935-10148 in mouse (Bibb et al. 1981). The identities of the genes are as in Brown (1985) and include the NADH dehydrogenase subunits (ND) (Chomyn et al., 1985).



RESULTS AND DISCUSSION

Comparison of Protein coding sequences

The trout amino acid and nucleotide sequences for the fragment are shown in Fig. 3. The organization of the genes in this trout fragment is identical to that found in other vertebrates. The amino acid sequences of each of the four genes are aligned with the homologous sequences from other animal mitochondrial genomes in Fig. 4. From these alignments the patchy nature of the variations is obvious. Several sequence segments show complete conservation while others show little identity among the animal groups. The simplest explanation for this observation is that the conserved sections reflect portions of the protein for which the amino acid sequence is critical. The extension of this conservation to the Drosophila sequence is consistent with this interpretation. An examination of Fig. 4 also reveals the differences in levels of accumulated amino acid replacements between protein coding sequences. I have determined the levels of amino acid identity for each pairwise comparison for each protein (Fig. 5). These pairwise comparisons show consistent and dramatic differences in the level of amino acid replacements in the four proteins. The CO III protein sequence is clearly less divergent in each pairwise comparison than the other three proteins. This conservation probably reflects greater constraints on the CO III protein sequence.

For each gene, the trout and Xenopus show an unexpectedly

FIGURE 3.

The sequence of the 2214 base pair Hind III fragment of trout. The single letter amino acid designations are presented above the protein coding regions. The first and last nucleotide of each tRNA gene is marked (*).

ATPase 6

S F F D Q F M S P T Y L G I P L I A V A
 AAGCTTCTTCGACCAATTTATGAGCCCCACATACCTAGGTATCCCACTTATCGCCGTAGC
 0-

L T L P W I L F P T P S A R W L N N R L
 ATTAACCCTCCCATGAATTCTTTTCCCTACCCCCTCTGCCGATGACTAAACAACCGCCT

I T L Q G W F I N R F T Q Q L L L P L N
 AATTACCCTGCAAGGGTGGTTCATCAACCGATTTACCCAGCAACTTCTTTTACCGCTAAA

L G G H K W A A L L T S L M L F L I T L
 TCTAGGCGGTACAAGTGAGCAGCTCTACTAACTTCCCTCATACTATTTCTTATTACCCT

N M L G L L P Y T F T P T T Q L S L N M
 AAATATACTTGGCCTACTTCCATATACATTCACCCCGACCACACAGCTCTCCCTAAATAT

G L A V P L W L A T V I I G M R N Q P T
 GGGCCTCGCAGTCCCACTGTGGCTTGCTACAGTAATTATCGGCATACGAAACCAACCTAC

A A L G H L L P E G T P V P L I P V L I
 GGCCGCCCTCGGCCATTTATTGCCTGAAGGAACCCCGTTCCACTGATCCCAGTACTGAT

I I E T I S L F I R P A L G V R L T A N
 CATTATCGAAACAATTAGCCTTTTATCCGCCCCGCCCTTGGCGTACGACTTACAGCCAA

L T A G H Q L I A T A A F V L L P M M P
 TCTCACAGCAGGCCACCAACTAATTGCTACAGCAGCCTTTGTTCTTCTACCTATAATACC

T V A I L T S I V L F L L T L L E I A V
 TACAGTAGCAATCCTAACTTCTATTGTCCTCTTTCTACTCACCTTCTCGAAATCGCCGT

A M I Q A Y V F V L L L S L Y L Q E N V
 AGCCATGATTC AAGCCTACGTTTTTGTCTTACTCCTAAGCCTCTATTTACAAGAAAACGT

CO III

U M A H Q A H A Y H M V D P S P W P L T
 TTAATGGCACACCAAGCACACGCATACCACATGGTTGACCCAAGCCCCTGACCTCTGACC
 -663

G A I A A L L L T S G T A V W F H F H S
GGCGCAATTGCCGCCCTTTACTTACATCAGGCACTGCAGTCTGATTCCATTTCCACTCG

L T L L T L G N I L L L L T M Y Q W W R
CTCACACTTCTTACCTTAGGTAACATTCTCTTACTTCTAACCATATAACCAATGATGGCGG

D I I R E G T F Q G H H T P P V Q K G L
GATATCATCCGAGAAGGTACCTTTCAAGGACACCACACGCCCCCGTCCAAAAAGGGCTA

R Y G M I L F I T S E V F F F L G F F W
CGATATGGCATAATCTTATTTATTACCTCCGAGGTATTCTTTTTCTTAGGTTTCTTCTGA

A F Y H A S L A P T P E L G G C W P P A
GCCTTCTACCACGCCAGCCTCGCCCCACACCTGAATTAGGAGGTTGCTGACCCCCCGCA

G I T T L D P F E V P L L N T A V L L A
GGTATTACTACTCTAGACCCCTTTGAGGTACCCCTTCTTAATACTGCAGTCCTTCTAGCA

S G V T V T W A H H S I M E G E R K Q T
TCTGGTGTCCCGTAACATGAGCCCACCACAGCATCATAGAAGGTGAACGAAAACAAACC

I Q A L T L T I L L G F Y F T F L Q G M
ATTCAAGCTCTTACTCTCACTATCTTACTGGGATTTTACTTCACTTTCTTACAAGGTATA

E Y Y E A P F T I A D G V Y G S T F F V
GAATACTACGAAGCCCCATTTACAATCGCTGACGGCGTATACGGCTCTACTTTCTTTGTC

A T G F H G L H V I I G S T F L A V C L
GCTACAGGATTCCATGGCCTACACGTAATTATTGGCTCTACCTTTCTGGCCGTTTGCCTT

L R Q V Q Y H F T S E H H F G F E A A A
CTACGACAAGTTCAATACCACTTTACATCTGAACATCATTTTGGCTTTGAAGCTGCTGCC

W Y W H F V D V V W L F L Y V S I Y W W
TGATATTGACACTTTGTAGACGTTGTATGGCTCTTCCCTATACGTCTCTATTTACTGATGA

trnA^{GLY}

G S U*
GGCTCATATCTTTCTAGTATTAATACGTATAAGTGACTTCCAATCACCCGGTCTTGGTT

ND 3

* M N L I T T I I T I T I T L
 AAAATCCAAGGAAAGATAATGAATTTAATCACAACAATCATCACTATTACCATCACATTA
 -1519

S A V L A T I S F W L P Q I S P D A E K
 TCCGCAGTACTAGCCACTATTTCTTTCTGATTACCACAAATCTCCCCAGACGCAGAGAAG

L S P Y E C G F D P L G S A R L P F S L
 TTATCCCCCTACGAATGTGGATTTGACCCCTTAGGGTCCGCCCGCCTGCCCTTCTCCTTA

R F F L I A I L F L L F D L E I A L L L
 CGCTTCTTTCTAATCGCCATCTTATTCCTCCTATTTGATCTAGAAATCGCCCTCCTTTTG

P L P W G D Q L H T P T L T L I W S T A
 CCCCTACCTTGAGGGGATCAACTCCACACCCCGACCCTGACACTCATCTGATCCACTGCC

V L A L L T L G L I Y E W T Q G G L E W
 GTTCTAGCCCTTCTTACTCTTGGCTTAATCTATGAATGAACCCAAGGAGGCTTAGAATGA

tRNA^{ARG}

A E U*
 GCCGAGTACGGAGTTAGTCCAAAACAAGACCCTTGATTTGGGCTCAAAGACCATGGTT
 -1876

ND 4L

* M T P V H F S F T S A F I L
 TAAGTCCATGACCGCCTTATGACACCAGTACACTTCAGCTTTACCTCAGCCTTTATTTTA
 -1939

G L M G L A F H R T H L L S A L L C L E
 GGGCTTATAGGACTCGCGTTTCACCGCACCCACCTTCTCTCAGCCCTTCTATGCCTAGAA

G M M L S L F I A L S L W A L Q M E A T
 GGAATAATACTCTCTTATTCATCGCCCTCTCCCTCTGAGCCCTCAAATGGAAGCGACT

G Y S V A P M L L L A F S A C E A S A G
 GGCTACTCAGTGGCCCCGATACTTCTCCTAGCGTTCTCAGCCTGTGAAGCCAGCGCAGG

L A L L V A T A R T H G T D R L Q S --
 TTAGCCCTACTAGTAGCAACTGCACGAACACACGGCACAGACCGCCTCCAAAGCTT
 -2214

FIGURE 4.

A comparison of the protein coding sequences: A, ATPase 6; B, CO III; C, ND 3 and D, ND 4L; of trout and other animal mitochondria. The comparisons include trout, Xenopus, human, mouse and Drosophila. The alignments of Xenopus, human bovine and mouse are as in Roe et al. (1985). The mammals were aligned with the Drosophila ATPase 6 and CO III genes as in Clary and Wolstenholme (1983). All trout genes and Drosophila ND 3 and ND 4L genes were aligned so as to maximize the homology of sequences conserved in the other genomes. Amino acid identities to rainbow trout are indicated by (.)

A
ATPase 6

RAINBOW 25
 XENOPUS W F P S L
 HUMAN P . . . G
 MOUSE L D F F .
 DROS. T L L L M F .
 L M I I T
 G A I I R
 V I L A L
 A . V V W
 I . A V N
 L . A I .
 P . . S
 I . L F L
 G . . N
 Y . L .
 I . M F
 T V . . A
 P . . S
 S . A T P
 M . I I D
 F . . .
 Q . S . V
 D . A A S
 F . . .
 S . L L L

RAINBOW 49
 XENOPUS I N H K L
 HUMAN I L . V L
 MOUSE . L . L I
 DROS. G S . Q H N
 Q . . W
 L . T F F
 T . . S I
 I . . H N
 L . . Y
 R . . .
 N . . S
 N . . P
 ? F I I M
 L G . .
 W N Y R .
 R S K K Y
 A Q S S I
 S I T . .
 P . . .
 T P W . .
 S W . .
 P T . .
 F I S W . .
 I . . .
 L . M I

RAINBOW 73
 XENOPUS L . . .
 HUMAN S . . .
 MOUSE T . . V V I
 DROS. L . . I F I
 L . . I F I
 L . M M I
 A L L L F
 A . . S T T
 --- ---
 W . . G S
 K . T T N
 H . R R .
 G . . .
 G P K K S
 L S T P P
 N T . T G
 L ? H H .
 P L M I L
 L Q T . T
 L Y M M K
 L F M M F
 Q I . E
 T K K K
 T S I H
 F T I L

RAINBOW 98
 XENOPUS L . M M .
 HUMAN S . . . T
 MOUSE L
 DROS. Q . . . H
 T . . . S
 T
 P . . . R
 T
 F
 T . S . I
 Y . H H .
 P . . . F
 L
 L G
 L
 M L L L F
 L . T T N
 T S . S F
 I M A G L
 L I I I
 F L . L
 L I M S

A (cont.)

ATPase 6

RAINBOW	N	M	G	L	A	V	P	L	W	L	A	T	V	I	I	G	M	R	N	Q	P	T	A	A	122	
XENOPUS
HUMAN	.	L	A	M	.	I	.	.	A	G	A	---	S	S	K	I	.	N	Y	.	
MOUSE	.	L	S	M	.	I	.	.	A	G	A	F	.	H	S	K	L	K	S	---	S	
DROS.	T	L	S	.	.	L	.	.	.	C	F	M	L	Y	.	W	I	.	H	T	Q	H	---	M		

RAINBOW	L	G	H	L	L	P	E	G	T	P	V	P	L	I	P	V	L	I	I	I	E	T	I	S	L	147
XENOPUS	T
HUMAN	.	A	.	F	.	.	Q	.	.	.	T	M	.	V	
MOUSE	.	A	.	F	.	.	Q	.	.	I	S	.	.	.	M	
DROS.	F	A	.	.	V	.	Q	.	.	A	I	.	.	M	.	F	M	V	R	N	

RAINBOW	F	I	R	P	---	A	L	G	V	R	L	T	A	N	L	T	A	G	H	---	---	---	Q	L	I	168
XENOPUS
HUMAN	.	.	Q	.	M	.	.	A	I	.	.	.	
MOUSE	.	.	Q	.	M	.	.	A	M	H	.	.	
DROS.	I	.	.	.	G	T	.	A	M	I	L	H	.	L	

RAINBOW	A	T	A	A	F	V	L	L	P	S	M	M	P	T	V	A	I	L	T	S	I	V	L	L	L	193
XENOPUS
HUMAN	G	S	.	T	L	A	M	.	S	T	I	I	
MOUSE	G	G	.	T	L	.	.	S	M	N	S	
DROS.	G	N	---	T	G	P	S	M	S	Y	L	L	L	V	T	F	L	.	.	A	Q	I	A	L	---	

A (cont.)

ATPase 6

RAINBOW T L L E I A V A M I Q A Y V F V L L L S L Y L Q E 218
 XENOPUS
 HUMAN . I
 MOUSE . I . . F
 DROS. --- V . . S

RAINBOW N V * 221
 XENOPUS
 HUMAN . T
 MOUSE . T
 DROS. V N

B

CO III

RAINBOW	M	-	A	H	Q	A	H	A	Y	H	M	V	D	P	S	P	W	P	L	T	G	A	I	A	A	L	L	26	
XENOPUS	.	.	T	.	.	S	K	V
HUMAN	.	.	T	.	.	T	N	L	S	
MOUSE	.	.	T	.	.	N	F	S	
DROS.	.	S	T	.	S	.	.	P	F	.	L	.	.	.	Y	G	.	M	T	.	

RAINBOW	L	T	S	G	T	A	V	W	F	H	F	H	S	L	T	L	L	T	L	G	N	I	L	L	L	L	53	
XENOPUS	L	.	M	.	.	G	.	.	.	M	I	L	L	T	M	V	.	.	.
HUMAN	M	.	.	.	L	.	M	M	L	L	T	N	T	.	.	.
MOUSE	L	V	M	.	.	N	.	.	.	I	L	L	T	N	T	.	.	.
DROS.	T	V	.	.	M	V	K	.	.	Y	Q	Y	D	I	S	.	F	L	T	I	.	.	.	

RAINBOW	M	Y	Q	W	W	R	D	I	I	R	E	G	T	F	Q	G	H	H	T	P	P	V	Q	K	G	L	80	
XENOPUS	.	I	V
HUMAN	V	T	.	.	S	.	.	Y	
MOUSE	V	Y	
DROS.	V	V	S	Y	.	L	.	.	Y	A	.	T	I	.	.	.	

RAINBOW	Y	G	M	I	L	F	I	T	S	E	V	F	F	F	L	G	F	F	W	A	F	Y	H	A	S	L	107	
XENOPUS	I	N	S	.	.	.
HUMAN	A	S	.	.	.	
MOUSE	V	A	S	.	.	.	
DROS.	W	L	.	.	.	L	.	.	V	R	F	.	R	.	.	S	

B (cont.)

CO III

RAINBOW P T P E L G G C W P P A G I T T L D P F E V P L L N T 134
XENOPUS . . . Y . Q . . . E . H . . . T . . . P . . . N . . . L
HUMAN H .
MOUSE H D .
DROS. A I .

RAINBOW A V L L A S G V T V T W A H H S I M E G E R K Q T I Q 161
XENOPUS .
HUMAN .
MOUSE .
DROS. I .

RAINBOW A L T L T I L L G F Y F T F L Q G M E Y Y E A P F T I 188
XENOPUS .
HUMAN L I .
MOUSE L I .
DROS. G F F .

RAINBOW A D G V Y G S T F F V A T G F H G L H V I I G S T F L 215
XENOPUS .
HUMAN .
MOUSE .
DROS. .

B (cont.)

CO III

RAINBOW	A	V	C	L	L	R	Q	V	Q	Y	H	F	T	S	E	H	H	F	G	F	E	A	A	A	W	Y	W	242	
XENOPUS	S	:	:	:	:	:	:	I	:	:	:	:	:	:	:	K	:	:	:	:	:	:	:	:	:	:	:	:	:
HUMAN	T	I	:	:	:	:	:	L	M	F	:	:	:	:	:	K	:	:	:	:	:	:	:	:	:	:	:	:	:
MOUSE	I	:	:	:	:	:	:	L	K	N	:	:	:	:	:	K	:	:	:	:	:	:	:	:	:	:	:	:	:
DROS.	L	:	:	:	:	:	H	L	N	:	:	:	S	K	N	:	:	:	:	:	:	:	:	:	:	:	:	:	:

RAINBOW	H	F	V	D	V	V	W	L	F	L	Y	V	S	I	Y	W	W	G	S	*	261
XENOPUS	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
HUMAN	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
MOUSE	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
DROS.	:	:	:	:	:	:	:	:	:	:	:	I	T	:	:	:	:	:	:	:	G

TROUT	M	N	L	I	T	T	I	I	T	I	T	I	T	I	A	T	I	A	L	L	20
XENOPUS	.	T	---	A	---	.	.	L	M	.	M	.	M	.	N	
HUMAN	.	.	F	A	---	L	.	L	M	.	M	.	M	
MOUSE	I	.	.	Y	---	.	V	.	F	.	F	
DROS.	I	---	-----	-----	-----	I	A	S	V	.	L	.	L	.	T	
TROUT	I	S	F	W	L	P	Q	I	S	P	D	A	E	K	L	S	P	Y	E	C	40
XENOPUS	L	M	T	.	.	M	
HUMAN	.	T	L	N	G	.	M	
MOUSE	V	A	.	.	.	K	.	M	N	L	Y	S	.	A	N	---	
DROS.	A	.	I	L	S	K	.	A	N	L	.	R	
TROUT	G	F	D	P	L	G	S	A	R	L	P	F	S	L	R	F	F	L	I	A	60
XENOPUS	M	
HUMAN	M	S	P	.	.	V	.	.	.	M	
MOUSE	T	S	M	K	
DROS.	K	S	.	S	M	K	

TROUT	I	L	F	L	L	F	D	L	E	I	A	L	L	L	P	P	L	F	P	W	G	D	80
XENOPUS	A	A
HUMAN	.	T	A	L
MOUSE	.	T	A	I
DROS.	.	I	.	.	V	I	.	.	M	.	I	.	.	I	L
TROUT	Q	L	H	T	L	T	L	T	L	I	W	S	T	A	V	L	A	L	L	L	L	T	100
XENOPUS	.	.	N	.	V	V	I	I	I	L	.	A	A	L	I	I	T	I	
HUMAN	.	T	T	N	M	M	.	M	.	M	S	L	L	L	L	I	I	
MOUSE	.	T	T	K	M	M	.	M	.	M	.	L	L	L	L	I	I	
DROS.	K	Y	I	N	W	T	.	T	.	T	.	A	L	L	L	I	I	
TROUT	L	G	L	I	Q	T	W	T	Q	G	G	L	E	W	A	E	*	116					
XENOPUS	.	S	.	A	.	L	.	L	
HUMAN	
MOUSE	.	.	.	A	
DROS.	I	.	.	A	

D

ND 4L

TROUT ---M T P V H F S F T S A F I L G L M G 18
 XENOPUS L I I T M T . . .
 HUMAN P L L I S L L . . .
 MOUSE P S T Y F W F C
 DROS. M I I L T Y S L P M I L F C

TROUT L A F H R T H L L S A L L L C L E G M M L 38
 XENOPUS L V M N Y P I L M L . . .
 HUMAN M L L M S S T V . . .
 MOUSE T L M F S S M F I V . . .
 DROS. F V S N K S S M T V V . . .

TROUT S L F I A L W L S L A L Q H L M E A T G Y S V 58
 XENOPUS M S M I D G M M T P L T H N S M N P Y
 HUMAN M A T V I T S L S . . .
 MOUSE M M Y T S N S N M N
 DROS. M F M M M L N Y

TROUT	A	L	L	A	C	E	A	S	A	G	L	A	L	L	78
XENOPUS	M	P	L	F	S	A	P
HUMAN	V	.	.	A	A	.	V	.	A	.	.	S	.	N	.
MOUSE	I	.	.	A	A
DROS.	F	.	.	R	V	.	.	G	V	.	.	S	.	I	.
		L
	
	

TROUT	V	A	R	H	D	L	Q	S	92
XENOPUS	S	D	T	.	.	.	F	.	.
HUMAN	.	S	.	Y	.	.	H	.	.
MOUSE	.	K	L
DROS.	.	.	.	Y

FIGURE 5.

Pairwise comparisons of the nucleotide and amino acid sequences for the homologous animal mitochondrial fragments aligned in Fig. 4. The observed percent nucleotide divergence is calculated using the algorithm in Queen and Korn (1984) based on alignments of the amino acid sequences. The protein similarity was determined using an algorithm by Dayhoff et al. (1972).

	TROUT		XENOPUS		MOUSE		HUMAN		DROSOPHILA	
	NA	AA	NA	AA	NA	AA	NA	AA	NA	AA
TROUT										
ATP 6			70.6	78.4	60.5	52.7	57.4	54.4	47.0	39.2
CO III			74.9	87.4	71.8	77.1	70.7	79.0	63.0	66.5
ND 3			68.7	72.6	59.3	58.5	60.8	54.6	51.8	43.0
ND 4L			60.6	50.5	56.8	48.9	57.8	44.7	47.7	41.1
XENOPUS										
ATP 6					64.2	51.5	62.8	54.3	54.7	38.4
CO III					74.6	78.6	73.3	79.8	66.8	64.8
ND 3					61.5	56.4	60.9	56.4	55.2	45.1
ND 4L					53.5	40.9	56.8	38.0	53.2	35.1
MOUSE										
ATP 6							71.0	75.4	52.4	37.4
CO III							76.0	86.2	66.3	65.0
ND 3							66.8	64.1	60.2	43.4
ND 4L							66.3	63.4	49.1	37.2
HUMAN										
ATP 6									51.3	39.9
CO III									62.6	62.9
ND 3									51.1	37.7
ND 4L									45.7	37.2

high level of identity when the classical phylogenetic relationship of these groups is considered (Dayhoff, 1972). The close relationship of trout with the previously sequenced groups is consistent with the phylogeny of the 5s rRNA sequences of these groups (Hori and Osawa, 1987). The relationship between the Xenopus and trout protein sequences is interesting in light of the low level of amino acid replacement found among salmonids (Section 2). The observed number of amino acid replacements among salmonid species much lower than that found among humans (Cann et al., 1984). The lack of divergence between trout and Xenopus and the increased bias against amino acid replacements in salmonid mitochondria suggests that "cold-blooded" mitochondrial systems may have greater functional constraints on their mitochondrial proteins. One might expect that because each of these proteins perform the same essential function in every animal group, the constraints on amino acid sequences may be greater on proteins which must perform the same functions at a variety of temperatures. Although the data presented here are only suggestive, effects of such a constraint could alter the dynamics of evolution in different animal groups.

Codon Usage and Nucleotide Composition

Table 2 is a presentation of the trout mitochondrial genetic code and codon usage for the protein coding sections of the segment sequenced. Although this segment is only a part of the entire mitochondrial genome it is probably representative of

Table 2: Codon usage in rainbow trout

TTT	Phe	25	TCT	Ser	9	TAT	Tyr	5	TGT	Cys	2
TTC	Phe	24	TCC	Ser	10	TAC	Tyr	14	TGC	Cys	3
TTA	Leu	23	TCA	Ser	6	TAA	End	3	TGA	Trp	19
TTG	Leu	2	TCG	Ser	1	TAG	End	0	TGG	Trp	4
CTT	Leu	31	CCT	Pro	8	CAT	His	5	CGT	Arg	0
CTC	Leu	28	CCC	Pro	15	CAC	His	19	CGC	Arg	6
CTA	Leu	36	CCA	Pro	12	CAA	Gln	21	CGA	Arg	9
CTG	Leu	9	CCG	Pro	4	CAG	Gln	2	CGG	Arg	1
ATT	Ile	20	ACT	Thr	16	AAT	Asn	6	AGT	Ser	0
ATC	Ile	27	ACC	Thr	23	AAC	Asn	6	AGC	Ser	10
ATA	Met	13	ACA	Thr	23	AAA	Lys	2	AGA	Arg	0
ATG	Met	8	ACG	Thr	2	AAG	Lys	2	AGG	Arg	0
GTT	Val	9	GCT	Ala	8	GAT	Asp	3	GGT	Gly	10
GTC	Val	9	GCC	Ala	34	GAC	Asp	8	GGC	Gly	20
GTA	Val	16	GCA	Ala	19	GAA	Glu	19	GGA	Gly	9
GTG	Val	1	GCG	Ala	3	GAG	Glu	4	GGG	Gly	6

the codon bias found throughout the genome. Evidence for this comes from the fact that the homologous fragments from Xenopus, mouse, human and Drosophila give codon usage profiles which are the same as those derived from all protein coding regions of their respective mitochondrial genomes. The amino acid composition among the species is also conserved to a great extent as can be seen in the comparison of protein coding sequences. The general appearance of the trout codon usage table is similar to that of other vertebrates (Brown, 1985).

The biased distribution of nucleotides in animal mitochondrial genomes is correlated with the distribution of coding sequences and is expected to be affected by complex functions such as codon anticodon interactions. A typical form of this bias is the G+T content difference between the light and heavy strands of vertebrate mitochondrial DNAs. The nucleotide composition of the sense strand for four vertebrate species and Drosophila is given in Fig. 6. The G+T content of this fragment in trout is intermediate to the mammalian and Xenopus G+T contents. The lower G+T bias in trout is the result of a decreased general bias against G in the sense strand and in the third position of codons.

A comparison of the percentage of codons ending in each of the four nucleotides is given in Fig 6. The data for the two-fold degenerate codons and the four-fold degenerate codons have been kept separate to eliminate the possibility of losing biases which are a result of the differences in mechanisms of

FIGURE 6.

Nucleotide composition and codon usage of animal mitochondrial DNA. The data are from the homologous sequences in trout, Xenopus (Roe et al., 1985), mouse (Bibb et al., 1981), human (Anderson et al., 1981) and Drosophila (Clary and Wolstenholme, 1986).

	% NUCLEOTIDE COMPOSITION OF SENSE STRAND				% CODONS ENDING IN 4 CODON BOXES				% CODONS FOR LEU		% LEU CODONS BEGINNING WITH				
	T	C	A	G	T	C	A	G	T	C	T	C			
TROUT	28.4	31.5	24.4	15.7	22.3	37.5	32.8	7.4	23.1	36.9	33.1	6.9	18.6	19.4	80.6
XENOFUS	32.4	24.9	29.4	13.3	29.0	28.7	37.9	4.4	29.6	17.1	50.4	2.9	18.6	34.1	65.9
MOUSE	30.9	25.8	31.7	11.6	33.0	30.3	34.6	2.1	20.9	19.6	57.7	2.5	16.2	16.1	83.9
HUMAN	27.2	32.0	28.4	12.4	12.1	41.9	39.3	6.7	23.5	37.3	33.9	5.3	18.2	15.1	84.9
DROS.	42.1	14.1	31.3	12.5	48.3	5.1	44.9	1.7	49.0	2.1	47.4	1.5	15.0	89.2	10.8

translation used by each of the two classes. In all species the four-fold degenerate codons exaggerate the bias against G found in the third position of two-fold degenerate codons.

The hydrophobic amino acid, leucine, makes up a significant proportion (15%-18.6%) of the amino acids coded for by this fragment. The species in Fig. 6 show variation in the usage of the CTN and TTPu tRNA species to code for leucine. At the extremes of this usage are the Drosophila and human sequences. Drosophila shows a bias toward the TTPu tRNA whereas human mitochondria is biased toward the CTN tRNA. The usage of TTPu and CTN is correlated at the extremes with the nucleotide composition and/or the percent codons ending in T and C. This does not hold for the variation found among the vertebrate species. Trout, which contains more C than T in both nucleotide composition and codon usage, is less biased against the TTPu codon than mouse which has more T than C. Trout is intermediate to the mammals and Xenopus in usage of TTPu and CTN. As pointed out in section 2 of this thesis, the TTPu and CTN codons can have dramatic effects on the potential for, and distribution of, variation in the mitochondrial protein coding sequences. Species which are biased toward either tRNA codon are limited in the variability of the first position of those codons. Those species such as Drosophila, which are biased in favour of the two codon family tRNA species, are limited to transitions in the third positions of their leucine codons.

A comparison of the nucleotide sequence composition of the

sense strand shows that each species has a unique sequence composition. The sequence composition profile is dominated by a universal bias against G in the sense strand. Trout appear to be the least biased against G and also have the highest G+C ratio at 47.2, compared to 26.2 in Drosophila. Trout nucleotide composition is most similar to the human nucleotide composition except that trout is more biased against A than is human. It is important to note that the sequence biases are not limited to the protein coding regions of the genome in species for which the entire mitochondrial sequence is known.

The biases in nucleotide composition are less extreme in trout than in other species. The biases in Drosophila are much more pronounced. The maintenance of this bias can be explained using a mechanism proposed by Wolstenholm and Clary, (1985), whereby once a sequence becomes biased, the transcription and replication machinery may become adapted to those particular conditions. An analogous mechanism involving replication is difficult to extend to the vertebrate mitochondrial sequence biases because of their strand asymmetry.

tRNA Genes

The two sequence blocks occupying the junctions between CO III/ND 3 and ND 3/ND 4L have been identified as the coding sequences for tRNA^{GLY} and tRNA^{ARG} respectively. They have been identified as tRNA sequences based on their ability to form typical cloverleaf secondary structures and by their primary

sequence homology to other known mitochondrial tRNA sequences. As no mitochondrial tRNAs have been shown to be imported from the cytoplasm (Aujame and Freeman, 1979), I assume that these are the only functional tRNAs in trout mitochondria. The tRNA genes punctuate the protein coding sequences in the same fashion as other species. Hence, the role of the tRNA sequences in the processing of the polycistronic message (Ojala et al., 1981) is probably conserved in salmonids.

Both trout tRNA^{GLY} and tRNA^{ARG} correspond to four-fold degenerate codons. The primary base in their anticodons is U. They must therefore read all four codons in their respective families via U:N wobble (Barrell et al., 1980).

The proposed secondary structures of these tRNAs are shown in Fig. 7 and 8. The nucleotide composition of these tRNAs is similar to that for the protein coding portions of the sequence, with a slightly higher proportion of A and less C. The higher proportion of A is due to the high frequency of A in the DHU and T ψ C loops and is typical of animal mitochondrial tRNAs.

In Fig. 9 the homologous tRNAs from trout, Xenopus, mouse, human, bovine and Drosophila are aligned. This figure reveals a high level of sequence similarity among the vertebrates which in many cases extends to Drosophila.

The aminoacyl (AA) stems are seven base pairs in length. The tRNA^{GLY} aminoacyl stems in mouse and Drosophila each have a single non-standard pair (C:A and A:G, respectively). Non-standard pairs in the AA stems of tRNA^{ARG} include three in trout

FIGURE 7.

Cloverleaf structure for trout tRNA^{GLY}. The structure was determined from its alignment with vertebrate mitochondrial tRNA^{GLY} (Roe et al., 1985). Standard base pairs in stems are designated (-).

TRNAGLY

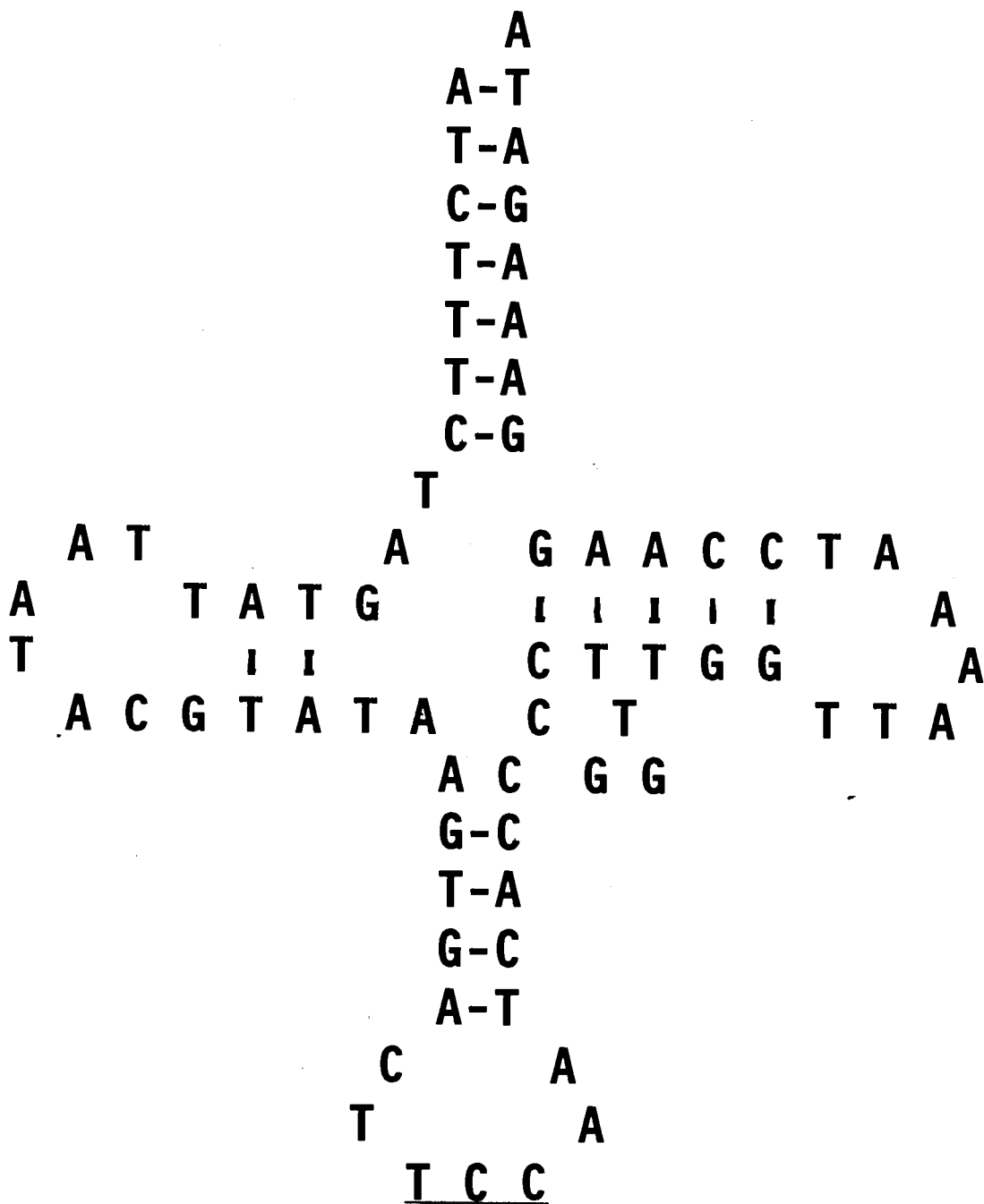


FIGURE 8.

Cloverleaf structure for salmon tRNA^{ARG}. The structure was determined from its alignment with vertebrate mitochondrial tRNA^{ARG} (Roe et al., 1985). Standard base pairs in stems are designated (-).

TRNAARG

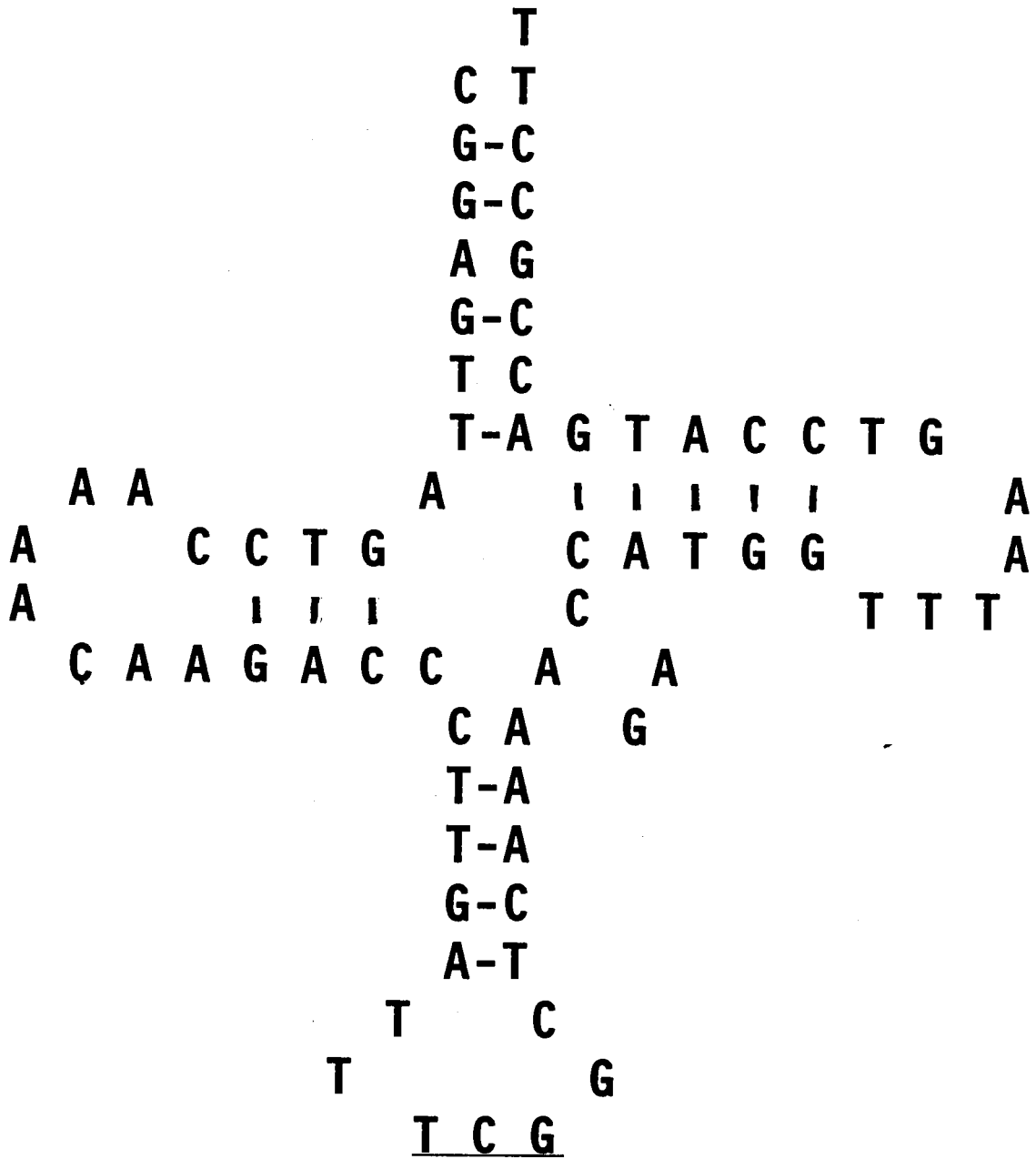


FIGURE 9.

The alignment of the homologous tRNA^{GLY} and tRNA^{ARG} gene sequences from the trout, Xenopus, mouse and Drosophila mitochondrial genomes (References in Fig. 6). The sequences are aligned as in Gauss et al. (1979). The nucleotides between the AA and DHU stems are omitted.

SPECIES	AA STEM	DHU STEM	DHU LOOP	DHU STEM	AC STEM	AC LOOP	AC STEM	VAR LOOP	TψC STEM	TψC LOOP	TψC STEM	AA STEM	3'
<u>tRNA^{GLY}</u>													
TROUT	ATCTTTC	GTAT	TAATAC	GTAT	AGTGA	CTTCCAA	TCACC	CGGTCTTGG	TTAAAAT	CCAAG	GAAAGAT	A	
XENOPUS	ACTTTCT	GTAT	TAACCA	GTAC	CGTGA	CTTCCAA	TCACA	AAGTCTTAG	TAGAAT	CTAAG	AGAAAAGT	A	
MOUSE	ACTCCCT	GTAT	AATTA	ATAT	ACTGA	CTTCCAA	TTAGT	AGAT TCTGA	ATAAAC	CCAGA	AGAGAGT	A	
DROS.	ATATATA	GTAT	AAAA	GTAT	TTTGA	CTTCCAA	TCATA	AGGTCTATT	AATA	AATAG	TATAGAT	A	
<u>tRNA^{ARG}</u>													
TROUT	CGGAGTT	GTCC	AAAACA	AGAC	CTTGA	TTTCGGC	TCAAA	AGACCATGG	TTTAAAGT	CCATG	ACCGCCT	T	
XENOPUS	GAGTTGT	GTCT	AAACA	AGAC	GTTGA	TTTCGGC	TCAAC	AAAT TATGG	TTAAAACC	CCATA	ATAACTC	T	
MOUSE	TGGTAAT	GTTT	AAAAA	AAAT	AATGA	TTTCGAC	TCATT	AGAT TATGA	TGCTGT	TCATA	ATTACCA	A	
DROS.	GAATATG	GCGA	TTAA	TTGC	GTTAG	TTTCGAC	CTAAC	CTTA GGTAT	TAT	ATACC	CTTATTT	T	

(2xC:T, 1xA:G), two in Drosophila (T:T,G:T), and one in Xenopus (G:T). The primary sequence homology of the AA stems is high and all species are related by a limited number of substitutions, additions or deletions. The dihydrouracil (DHU) stems show an interesting conservation of their 5' strands in tRNA^{GLY}.

As in most tRNAs the salmon AA and AC stems are asymmetric with respect to the distribution of purines and pyrimidines. This asymmetry maintains the rigidity of these stems (Dickerson and Drew, 1981). The DHU and T ψ C stems of trout and other species do not show this asymmetry, perhaps due to the tertiary structure of the DHU and T ψ C stems. The predominant non-standard pairs found in mitochondria are G:U and A:C. Both pairs are allowable under wobble and allow for the asymmetry discussed above. The predominance of these pairs may reflect their ability to maintain this asymmetry.

The DHU stems are four bp long with 1-2 nonstandard pairs. The non-standard pairs include six T:G and one T:A pair, all at external positions. The DHU loop shows little primary sequence similarity aside from being A+T rich. The size of the loop decreases from trout with six bases to Drosophila with four bases. In eukaryotic nuclear tRNA genes the DHU and T ψ C loops contain the A and B blocks of conserved sequences. These sequences act as the RNA polymerase III promoters for these genes (Hall et al., 1982; Ciliberto et al., 1983; Baker et al., 1986).

The anticodon (AC) stems are five base pairs long. The tRNA^{GLY} AC stem has a single non-standard pair in each species

(2xC:A, 1xG:T, 1xT:T). The tRNA^{ARG} AC stem shows a single non-standard pair (C:A) in trout. Significant sequence identity exists among the AC stems in the first three bases proximal to the AC loop. The AC loops are seven base pairs in length with, perfect identity in tRNA^{GLY} and near-perfect identity in tRNA^{ARG}. The size and conservation of the AC loop is consistent with the model of codon-anticodon interaction proposed by Curran and Yarus, (1987). The only change is an A rather than a G at the Pu37 position in mouse and Drosophila tRNA^{ARG}. The variable loops are four bases long and show no convincing conservation among the vertebrate species. The T ψ C stems are five base pairs long with a single non-standard pair (C:A) in mouse tRNA^{GLY}. The T ψ C loops show length variation similar to the DHU loops, with trout the largest at seven bases and Drosophila the smallest with four and three base loops in tRNA^{GLY} and tRNA^{ARG} respectively.

Trout tRNAs lack most of the invariant bases found in other non-mitochondrial tRNAs (Kim et al., 1974; Gauss and Sprinzl, 1979). Trout tRNA^{GLY} and tRNA^{ARG} lack all of the invariant bases of the DHU loop and the C₅₆G A of the T ψ C loop. This is also generally true for the other mitochondrial tRNA sequences known. The invariant bases found in the trout tRNAs include the T₃₃ and Pu₃₇ bases flanking the anticodon, and the bases G₅₃ and C₆₁, the last pair in the T ψ C stem. It is interesting to note that none of these bases are involved in tertiary interactions. The only bases involved in tertiary interactions that are conserved in trout tRNAs are the bases T₅₄ and A₅₈, which may assist in

maintaining the helicity of the loop with respect to G₅₃ and C₆₁ of the stem, also conserved. Both of the trout tRNAs show the potential to form some of the hydrogen bonds involved in the tertiary interactions described by Kim et al. (1974). This potential is particularly evident in tRNA^{GLY}, but none of the interactions between the DHU, variable and T ψ C loops are conserved in all trout sequences. This observation suggests that the tertiary interactions of the trout mitochondrial tRNAs as in other animal mitochondrial tRNAs are weaker or different in form.

Among the mitochondrial tRNAs aligned in Fig. 9, the most conserved region is the anticodon loop. This conservation is extraordinary in light of the lack of primary sequence homology in the other loops and stems. The sizes of the DHU and T ψ C loops are highly variable, suggesting a lack of consistent interactions between DHU and T ψ C loops. The trout have six base DHU loops and seven base T ψ C loops in both mitochondrial tRNAs. The sizes of the homologous tRNAs in the other species seem to show a consistent pattern of size variation. In each case the mouse and Drosophila loops are smaller than the trout and Xenopus. In comparison to the DHU and T ψ C loop sizes, the lengths of the stems are highly conserved and the 5' DHU stem is also conserved in primary sequence. These consistent features of mitochondrial tRNAs suggest that beyond the maintenance of secondary and tertiary structure, the primary sequence of the tRNAs outside of the AC loop is not important.

Junction Sequences

I have presented the results of comparisons of the protein coding and tRNA regions of the trout fragment with other animal mitochondrial sequences. Aside from the more obvious coding functions of the mitochondrial DNA sequences there are sequence directed functions associated with other processes in the mitochondria which would be expected to impose constraints on the sequence.

The transcription of animal mitochondrial DNA occurs through the production of a polycistronic message from promoters in the D-loop region (Clayton, 1984; Chang and Clayton, 1984). Individual transcripts of rRNAs, mRNAs, and tRNAs result from endonucleolytic cleavage at specific junctions in these primary transcripts. The signals involved in these events are poorly understood. The trout mitochondrial fragment sequenced has a variety of potential cleavage junctions representing a range of processing mechanisms in mitochondria. Although most of the sequences currently implicated in non-coding functions are stem-loop structures, there are clearly other sequence directed mechanisms which will affect the potential variability of these regions.

tRNA sequences have been implicated in the processing of primary mitochondrial transcripts because they flank rRNA genes and most of the protein coding genes. The trout fragment sequenced contains two tRNA sequences, tRNA^{GLY} and tRNA^{ARG}. A comparison of the homologous stem structures formed by tRNA^{GLY} in

Drosophila, mouse, Xenopus, and trout are given in Fig 10. The comparison shows that a cleavage of the RNA at the 5' primary nucleotide of the stem will result in a T or TA at the end of the CO III coding sequence. The final maturation and completion of the stop codon of the message are accomplished through polyadenylation. Cleavage at the 3' end of the stem appears to occur one nucleotide beyond the stem. In all cases for tRNA^{GLY} this results in a complete tRNA, with the A of the ATG start codon as the first nucleotide of the following message (ND 3). Manam and Van Tuyle (1987) have recently isolated and characterized the 5' and 3' processing nucleases from rat liver mitochondria believed to be responsible for this event. The 5' nuclease cleaves the primary transcript and the 3' nuclease requires the transcript to be processed at the 5' end.

The putative cleavage site formed by the tRNA^{ARG} stems from Drosophila, mouse, Xenopus, and trout are presented in Fig. 11. The tRNA^{ARG} stem that may connect the ND 3 and ND 4L genes contains three non-standard pairs. One of these mispairs is located in the primary position of the proposed cleavage site. An examination of the vertebrate mitochondrial tRNAs shows that the stems that are involved in precise processing events are without non-standard primary base pairs and the few non-standard pairs present are all G:T or A:C. The lack of a good aminoacyl stem in the tRNA^{ARG} sequences in trout and Drosophila may suggest that the constraints on that sequence in these taxa have changed.

FIGURE 10.

A comparison of the potential stem structures involved in the tRNA^{GLY} processing. The comparison includes trout, Xenopus, mouse, and Drosophila (References as in Fig. 6). Standard base pairs are indicated by a (-). The suspected cleavage site is shown as (/). The dashed loop at the bottom of each structure corresponds to the remainder of the tRNA^{GLY}.

POTENTIAL PROCESSING SITES

TRNAGLY

<u>DROSOPHILA</u>	<u>MOUSE</u>	<u>XENOPUS</u>	<u>TROUT</u>
CAT/A-TA/ATT	CTT/A-TA/ATC	CAT/A-TA/ATG	AT/A-TA/ATG
T-A	C-G	C-G	T-A
C-G	T-A	T-A	C-G
T-A	C-G	T-A	T-A
A-T	C A	T-A	T-A
T-A	C-G	C-G	T-A
A-T	T-A	T-A	C-G

FIGURE 11.

A comparison of the potential stem structures involved in the tRNA^{ARG} processing. The comparison includes trout, Xenopus, mouse, and Drosophila (References as in Fig. 6). Standard base pairs are indicated by a (-). The suspected cleavage site is shown as (/). The dashed loop at the bottom of each structure corresponds to the remainder of the tRNA^{ARG}.

POTENTIAL PROCESSING SITES

TRNAARG

<u>DROSOPHILA</u>	<u>MOUSE</u>	<u>XENOPUS</u>	<u>TROUT</u>
G-T	TAAA/T-AA/TATG	GAAT/G-CT/ATG	AATA/C TT/ATG
A-T	G-C	A-T	G-C
A-T	G-C	G-C	G-C
T-A	T-A	T-A	A G
A-T	A-T	T-A	G-C
T T	A-T	G T	T C
G-C	T-A	T-A	T-A
/ \	/ \	/ \	/ \
\ /	\ /	\ /	\ /

A different type of processing site and presumably sequences to direct the processing are found at the junction of the ATPase 6 and CO III genes. This is the only non-overlapping junction between protein coding genes not punctuated by a tRNA sequence in these taxa. As first noted by Bibb et al. (1981), the junction between the coding sequences for ATPase 6 and CO III contains a potential stem loop structure in the 3' end of ATPase 6. I have found an analogous structure in the trout junction. The analogous vertebrate and Drosophila structures are probably all cleaved at the 3' end of the stem structure. The trout cleavage site is 3 bases 3' from the stem. The cleavage site of yeast is also 3' from the stem. This variation in the distance of the cleavage site from the stem may reflect a divergence of cleavage mechanisms or a tolerance of the endonuclease. It is interesting to note that a deletion in the ATPase 6 sequence of the pink salmon alters the position of the stem relative to the cleavage site. If in fact the nucleases are tolerant of slight distance changes, the lack of non-coding nucleotides at most junctions may not result from constraints imposed by the cleavage mechanism.

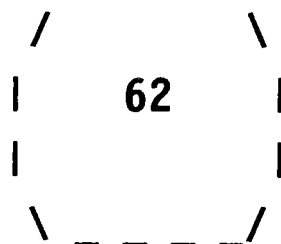
The ATPase 6 gene is also novel with respect to its translation. The ATPase 8 and ATPase 6 genes are present on a single transcript, and both genes are coded for by this transcript (Mariottini et al., 1983). This is one of only two cases where mitochondrial mRNAs overlap and presents a question concerning their translation. Most mRNA sequences contain little

or no 5' untranslated sequence and thus no ribosome binding sites. This fact has led Montoya et al. (1981) to suggest a mechanism whereby ribosomes attach at or near the 5' end of sequences and recognize the initiator codon after fine adjustment. These investigators also suggest that secondary structure of the mRNA may play a role in the exclusion of alternate AUG codons. Because the ATPase 6 transcript does not have a normal 5' end it must be translated by a different mechanism. Several possible mechanisms have been proposed (Anderson et al., 1982). These mechanisms include alternative splicing, ribosomal frameshifts and the requirement of expression of ATPase 8 for expression of ATPase 6.

Using the method of Tinoco et al., (1973) I have searched the trout mitochondrial fragment for other stem loop structures which may be functional in the mitochondrial system. I have identified eleven stem loop structures in the trout mitochondrial fragment. These stem loop structures all have free energies lower than those of the stem loops discussed above, including the tRNAs and ATPase 6/CO III junction. A stem loop structure was found in the ATPase 6 sequence 112 bases 3' of the ATPase 8 stop codon and 121 bases 3' to the probable AUG of ATPase 6 (Fig. 12). If this stem structure were to act as an endonuclease cleavage

FIGURE 12.

The secondary structure found in the 5' end of ATPase
6. The sequence extends from nucleotide 108 to 215
in the trout fragment. Standard base pairs are designated (-).

SECONDARY STRUCTURE IN ATPase 6**5' -AACAA-TCACAA-3'****C-G****C-G****G-C****C-G****C-G****T-A****TTAA-TC**

site it would have to act at a distance and would eliminate either ATPase 8 or ATPase 6 coding sequence. The conservation of the ATPase 6 and 8 sequences suggests that they are both expressed as proteins. A possible role for this structure involves orientation of the ribosomes for initiating translation at the AUG of ATPase 6. An alternative role might be that this structure eliminates the accessibility of alternate AUG sites in ATPase 6. A role for mRNA secondary structure in the control of ribosome binding has been found in bacteriophage MS2 (Kastelein et al., 1983).

Another interesting stem loop in the trout sequence occurs near the ND 3/tRNA^{ARG} junction. The alternative structure shown in Fig. 13 has a 9 bp stem and a 31 base loop. This structure is theoretically much more stable than the tRNA^{ARG} stem. Cleavage at the position corresponding to those found for tRNA junctions would result in a loss of the 3' ND 3 sequence and cleavage in the middle of the tRNA^{ARG} gene. Conservation between the trout ND 3 and other species suggest that the 3' end of ND 3 has not lost its protein function. Alternative splicing of this sequence may be possible.

In light of these potential stem-loop structures it is interesting to note that the mRNA sequences of other known animal mitochondria also contain numerous and complex patterns of potential stem loop structures. A comparison of the potential stem loop structures in other animal mitochondria with those found in trout reveals some similarities. In most cases the 5'

FIGURE 13.

Alternate structure for the processing site between ND 3 and tRNA^{ARG}. The stem loop structure extends from nucleotide 337 of ND 3 to nucleotide 40 of tRNA^{ARG}. Standard base pairs are designated (.).

**ALTERNATE SECONDARY STRUCTURE AT
ND 3-TRNA^{ARG} JUNCTION**

5' -GAAT-AAA-3'

G-C

A-T

G-C

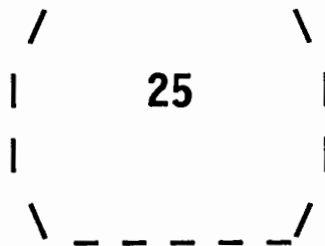
C-G

C-G

G-C

A-T

CATA-TAGT



end of the putative mRNA is not involved in a stem loop and the extreme 3' end of most genes are involved in stable stem loop structures. The exact position of the stems in the sequences involved varies, suggesting that the constraints imposed by these putative structures may be slight and that their functions are more general in nature.

Conclusions

In this section I have attempted to determine some of the constraints on the evolution of the animal mitochondrial genome. Under the assumption that sequences and structures which are conserved over long periods of time are constrained by important functions, I have compared the sequence of a trout mitochondrial fragment to the homologous sequences from other animals.

The constraints on sequence evolution in the animal mitochondrial genome range from absolute primary sequence conservation in the anti-codon loops of tRNAs, to maintenance of nucleotide composition biases. By comparing distantly related species we can identify functionally important sequences in the mitochondrial genome. The conserved sequences I have identified in this comparison include both stretches of amino acids in of the four proteins examined and the AC loops of the two tRNA genes. I have also found that the nucleotide composition and codon usage in trout is generally less biased than in other vertebrates. In addition I have examined the conservation of potential secondary structures of "known" and unknown function.

One would expect sequences and structures conserved among these relatively divergent group to also be constrained in comparisons of more closely related species. An understanding of the effect of these constraints on the dynamics of the evolution of the mitochondrial molecule requires comparisons between more closely related groups and an understanding of the basic mechanisms of mutation in this molecule. In the following two sections I investigate the distribution of variation among closely related salmonid species and the spectrum of substitutions among them.

SECTION II

DISTRIBUTION OF VARIATION
IN
THE SALMONID MITOCHONDRIAL GENOME

INTRODUCTION

In this report I investigate the distribution of variation in a 2214 base pair Hind III fragment from six closely related salmonid species. The fragment contains the genes for ATPase subunit 6 (ATPase 6), cytochrome oxidase subunit 3 (CO III), NADH dehydrogenase subunit 3 (ND 3), NADH dehydrogenase subunit 4L (ND 4L), and two tRNAs, tRNA^{GLY} and tRNA^{ARG}. By comparing the sequences of homologous fragments from six salmonid species I am able to examine the distribution of variation in this fragment. In combination with the comparison made in section I, the comparisons made here will allow for an examination of the dynamics of the evolution of this fragment. The variation observed in this fragment includes 478 substitutions at 351 positions and a single 3 base pair deletion. Of these, 329 are silent positions in protein coding sequences, 12 are non-silent positions and 10 are in tRNA sequences. I have found the frequency of multiple substitutions to be significantly high. I have also found that the level of divergence observed between species varies among protein coding sequences as well as tRNAs. Further investigation of the variation within protein coding sequences has shown a complex distribution correlated with the distribution of potential variable sites. The variation within the tRNA genes is limited to the DHU loop. These findings have important implications for both the mechanism of evolution, and

the use of restriction enzymes in sampling variation in the mitochondrial genome. In addition I use the different levels of divergence among these species to investigate the dynamics of observed changes. Because of the lack of addition/deletion events and non-silent changes this analysis is limited to comparisons of the transition and transversion levels. The substitution dynamics in salmon are compared to those found in the protein coding regions (Brown et al. 1982) and rRNA sequences (Hixson and Brown, 1986) of primates.

RESULTS AND DISCUSSION

SEQUENCE VARIATION

The sequences of the homologous Hind III fragments from the six salmonid species are aligned in Fig. 14. The sequence extends from the ninth nucleotide of the ATPase 6 coding sequence to within eighteen nucleotides of the 3' end of the ND 4L gene (positions 10039 to 12251 of Xenopus laevis, Roe et al., 1985). The gene order within this fragment is identical to the order found in other vertebrate species and a comparison of the salmonid sequence with other organisms has been made elsewhere (Section I). Among the different salmonid species analyzed there are 771 positions within the protein coding regions which have the potential for silent variation. Of these sites 329 are variable among the six species. The spectrum of substitutions at these positions is described in detail elsewhere (Section III). The variation also includes a single 3 base pair deletion in a protein coding region, 14 non-silent substitutions at 12 positions and 13 substitutions at 10 positions in tRNA genes.

The distribution of the 478 substitutions among the 351 variable sites is shown in Fig. 15. There are two ways by which observable multiple substitutions can occur in these silent positions: 1) parallel changes in separate lineages; 2) different changes in separate lineages. Other multiple changes which

FIGURE 14.

The nucleotide sequences for the six salmonid species compared for each of six genes: A) ATPase 6; B) CO III; C) tRNA^{GLY}; D) ND 3; E) tRNA^{ARG}; and F) ND 4L. The sequences are aligned assuming only a single length variation in pink salmon which is missing nucleotides 616-618 designated (---). The amino acids coded for by each triplet are given above the nucleic acid sequences following the salmon mitochondrial genetic code. The complete sequence for rainbow trout is shown on the first line. The nucleotides which differ are given for each of the other five species. Also included are the minimum number of substitutions (SUBS) which have occurred at each variable position assuming the parsimony relationship given in Fig. 22 (Section III). The class of site (A, B, C, D) at which each change occurs refers to A) first position of a leucine codon; B) silent in two-fold degenerate codons (C×T); C) silent in two-fold degenerate codons (A×G); D) silent change in a four-fold degenerate codon; S) non-silent changes.

ATPASE SUBUNIT SIX

A

Ser	Phe	Asp	Gln	Phe	MET	Ser	Pro	Thr	Tyr	Leu	Gly	Ile	Ala	Val	Gly	60
RAINBOW	AGC	TTC	GAC	CAA	TTT	ATG	AGC	CCC	ACA	TAC	CTA	CCA	CTT	ATC	GCC	GTA
PINK
COHO
SOCKEYE
CHINOOK
CUTTHROAT
CLASS	B
SUBS	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Leu	Thr	Leu	Pro	Trp	Ile	Leu	Phe	Pro	Thr	Pro	Ser	Ala	Arg	Trp	Leu	Asn	Asn	Arg	Leu	120
RAINBOW	TTA	ACC	CTC	CCA	TGA	ATT	CTT	TTC	CCT	ACC	CCC	TCT	GCC	TGA	CTA	AAC	AAC	CGC	CTA	120
PINK
COHO
SOCKEYE
CHINOOK
CUTTHROAT
CLASS	D	D	D	D	S	D	D	D	D	D	S	D	S	C	A	B	B	D	D	...
SUBS	1	2	1	1	1	2	3	1	1	1	1	1	1	2	1	1	1	1	1	1

Ile	Thr	Leu	Gln	Gly	Trp	Phe	Ile	Asn	Arg	Phe	Thr	Gln	Gln	Leu	Leu	Pro	Leu	Asn	180	
RAINBOW	ATT	ACC	CTG	CAA	GGG	TGG	TTC	ATC	AAC	CGA	TTT	ACC	CAG	CAA	CTT	TTA	CCG	CTA	AAT	180
PINK
COHO
SOCKEYE
CHINOOK
CUTTHROAT
CLASS	B	D	D	D	D	C	B	D	D	D	S	C	C	D	A	D	D	D	D	...
SUBS	2	1	4	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1

ATPASE SUBUNIT SIX

A (cont.)

Leu	Gly	Gly	His	Lys	Trp	Ala	Ala	Ala	Leu	Leu	Thr	Ser	Leu	MET	Leu	Phe	Leu	Ile	Thr	Leu	240
RAINBOW	CTA	GGC	GGT	CAC	AAG	TGA	GCA	GCT	CTA	CTA	ACT	TCC	CTC	ATA	CTA	TTT	CTT	ATT	ACC	CTA	
PINK	T..	...C	...A	...A	...G	...G	...C	...C	...G	...G	...C	..T	..A	...A	...C	...C	...C	...C	...C	...C	
COHO	T..	...C	...A	...A	...G	...G	...C	...C	...G	...G	...C	..C	..A	...A	...G	...C	...C	...C	...C	...C	
SOCKEYE	T..	...C	...A	...A	...G	...G	...C	...C	...G	...G	...C	..C	..A	...A	...G	...C	...C	...C	...C	...C	
CHINOOK	T..	...C	...A	...A	...G	...G	...C	...C	...G	...G	...C	..C	..A	...A	...G	...C	...C	...C	...C	...C	
CUTTHROAT	T..G	...C	...A	...A	...G	...G	...C	...C	...G	...G	...C	..C	..A	...A	...G	...C	...C	...C	...C	...C	
CLASS	A	D	D	C	C	C	D	D	A	D	D	D	D	D	D	D	D	D	D	D	
SUBS	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	

Asn	MET	Leu	Gly	Leu	Leu	Pro	Tyr	Thr	Thr	Phe	Thr	Pro	Thr	Thr	Gln	Leu	Ser	Leu	Asn	MET	300
RAINBOW	AAT	ATA	CTT	GCC	CTA	CTT	CCA	TAT	ACA	TTC	ACC	CCG	ACC	ACA	CAG	CTC	TCC	CTA	AAT	ATG	
PINK	..C	..G	..A	..G	..G	..G	..G	..G	..G	..G	..G	..C	..C	..G	..A	..A	..A	..A	..C	..A	
COHO	...C	...A	...A	...G	...G	...G	...G	...G	...G	...T	...T	...C	...C	...G	...A	...A	...A	...A	...C	...A	
SOCKEYE	...C	...A	...A	...G	...G	...G	...G	...G	...G	...T	...T	...C	...C	...G	...A	...A	...A	...A	...C	...A	
CHINOOK	...C	...A	...A	...G	...G	...G	...G	...G	...G	...T	...T	...C	...C	...G	...A	...A	...A	...A	...C	...A	
CUTTHROAT	...C	...A	...A	...G	...G	...G	...G	...G	...G	...T	...T	...C	...C	...G	...A	...A	...A	...A	...C	...A	
CLASS	B	C	D	D	D	D	D	D	B	B	D	D	D	D	C	D	D	D	B	C	
SUBS	1	1	2	1	1	1	1	1	2	2	2	3	2	2	2	1	2	1	2	1	

Gly	Leu	Ala	Val	Pro	Leu	Trp	Leu	Ala	Thr	Val	Ile	Ile	Gly	MET	Arg	Asn	Gln	Pro	Thr	360	
RAINBOW	GGC	CTC	GCA	GTC	CCA	CTG	CCA	CTT	GCT	ACA	GTA	ATT	ATC	GGC	ATA	CGA	AAC	CAA	CCT	ACG	
PINK	..T	...C	...C	...G	...G	...A	...A	...C	...C	...C	...C	...C	...C	...C	...C	...C	...C	...C	...C	...C	
COHO	..T	...C	...C	...G	...G	...A	...A	...C	...C	...C	...C	...C	...C	...C	...C	...C	...C	...C	...C	...C	
SOCKEYE	..T	...C	...C	...G	...G	...A	...A	...C	...C	...C	...C	...C	...C	...C	...C	...C	...C	...C	...C	...C	
CHINOOK	..T	...C	...C	...G	...G	...A	...A	...C	...C	...C	...C	...C	...C	...C	...C	...C	...C	...C	...C	...C	
CUTTHROAT	..T	...C	...C	...G	...G	...A	...A	...C	...C	...C	...C	...C	...C	...C	...C	...C	...C	...C	...C	...C	
CLASS	D	A	D	D	A	D	C	D	D	D	D	B	B	B	B	C	C	D	D	D	
SUBS	1	1	1	2	2	2	2	2	2	2	2	1	1	1	2	2	1	1	1	2	

ATPASE SUBUNIT SIX

A (cont.)

Ala	Ala	Leu	Gly	His	Leu	Leu	Pro	Glu	Gly	Thr	Pro	Val	Pro	Leu	Ile	Pro	Val	Leu	Ile
RAINBOW	GCC	CTC	GGC	CAT	TTA	TTG	CCT	GAA	GGA	ACC	CCC	GTT	CCA	CTG	ATC	CCA	GTA	CTG	ATC
PINKA	..T	..C	C.CC	..G	..ATT
COHOT	C.CC	..G	..AG	..G	T.A	...
SOCKEYEAC	C.CC	..C	..ATA	..T
CHINOOKT	C.CC	..C	..AA	...
CUTTHROATA	..T	...	C..	C.AC	..C	..ATA	...
CLASS	D	D	D	B	A	D	A	D	D	D	D	D	D	D	D	D	D	A	D
SUBS	1	2	2	1	1	1	1	1	1	3	1	1	3	1	1	3	1	1	2

Ile	Ile	Glu	Thr	Ile	Ser	Leu	Phe	Ile	Arg	Pro	Ala	Leu	Gly	Val	Arg	Leu	Thr	Ala	Asn
RAINBOW	ATT	ATC	ACA	ATT	AGC	CTT	TTT	ATC	CGC	CCC	GCC	CTT	GGC	GTA	CGA	CTT	ACA	GCC	AAT
PINK
COHO
SOCKEYE
CHINOOKCC
CUTTHROAT
CLASS	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
SUBS	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Leu	Thr	Ala	Gly	His	Gln	Leu	Ile	Ala	Thr	Ala	Ala	Phe	Val	Leu	Leu	Pro	MET	MET	Pro
RAINBOW	CTC	ACA	GCA	GGC	CAC	CAA	CTA	ATT	GCT	ACA	GCA	GCC	TTT	GTT	CTT	CTA	CCT	ATA	ATA
PINKC	..CC	..C	T..	...	T.G	...
COHO	..A	..G	..GC	..CC	..GC
SOCKEYEG	..AC	..CC	T..C
CHINOOKTA	T..C
CUTTHROATC	..C	..C	..CC
CLASS	D	D	D	D	B	D	D	B	D	B	D	D	D	D	D	D	D	S	C
SUBS	1	2	2	2	2	1	1	2	1	1	1	1	1	3	2	1	1	2	1

B

MET	Ala	His	Gln	Ala	His	Ala	Tyr	His	MET	Val	Asp	Pro	Ser	Pro	Trp	Pro	Leu	Thr	Gly	60
RAINBOW	ATG	GCA	CAC	CAA	GCA	GCA	TAC	CAC	ATG	GTT	GAC	CCA	AGC	CCC	TGA	CCT	CTG	ACC	GGC	
PINK	
COHO	
SOCKEYEA	
CHINOOK	
CUTTHROAT	
CLASS																				D
SUBS																				1

Ala	Ile	Ala	Ala	Leu	Leu	Leu	Thr	Ser	Gly	Thr	Ala	Val	Trp	Phe	His	Phe	His	Ser	Leu	120
RAINBOW	GCA	ATT	GCC	GCC	CTT	TTA	CTT	ACA	TCA	GGC	ACT	GCA	GTC	TGA	TTC	CAT	TTC	CAC	TCG	CTC
PINKC	..GT	..C	
COHO	
SOCKEYEGCA	
CHINOOKCA	
CUTTHROATA	..T
CLASS						D	D							B	B				D	D
SUBS						1	1							1	2				3	1

Thr	Leu	Leu	Leu	Thr	Ala	Ala	Ile	Leu	Leu	Leu	Leu	Thr	MET	Tyr	Gln	Trp	Trp	Arg	Asp	180
RAINBOW	ACA	CTT	..A	..A	..C	..C	..G	..T	..G	..T	..A	..C	..C	..T	..T	..C	..G	..G	..C	
PINK	..T	..A	..A	..A	..C	..G	..T	..G	..T	..C	..C	..C	..C	..T	..T	..C	..G	..G	..C	
COHOA	..CC	..G	..T	..G	..T	..C	..C	..C	..C	..T	..T	..C	..G	..G	..C	
SOCKEYEA	..CC	..G	..T	..G	..T	..C	..C	..C	..C	..T	..T	..C	..G	..G	..C	
CHINOOK	..G	T.AC	..G	..T	..G	..T	..C	..C	..C	..C	..T	..T	..C	..G	..G	..C	
CUTTHROATAC	..G	..T	..G	..T	..C	..C	..C	..C	..T	..T	..C	..G	..G	..C	
CLASS	D	A	D	D	A	D	B	S	D	A				C	B	C	C	C	D	B
SUBS	2	1	1	2	1	1	1	1	3	1			2	1	1	1	1	3	2	

CYTOCHROME OXIDASE SUBUNIT 3

B (cont.)

Ile	Ile	Arg	Glu	Gly	Gln	Phe	Thr	Phe	Gln	Gly	His	His	Thr	Pro	Pro	Val	Gln	Lys	Gly	Leu	Arg
ATC	ATC	CGA	GAA	GGT	ACC	TTT	CAA	GGA	GGA	CAC	CAC	CAC	ACG	CCC	CCA	GTC	CAA	AAA	GGG	CTA	CGA
240																					
PINK	..TGGAG	..T	T..	...
COHO	..TCGAAAA	..AG	T..	...
SOCKEYEA
CHINOOK	..TCGAAAA
CUTTHROATGAAT	..GA
CLASS	B	C	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	C	D	D	A
SUBS	2	1	1	2	2	1	1	2	2	2	1	1	2	2	1	1	1	1	2	2	2

Tyr	Gly	MET	Ile	Leu	Phe	Ile	Thr	Ser	Glu	Val	Phe	Phe	Phe	Phe	Leu	Gly	Phe	Phe	Trp	Ala	
TAT	GGC	ATA	ATC	TTA	TTT	ATT	ACC	TCC	GAG	GTA	TTC	TTT	TTC	TTC	TTA	GGT	TTC	TTC	TGA	GCC	
300																					
PINKTT	..A	C.C	..C	
COHO	..CT	..A	C..	
SOCKEYE	..CT	..A	C.T	
CHINOOK	..CA	C.G	
CUTTHROAT	..CA	C..	
CLASS	B	B	D	D	D	D	D	D	C	C	D	D	D	D	A	D	D	D	D	D	D
SUBS	2	1	1	2	1	2	1	1	2	2	1	3	1	1	1	3	1	1	1	1	1

Phe	Tyr	His	Ala	Ser	Leu	Ala	Pro	Thr	Pro	Glu	Leu	Gly	Gly	Gly	Cys	Trp	Pro	Pro	Ala	Gly	
TTC	TAC	CAC	GCC	AGC	CTC	GCC	CCC	ACA	CCT	GAA	TTA	GGA	GGT	GGT	TGC	TGA	CCC	CCC	GCA	GGT	
360																					
PINKTA	..TCC	A..	..C	
COHOTAG	C..	A..	..C	
SOCKEYETAG	C..	A..	..C	
CHINOOKTA	..TG	C..	A..	..C	
CUTTHROATTA	..TG	C..	A..	..C	
CLASS	D	D	D	D	D	D	D	D	D	C	C	D	D	D	D	D	D	D	S	D	D
SUBS	1	1	1	2	2	2	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1

B (cont..)

Ile	Thr	Thr	Leu	Asp	Pro	Phe	Glu	Val	Pro	Leu	Leu	Asn	Thr	Ala	Val	Leu	Leu	Ala	Ser	420
RAINBOW	ATT	ACT	CTA	GAC	CCC	TTT	GAG	GTA	CCC	CTT	CTT	AAT	ACT	GCA	GTC	CTT	CTA	GCA	TCT	
PINK	T...C	..AG	..T	
COHOA	..C	
SOCKEYEGG	T...	
CHINOOK	..CA	..C	
CUTTHROATGCG	
CLASS	B	A	C	D	D	B	...	D	D	...	A	
SUBS	1	1	1	3	1	1	...	2	1	...	1	

Gly	Val	Thr	Val	Thr	Trp	Ala	His	His	Ser	Ile	MET	Glu	Gly	Glu	Arg	Lys	Gln	Thr	Ile	480
RAINBOW	GGT	GTC	ACC	GTA	ACA	TGA	GCC	CAC	AGC	ATC	ATA	GAA	GGT	GAA	CGA	AAA	CAA	ACC	ATT	
PINKTTG	..GT	...	
COHOTC	
SOCKEYETG	
CHINOOKTTCG	
CUTTHROATTC	..C	
CLASS	D	D	B	B	C	D	C	D	...	D	
SUBS	2	3	1	1	1	3	1	1	...	1	

Gln	Ala	Leu	Thr	Leu	Thr	Ile	Leu	Gly	Phe	Tyr	Phe	Thr	Thr	Phe	Leu	Gln	Gly	MET	Glu	540
RAINBOW	CAA	GCT	CTT	ACT	CTC	ACT	ATC	TTA	CTG	GGA	TTT	TAC	TTC	ACT	TTC	CTA	GGT	ATA	GAA	
PINKC	..CT	..CCG	..C	
COHOCTACCCG	
SOCKEYETG	..CG	
CHINOOKTAC	
CUTTHROATT	
CLASS	D	D	D	D	D	B	B	D	D	B	D	...	D	C	D	...	C	
SUBS	1	1	1	1	1	3	1	1	1	1	1	...	1	1	2	

B (cont.)

Tyr	Tyr	Glu	Ala	Pro	Phe	Thr	Ile	Ala	Asp	Gly	Val	Tyr	Gly	Ser	Thr	Phe	Val	Ala	600
RAINBOW	TAC	GAA	GCC	CCA	TTT	ACA	ATC	GCT	GAC	GGC	GTA	TAC	GGC	TCT	ACT	TTT	GTC	GCT	
PINKA	..CA	..C	
COHOGC	
SOCKEYET	..C	
CHINOOKGC	
CUTTHROAT	
CLASS								D			D	D	D	D			D	D	
SUBS								1			1	1	1	1			2	1	

Thr	Gly	Phe	His	Gly	Leu	His	Val	Ile	Ile	Gly	Ser	Thr	Phe	Leu	Ala	Val	Cys	Leu	660
RAINBOW	ACA	GGA	TTC	CAT	GGC	CTA	CAC	GTA	ATT	GGC	TCT	ACC	TTT	CTG	GCC	GTT	TGC	CTA	
PINKC	..CC	..C	T..C	
COHOC	..CC	..A	
SOCKEYE	..GC	..A	..A	
CHINOOK	..CGA	..AA	..AG	
CUTTHROAT	
CLASS	D				A	D	B	B	D	D	B	A	D	D	D			D	
SUBS	2				1	1	1	2	1	1	1	2	1	1	1			1	

Arg	Gln	Val	Gln	Tyr	His	Phe	Thr	Ser	Glu	His	His	Phe	Gly	Phe	Glu	Ala	Ala	Ala	720
RAINBOW	CGA	CAA	GTT	CAA	TAC	CAC	TTT	ACA	TCT	GAA	CAT	CAT	TTT	GGC	TTT	GAA	GCT	GCC	
PINKACGCT	
COHOCT	
SOCKEYEACT	
CHINOOK	..G	..CCT	
CUTTHROATT	
CLASS	C	D				B	D	D	C	D	D	D	B	A	D	D		D	
SUBS	1	1	1			1	1	1	2	1	1	1	1	1	1			1	

B (cont.)

Tyr	Trp	His	Phe	Val	Asp	Val	Val	Val	Trp	Leu	Phe	Leu	Tyr	Val	Ser	Ile	Tyr	Trp	Trp	Gly
RAINBOW	TAT	TGA	TTT	GTA	GAC	GTT	GTA	GTA	TGG	CTC	TTC	CTA	TAC	GTC	TCT	ATT	TAC	TGA	TGA	GGC
PINKGA
COHOA
SOCKEYEGA
CHINOOKGA
CUTTHROATA
CLASS							D		C											
SUBS							2		1											780

Ser	**																			
RAINBOW	TCA	TAA																		786
PINK
COHO
SOCKEYE
CHINOOK
CUTTHROAT
CLASS																				
SUBS																				

trNA GLY

C

RAINBOW ATCTTTCTAGTATTAAATACGTATAAGTGACTTCCAAATCACCCGGTCTTGGTTAAAAATCCAAGGAAAAGATA 70
PINKTA...CG.....T.....
COHOCG.....
SOCKEYEC...C...GA.....
CHINOOKC.....
CUTTHROATC.....
SUBS1..231...11.....

NADH DEHYDROGENASE SUBUNIT 3

D

MET	Asn	Leu	Ile	Thr	Thr	Ile	Ile	Thr	Ile	Thr	Ile	Thr	Leu	Ser	Ala	Val	Leu	Ala	Thr	60
RAINBOW	ATG	AAT	TTA	ATC	ACA	ATC	ATC	ACT	ATT	ACC	ATC	ACA	TTA	TCC	GCA	GTA	CTA	GCC	ACT	
PINK	..C	..C	..T	..G	..T	..T	..T	..T	..C	..T	..T	..C	..G	..T	..T	..T	..T	..T	..T	
COHO	..C	..C	..T	..T	..T	..T	..T	..T	..C	..T	..T	..C	..G	..T	..T	..T	..T	..T	..T	
SOCKEYE	..C	..C	..T	..T	..T	..T	..T	..T	..C	..T	..T	..C	..G	..T	..T	..T	..T	..T	..T	
CHINOOK	..C	..C	..T	..T	..T	..T	..T	..T	..C	..T	..T	..C	..G	..T	..T	..T	..T	..T	..T	
CUTTHROAT	..C	..C	..T	..T	..T	..T	..T	..T	..C	..T	..T	..C	..G	..T	..T	..T	..T	..T	..T	
CLASS	B	A	S	B	D	B	B	B	B	B	B	A	D	D						
SUBS	1	1	1	1	1	1	1	1	2	1	1	2	1	1	1	1	1	1	1	

Ile	Ser	Phe	Trp	Leu	Pro	Gln	Ile	Ser	Pro	Asp	Ala	Glu	Lys	Leu	Ser	Pro	Tyr	Glu	Cys	120
RAINBOW	ATT	TCT	TGA	TTA	CCA	CAA	ATC	TCC	CCA	GAC	GCA	GAG	AAG	TTA	TCC	CCC	TAC	GAA	TGT	
PINK	..C	..C	..C	..G	..C	..C	..T	..C	..C	..C	..C	..A	..A	..C	..C	..C	..C	..C	..C	
COHO	..G	..C	..C	..C	..C	..C	..T	..C	..C	..C	..C	..A	..A	..C	..C	..C	..C	..C	..C	
SOCKEYE	..C	..C	..C	..C	..C	..C	..T	..C	..C	..C	..C	..A	..A	..C	..C	..C	..C	..C	..C	
CHINOOK	..C	..C	..C	..C	..C	..C	..T	..C	..C	..C	..C	..A	..A	..C	..C	..C	..C	..C	..C	
CUTTHROAT	..C	..C	..C	..C	..C	..C	..T	..C	..C	..C	..C	..A	..A	..C	..C	..C	..C	..C	..C	
CLASS	S	B		A	D	D	B	D	D	D	C	C	A	D						
SUBS	1	1	2	1	1	1	1	2	2	1	1	1	1	1	1	1	1	1	1	

Gly	Phe	Asp	Pro	Leu	Gly	Ser	Ala	Arg	Leu	Pro	Phe	Ser	Leu	Arg	Phe	Phe	Leu	Ile	Ala	180
RAINBOW	GGA	TTT	GAC	CCC	TTA	GGG	TCC	GCC	CGC	CTG	CCC	TTC	TCC	TTA	CGC	TTT	CTA	ATC	GCC	
PINK	..C	..C	..T	..A	..G	..T	..T	..T	..T	..A	..C	..T	..T	..A	..C	..C	..C	..C	..C	
COHO	..C	..C	..T	..A	..G	..T	..T	..T	..T	..A	..C	..T	..T	..A	..C	..C	..C	..C	..C	
SOCKEYE	..C	..C	..T	..A	..G	..T	..T	..T	..T	..A	..C	..T	..T	..A	..C	..C	..C	..C	..C	
CHINOOK	..C	..C	..T	..A	..G	..T	..T	..T	..T	..A	..C	..T	..T	..A	..C	..C	..C	..C	..C	
CUTTHROAT	..C	..C	..T	..A	..G	..T	..T	..T	..T	..A	..C	..T	..T	..A	..C	..C	..C	..C	..C	
CLASS	B	B	D	A	D	D	D	D	D	D	D	D	D	D	B	B				
SUBS	1	1	1	1	1	2	1	1	1	2	1	1	1	1	1	1	1	1	1	

NADH DEHYDROGENASE SUBUNIT 3

D (cont.)

Ile	Leu	Phe	Leu	Phe	Asp	Leu	Glu	Ile	Ala	Leu	Leu	Leu	Pro	Leu	Pro	Trp	Gly	Asp	240
RAINBOW	ATC	TTC	CTC	CTA	TTT	GAT	CTA	GAA	ATC	GCC	CTC	CTT	TTG	CCC	CTA	CCT	TGA	GGG	GAT
PINK
COHO
SOCKEYE
CHINOOK
CUTHROAT
CLASS	A	D	D	D	B	B	B	B	B	D	D	D	A	D	D	D	C	D	
SUBS	1	2	1	1	1	2	1	2	2	1	1	1	3	1	3	2	1	1	1

Gln	Leu	His	Thr	Pro	Thr	Leu	Thr	Leu	Ile	Trp	Ser	Thr	Ala	Val	Leu	Ala	Leu	Leu	Thr	300
RAINBOW	CAA	CTC	CAC	ACC	CCG	ACC	CTG	ACA	CTC	TGA	TCC	ACT	GCC	GTT	CTA	GCC	CTT	CTT	ACT	
PINK	
COHO	
SOCKEYE	
CHINOOK	
CUTHROAT	
CLASS	C	S	B	S	D	D	A	D	D	S	B	C	D	D	D	D	C	D		
SUBS	1	2	1	1	3	1	1	1	3	2	2	1	1	1	1	1	1	1	1	

Leu	Gly	Leu	Ile	Tyr	Glu	Trp	Thr	Gln	Gly	Gly	Leu	Glu	Trp	Ala	Glu	***				
RAINBOW	CCT	GGC	TTA	ATC	TAT	GAA	TGA	ACC	CAA	GGA	GGC	TTA	GAA	TGA	GCC	GAG	TAA			351
PINK
COHO
SOCKEYE
CHINOOK
CUTHROAT
CLASS	D	A	B	B	C	C	D	A	D	D	S	B	C	D	D	D	C	C		
SUBS	1	1	1	1	1	1	1	1	1	3	2	2	1	1	1	1	1	1	1	1

tRNA ARG

E

RAINBOW CCGAGTTAGTCCAAAACAAGACCCTTGATTTTCGGCTCAAAAAGACCATGGTTTAAAGTCCATGACCGCCTT 69

PINK
COHO
SOCKEYET.....
CHINOOK
CUTTHROATA.....
SUBS1.....

NADH DEHYDROGENASE SUBUNIT 4L

F

MET	Thr	Pro	Val	His	Phe	Ser	Phe	Thr	Ser	Ala	Phe	Ile	Leu	Gly	Leu	MET	Gly	Leu	Ala	60
RAINBOW	ATG	ACA	CCA	GTA	CAC	TTC	AGC	TTT	ACC	TCA	GCC	TTT	ATT	TTA	GGG	CTT	ATA	GGA	CTC	GCG
PINK	C..
COHO	C..C
SOCKEYET	C..
CHINOOK	C..C
CUTTHROATT	C..
CLASS	D				B								A		D					
SUBS	1				1								1		1					

Phe	His	Arg	Thr	His	Leu	Leu	Leu	Ser	Ala	Leu	Leu	Cys	Leu	Glu	Gly	MET	MET	Leu	Ser	Leu	120
RAINBOW	TTT	CAC	CGC	ACC	CAC	CAC	CTT	CTC	TCA	GCC	CTT	CTA	TGC	CTA	GAA	GGA	ATA	ATA	CTC	TCT	CTA
PINKTTG
COHOTTG
SOCKEYEG
CHINOOK
CUTTHROATT
CLASS	B				B			D			D					C	C				
SUBS	3				1			1			1					1	1				

Phe	Ile	Ala	Leu	Ser	Leu	Trp	TGA	Trp	Ala	Leu	Gln	MET	Glu	Ala	Thr	Gly	Tyr	Ser	Val	Ala	Pro	180
RAINBOW	TTC	ATC	GCC	CTC	TCC	CTC	TGA	CTC	GCC	CTC	CAA	ATG	GAA	GCG	ACT	GGC	TAC	TCA	GTG	GCC	CCG	
PINK	..TTTAAT	..T	
COHO	..T	..TGAATT	..T	
SOCKEYE	..T	..TT.AGAACT	..T	
CHINOOK	..T	..TT.ATTAATT	..T	
CUTTHROAT	..T	..TTTTAATT	..T	
CLASS	B	B			A	D			D	D	C	C		D		D		D		D	D	
SUBS	1	2			2	3			1	2	1	1		1		1		1		1	1	

NADH DEHYDROGENASE SUBUNIT 4L

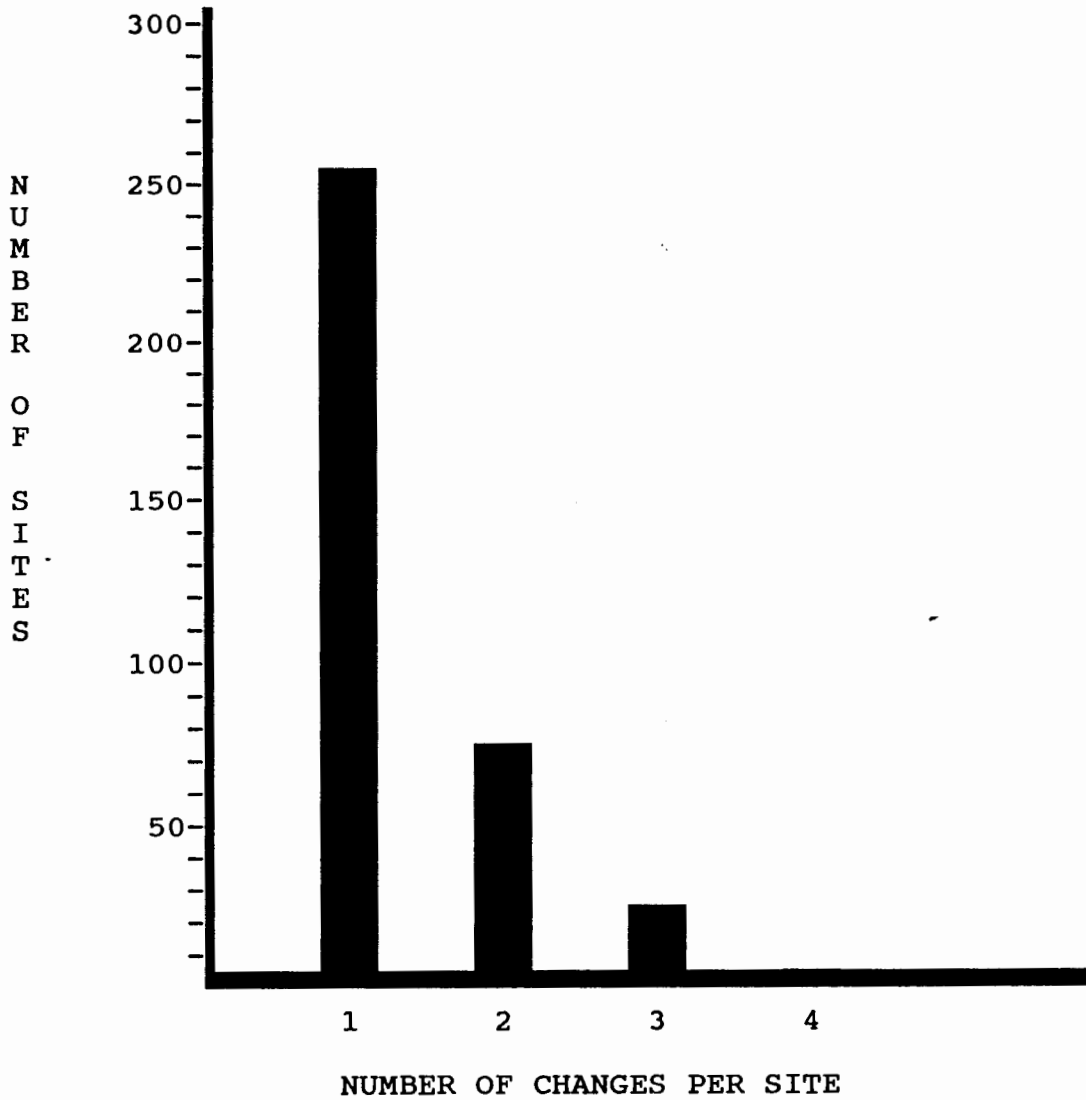
F (cont.)

MET	Leu	Leu	Ala	Phe	Ser	Ala	Cys	Glu	Ala	Ser	Ala	Gly	Leu	Ala	Leu	Val	Ala	240
RAINBOW	ATA	CTT	CTA	GCG	TTC	TCA	GCC	TGT	GAA	GCC	AGC	GCA	GGG	TTA	GCC	GTA	GCA	
PINK	..GAA	
COHO	..CAT	
SOCKEYEAT	
CHINOOKGTT	
CUTTHROATT	..C	..TGTTG
CLASS	C	D	D	B	B	C						D	D	D	D			
SUBS	1	1	3	1	2	2						2	1	1	1			1

Thr	Ala	Arg	Thr	His	Gly	Thr	Asp	Arg	Leu	Gln	Ser	--
RAINBOW	ACT	GCA	ACA	CAC	GGC	ACA	GAC	CGC	CTC	CAA	AGC	TT
PINK	..CTG	..T
COHO	..CTG	..TA
SOCKEYE	..C	..G	..T	..TT	..T
CHINOOK	..C
CUTTHROATG	..T
CLASS	D	D	B	B	D	B	B	D	D			
SUBS	1	1	3	1	1	1	1	1	1			1

FIGURE 15.

This graph shows the number of sites with one, two, three, and four changes per variable site. The number of changes per site is derived from Fig. 14.



involve back and forth or multiple changes in a single lineage are not observable. Most of the multiple changes among the salmonid species are of the parallel type. This is consistent with the substitution bias and the fact that parallel changes can occur at all silent positions whereas different substitutions in separate lineages must involve at least one transversion and therefore can only occur in the third position of four codon families.

The comparison of mitochondrial DNA sequences eliminates some of the problems implicit in the use of restriction enzymes to estimate levels of divergence (Wilson et al., 1985). The problem of multiple substitutions per site has been addressed (Brown et al., 1982; Cann et al., 1984; Templeton, 1983). The 100 sites with multiple substitutions in the 351 variable sites found among the salmon sequences demonstrates the problem of interpreting restriction enzyme data among even moderately divergent groups (Fig. 15).

DISTRIBUTION OF VARIATION AMONG GENES

I have examined the variation observed in each functional sequence. Figure 16 gives the number of observed variable sites and the total observed variation as a percentage of the sequence length. It is clear that both the ATPase 6 and ND 3 coding sequences have higher levels of variation than the CO III and ND 4L genes. The pattern of observed variation is consistent with the number of potential variable sites for each gene. The

FIGURE 16.

The potential and observed variation is divided into each of the four possible site types of silent changes as described in Fig. 14. The results are presented for each of the protein coding sequences separately. The tRNA sequences are not included. The potential for variation (upper panel), is the total number of silent positions in the fragment or positions at which substitutions do not change the amino acid sequences of the proposed proteins. The observed variable sites (lower panel), are the silent changes that are observed among the six species. These do not include multiple changes per site.

CLASS**GENE****POTENTIAL VARIABLE SITES**

	ATPase 6	CO III	ND 3	ND 4L
1st Pos. Leu (C<=>T)	30	19	17	8
3rd Pos. (C<=>T)	51	77	28	19
3rd Pos. (A<=>G)	39	46	23	14
3rd Pos. (C,T,A,G)	135 ---	140 ---	66 ---	59 ---
Totals	255	282	134	100
	38.5	35.9	38.2	36.0
			Total =	771

OBSERVED VARIABLE SITES

1st Pos. Leu (C<=>T)	15	10	9	2
3rd Pos. (C<=>T)	20	18	16	7
3rd Pos. (A<=>G)	17	16	8	4
3rd Pos. (C,T,A,G)	75 ---	62 ---	35 ---	24 ---
Totals	127	106	68	39
	19.1	13.5	19.4	14.0
			Total =	340

distribution of variation is not related to functional constraints on the proteins coded for by the sequences because the variation being examined is silent. Differences in the number of potential variable sites among genes is due to different frequencies of leucine codon in each gene. Although the observed and potential variation correlate, the differences in observed variation among genes is not limited to variation incurred in the first position of leucine codons. One possible explanation could be that there are differences in potential variability due to different levels of four-fold degenerate codons. The numbers of four-fold degenerate codons do not differ among the genes. It should also be noted that the differences in observed variation are reflected in all site types (Fig. 16). Differences in variation have been found among primate protein coding genes (Brown et al., 1982) and among humans (Cann et al., 1984). Among the human genes, Cann et al. found the ATPase 6 gene to have the highest number of variable restriction sites. With the exception of the ND 4L sequence for which there was insufficient data, the relative levels of restriction site variability among the human genes is the same as that found in our comparison of the salmonid mitochondrial sequences.

INTRAGENIC VARIATION

In order to investigate the distribution of variation more closely I have determined the distribution of variation within the protein coding sequences. I have divided the protein coding

sequences into 50 base pair blocks and determined the number of variable sites (Fig. 17) and total number of substitutions (Fig. 18) within each block. Because of the inequity of the constraints on the tRNA and protein coding genes I have not included the tRNA genes in these graphs. The tRNA genes are examined separately below.

It is obvious from the graph that the distribution of variation in these sequences is complex and highly variable. Our division of the sequence into 50 base pair blocks in fact eliminates some of the complexity. These graphs show that the distribution of variation among genes is the result of conserved and/or highly variable regions within genes. In order to interpret the variation I have included an analogous examination of the number of potentially variable sites, Fig. 19. The patterns of potential variation are similar to the observed pattern but are not nearly as dramatic. As pointed out for the comparisons between genes above, the number of potentially variable sites does not account for the differences in variability within the genes. The most prominent pattern of observed variation is the reduction of variation at the junctions of genes. Closer examination shows that this is a result of reduced variation in the 5' ends of ATPase 6, CO III, and ND 4L, and the 3' end of ND 3. Because the variation is silent, explanation for differences in variation are limited to two general categories. One possibility is that the non-random distribution of amino acid codons leads to a non-random

FIGURE 17.

Observed number of variable positions within the protein coding regions of the cloned fragment. For each block of fifty base pairs the ordinate gives the total number of variable positions within that block. The base pair positions are numbers such that the first position is the first base in Fig. 3. The punctuating tRNA sequences are not included. A functional map of the sequence is displayed below the graph.

DISTRIBUTION OF VARIABLE SITES

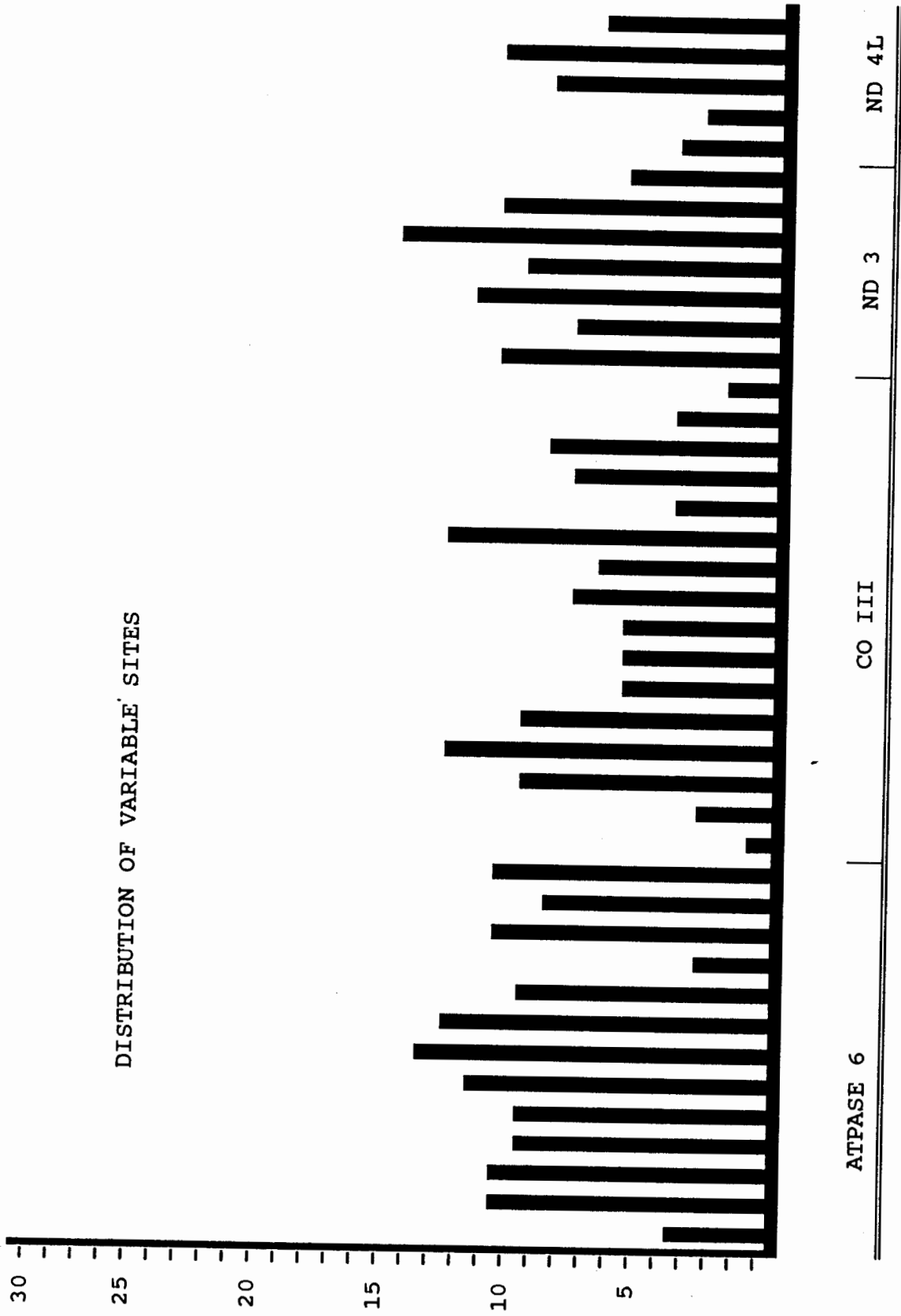


FIGURE 18.

Observed variation within the protein coding sections of the cloned fragment. For each block of fifty base pairs the ordinate gives the total variation within that block. The total variation includes the number of changes as given in Fig. 14. Numbering is as in Fig 17.

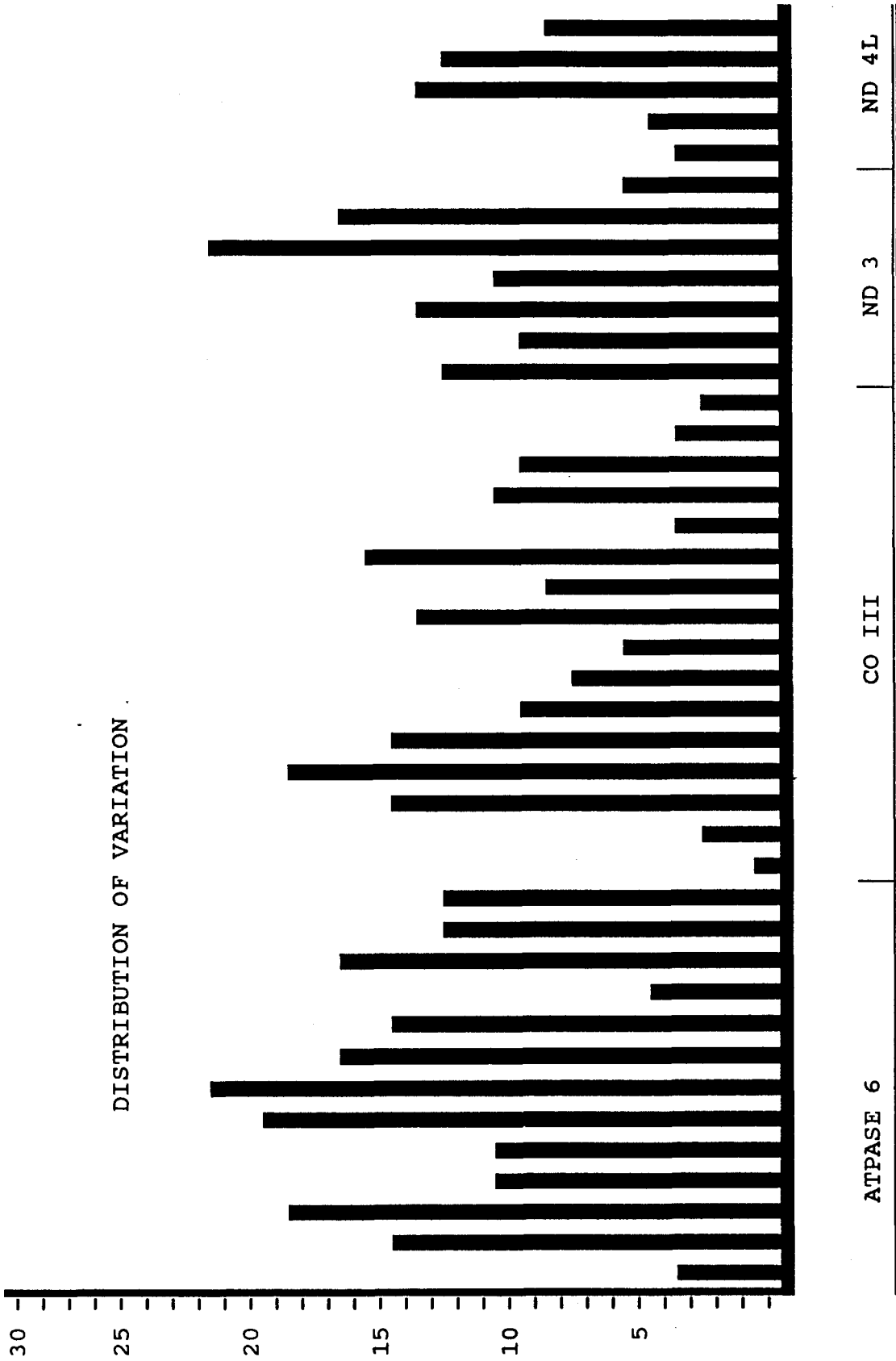
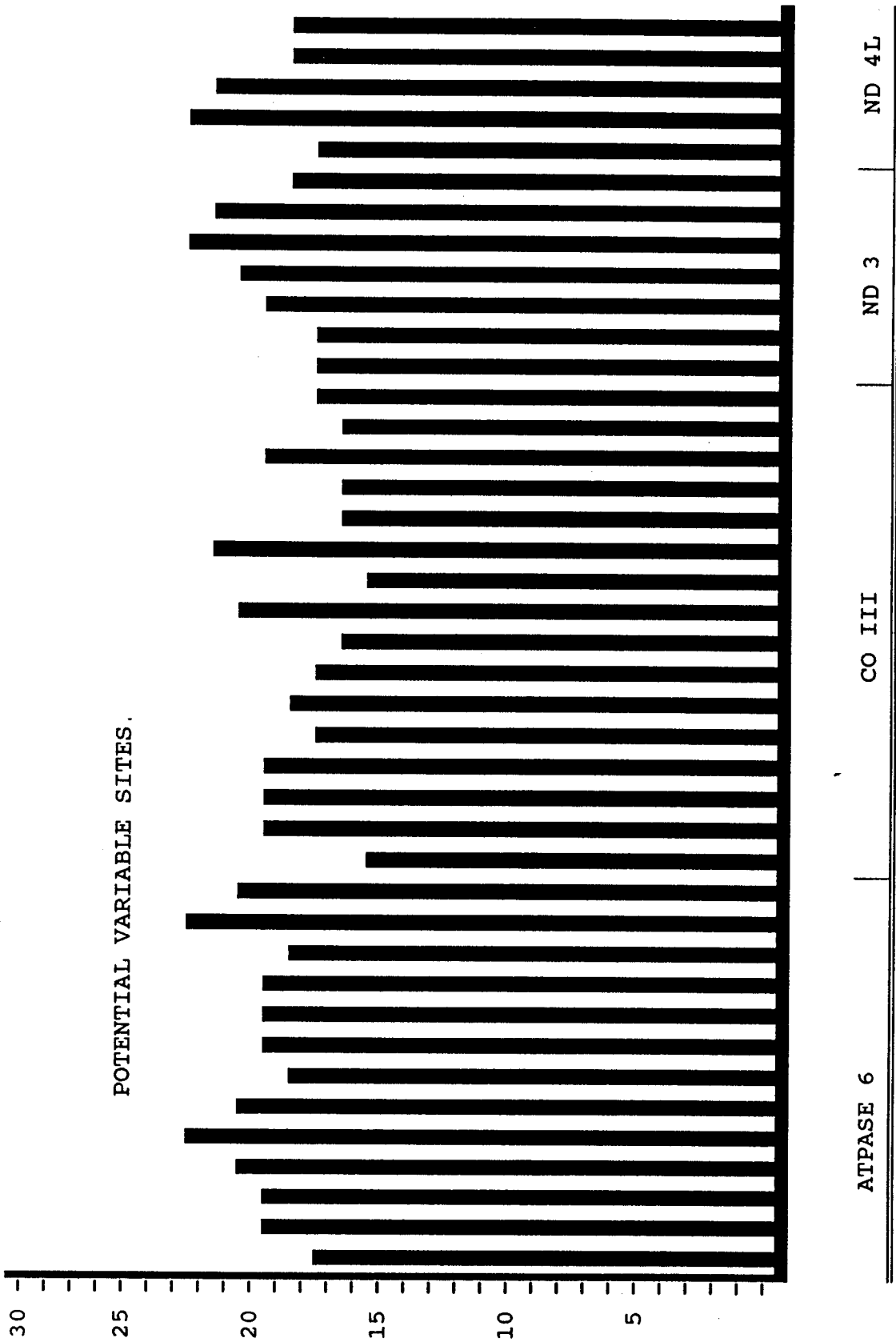


FIGURE 19.

The number of potential variable sites within the protein coding sections of the cloned fragment. For each block of fifty base pairs the ordinate gives the total number of positions at which variation can take place without changing the amino acid sequence. Numbering is as in Fig. 17.

FIG 19



distribution of the nucleotide composition which in turn may have different potentials for mutation. An alternative suggestion is that the nucleic acid sequences serve not only a protein coding function but functions related to the replication, transcription, processing and/or translation of the mitochondrial genome.

Below, I examine the distribution of variation among the amino acid codons and the potential for alternative functions of the mitochondrial sequence.

VARIATION AND AMINO ACIDS

In an attempt to associate the distribution of variation and amino acid composition I have determined the codon frequencies for variant and invariant codons. These are given in Fig 20. The codon frequencies are divided into six-, four- and two-fold degenerate codons. The two-fold degenerate families show a lower aggregate ratio of variable to conserved codons than the four-fold degenerate codons. This is consistent with the increased potential for variation by transversions in the four-fold degenerate codons. It is also true that the variation within groups is large. The ratio of variable to conserved sites in the leucine codons is consistent with the two variable sites found in the only six-fold degenerate codons when compared to the two- and four-fold degenerate codons. Taken individually the variation in some amino acids is higher than in others but, with the exception of the leucine codons we have not been able to correlate distribution of variability among amino acids and the

FIGURE 20.

The amino acid codons represented in Fig. 14 are divided into conserved and variable codons. The codons are separated into the six member family (LEU), the four member families, and the two member families. For each family the total variable codons and the total conserved codons are given. Also included is the ratios of variable to conserved codons for each amino acid codon and for the total conserved and variable codons for each of the three families.

AMINO ACID	# CONSERVED	# VARIABLE	RATIO V/C
		SIX CODON FAMILY	
LEU	44	86	1.95
		FOUR CODON FAMILIES	
VAL	16	19	1.19
SER	15	11	0.73
PRO	16	23	1.44
THR	41	23	0.56
ALA	38	26	0.68
ARG	10	6	0.60
GLY	24	21	0.88
TOTAL	160	129	0.81
		TWO CODON FAMILIES	
PHE	34	15	0.44
ILE	19	28	1.47
MET	12	8	0.67
TYR	14	5	0.35
HIS	17	7	0.41
GLN	13	10	0.77
ASN	5	7	1.40
LYS	1	3	3.00
ASP	7	4	0.57
GLU	10	13	0.77
CYS	5	0	0.00
TRP	12	10	0.83
SER	10	0	0.00
TOTAL	159	110	0.69

distribution of variation in proteins.

There is precedence for non-protein coding functions in these sequences. One possibility is that the sequence conservation is due to processing functions attributed to junction sequences. There are several problems with this interpretation. Firstly, the junction of ATPase 8 and ATPase 6 is not cleaved (Mariottini et al., 1983). It should also be pointed out that in salmon as in Xenopus the ATPase 8 gene does not overlap into the sequence presented here and no additional conservation due to overlapping reading frames is expected. The A-A in the 5' Hind III site A-A-G-C-T-T is predicted to be the A-A in the T-A-A stop codon of ATPase 8. Secondly, the tRNA genes which punctuate the CO III, ND 3 and ND 4L genes possess the secondary structure implicated in the processing of the primary mitochondrial transcript at those junctions. The only junction that is included in the sequence presented in Figs. 5 and 6 is the ATPase 6-CO III junction. This junction, associated with a stem loop structure, was first identified in mouse mitochondrial DNA (Bibb et al., 1981). This structure is conserved in the salmonid sequence but, as in mouse, it is located in the 3' end of ATPase 6 and therefore not associated with the region of primary sequence conservation. The lack of correlation between the processing site junctions and the conservation of the sequences does not discount the possibility that the conservation is due to a processing function, but does show that it is not related to the stem loop structure implicated in the processing.

Although less well understood, the sequences of the 5' ends of mitochondrial genes may be involved in functions related to ribosome binding. The potential for secondary structure in each of the predicted mRNA sequences is interesting (Section I). I have not been able to correlate the presence of these structures with the conserved regions.

tRNA GENES

The distribution of variation in tRNA genes is directly associated with the functional regions of the tRNA molecules. In Fig. 21 two things are apparent. One, the variation is not equivalent in the two tRNA genes and two, the DHU loop is the most variable region in these tRNAs.

Among human mitochondrial tRNAs, the tRNAs for four-fold degenerate codons are in general more variable than the tRNAs for two-fold degenerate codons (Cann et al., 1984). Both the tRNA^{GLY} and tRNA^{ARG} examined here recognize four-fold degenerate codons. Our results demonstrate that among the tRNAs for four-fold degenerate codons there can be great differences in variation.

Comparisons among several vertebrate species show that the mitochondrial tRNAs do not contain many of the invariant bases found in nuclear and bacterial tRNAs (Anderson et al., 1981, 1982; Bibb et al., 1981; Clary and Wolstenholm 1986; Roe et al., 1985; Section I). This is especially true with respect to the DHU and T ψ C loops. These loops in mitochondria

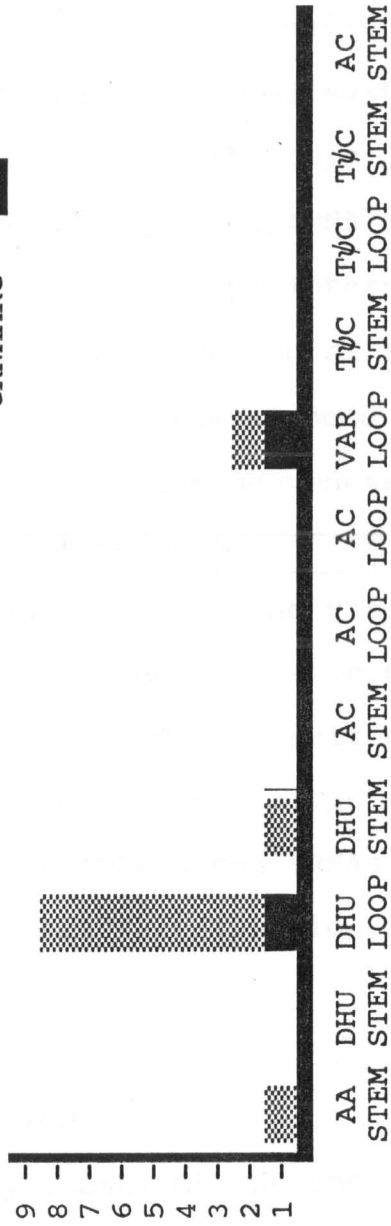
FIGURE 21.

Observed variation within the tRNA genes of the cloned fragment. The tRNA genes tRNA^{GLY} and tRNA^{ARG} are divided into the twelve functional regions (Gauss et al., 1979). For each region the ordinate gives the total observed variation. A single change between designated functional regions in tRNA^{GLY} is designated by (|).

VARIATION IN tRNA GENES

tRNA^{GLY} - [checkered box]

tRNA^{ARG} - [solid black box]



show little if any primary sequence conservation among species (Section I). In all vertebrate species, tRNA^{SER} is missing the DHU loop, and the T ψ C loops are absent from the tRNAs in Ascaris suum and Caenorhabditis elegans (Wolstenholme et al., 1987). The absence of primary sequence conservation and in some cases the DHU and T ψ C structures, is an indication of a loss of their function in animal mitochondrial systems. In yeast nuclear tRNAs, the sequences of these loops have been attributed to regulation of transcription of tRNA genes (Hall et al., 1982). Like all RNA polymerase III promoters, the tRNA promoters are located within the coding region of the gene, specifically in the DHU and T ψ C loops. The altered mode of transcription eliminates this as a constraint on the sequences in the DHU and T ψ C loops in animal mitochondrial genomes.

The data presented here and the data of Brown et al. (1982) show that neither of these loops or the variable loop are consistently variable. This result suggests that the sequences in these loops are not simply silent sites, but that the constraints are variable among tRNAs and that the variation reflects complex structure/function relationships in the tRNA sequences.

NON-SILENT VARIATION

There are 14 substitutions and one deletion which result in amino acid changes in the proteins coded for by this fragment. This level of non-silent substitutions is low when compared to

that detected in human populations (Cann et al., 1984). Although the small number of amino acid substitutions found in this fragment precludes an analysis of their distribution, we can analyse the types of changes which occur. Of the eleven positions at which non-silent substitutions occur only one position involves a change of amino acid charge and hydrophobicity. Six of the substitutions are Ile×Val. Both of these amino acids have aliphatic side chains. This is also the case for the Ile×Leu substitution. Thr and Ser have aliphatic hydroxyl side chains. Two other changes (Thr×Ala) are between the aliphatic hydroxyl and aliphatic side chains. In summary twelve of the fourteen substitutions which result in amino acid replacements are among uncharged aliphatic amino acids. It therefore appears that there is a strong pressure within the salmonid species to maintain the physical properties of the proteins.

The amino acid substitutions among more divergent taxa vary greatly from protein to protein (Anderson et al., 1982; Roe et al., 1985; Section I). It is also clear that the amino acid variation among vertebrate species is not distributed randomly within protein sequences.

Dynamics of Sequence Evolution

I have already pointed out that the salmonid mitochondrial genome appears to have a lower rate of replacement substitution than that found among primates or even within the humans. In

addition the dynamics of the substitution process may be different. In a comparison of the levels of divergence with the percentage of the substitutions that are transitions, the salmon mitochondria have a lower proportion of transitions at comparable levels of divergence when compared to protein coding sequences in primates (Brown et al., 1982) and rRNA sequences in primates (Hixson and Brown, 1986). Hixson and Brown (1987) have shown the substitution dynamics in the rRNA coding sequences to be much different from those in the protein coding regions. They propose that this is due to constraints placed on the sequence by the secondary structure of the rRNA. This finding may also suggest that an accurate comparison of dynamics requires a comparison of homologous sequences. If in fact the substitution dynamics are different between the salmon and primates this suggests that either the mechanisms which lead to the production of transversions are higher in salmon or that the mutational mechanisms are the same but the saturation point for mutations is lower in salmon than in the primates. Precedence for such a difference in the dynamics of evolution in mitochondrial DNA can be found in the Drosophila mitochondrial genome (DeSalle et al., 1987). The critical test of the alternatives is the sampling of more closely related salmon mitochondria. If the initial slope of the salmonid transition curve is high then the low level of transitions in the salmonids sampled here is due to the saturation of the salmonid mitochondrial genome and the erasure of transitions by transversions. Because different mitochondrial

genes have different rates of divergence it is critical that homologous sequences are used in comparisons.

Conclusions

Understanding the distribution of variation in mitochondrial genomes has important implications in understanding the mechanisms of molecular evolution of the mitochondrial genome and in the use of restriction enzymes to estimate the variability among groups. Both the high level of multiple substitution and the non-random distribution of variation in these sequences show the need to account for both problems when interpreting restriction site data. The non-random distribution of variation is also important in interpreting the expected level of saturation for mitochondrial genomes. If in fact the conserved regions within genes identified in this analysis represent functional constraints, this can have important effects on the ceiling of variation similar to those found in Drosophila sp. (DeSalle et al., 1987). Differences between salmon and primates in the levels of transitions at comparable divergence levels requires further investigation. In the first two sections I have described the distribution of variation in this mitochondrial fragment in an attempt to understand the functional constraints on the mitochondrial genome. In the next section I examine the spectrum of mutations that has accumulated among the salmonid species in an effort to determine the mutational pressures on the mitochondrial genomes.

SECTION III

**THE SPECTRUM OF SUBSTITUTIONS
AND THE OCCURRENCE OF ADDITIONS/DELETIONS
IN
THE SALMONID MITOCHONDRIAL GENOME**

Introduction

The rapid rate of evolution of animal mitochondrial DNA (mtDNA) when compared to nuclear DNA (Brown et al. 1982; Wilson et al. 1985) is probably due to an increased mutation rate as a result of inefficient (or non-existent) repair of DNA damage and/or replication errors (Brown et al. 1979, 1982; Wilson et al. 1985). Previous studies looking at the naturally occurring variation in animal mtDNA have shown that it includes a high incidence of length variation and transitions in relation to transversions (Cann and Wilson 1983, 1984, 1987; Aquadro and Greenberg, 1983; and Brown, 1983). This mutational spectrum is similar to that found in bacteria deficient in repair mechanisms (Wilson et al., 1985; Fowler et al. 1986). Although evidence for recombination of mtDNA was obtained in an early study with hybrid somatic cells (Horak et al., 1974), the results of subsequent studies all indicate that animal mitochondria appear to be without mechanisms of recombination (Lansman and Clayton, 1975). There is also some direct evidence that some mechanisms of post-replicative repair are absent (Clayton et al., 1974). If animal mitochondria lack post-replicative repair mechanisms and exonuclease associated proofreading, naturally occurring "silent" variation in mitochondrial DNA (mtDNA) should reflect the spontaneous substitution spectrum and provide a method to investigate the basic mechanisms of mutation.

There are three means by which substitutions can arise in a

system without repair mechanisms: (1) formation of mispairs following the model of Topal and Fresco (1976); (2) differential DNA polymerase selectivity; (3) unrepaired DNA damage. In this study I compare silent substitutions in a 2214 base pair segment of mitochondrial DNA from 6 closely related salmon species to the substitution spectra predicted by these three mechanisms.

Because the transition to transversion ratio is so high, past studies of mitochondrial variation have been limited to an analysis of transition frequencies. In order to address the question of all substitution frequencies I have undertaken an analysis of a large fragment from a group of species which have high enough levels of divergence to show sufficient numbers of transversions without eliminating specificity through subsequent transitions. The coding function of the sequence is known and the analysis can be limited to silent substitutions.

RESULTS AND DISCUSSION

Substitution Frequencies

The phylogenetic relationships derived from the sequence variation is shown in Fig. 22. These relationships are used to infer the substitution frequencies given in Fig. 23. There are 771 potential silent sites in the sequence compared. 329 of these sites show variation among the six species. These results show the high transition/transversion ratio typical of mtDNA variation. Because of the obvious bias in favour of transitions, we have calculated the expected numbers of transitions and transversions separately. The expected values closely match the observed values for the transitions but not for the transversions. In both transitions and transversions the bases which are infrequent in the silent positions have greatly reduced observed numbers.

Mispairing

For a model based on mispairing it is difficult to develop strong predictions for the individual substitution frequencies. If the model is based on the assumption in Topal and Fresco (1976) Pur-Pyr and Pyr-Pur mispairs that lead to transitions will be the most prevalent class of substitution. Pur-Pur mispairs will be the only transversion mispairs. Based on the geometries and the relative frequencies of the two purine bases, the G-G and G-A mispairs will occur more often than the A-A and A-G mispairs. That not all Pur-Pur mispairs have the same geometry is supported

FIGURE 22.

Parsimony tree relating mitochondrial DNA sequences from the six salmonid species. Following the principles of parsimony, the variation at silent positions was used to infer a branching pattern for the six taxa. The parsimony tree was derived using the PHYLIP phylogeny inference package supplied by Joe Felsenstein. The nodes 2,3,4,5 represent branch points for which hypothetical taxonomic units based on the most parsimonious order of substitutions can be determined. The minimum number of substitutions in this tree is 477.

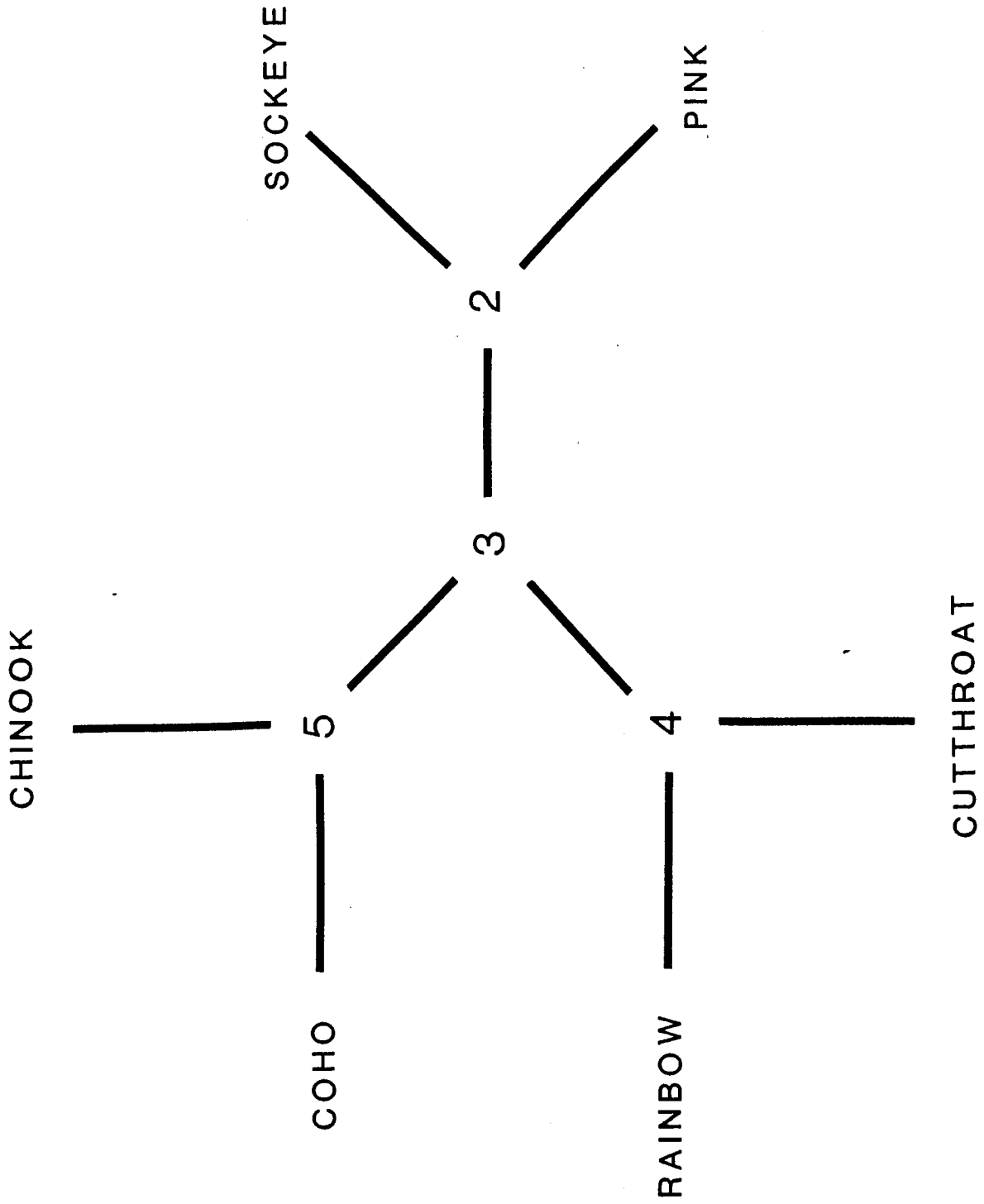


FIGURE 23.

Substitution frequencies for each of the twelve possible substitutions in each strand of the mitochondrial genome. The mispairs that result from the substitutions are presented. The template (or damaged) base is given in bold. I have constructed a hypothetical ancestral sequence based on comparisons of the present sequences of these species and the phylogeny in Fig. 22. I have calculated the frequency of each kind of substitution based on (1) a comparison of the hypothetical ancestral sequences and the sequences from extant taxa and (2) the phylogeny of Fig. 22. The expected number of each substitution is the product of the total number of transitions or transversions and the frequencies of the base changing at all silent positions. Each substitution is ranked according to each of the three proposed mechanisms of substitution. In each case 1 is the substitution predicted to occur most often for that mechanism.

L-strand Mispair Substitution	H-strand Mispair Substitution	# Observed	# Expected	Mispair Rank	Polymerase Rank	Damage Rank
Transitions						
A:C	G > A	21	22	1	2	4
G:T	A > G	111	105	1	4	5
T:G	C > T	110	117	1	2	4
C:A	A > C	94	73	1	4	5
Transversions						
T:C	A:G	0	2	2	3	1
C:C	G:G	2	2	2	1	2
G:G	C:C	2	12	2	1	2
A:G	T:C	26	12	3	3	1
C:T	G:A	7	10	2	4	3
T:T	A:A	11	10	3	4	2
A:A	T:T	12	7	3	4	2
G:A	C:T	2	7	2	4	3

indirectly by the ability of mismatch repair systems to differentiate between different Pur-Pur pairs. The Pyr-Pyr mispairs are predicted to be extrahelical and not detectable at the limits of this analysis.

With respect to the preponderance of transitions over transversions, the mispair model fits well with the data in Fig. 23. It does not fit well when the transversion mispairs are considered.

Polymerase selection

Recent in vitro fidelity studies of eukaryotic DNA polymerases, including a mitochondrial polymerase, have shown that base substitution spectra are a result of polymerase selectivity (Kunkel and Alexander, 1986). The frequencies of each of the twelve substitutions in their study allow for strong comparisons with the substitution spectra in Fig. 23. I have ranked the substitutions into four categories based on the frequencies of substitutions found in vitro with mitochondrial polymerase. These categories are not consistent with the data in Fig. 23 and are not consistent with the general predominance of transitions over transversions in mtDNA.

DNA Damage

The third hypothesis for the production of transversions can be based on damage to the mitochondrial genome. The damage to mtDNA which will lead to substitutions is limited by the lack of

repair. The most frequent spontaneous alteration in DNA in vivo is depurination (Lindahl and Nyberg, 1972). Depurination results in abasic lesions which have been shown to allow polymerization for several eukaryotic and prokaryotic polymerases (Loeb and Preston, 1986; Randall et al., 1987; and Takeshita et al., 1987). The propensity toward depurination and the possibility of polymerization through abasic sites without repair suggest that depurination is the most likely damage to affect the animal mitochondrial genome. Depurinations occur 100-500 times as frequently as depyrimidizations (Lindahl and Karlstrom, 1973, Lindahl and Nyberg, 1972 and Schaaper and Loeb, 1981). This predicts that depurinations are the only events which would affect the transversion frequencies in this study. In addition, the N-glycosylic bond of deoxyguanosine is hydrolysed more frequently than that of deoxyadenosine. Recent in vitro analyses (Randall et al., 1987; Takeshita et al., 1987) have shown that the incorporation of bases opposite abasic sites occurs such that $A > G > T = C$. Although the overall frequency of incorporation varies with the context of the template, the relative frequencies remain fairly consistent. With the above information it is possible to predict that the most frequent class of substitutions at those sites would be transversions that occur via A-G mispairing following the convention that G is the template or depurinated base and A is the incorporated base. The second most frequent class should be transversions via G-G or A-A mispairs. Although the two A-A products should have equal frequencies, as

should the two G-G products, the relative frequencies of the A-A and G-G products depend on the relative contributions of depurination and polymerase fidelity. The third most frequent class would be transversions via the G-A mispairs. The fourth and fifth classes would be transitions. The model predicts no detectable contribution from depyrimidizations. The spectrum of mutations does not indicate any contribution from deamination of C which should increase the proportion of C to T transitions.

The mtDNA damage model alone cannot account for the substitution spectrum in mitochondria, since the transitions, which occur at a high frequency, are predicted to be only the fourth and fifth most frequent classes. This does not preclude the role of damaged DNA in the production of transversions, but simply suggests that transitions are not consistent with this model and in that way supports the mispairing model for the production of transitions. In comparing the transversion frequencies with their ranking in the model (see Fig. 23) it is clear that there is a reduced proportion of C to A transversions from an A-G mispair on the H-strand, due to the low frequencies of C in that strand. The high frequency of the other A-G product and the high level of the A-A product fits well with this model. The only other inconsistency with this model concerns the higher than predicted level of T to G in the H-strand when compared to the other transversion via G-A mispairing and the two G-G mispair transversions. One explanation could involve alternative pathways, which involve transitions subsequent to transversions.

If the T to G transversion is over-represented because of a subsequent A to G transitions, the A to C transversions should also be over-represented. In light of the effect of the under-represented bases in both transitions and transversions it is interesting to look at the transversion frequencies in Drosophila mitochondrial DNA (Wolstenholme and Clary, 1985) where C and G are both under-represented. In a comparison of the sequences from protein coding regions in Drosophila yakuba and Drosophila melanogaster the prominent transversion class is T-A, which is consistent with the depurination model because both C to A pathways would be much reduced.

Deletion

Among the six salmonid species there is a single length variation. This is a deletion of a single codon in the ATPase 6 gene. Non-protein coding regions of mitochondria show a high level of small additions and deletions (Hixson and Brown, 1986; Wrischnik et al., 1987). Bacteria without repair also show additions and deletions in their spectrum of mutations. The single deletion in pink salmon does not allow for a calculation of rates of deletions in salmon. Considering the fact that the deletion in pink salmon is a non-silent change and that only a fraction of the addition/deletion mutational events will result in loss or gain of a complete codon, a single deletion is not inconsistent with a high level of addition/deletion mutational events.

Conclusions

It is clear that among the three mechanisms of substitution proposed here, the mispairing model is the most consistent with the propensity toward transitions found in mtDNA. Because the relative frequencies of the four possible Pur-Pur mispairs is difficult to predict based on current knowledge, it is difficult to completely rule out mispairing as the mechanism by which mitochondrial transversions arise. However, it is interesting that the relative frequencies of transversions predicted by the depurination model match the frequencies observed in mtDNA. The mispairing and depurination models are not mutually exclusive. If the effect of depurination falls between the effects of Pyr-Pur and Pur-Pur mispairing then both models could result in the transition and transversion substitutions separately. If this is the case, the substitution spectrum is best explained by a combination of the mispairing and damage models. The finding that different transversions occur at different rates is important in light of the recent advances in the production of phylogenies based on sequence comparisons. The method of evolutionary parsimony (Lake, 1987) assumes that the transversions occur at equal rates. Modifications of this method will be necessary for analysis of mitochondrial data.

SECTION IV

PHYLOGENY OF THE SALMONID
MITOCHONDRIAL GENOME

INTRODUCTION

Evolutionary relationships among the various salmonid species have been constructed on the basis of many different characters (Thomas et al., 1986). Successful experimental hybridizations among many members of the genus Oncorhynchus and subgenus Parasalmo indicate that the Pacific salmonids form a closely related evolutionary group. Although Regan (1914) and Benhke (1962) proposed that members of the Oncorhynchus and Parasalmo actually constitute a single genus, differences between these species have been emphasized. The mitochondrial DNA of several salmonid species have been compared (Thomas et al., 1986; Berg and Ferris, 1984 and Gyllensten and Wilson 1986). These studies have shown consistent levels of divergence. Where the same species were compared in different laboratories using different sets of restriction enzymes, the levels of divergence estimated were very similar. The relationships among the salmonid species are consistent with the close relationship of Parasalmo and Oncorhynchus (Thomas et al., 1986) and a distant relationship between Parasalmo and Salmo (Berg and Ferris, 1984; Gyllensten and Wilson, 1986).

In this section I address the evolutionary relationships among these six salmonid species and compare the relationships as determined by direct sequencing of a specific fragment of mitochondrial DNA with those determined previously using restriction enzymes to sample the mitochondrial sequence

divergence.

The fossil record for the salmonids is not complete. I have attempted to compare the molecular relationship among the extant species with the currently known fossils. This is done both to place limits on the rate of mitochondrial divergence and to begin a necessary synthesis of the molecular and fossil data.

RESULTS AND DISCUSSION

The observed level of nucleotide divergence for each pairwise comparison among the six species is given in Fig. 24. The levels of observed divergence range from a low of 4.74 percent between the rainbow and cutthroat trouts to a high of 8.76 percent between the pink salmon and rainbow trout. A comparison of these levels with those previously calculated using restriction site divergence estimates (Thomas et al., 1986; Berg and Ferris, 1984; Gyllensten and Wilson, 1986) show that in all cases the estimates of divergence from restriction site data significantly underestimate the levels observed in the sequences of this specific fragment. The first factor which must be noted is that the restriction site analysis samples sequences representing the entire mitochondrial genome and potentially representing all classes of functional sequences. The sequences represented in the 2214 base pair mitochondrial fragment analyzed in this thesis do not include the rRNA sequences which typically evolve at half the rate of the protein coding sequences, nor do they include the control region. The mitochondrial control region may evolve as much as 5 times faster than protein coding regions (Aquadro and Greenberg, 1983; Brown, 1985). Because of the spectrum and distribution of changes in the control region, the rates observed using restriction enzyme methods such as those used in (Thomas et al., 1986) will drastically underestimate the level of sequence divergence in the control region. Both of

FIGURE 24.

Pairwise nucleotide sequence divergence among the six salmonid species. The observed number of mismatches between each pair of species is given above the diagonal. The percentage of the 2214 base pairs not matching is given below the diagonal. The mismatches include all types of substitutions. The single deletion found in the pink salmon is not included.

	RAINBOW	PINK	COHO	SOCKEYE	CHINOOK	CUTTHROAT
RAINBOW	---	194	160	167	157	105
PINK	8.76	---	187	166	187	191
COHO	7.23	8.45	---	170	114	159
SOCKEYE	7.54	7.50	7.68	---	165	158
CHINOOK	7.09	8.45	5.15	7.45	---	150
CUTTHROAT	4.74	8.63	7.18	7.14	6.78	---

these factors suggest that this 2214 base pair fragment will have a higher actual sequence divergence than that which would be estimated from restriction site variation across the entire genome.

As pointed out in previous sections the variation in the sequence of this fragment is predicted to be greater than that predicted by restriction site estimates, due to the unequal distribution of variation in this mitochondrial fragment. The graph in Fig. 15 demonstrates that over 30 percent of the variable sites have undergone more than one substitution. The pairwise divergence shown in Fig. 24 includes those multiple substitutions which involve different changes at the same site in different lineages.

Another observation of major importance is the non-random distribution of variation within mitochondrial genes. One of the major assumptions of estimating divergence by sampling restriction sites is that the variation is randomly distributed. The fact that some sequences have a higher rate of divergence than others has been known for some time (Brown et al., 1982). In this salmon fragment I have found that the variation is not randomly distributed even within functional units. This clumping of variation is especially evident in the tRNA^{GLY} sequence. In this case the variation is limited to the DHU loop. Mitochondrial tRNA sequences make up almost 10 percent of the mitochondrial genome. If the variation in tRNA^{GLY} is typical of the variation in salmon mitochondrial tRNAs the lack of variation

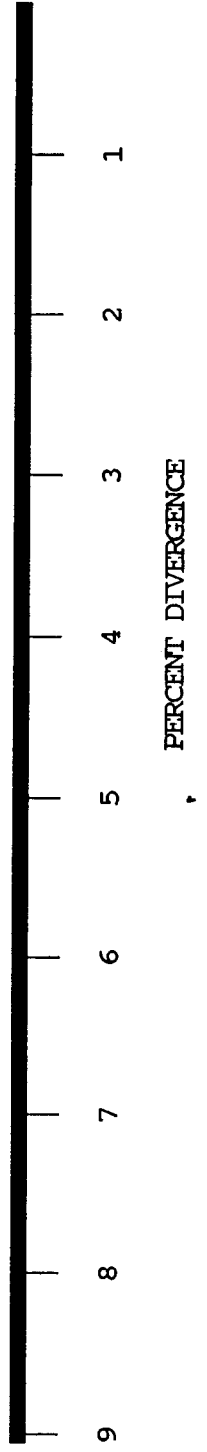
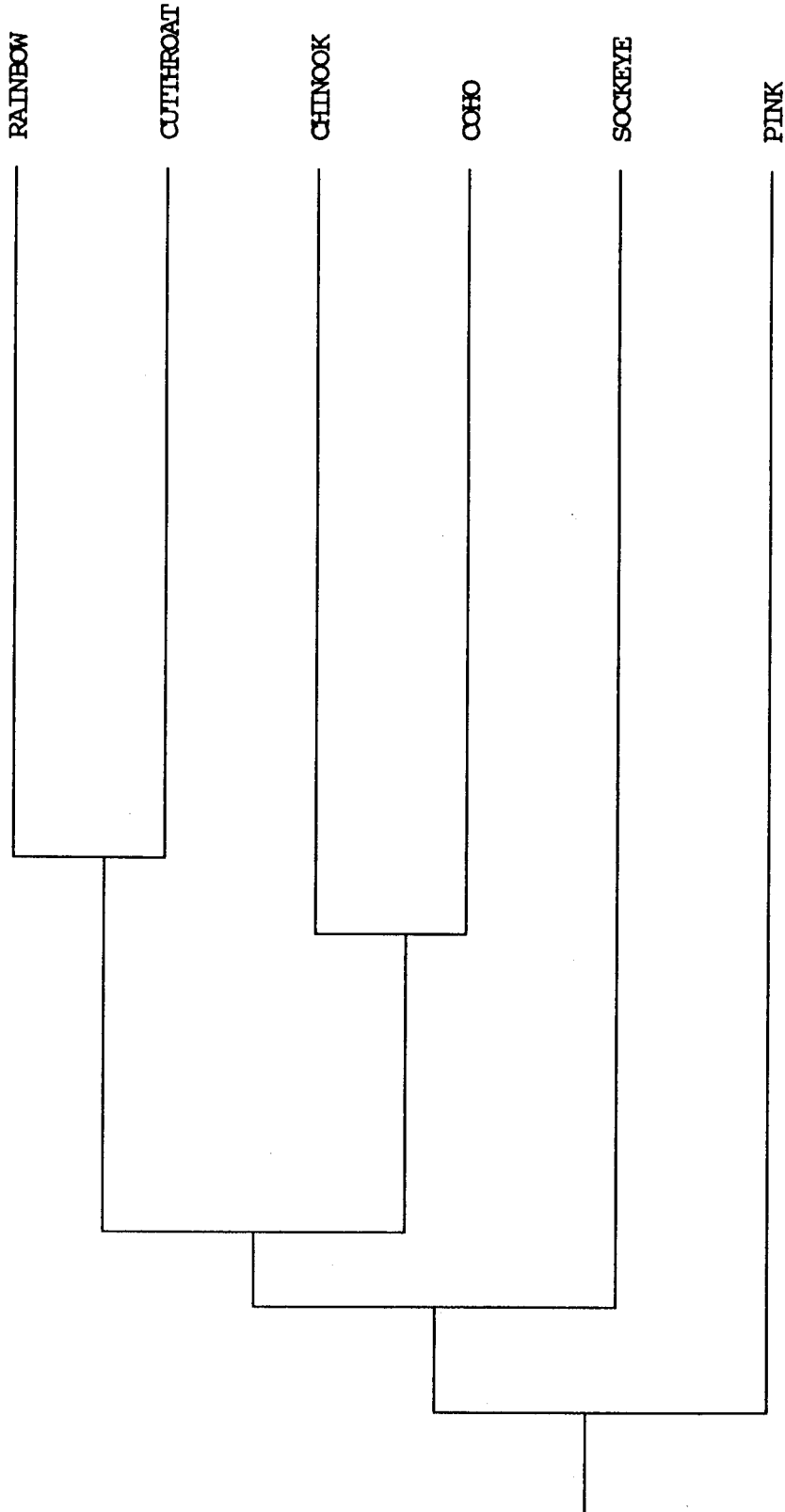
in most of the tRNA and the distribution of variation in such a small region could have a dramatic effect on the ability to predict divergence levels with restriction enzymes. In the protein coding regions which make up approximately 70 percent of the mitochondrial genome the variation also shows a clumped distribution (Fig. 18). This is not as drastic as the distribution in the tRNAs. Correlated with the distribution of variation in protein coding regions is the distribution of leucine codons. As noted in section II, leucine codons make up a significant proportion of the mitochondrial amino acids. The leucine codons present a particular problem for sampling with restriction enzymes. Variation in all other mitochondrial codons is limited to the third position. This is consistent with the finding that the majority of the variation observed in mitochondrial genomes is silent. In the leucine codons both the first and third positions are silent. This increases the maximum number of potential silent sites in a six base restriction site in mitochondria from two to four.

A UPGMA evolutionary tree based on the observed nucleotide substitutions in the 2214 base pair fragment is given in Fig. 25. The relationship among these six species predicted in this tree is the same as that predicted by restriction site estimates. It is only the level of divergence that varies between the two methods. The relationship in Fig. 25 is also consistent with the parsimony tree presented in Fig. 22. These results are in accordance with the proposal that Oncorhynchus and Parasalmo

FIGURE 25.

An unweighted pair-group method with arithmetic mean (UPGMA) evolutionary tree relating the six salmonid species analyzed. Data used for this tree are from Fig. 24.

UPGMA EVOLUTIONARY TREE



constitute a single genus.

RATES

Any attempt to determine a rate of substitution for the salmonid mitochondrial genome requires a well defined fossil record. The salmonid fossil record is not complete, but the presence of several fossil salmonids representing North American faunas throughout the late Cenozoic does allow for at least some comparison of the molecular and paleontological data. Aside from the general lack of fossil salmonids, one of the problems with the salmonid fossil record is the lack of a clear evolutionary relationship of the extant taxa. As pointed out by Calvender and Miller (1982) in their description of the fossil species Salmo australius, there is a difficulty in the alignment of the salmo species based on the current use of the generic name Salmo. This problem does not affect higher level taxonomic comparisons or zoogeographical proposals within species.

The earliest known salmonid fossil is the Eocene Eosalmo driftwoodensis (Wilson, 1977). This species is morphologically intermediate to the recent Salmoninae and Thymallinae salmonid subfamilies. No comparisons of mitochondria from these subfamilies have been published. A fossil Salvalinus is known from the late Miocene (Calvender, 1986). The divergence of Salvalinus fontinalis, the brook trout, from the Salmo and Oncorhynchus species is approximately 13 percent (Gyllensten and Wilson, 1986). A rate of 2 percent per million years is

consistent with a late Miocene origin of the Salvalinus group. The extant Pacific salmon species analyzed here and previously (Thomas et al., 1986) are not found in the fossil record prior to the Pliocene (Smith, 1981). It is interesting to note that Salmo clarki, the cutthroat trout, has the earliest known fossils of any of the extant species and is also the species with the greatest level of intra-specific divergence, 2.0% (Gyllensten and Wilson 1986), as compared with 1.5% in rainbow trout (Wilson et al., 1985) and less than 1.0% among chinook salmon individuals (Wilson et al., 1987).

The zoogeography of the great basin trouts is an additional source of reference for the timing of divergence within the cutthroat trout species. In his analysis of the great basin cutthroat trouts, Behnke (1981) has suggested that the three distinct cutthroat trout subspecies diverged after the start of the Pleistocene. The divergence between Salmo clarki bouvieri, the Yellowstone subspecies and Salmo clarki lewisi, Westslope subspecies is 2% (Gyllensten and Wilson, 1986), consistent with a Pleistocene origin for the cutthroat subspecies.

Conclusions

The evolutionary relationships determined by both parsimony and UPGMA methods are consistent with a close relationship between the Parasalmo and Oncorhynchus taxa. Although this relationship is the same as that estimated with restriction enzymes, the levels of divergence between taxa estimated from

restriction site variation were consistently lower than that observed in this fragment. The fossil record for salmonids is completely consistent with a rate of 2% sequence divergence per million years.

CONCLUSIONS

In the first section of this thesis I have found that the levels of amino acid identity between mitochondrially encoded proteins from trout and other animals are consistent with those previously found among other animal groups studied. A low level of amino acid sequence divergence between trout and Xenopus may suggest that the proteins in these two "cold-blooded" species are under similar constraints. I have found that the tRNA sequences among animal mitochondria have a complex set of constraints ranging from almost complete conservation in the AC loop to a complete lack of primary sequence conservation and lengths in the other loops. Other factors, such as nucleotide strand composition and codon usage, appear to be less biased in trout than in mammals or Drosophila. These comparisons among relatively divergent taxa probably reflect sequences that have been saturated with mutational events. Sequences and structures that are conserved among these groups are assumed to be critical to the function of this molecule.

The comparisons among salmonid species have revealed a complex pattern of variation both between and within protein coding and tRNA coding sequences. A much lower proportion of the variation among salmonids is non-silent when compared to that found among primates and within humans (Brown et al., 1982; Cann et al., 1984). These non-random patterns of variation and high levels of multiple substitution are consistent with the

difference in divergence estimated from the restriction site variation in the whole molecule and that found in this sequence.

The spectrum of mutations among salmonids is consistent with a combination of mispairing and DNA damage. This assumes that there is no repair in animal mitochondrial DNA, an assumption which requires more direct testing.

The phylogeny of the Pacific salmonid mitochondrial DNA examined here is consistent with a close relationship of the Parasalmo and Oncorhynchus groups. The fossil record is consistent with a rate of divergence of 2 percent per million years for these extant taxa. The set of oligonucleotides used in this analysis should facilitate the examination of variation within each of these species and allow the inclusion of other closely related groups.

LIST OF REFERENCES

- Anderson, S., A. T. Bankier, B. G. Barrell, M. H. L. de Bruijn, A. R. Coulson, J. Drouin, I. C. Eperon, D. P. Nierlich, B. A. Roe, F. Sanger, P. H. Schreier, A. J. H. Smith, R. Staden and I. G. Young. 1981. Sequence and organization of the human mitochondrial genome. *Nature*. 290:457-465.
- Anderson, S., M. H. L. deBruijn, A. R. Coulson, I. C. Eperon, F. Sanger and I. G. Young. 1982. Complete sequence of bovine mitochondrial DNA. *J. Mol. Biol.* 156:683-717.
- Aquadro, C. F. and B. D. Greenberg. 1983. Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. *Genetics*. 103:287-312.
- Atkinson, T. and M. Smith. 1985. Chapter 3. In Oligonucleotide synthesis: a practical approach. Edited by M. J. Gait, IRL press, New York. pp. 35-81.
- Aujame, L. and K. B. Freeman. 1979. Mammalian mitochondrial transfer RNAs: chromatographic properties, size and origin. *Nucleic acids Res.* 6:455-469.
- Baker, R. E., O. Gabrielsen, and B. D. Hall. 1986. Effects of tRNA^{TYR} point mutations on the binding of yeast RNA polymerase III transcription factor C. *J. Biol. Chem.* 261:5275-5282.
- Barrel, B. G., S. Anderson, A. T. Bankier, M. H. L. de Bruijn, E. Chen, A. R. Coulson, J. Drouin, I. C. Eperon, D. P. Nierlich, B. A. Roe, F. Sanger, P. H. Schreier. 1980. Differential pattern of codon recognition by mammalian mitochondrial tRNAs. *Proc. Natl. Sci. USA*. 77:3164-3166.
- Batley, J. and D. A. Clayton. 1980. The transcription map of mitochondrial DNA implicates transfer RNA excision as a major processing event. *J. Biol. Chem.* 255:11599-11606.
- Behnke, R. J., T. P. Koh, and P. R. Needham. 1962. Status of the landlocked salmonid fishes of Formosa with a review of Oncorhynchus masou. (Brevoort). *Copea*. 1962:400-407.
- Behnke, R. J. (1981). Systemic and Zoogeographical Interpretation of Great Basin trouts. In Fishes in North American deserts. Edited by R. J. Naiman and D. L. Soltz. John Wiley & Sons, New York. pp. 95-124.
- Berg, W. J. and S. D. Ferris. 1984. Restriction endonuclease analysis of salmonid mitochondrial DNA. *Can. J. Fish. Aquatic Sci.* 41:1041-1047.

- Bibb, M. J., R. A. Van Etten, C. T. Wright, M. W. Walberg and D. A. Clayton. 1981. Sequence and gene organization of mouse mitochondrial DNA. *Cell*. 26:167-180.
- Birnboim, C. and J. Doly. 1979. Rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* 7:1513-1523.
- Brown, W. M., M. George Jr., and A. C. Wilson. 1979. Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. USA*. 76:1967-1971.
- Brown, W. M., E. M. Prager, A. Wang, and A. C. Wilson. 1982. Mitochondrial DNA sequences of primates: Tempo and mode of evolution. *J. Mol. Evol.* 18:225-239.
- Brown, W. M. 1983. Evolution of animal mitochondrial DNA. *In* Evolution of genes and proteins. Edited by M. Nei and R. K. Koehn, Sinauer associates Inc., Sunderland, MA. pp. 62-88.
- Brown, W. M. 1985. The mitochondrial genome of animals. *In* Molecular evolutionary genetics. Edited by R. H. MacIntyre, Plenum, New York. pp. 95-130.
- Cann, R. L. and A. C. Wilson. 1983. Length mutations in human mitochondrial DNA. *Genetics*. 104:699-711.
- Cann, R. L., W. M. Brown and A. C. Wilson. 1984. Polymorphic sites and the mechanism of evolution in human mitochondrial DNA. *Genetics*. 106:479-499.
- Cann, R. L., M. Stoneking, and A. C. Wilson. 1987. Mitochondrial DNA and human evolution. *Nature*. 325:31-36.
- Cavender, T. M. 1986. Review of the fossil history of North American freshwater fishes. *In* The zoogeography of North American freshwater fishes. Edited by C. H. Hocutt and E. O. Wiley. John Wiley and Sons, New York. pp. 699-724
- Cavender, T. M. and R. R. Miller. 1982. Salmo australis, A new species of fossil salmonid from southwestern Mexico. *Cont. Mus. Paleo.*, Univ. Mich. 26:1-17.
- Chang, D. D. and D. A. Clayton. 1984. Precise identification of individual promoters for transcription of each strand of human mitochondrial DNA. *Cell*. 36:635-643.
- Ciliberto, G., L. Castagnoli, and R. Cortese. 1983. Transcription by RNA polymerase III. *Curr. Top. Dev. Biol.* 18:59-87.
- Clary, D. O. and D. R. Wolstenholm. 1983. Nucleotide sequence of

- a segment of Drosophila mitochondrial DNA that contains the genes for cytochrome oxidase subunits II and III and ATPase subunit 6. *Nucleic Acids Res.* 11:4211-4227.
- Clary, D. O. and D. R. Wolstenholm. 1985. The mitochondrial DNA molecule of Drosophila yakuba: Nucleotide sequence, gene organization and genetic code. *J. Mol. Evol.* 22:252-271.
- Clayton, D. A. 1984. Transcription of the mammalian mitochondrial genome. *Ann. Rev. Biochem.* 53:573-594.
- Clayton, D. A., J. N. Doda, and E. C. Friedberg. 1974. The absence of a pyrimidine dimer repair mechanism in mammalian mitochondria. *Proc. Natl. Acad. Sci.* 71:2777-2781.
- Chomyn, A., P. Mariottini, M. W. J. Cleeter, C. I. Ragan, A. Matsuno-Yagi, Y. Hatefi, R. F. Doolittle and Giuseppe Attardi. 1985. Six unidentified reading frames of human mitochondrial DNA encode components of the respiratory-chain NADH dehydrogenase. *Nature.* 314:592-597.
- Curran, J. F. and M. Yarus. 1987. Reading frame selection and transfer RNA anticodon loop stacking. *Science.* 238:1545-1550.
- de Bruijn, M. H. L. 1983. Drosophila melanogaster mitochondrial DNA, a novel gene organization and genetic code. *Nature* 304:234-241.
- DeSalle, R., T. Friedman, E. M. Prager and A. C. Wilson. 1987. Mitochondrial DNA sequences of Hawaiian Drosophila. *J. Mol. Evol.* 26:157-164.
- Dickerson, R. E., and H. R. Drew. 1981. Structure of a B-DNA dodecamer II. Influence of base sequence on helix structure. *J. Mol. Biol.* 149:761-786.
- Fowler, R. G., R. M. Schaaper and B. W. Glickman. 1986. Characterization of mutational specificity within the lac I gene for a mutD5 mutator strain of Escherichia coli defective in 3' to 5' exonuclease (proofreading) activity. *J. Bacteriology.* 167:130-137.
- Gauss, D. H., F. Gruter and M. Sprinzl. 1979. Compilation of tRNA sequences. *Nucleic Acids Res.* 6:r1-r19.
- Gyllensten, U., R. F. Leary, F. W. Allendorf, and A. C. Wilson. 1985. Introgression between two cutthroat trout subspecies with substantial karyotypic, nuclear and mitochondrial genomic divergence. *Genetics.* 111:905-915.
- Gyllensten, U. and A. C. Wilson. 1986. Mitochondrial DNA of

- salmonids: intraspecific variability detected with restriction enzymes. In. The Application of Population Genetics to Fisheries Management. Edited by N. Ryman and F. M. Utter, University of Washington press, Seattle.
- Hall, B. D., S. G. Clarkson, and G. Tocchini-Valentini. 1982. Transcription initiation of eucaryotic transfer RNA genes. *Cell*. 29:3-5.
- Hattori, M. and Y. Sakaki. 1985. Dideoxy sequencing method using denatured plasmid templates. *Anal. Biochem.* 152:232-238
- Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene*. 28:351-359.
- Henikoff, S. 1987. Unidirectional digestion with Exonuclease III in DNA sequence analysis. *Methods in Enzymology*. (In Press)
- Hixson, J. E., and W. M. Brown. 1986. A comparison of the small ribosomal RNA genes from the mitochondrial DNA of the great apes and humans: sequence, structure, evolution, and phylogenetic implications. *Mol. Biol. Evol.* 3:1-18.
- Holmes, D. S. and M. Quigley. 1981. A rapid boiling method for the preparation of bacterial plasmids. *Anal. Biochem.* 114:193-197.
- Horak, I., G. C. Hayden and I. B. Dawid. 1974. Interspecific recombination of mitochondrial DNA molecules in hybrid somatic cells. *Proc. Natl. Acad. Sci. USA*. 71:1828-1832.
- Hori, H. and S. Osawa. 1987. Origin and evolution of organisms as deduced from 5S ribosomal RNA sequences. *Mol. Biol. Evol.* 4:445-472.
- Kastelein, R. A., B. Berkhout, and J. van Duin. 1983. Opening the closed ribosome-binding site of the lysis cistron of bacteriophage MS2. *Nature*. 305:741-743.
- Kim, S. H., F. L. Suddath, G. J. Quigley, A. McPherson, J. L. Sussman, A. H. J. Wang, N. C. Seeman and A. Rich. 1974. Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. *Science*. 185:435-439.
- Kunkel, T. A. and P. S. Alexander. 1986. The base substitution fidelity of eucaryotic DNA polymerases. *J. Biol. Chem.* 261:160-166.
- Lake, J. A.. 1987. A rate-independent technique for analysis of nucleic acid sequences: evolutionary parsimony. *Mol.*

Biol. Evol. 4:167-191.

- Lansman, R. A., and D. A. Clayton. 1975. Selective nicking of mammalian mitochondrial DNA in vivo: photosensitization by incorporation of 5-bromodeoxyuridine. J. Mol. Biol. 99:761-776.
- Lansman, R. A., R. O. Shade, J. F. Shapira, and J. C. Avise. 1981. The use of restriction endonucleases to measure mitochondrial DNA sequence relatedness in natural populations. J. Mol. Evol. 17:214-226.
- Lindahl, T. and O. Karlstrom. 1973. Heat-Induced depyrimidination of Deoxynucleic acid in neutral solution. Biochem. 12:5151-5154.
- Lindahl, T. and B. Nyberg. 1972. Rate of depurination of native deoxyribonucleic acid. Biochem. 11:3610-3618.
- Loeb, L. A. and B. D. Preston. 1986. Mutagenesis by apurinic/apyrimidinic sites. Ann. Rev. Genet. 20:201-230.
- Manam, S. and G. C. Van Tuyle. 1987. Separation and characterization of 5'-and 3'- tRNA processing nucleases from rat liver mitochondria. J. Biol. Chem. 262:10272-10279.
- Maniatis, T., E. F. Fritsch and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratories.
- Mariottini, P., A. Chomyn, G. Attardi, D. Trovato, D. D. Strong and R. F. Doolittle. 1983. Antibodies against synthetic peptides reveal that the unidentified reading frame A6L, overlapping the ATPase 6 gene, is expressed in human mitochondria. Cell. 32:1269-1277.
- McAllister, D.E. and C.R. Harington. 1969. Pleistocene grayling, Thymalis, from Yukon, Canada. Can. J. Earth Sci. 6:1185-1190.
- Mclaughlin, P. J. and M. O. Dayhoff. 1972. Evolution of Species and Proteins: a time scale. In Atlas of Protein Sequence and Structure. Edited by M. O. Dayhoff, National biomedical research foundation, Washington, D. C.
- Montoya, J., D. Ojala and G. Attardi. 1981. Distinctive features of the 5'-terminal sequences of the human mitochondrial genome. Nature. 290:465-470.
- Ojala, D. J. Montoya and G. Attardi. 1981. tRNA punctuation model

of RNA processing in human mitochondria. *Nature*. 290:470-474.

Queen, C., and L. J. Korn. 1984. A comprehensive sequence analysis program for the IBM personal computer. *Nucleic Acids Res.* 12:581-599.

Rajput, B., L. Duncan, D. Demille, R. C. Miller and G. B. Spiegelman. 1982. Transcription of cloned transfer RNA genes from *Drosophila melanogaster* in a homologous cell free extract. *Nucleic Acids Res.* 10:6541-6550.

Randall, S. K., R. Eritja, B. E. Kaplan, J. Petruska and M. F. Goodman. 1987. Nucleotide insertion kinetics opposite abasic lesions in DNA. *J. Biol. Chem.* 262:6864-6870.

Regan, C. T. 1914. The systematic arrangement of the fishes of the family Salmonidae. *Ann. Mag. Nat. Hist. Ser. 8*, 13:405-408.

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

Schaaper, R. M. and L. A. Loeb. 1981. Depurination causes mutations in SOS-induced cells. *Proc. Natl. Acad. Sci. USA.* 78:1773-1777.

Smith, G. R. 1981. Late Cenozoic freshwater fishes of North America. *Ann. Rev. Ecol. Syst.* 12:163-93.

Takehita, M., C. Chang, F. Johnson, S. Will and A. P. Grollman. 1987. Oligodeoxynucleotides containing synthetic abasic sites: model substrates for DNA polymerases and apurinic/apyrimidinic endonucleases. *J. Biol. Chem.* 262:10171-10179.

Templeton, A. R. 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and apes. *Evolution.* 37:221-244.

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

Tinoco, I., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla. 1973. Improved estimation of secondary structure in ribonucleic acids. *New Nature Biol.* 246:40-41.

Topal, M. D. and J. R. Fresco. 1976. Complementary base pairing and the origin of substitution mutations. *Nature.* 263:285-

289.

- Tsuyuki, H. and E. Roberts. 1962. Species differences of some members of Salmonidae based on their muscle myogen patterns. J. Fish. Res. Bd. Canada. 20:101-104.
- Utter, F. M., F. W. Allendorf and H. O. Hodgins. 1973. Genetic variability and relationships in Pacific salmon and related trout based on protein variations. Syst. Zool. 22:257-270.
- Wilson, A. C., R. L. Cann, S. M. Carr, M. George, U. B. Gyllensten, K. M. Helm-Bychowshi, R. G. Higuchi, S. R. Palumbi, E. M. Prager, R. D. Sage, and M. Stoneking. 1985. Mitochondrial DNA and two perspectives on evolutionary genetics. Biol. J. Lin. Soc. 26:375-400.
- Wilson, G. M., W. K. Thomas, and A. T. Beckenbach. 1987. Mitochondrial DNA analysis of Pacific northwest populations of Oncorhynchus tshawytscha. Can. J. Fish. Aquat. Sci. 44:1301-1305.
- Wilson, G. M., W. K. Thomas, and A. T. Beckenbach. 1985. Intra- and Inter-specific mitochondrial DNA sequence divergence in Salmo: rainbow, steelhead, and cutthroat trouts. Can. J. Zool. 63:2088-2094.
- Wilson, M. V. H. 1977. Middle Eocene freshwater fishes from British Columbia. Life Sci. Contrib. Royal Ont. Mus. 113:1-61
- Wolstenholme, D. R. and D. O. Clary. 1985. Sequence evolution of Drosophila mitochondrial DNA. Genetics. 109:725-744.
- Wolstenholme, D. R., J. L. Macfarlane, R. Okimoto, D. O. Clary, and J. A. Wahleithner. 1987. Bizarre tRNAs inferred from DNA sequences of mitochondrial genomes of nematode worms. Proc. Natl. Acad. Sci. USA. 84:1324-1328.
- Wrischnik, L. A., R. G. Higuchi, M. Stoneking, H. A. Erlich, N. Arnheim, and A. C. Wilson. 1987. Length mutations in human mitochondrial DNA: direct sequencing of enzymatically amplified DNA. Nucleic Acids Res. 15:529-542.