SEQUENCE AND EVOLUTION OF APOLIPOPROTEIN A-I IN THREE SALMONIDS

VICTOR A.B. DROVER

Sequence and Evolution of Apolipoprotein A-I in Three Salmonids

by

Victor A.B. Drover

A thesis submitted to the
School of Graduate Studies
in partial fulfilment of the
requirements for the degree of
Master of Science

Department of Biochemistry
Memorial University of Newfoundland

January 1995

ISBN 0-612-17590-1

Abstract

A cDNA library was constructed from brown trout liver tissue and one clone, BTLB27, containing the processed mRNA transcript of apolipoprotein A-I was isolated and sequenced. This sequence was then compared to all other known apoA-I cDNA/genomic DNA sequences and the phylogeny of the represented species inferred. The sequences isolated from the fish species grouped outside both avian and mammalian sequences while the avian sequence was an outgroup to the mammals. The phylogenetic tree also revealed that rodents diverged from the mammals before lagomorphs, carnivores, artiodactyls, and primates. Using the cDNA/genomic DNA comparisons, the insertion sites of apolipoprotein A-I introns II and III were predicted and amplified using the polymerase chain reaction and subsequently sequenced. We find that there are two loci for apolipoprotein A-I, one of which has undergone major evolutionary changes. Also, phylogenetic inference using intron II sequences support the findings from the cDNA sequences. Molecular clocks were constructed from the phylogenetic data and the accepted fossil record to estimate the time of various evolutionary events. We find that the accepted time for the divergence of rat and mouse is much too recent. Estimates of the divergence times of three salmonid species as well as a genome duplication event which preceded salmonid speciation are also greater than current accepted values, although the differences are not as great as that observed for the rodent lineage.

Acknowledgments

Although it is not possible to thank everyone who has contributed to this project, there are a number of people to whom I am greatly indebted. Willie Davidson has been a constant source of guidance and support throughout this project as well as my Honours dissertation. Willie was in fact the driving force which led the path of my career towards research and subsequently allowed me to think and act with relative freedom. It is this later point for which I am most grateful. Sylvia Bartlett has also contributed tremendously. Her technical expertise and ever-cheerful personality has made the whole experience both educational and enjoyable. I am also thankful for the helpful words, laughs, arguments, and beers supplied in abundant quantities by people such as Colin McGowan, John Goodier, Caroline Hurst, Blair Pritchett, Tony Nakla, Dorothy Crutcher, Annette Greenslade, and Michelle Normore. The continued moral support by Christie Dean and my parents, Bruce and Linda Drover, is also greatly appreciated. Finally, I would like to thank the Biochemistry Department at Memorial University of Newfoundland.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

apoA-I, apolipoprotein A-I

CETP, cholesteryl ester transfer protein

HDL, high density lipoprotein(s)

I3M, the microsatellite contained within IVS III of apoA-I in salmonids

IDL, intermediate density lipoprotein(s)

IVS, intervening sequence (intron)

LAAT, lecithin alcohol acyltransferase

LCAT, lecithin-cholesterol acyltransferase

LDL, low density lipoprotein(s)

MYA, millions of years

nHDL, nascent HDL

OTU, operational taxonomic unit

RAPD, randomly amplified polymorphic DNA

SINE, short interspersed nucleotide element(s)

UPGMA, unweighted pair group method using arithmetic averages

VLDL, very low density lipoprotein(s)

Chapter I

General Introduction

1.1 Composition and metabolism of high density lipoproteins.

Lipoproteins are macromolecular complexes consisting of triacylglycerol, phospholipid, cholesterol, and protein. They are found circulating in the plasma and act as the main transporters of lipids between tissues. In mammals there are five types of plasma lipoproteins. Chylomicrons are transient lipoproteins produced in the intestine following a meal high in fat. The other four lipoproteins are distinguished by their hydrated densities as high, intermediate, low and very low density lipoproteins (HDL, IDL, LDL, and VLDL, respectively; Babin, 1987). Besides density, the lipoproteins can be distinguished by their chemical compositions (Table 1.1). One of the most important differences between these lipoproteins is their apolipoprotein content. The distribution of the apolipoproteins gives each lipoprotein a distinct function and thus plays a key role in lipoprotein metabolism. This research focuses on apolipoprotein A-I (apoA-I), the main protein component of HDL.

HDL precursors (nascent HDL; nHDL) are produced in the liver, the intestine, and through the lipolysis of chylomicrons and VLDL (Eisenberg, 1984). Although each of these nHDL has a slightly different composition, they all contain low levels of triglyceride and unesterified cholesterol and high levels of phospholipid and apoA-I. Because these precursor particles are low in cholesterol they are excellent acceptors of excess cholesterol from peripheral tissues. Once the nHDL particle associates with the cell membrane, apoA-I activates lecithin-cholesterol acyltransferase (LCAT; Aron *et al.*, 1978) which can then esterify the free cholesterol within the cell. The hydrophobic

Table 1.1. Chemical and physical characteristics of some apolipoproteins.

Lipoprotein	VLDL	LDL	HDL
Molecular Weight (10^{-6})	5-6	2.3	.018-0.36
Density (g/L)	0.95-1.006	1.006-1.063	1.063-1.210
Component	Percentage Composition		
Triacylglycerol	50	10	4
Free Cholesterol	7	8	2
Esterified Cholesterol	12	37	15
Phospholipid	18	20	24
Protein	10	23	55
Major apolipoprotein	C-I, C-II, C-III, E	B	A-I, A-II

cholesteryl ester can then be transported to the HDL core through the cholesteryl ester transfer protein (CETP; Francone *et al.*, 1989). nHDL contains both LCAT and CETP. Together, they raise the cholesteryl ester content of the lipoprotein thereby producing mature HDL (HDL). Once formed, HDL is directed to the liver where excess cholesterol ester is extracted by at least four mechanisms (Assmann *et al.*, 1992) and then excreted through bile acid synthesis and secretion. These processes make up the reverse cholesterol transport pathway (Schmitz *et al.*, 1985; Fig 1.1) and are thought to be responsible for the inverse relationship between HDL serum cholesterol concentrations and the risk of developing ischaemic heart disease (Miller and Miller, 1975). The high incidence of coronary heart disease has thus generated much interest in HDL and, of course, apoA-I.

1.2 Structure and function of apoA-I.

ApoA-I is a soluble, group I apolipoprotein and contains between 238 (Atlantic salmon; Powell *et al.*, 1991) and 243 (human; Karathanasis *et al.*, 1993) amino acids. A short leader sequence of 23 to 24 amino acids tags apoA-I for secretion and is thus removed before the protein enters the plasma (Pownall and Gotto, Jr., 1992). Once in the plasma, apoA-I can spontaneously associate with lipid (Chan and Li, 1992), a phenomenon that is accompanied by an increase in the μ-helical content of apoA-I (Morrisett *et al.*, 1977). Although three-dimensional crystal structures are not available for apolipoproteins, Chou-Fasman analysis predicts a repetitive helical content for apoA-I. Furthermore, the hydrophobic moment algorithm indicates that apoA-I is more amphipathic than the typical globular protein (Pownall *et al.*,

1983). The occurrence of prolines in the secondary structure of apoA-I is also quite intriguing. In most cases, they occur at sites between repeated helical segments and exhibit a minima in the helical moment (Pownall *et al.*, 1986). These pieces of evidence point toward a hypothetical model describing the mechanism whereby apoA-I associates with the lipid molecules to form HDL. The hydrophobic side of each amphipathic μ-helix can interact with the lipid molecules. The prolines between each of these helices allow adjacent helices to interact with the lipids in a different orientation, thereby compensating for the spherical or discoidal nature of HDL. In this arrangement, the hydrophilic side of the helices remains exposed to the plasma to solublize HDL and to interact with other proteins such as receptors and enzymes.

1.3 Organization and expression of the apoA-I gene.

Although much research has been devoted to the structure and function of the apoA-I protein, valuable information has also been obtained from the structure and organization of the apoA-I coding region. The high level of expression of this gene facilitates the isolation and cloning of its cDNA. Consequently, cDNA sequences have been determined for five mammals and two salmonid species (Table 1.2). When the inferred amino acid sequences were compared, a number of similarities were noticed. For instance, the proline residues which separate the proposed μ-helices are highly conserved. The region between these prolines consists of either 11 or 22 amino acids which are themselves conserved throughout apoA-I. Furthermore, when apoA-I was compared to other apolipoproteins, this internal repeat was still observed, as was a 33 amino acid region at the amino

terminus. It has thus been proposed that apoA-I arose through multiple intragenic duplications of a common ancestral gene (Kimura, 1983). Fig. 1.2 illustrates the evolution of various apolipoproteins based upon the number of 11/22 amino acid repeats[1]. It was also noticed that apoA-I was a rapidly evolving gene. When compared to the ß-globin gene, which evolves at the average rate for 35 mammalian genes (Li *et al.*, 1985), apoA-I has a 25% higher rate of non-synonymous substitutions (O'hUigin *et al.*, 1990). However, the overall structure of apoA-I appears to have been retained in all species studied to date.

The gene structure of apoA-I has also been studied in a number of species (Table 1.2). A comparison of these sequences revealed three introns or intervening sequences (IVS) which are spliced from the primary transcript before translation. The first intron is located in the 5'-untranslated region and the other two occur in the coding region. However, the placement of the two latter IVS appears conspicuously non-random. IVS II separates most of the leader peptide from the rest of the protein while IVS III separates the initial 33 amino-acid block from the 11/22 amino acid repeats. As the latter is the proposed lipid-binding domain, the IVS seem to separate apoA-I into

---

[1] Recent comparisons between apoA-I nucleotide sequences from various species and human apoE/apoA-IV suggest that the apoA-I gene is ancestral to both apoE and apoA-IV. Protein products coresponding to these genes would not be present in salmonids in this scenario (Powell et al., 1991).

Table 1.2. A comprehensive reference list of all known apoA-I cDNA and genomic sequences.

Species	cDNA sequence	Genomic Sequence
Human	Law *et al.*, 1983	Karathanasis *et al.*, 1993
Baboon	Hixson *et al.*, 1988	
Monkey		Murray and Marotti, 1992
Cow	O'hUigin *et al.*, 1990	
Pig		Birchbauer *et al.*, 1993
Dog	Luo *et al.*, 1989	
Rabbit		Pan *et al.*, 1987
Rat		Haddad *et al.*, 1986
Mouse	Stoffel *et al.*, 1992	
Chicken		Bhattacharyya *et al.*, 1991
Atlantic salmon	Powell *et al.*, 1991	
Rainbow trout	Delcuve *et al.*, 1992	

regions with distinct functions (Pownall and Gotto, Jr., 1992).

1.4 Lipid metabolism in fish.

The composition and metabolism of plasma lipoproteins in fish[2] has been well characterized (for review see Babin and Vernier, 1989). Unlike mammalian systems, these poikilothermic or cold-blooded vertebrates preferentially utilize lipid (as opposed to carbohydrate) as the primary source of energy. This characteristic may account for the hyperlipidemic nature of fish serum as demonstrated for rainbow trout (three-fold increase in comparison to rat plasma lipid levels). Cholesterol levels in this species are also elevated (as high as twelve times rat levels). Most fish species studied exhibit the standard apolipoprotein classes whose main apolipoprotein and lipid components are homologous to those found in mammals. Further, the most abundant apolipoprotein particle in both mammals and fish (rainbow trout) is HDL. In addition to the mammalian apolipoprotein classes, egg-laying fish (and other oviparous species of reptiles and birds) produce vitellogenin, the major yolk protein. This ancient protein associates with lipids, phosphate, and various metal ions and is used as a food source during embryogenesis. Recent studies have revealed a similarity between vitellogenin and human apoB-100 (the primary apolipoprotein of mammalian LDL). This raises the possibility that vitellogenin has additional functions (beyond an embryonic food source) and that the vitellogenin gene

---

[2]The term 'fish' is used to represent, in general, the approximately 20 000 species of Teleosts. However, most of the research has centred on the Class Actinopterygii (Order Salmoniformes).

may have been the ancestor to the present-day apoB-100 gene (Barber *et al.*, 1991; Steyrer *et al.*, 1990; Baker, 1988).

Enzyme activities common in mammalian lipid transport systems are also present in many fish species. The presence of LCAT- and CETP-like activities as well as lipoprotein lipase activity implies that the major pathways of cholesterol metabolism have been conserved. However, lecithin-alcohol acyltransferase (LAAT, similar to LCAT) has been identified in carp but is not detectable in mammals. LAAT catalyzes the transfer of fatty acids in lecthin to the acyl moiety of wax esters and may be responsible for some of the plasma wax esters found in this species.

In salmonids, apoA-I is the most abundant plasma protein and its cDNA has been isolated from two species spanning two genera (see Table 1.2). Atlantic salmon (*Salmo* genus) apoA-I cDNA has been sequenced from liver and its distribution of tissue expression examined. Like mammalian apoA-I, salmon apoA-I is highly expressed in the liver and intestine. Trace levels of apoA-I mRNA were also detected in salmon muscle (Powell *et al.*, 1991). In rainbow trout (*Oncorhynchus* genus) two cDNA sequences have been isolated. One copy (apoA-I-1) is the major transcript of normal liver cells while the other (apoA-I-2) is expressed only in hepatocellular carcinoma cells. The induction of neoplasia by aflatoxin-B_1 selectively increases the expression of apoA-I-2 above and beyond that of apoA-I-1 (Delcuve *et al.*, 1992). To date, no genomic sequence has been obtained for any species of salmonid or any other fish.

1.5 Evolution of salmonids.

Within the Salmonidae family there are three subfamilies; Coregoninae (whitefish and cisco), Thymallinae (graylings), and Salmoninae (commonly referred to as salmonids or trouts, char, and salmon). The latter includes four genera (*Hucho*, *Salvelinus*, *Oncorhynchus*, and *Salmo*) with some 68 species (Allendorf and Thorgaard, 1984) and has been the subject of numerous taxonomic/systematic studies using a diverse array of markers such as morphological, allozyme, and nucleotide sequence data (for a discussion see Phillips and Pleyte, 1991). For example, short interspersed nucleotide elements elements (SINEs) were used to reconstruct the phylogeny of some salmonids (Murata *et al.*, 1993; Kido *et al.*, 1994; Fig. 1.3). However, many questions still remain concerning the classification of some species and even the existence of additional genera such as *Salmothymus* and *Brachymystax* (Phillips and Pleyte, 1991). Much of this uncertainty arises from the fact that all salmonids apparently share a common tetraploid ancestor (Allendorf and Thorgaard, 1984). Thus, many genetic loci exist in duplicate, which can confound studies directly based on the genetic sequence (i.e. allozyme analysis, restriction enzyme analysis, and sequence comparison). It is estimated that 30% of these duplicate loci have been silenced, 46% have diverged giving rise to paralogous loci, while 24% have been conserved producing isoloci (Phillips and Pleyte, 1991). Consequently, it is important to compare orthologous genes (genes that have evolved directly from an ancestral locus) rather than paralogous (genes which have arisen from a gene duplication event) when inferring phylogenies. Alternatively, both loci may be used in the case of amino acid or nucleotide sequence comparison (if

sequences are available from each locus) to estimate both the phylogenetic relationship of the organisms and when the genome duplication occured.

Svardson (1945) first proposed the notion that salmonids were polyploid (i.e. having more than two sets of chromosomes; Svardson, 1945). Although his hypothesis that the basic chromosome number was n=10 was incorrect, later work supported his ideas as salmonids were found to be tetraploid. This was first suggested by Ohno *et al.* (1968) and is based on three lines of evidence: (1) Salmonid fish have about twice as much DNA per cell and twice as many chromosome arms as closely related fish, (2) Multivalents (i.e. more than two chromosomes whose homologous regions are synapsed by pairs) have been commonly observed in meiotic preparations from salmonid species, and (3) Salmonids exhibit many duplicated enzyme loci (Allendorf and Thorgaard, 1984).

1.6 Evolution and population genetics using variable genetic elements.

The study of evolution can be separated into two broad categories: macroevolution and microevolution (Funk and Brooks, 1990). Macroevolution focuses on the evolution among lineages and operates over very long periods of time (i.e. tens to hundreds of millions of years). Thus, conserved sequences must be used to study macroevolution (a hypervariable region would almost certainly lose its genetic identity altogether, making it of little use for long range evolutionary comparisons). The genes coding for conserved proteins and ribosomal RNAs are commonly used to infer phylogeny among lineages. Although the structure and function of each of these types of molecules have been retained, there is enough sequence

variation over time such that an accurate pattern of evolutionary relationships can be determined. For example, the amino acid and nucleotide sequence of conserved proteins such as myoglobin (Romero-Herrera *et al.*, 1973) and cytochrome c (McLaughlin and Dayhoff, 1972) can be used to examine deep evolutionary relationships.

Microevolution occurs within a lineage and thus operates within the time-scale of an individual species or a number of closely related species. The goal of population genetics is to identify microevolution within a species to determine the existence/absence of any sub-species or distinct populations. Typically, hypervariable regions such as repetitive elements are used to observe microevolution. The abundance of repetitive DNA within the genomes of eukaryotes was first observed in 1968 (Britten and Khone) by observing the rates of reassociation of denatured DNA. More recently, repetitive elements of various types have been identified (by DNA cloning and sequencing) and the study of their variability has become quite common in population genetics and species/individual identification. A good example of this type of repetitive element is microsatellite DNA: variable genetic elements consisting of short DNA sequences (1-5bp) repeated in tandem (Rassman *et al.*, 1991). Two important characteristics of microsatellites make them suitable for individual identification and pedigree analysis:

1. the length of the microsatellite (i.e. the number of repeats) is hypervariable and thus can be observed within a population (or a number of populations) of one species,

2. the regions flanking the microsatellites are unique and conserved such that oligonucleotide primers can be designed for these sites. This allows rapid analysis of many individuals using the polymerase chain reaction.

Thus, by amplifying unique microsatellites from a population, it is possible to extract information about the population structure based on the sizes of the microsatellites and their distribution. This procedure has been used in a wide variety of species including human (Litt and Luty, 1989), pig (Wintero *et al.*, 1992), brown trout (Estoup *et al.*, 1993) and *Drosophila* (Tautz, 1989).

1.7 Goals and Objectives.

In general, the objective of this research was to further characterize the apoA-I gene in salmonids and to use this information to examine the evolution of mammals, birds, and fish. To this end, three major lines of research were explored: (1) the gene structure of apoA-I was examined. The cDNA sequence of brown trout and the position/sequence of two apoA-I introns from three salmonids (including brown trout) were obtained, (2) using the apoA-I sequences obtained above and all other known apoA-I sequences, phylogenetic inference methods were used to investigate the evolutionary relationships between a number of species of mammals, fish, and birds. This information was also used to estimate rates of evolution and divergence times of various lineages, (3) duplicate apoA-I loci in salmonids were compared to determine the changes that have occurred since the genome duplication.

Fig. 1.1. Schematic representation of the reverse cholesterol transport pathway. A, apoA-I is synthesized in the liver. nHDL particles are then asssembled and secreted into the plasma. B, the nHDL particles interact with the peripheral tissues. ApoA-I activates LCAT which esterifies free cellular cholesterol. These esters are then transferred to nHDL via CETP. Mature HDL is thus formed. C, HDL particles return to the liver where the cholesteryl esters are extracted and catabolized. Although it is likely that this final step is receptor-mediated, the mechanism of cholesteryl ester uptake by the liver from HDL and the subsequent fate of the HDL particle itself is not well understood.

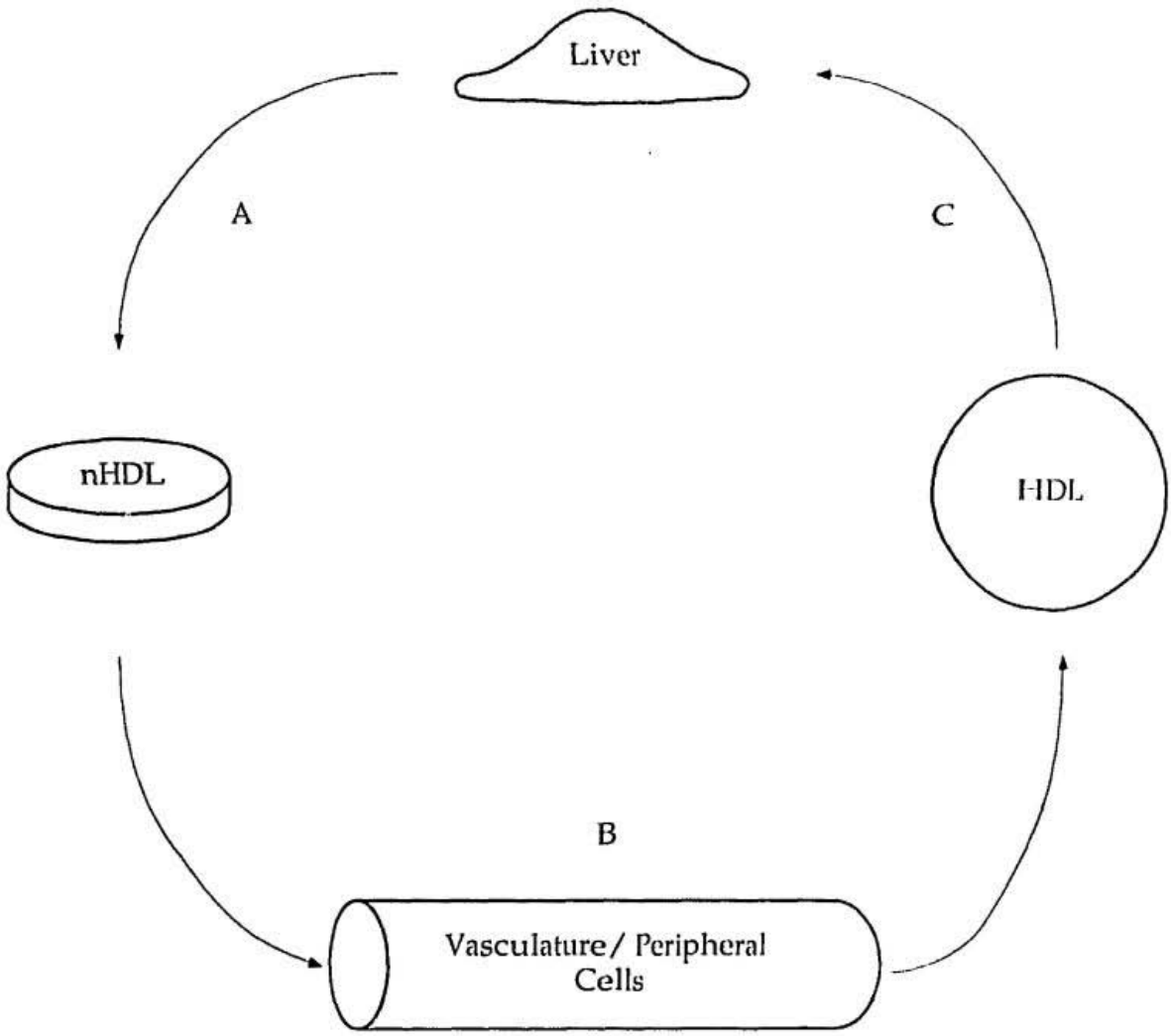
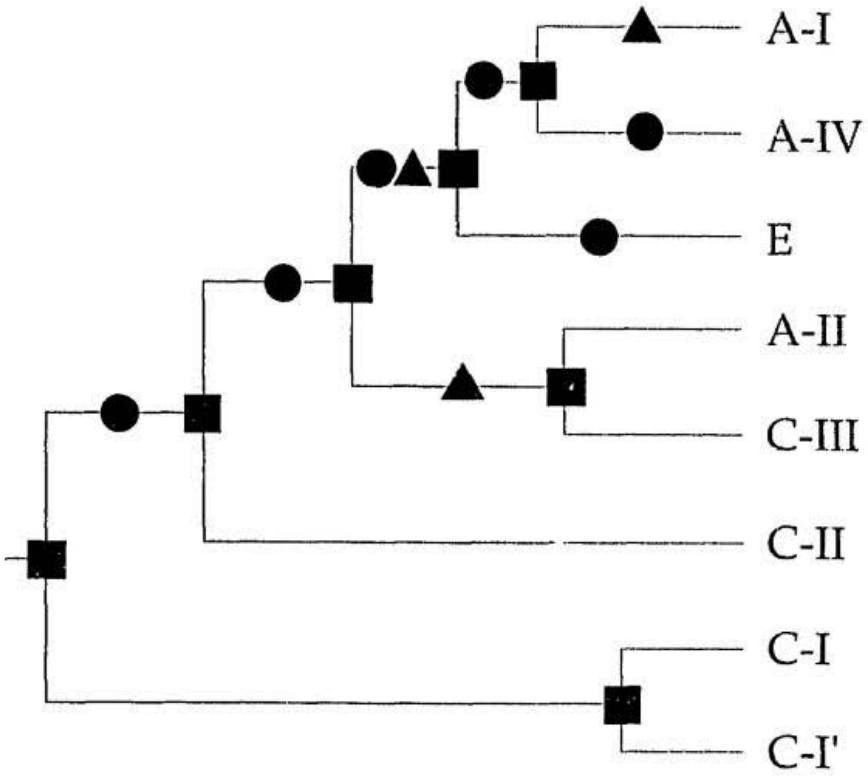

Liver
A
C
nHDL
HDL
B
Vasculature/ Peripheral
Cells

Fig. 1.2. A hypothetical scheme for the evolution of apolipoprotein genes (adapted from Chan and Li, 1992).

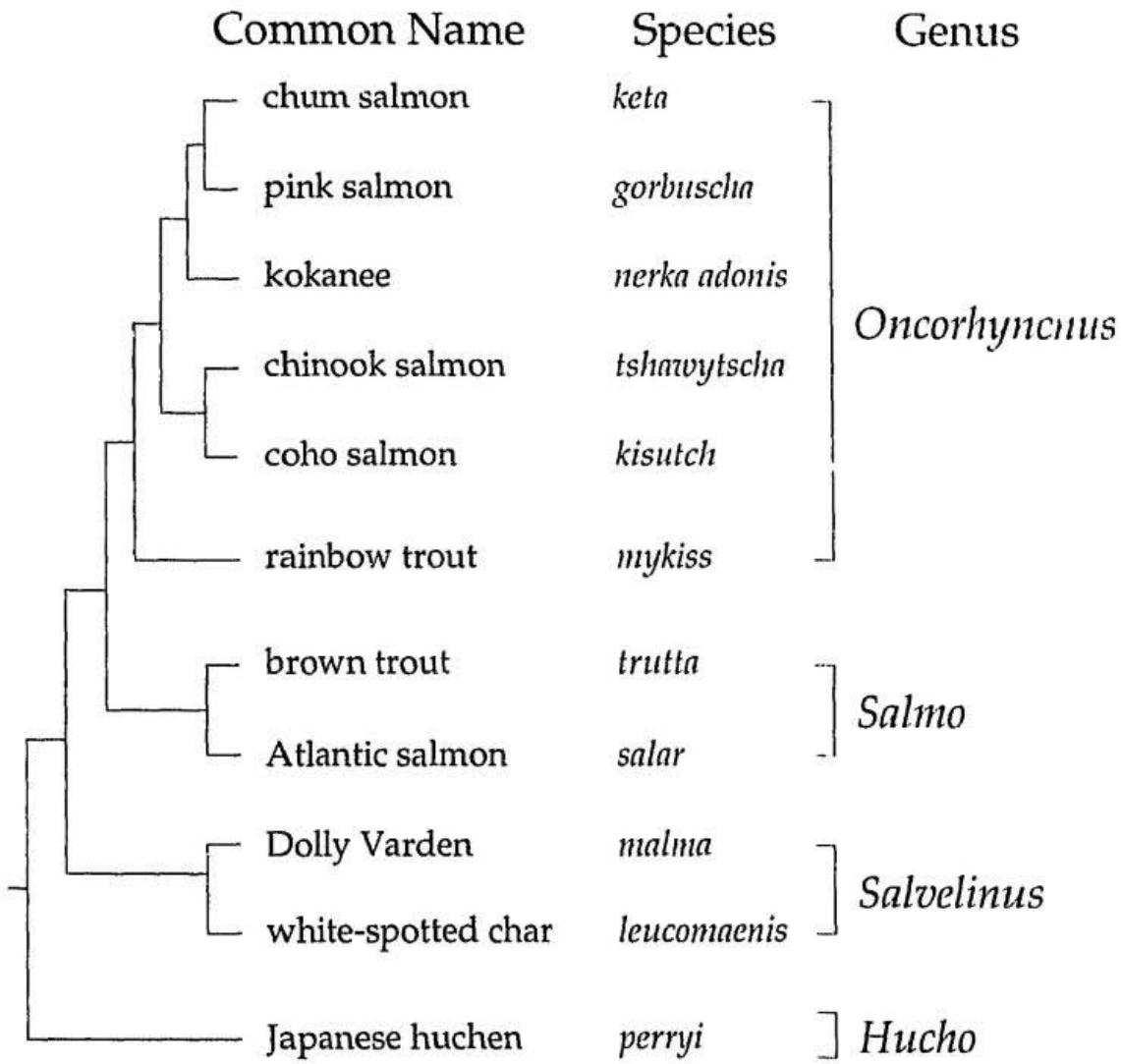


■ Gene duplication event

● Duplication of 33 or 66 codons

▲ Deletion of 33 codons

Fig. 1.3. Phylogenetic relationships within the family Salmonidae (adapted from Murata *et al.*, 1993).


Common Name
Species
Genus
chum salmon
keta
pink salmon
gorbuscha
kokanee
nerka adonis
chinook salmon
tshawytscha
coho salmon
kisutch
rainbow trout
mykiss
Oncorhynchus
brown trout
trutta
Atlantic salmon
salar
Salmo
Dolly Varden
malma
white-spotted char
leucomaenis
Salvelinus
Japanese huchen
perryi
Hucho

Chapter II

Isolation and Characterization of apoA-I cDNA and Mature Protein

2.1 INTRODUCTION

2.1.1 Preparation of cDNA

A cDNA library is a collection of DNA molecules which represent a population of mRNA molecules. Essentially, a cDNA library represents the expressed genes of the tissue from which the mRNA was isolated. The cDNA is propagated in a cloning vector and typically maintained in *Escherichia coli*. A good cDNA library must exhibit the following two characteristics:

1. It must contain enough individual clones such that low-abundance mRNAs are represented.
2. The cDNA molecules are full-length, thus representing complete messages.

If either of these conditions is not met, the isolation and sequencing of a particular message will be impaired.

There are three general steps involved in constructing a cDNA library. First, the mRNA must be isolated. The mRNA is then used as the template with reverse transcriptase to produce the cDNA. Finally, the cDNA is ligated into the appropriate cloning vector. Variation in this last step produces two broad categories of cDNA libraries. The simplest type of library to construct is a random library. In this case, the orientation of the cDNA insert within the vector is not constant. If a DNA probe is being used, the orientation of the insert does not affect the isolation of a particular clone from the library. However, immunological detection methods require the presence of an antigenic determinant on the protein produced from the cDNA. In this case, only 1/6 of the clones will be in the proper orientation and reading frame to express the protein. Thus, six times as many clones must be screened. Further,

sequence analysis can be troublesome particularly if a poly-A tail is not present to indicate the 3' end of the message. In directional libraries, all cDNA inserts are cloned in a specific orientation. Besides more efficient screening and sequence analysis, directional cDNA cloning also facilitates the construction of subtracted libraries. This type of library is the end-product of a comparison between two cDNA populations. The resultant library is thus enriched in messages which are present in one population but not the other. Although there are various methods used to produce subtracted libraries, each uses hybridization to remove sequence information common to both populations.

The cDNA library produced for this research was constructed using the SuperScript Plasmid System for cDNA Synthesis and Plasmid Cloning (Gibco BRL). This system has been optimized in a number of ways to maximize efficiency and versatility of the library. For instance, reverse transcriptase (RT) exhibits an RNAse activity. Superscript RT has been cloned and engineered such that the RNAse activity has been abolished. This decreases mRNA degradation thereby increasing both the proportion of full-length cDNAs as well as first-strand yields. Another enhancement used to produce this directional library is the use of a primer-adapter (Fig. 2.1). In this case, a poly dT primer is combined with a double-stranded adapter which encodes a *Not* I restriction endonuclease site. The poly dT region anneals with the poly A tail of the mRNA and acts as a primer for first-strand synthesis. The final double-stranded product will now contain a *Not* I restriction site. The addition of a *Sal* I adapter followed by a *Not* I restriction digest introduces asymmetry into the cDNA and allows for directional cloning (*Not* I restriction sites are

relatively rare in vertebrate genomes). The non-complementary *Not* I/*Sal* I restriction sites on the cloning vector also prevent the formation of empty clones which do not contain a DNA insert and are common in random libraries (for a discussion on cDNA libraries, see the instructional manual included with the SuperScript Plasmid System for cDNA Synthesis and Plasmid Cloning (Gibco BRL)).

2.1.2 Apolipoprotein purification using sequential flotation.

Sequential flotation is the most common method used to separate plasma lipoprotein particles. The procedure exploits the fact that the density of each type of lipoprotein falls within a well defined range (Table 1.1) and involves adjusting the density of the plasma/serum followed by ultracentrifugation. For instance, the density of VLDL is not greater than 1.006 g/mL. If the density of the plasma sample is adjusted to 1.006 and centrifuged, all of the particles with a density less than 1.006 will float to the top of the plasma while more dense particles will pellet at the bottom of the centrifuge tube. The uppermost VLDL fraction can then be removed and the process repeated at a different density. This allows the specific separation of VLDL, LDL, and HDL. A simple extraction to remove the lipids allows the protein component to be analyzed separately.

2.1.3 The polymerase chain reaction (PCR).

The polymerase chain reaction (PCR; Mullis and Faloona, 1987) is an *in vitro* procedure that uses short oligonucleotides (primers) to amplify specific DNA fragments from a DNA template. Initially, all the components required

for DNA replication are mixed with both the primers and the desired template. These components include a heat stable DNA polymerase (e.g. *Taq* polymerase isolated from *Thermus aquaticus*, a bacterium that grows in hot water springs). Two characteristics of *Taq* polymerase make it useful for PCR: it is resistant to denaturation at high temperatures and its optimal temperature for DNA replication is 74°C.

Once all the components are mixed, they are heated to 95°C to denature any double stranded DNA present. The mixture is then cooled to allow the primers to anneal to the appropriate complementary positions on the template DNA. The short length and relatively high concentration of the primers promotes primer annealing over reannealing of the complementary template strands. The actual temperature at which annealing occurs is usually between 40°C and 55°C and is empirically determined for each pair of primers. When the primers have annealed, the temperature of the reaction is raised to 72°C. *Taq* polymerase will then replicate the templates to which primers have annealed. For PCR to be successful, the primers must anneal to regions which flank the DNA fragment of interest, but always on the complementary strand. In this way, the fragment of interest will be amplified by both primers, but in opposite directions. These two products will be complementary to each other. Thus, the product of one primer becomes the template for the other primer in subsequent amplification cycles and leads to an exponential increase in the amount of amplified DNA. In practice, 25 to 35 cycles of amplification (Fig. 2.2) are sufficient to produce products which can be seen on an agarose gel after ethidium bromide staining.

2.2 EXPERIMENTAL and DISCUSSION

2.2.1 Isolation and sequencing of an apoA-I clone from a brown trout liver, cDNA library.

A cDNA library was constructed from mRNA extracted from the liver of a brown trout using the SuperScript Plasmid System for cDNA Synthesis and Plasmid Cloning (Gibco BRL; C. McGowan, unpublished results). In an attempt to identify highly expressed messages, random clones were chosen for sequencing and a 10mL LB/ampicillin liquid culture prepared for each one. Plasmid DNA (pDNA) was then prepared from 3mL of this culture using the Magic MiniPrep plasmid preparation kit (Promega) in a final volume of 50µL. An aliquot of 9µL was removed from each plasmid preparation and mixed with 1µL of 2M NaOH?2mM EDTA. The reaction was incubated at 37°C for five minutes and 10µL of 1.6M ammonium acetate (pH 4.8) was then added. This was followed by 40µL of 95% ethanol (-20°C) and an incubation at -70°C for 30 minutes. The precipitated plasmid was pelleted through centrifugation at 12 000 rpm for 15 minutes (4°C) and the supernatant carefully removed. The pellet was then washed with 500µL of 70% ethanol (-20°C) and centrifuged as above for five minutes. The supernatant was again removed and the pellet dried under vacuum for 30 minutes at room temperature. The pellet was finally suspended in 10µL of water and sequenced using the Sequenase DNA sequencing kit (United States Biochemical). Fractionation of the reactions through a standard sequencing gel (6% polyacrylamide/7M Urea/1XTBE (89mM Tris, 112mM boric acid, 2mM EDTA, pH 8.3) for two hours at 40 watts yielded approximately 200bp of 5' sequence. Each sequence was compared to entries in the GenBank nucleotide sequence database using

the FASTA program (Lipman and Pearson, 1985). The sequence obtained from the plasmid pBTLB27 showed a high level of sequence similarity to the apoA-1 cDNA sequence reported for Atlantic salmon (Powell *et al.*, 1991). Longer fractionation times and sequencing with the Universal primer yielded most of the remaining sequence including the 3' poly-A tail. Although the complete sequence had not been obtained, the presence of the poly-A tail and the ATG initiation codon implied that the complete apo A-1 coding region had been isolated.

To obtain the remaining sequence, an oligonucleotide primer was designed complementary to the 3' sequence obtained above (Fig. 2.3). This primer was then used to sequence pBTLB27. The data obtained from these sequencing reactions overlapped the 5' and 3' sequences obtained above, thus giving the complete sequence of the pBTLB27 cDNA insert. When this brown trout sequence was compared to rainbow trout (Delcuve *et al.*, 1992) and Atlantic salmon cDNA sequences, two things were immediately observed:

1. ApoA-I is highly conserved among the three Salmonid species,
2. An 87 bp region present in both rainbow trout and Atlantic salmon was not present in brown trout.

To determine if this deletion produced a corresponding phenotype (i.e. a truncated protein), an examination of the apoA-I protein in brown trout was initiated.

2.2.2 Isolation of HDL from brown trout serum and partial purification of apoA-I.

Human plasma (P. Davis, 27.5 mL) and brown trout serum (C. McGowan, 29.0 mL) were dialyzed exhaustively against a 0.85% (g/100 mL) solution of KBr (d=1.006 g/mL, mock solution A). Each sample was then transferred to a 30 mL ultracentrifuge tube and filled to volume/balanced with mock solution A. Both tubes were then centrifuged at 37 500 rpm/14°C/18 hours in a 70TI, fixed-angle rotor. The VLDL fraction was visible as a cloudy pellicle at the top of the tube. This fraction was removed with a Pasteur pipette and discarded. To obtain the LDL fraction, the remaining plasma/serum was adjusted to d=1.063 g/mL by the addition of 0.083 g/mL of KBr to each tube. Both tubes were then balanced with mock solution B (8.22% KBr, d=1.063) and then centrifuged at 37 500 rpm/14°C/24 hours. The clear, orange layer containing LDL was removed and discarded. The density of each sample was then adjusted to d=2.10 g/mL through the addition of 0.235 g/mL of KBr to each tube and centrifuged at 37 500 rpm/14°C/40 hours. Both the human and brown trout samples contained an orange layer similar to the LDL fraction. These HDL fractions (3mL and 8 mL, respectively) were then dialyzed exhaustively against a solution of 0.15M NaCl/50mM TrisHCl (pH 7.5)/5mM EDTA and stored at 4°C.

To analyze the protein content of each HDL fraction, the lipoprotein particles were delipidated as outlined in Fig. 2.4. The dried protein pellets were then resuspended in 1.0mL (human) and 3.0mL (brown trout) of 1X UTE (8M urea/10mM TrisHCl (pH 7.5)/1mM EDTA). A 10µL aliquot (in duplicate) of each protein sample was then fractionated through a sodium dodecyl sulfate (SDS), polyacrylamide gel (stacking gel-3% polyacrylamide; separating gel-15% polyacrylamide) at 80V (stacking gel) and 180V (separating gel) until

the blue dye reached the bottom of the gel (approximately 1 hour). Standard protein markers as well as human apoA-I (Sigma) were also fractionated as markers. To visualize the proteins the gel was stained for 30 minutes in 0.2% (w/v) Coomassie Brilliant Blue R-250 (Kodak) in 50% (v/v) ethanol, 10% (v/v) acetic acid and then destained in 20% (v/v) ethanol, 10% (v/v) acetic acid. As shown in Fig. 2.5, the major protein in both preparations migrates at or near the same rate as the human apoA-I standard. However, both samples show considerable contamination. One contaminant of relatively high concentration in the brown trout sample migrates just below 14 400Da. The contaminant is most likely to be apoA-II. This apolipoprotein is the other main protein component of HDL (Table 1.1) and migrates at approximately 13 000Da (Babin and Vernier, 1989).

In an attempt to further purify apoA-I from brown trout, semi-quantitative ion-exchange chromatography was performed a batch adsorption method[3]. DEAE-Sephadex (A-50, Pharmacia) was used since the inferred amino acid sequences from Atlantic salmon and rainbow trout cDNA sequences predict that apoA-I will be negatively charged at neutral pH. The binding or elution of apoA-I from the resin under various conditions was assayed by SDS-PAGE as described above.

Optimal binding of apoA-I in to the resin occurred at pH7.5 in 1X UTE (data not shown) while maximum elution occurred between 0.25M and 0.30M NaCl (Fig. 2.6). Although some low molecular weight contamination can be seen at/below 0.25M NaCl, most of the high molecular weight contamination remains bound to the resin. To purify this protein to homogeneity, column

---

[3]Ion Exchange Chromatography: Principles and Methods, 3rd edition. Pharmacia LKB Biotechnology, Uppsula, Sweden, 1991, p. 49.

chromatography using a salt gradient could be used. However, our intention was to determine the size of the major protein species of HDL. It is clear that this protein is similar to human apoA-I in both size and charge. Assuming that it is indeed brown trout apoA I, it is also clear that the 87bp deletion in the cDNA is not an actual phenotype. However, a truncated protein may have a decreased stability, may not be secreted, or may not properly assemble into HDL particles. In either of these scenarios, the purification protocol used for apoA-I would not detect a truncated protein. Western blotting of total liver proteins using an apoA-I-specific antibody could be used to detect different protein isoforms. PCR using apoA-I locus-specific primers was used to investigate this matter of the 87bp deletion further (see Chapter III).

2.2.3 Amplification and sequencing of genomic DNA to investigate the cDNA deletion observed in the pBTLB27 sequence.

To determine the genomic sequence of the cDNA deletion, oligonucleotide primers were designed to flank the region containing the 87bp deletion. These 20-mers, labeled 63 and 65 (see Fig. 2.8), were designed over regions that are 100% identical in the three salmonid species. Each PCR reaction was performed in a 10µL solution containing 50mM KCl, 10mM Tris-HCl (pH 9.0), 1.5mM $MgCl_2$, 0.2mM of each dNTP, 0.5mM of each primer, 0.25 units[4] of AmpliTaq DNA polymerase (Perkin Elmer), and 50ng of template DNA. Amplifications were performed in a Genamp PCR System 9600 thermal cycler (Perkin Elmer) using the following method: one cycle of 95°C/3 minutes followed by 30 cycles of 95°C/30 sec. (denaturation), 56°C/30sec.

[4]One unit is defined as the amount of enzyme that will incorporate 10nmol of dNTPs into an acid insoluble material in 30 minutes at 74°C.

(annealing), and 72°C/45sec. (extension). Plasmid DNA (pBTLB27) and genomic DNA from Atlantic salmon, brown trout, and rainbow trout were used as templates. Each 10µL reaction was then diluted with 2µL of 6X tracking dye (30% (w/v) glycerol, 0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol FF) and fractionated through a 100mL 3% Nuseive agarose/1X TA (40mM Tris HCl (pH 7.5), 20mM sodium acetate) gel. Ethidium bromide staining and U.V. illumination allowed the PCR products to be detected and photographed. The results of this experiment are shown in Fig. 2.7. The positive control, using pBTLB27 as the template, gives a single amplification product a little larger than the 603bp marker. This is expected as the cDNA sequence predicts a fragment of 613bp. When genomic DNA from the salmonids was used, two bands were produced for each template. The largest band produced in the Atlantic salmon and brown trout reactions migrated a little higher than the 872bp marker (~900bp). The cDNA deletion is, however, only 87bp in length. The additional 100bp (900-613-87≈200) is due to the presence of an intron which is also flanked by primers 63 and 65 (see Chapter III). Surprisingly, both the Atlantic salmon and brown trout templates also produced a band that was the same size as the positive control. Although contamination of the reaction mix with pBTLB27 is one possible explanation, these smaller bands may represent the presence of a pseudogene in the Atlantic salmon and brown trout genomes which lacks both the 87bp deletion region as well as the intron.

To obtain the sequence of the 87bp deletion region of the cDNA sequence, the 900bp fragment amplified from brown trout genomic DNA was excised from the gel and purified in 50µL of water using the Wizard PCR

Preps DNA Purification System (Promega). Using 9.5µL of this solution, the fragment was sequenced from the 3' end using ^{32}P end-labeled primer 65 as per the manufacturer's specifications included with the *fmol* DNA Sequencing System (cycle sequencing kit, Promega). All reactions were then fractionated through a standard sequencing gel (above) for ~2 hours. The 87bp sequence was obtained along with 3' and 5' flanking sequence. The composite apoA-I cDNA sequence[5] from brown trout is shown in Fig. 2.8.

The rainbow trout also produced two PCR products, both of which were larger than the positive control. The smaller is ~850bp in length. The difference in size between this product and the largest product produced from the other two species is due to the presence of a shorter intron (see Chapter III). Another band of ~1050bp was also produced from the rainbow trout template. The origins of this band are uncertain. Thus, each genomic DNA template produced one band of a similar size and one that could not be easily explained. In hopes of clarifying this issue, a more detailed study of the apoA-I intron structure/sequence was undertaken (see Chapter III).

2.2.4 Comparison of human and brown trout apoA-I sequences.

Using the alignment program Clustal V (Higgins *et al.*, 1992), the brown trout apoA-I inferred amino acid sequence was compared to the human apoA-I sequence (Fig. 2.9). Although the two sequences share only 26% nucleotide identity, many of the substitutions are synonymous and the overall structure of the protein has been conserved. This is best demonstrated in the conservation of an 11 or 22 amino acid repeat separated by proline

---

[5]Genbank Accession #L49383.

residues. Within each repeat we can also observe the repetitive nature of the placement of charged residues. In particular, a conserved glutamate or aspartate residue occurs every 7-8 residues (or approximately two turns of an μ-helix). Together, these data imply the conservation of a repeated amphipathic μ-helix and thus the conservation of function as a lipid-binding domain. The conservation of molecular weight and charge structure, as demonstrated above (, also suggest that the overall structure and function of apoA-I is similar in all vertebrate species.

Fig. 2.1. A schematic representation of cDNA synthesis using the Superscript Plasmid System for cDNA Synthesis and Plasmid Cloning. The production of asymmetric restriction sites on each end of the cDNA, to allow the production of a directional library, is illustrated.

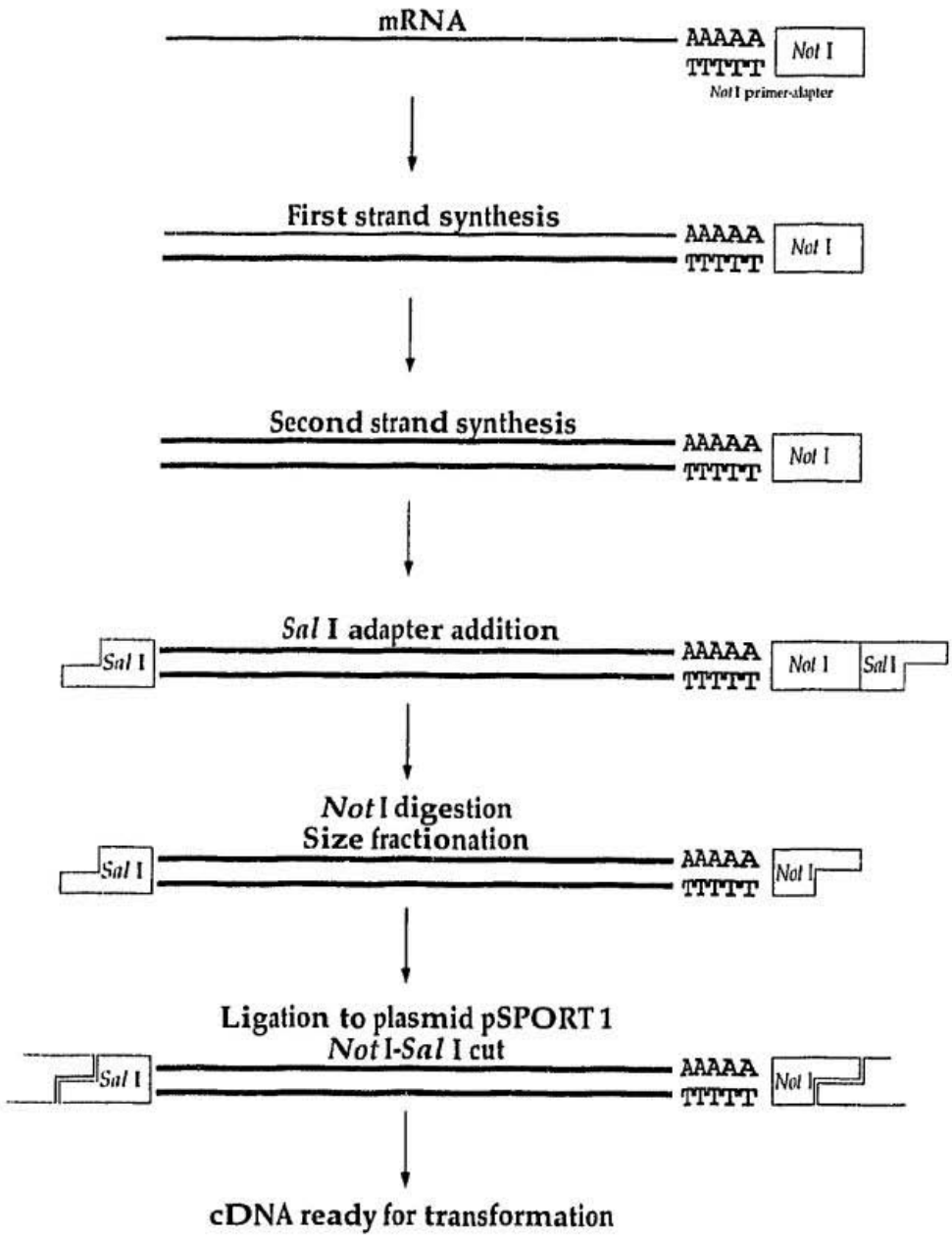

mRNA
AAAAA
TTTTT
Not I
Not I primer-adapter
First strand synthesis
AAAAA
TTTTT
Not I
Second strand synthesis
AAAAA
TTTTT
Not I
Sal I adapter addition
Sal I
AAAAA
TTTTT
Not I
Sal I
Not I digestion
Size fractionation
Sal I
AAAAA
TTTTT
Not I
Ligation to plasmid pSPORT 1
Not I-Sal I cut
Sal I
AAAAA
TTTTT
Not I
cDNA ready for transformation

Fig. 2.2. Schematic representation of a standard amplification cycle used in the polymerase chain reaction. The temperatures and times shown for each step are unique for each set of primers and template.

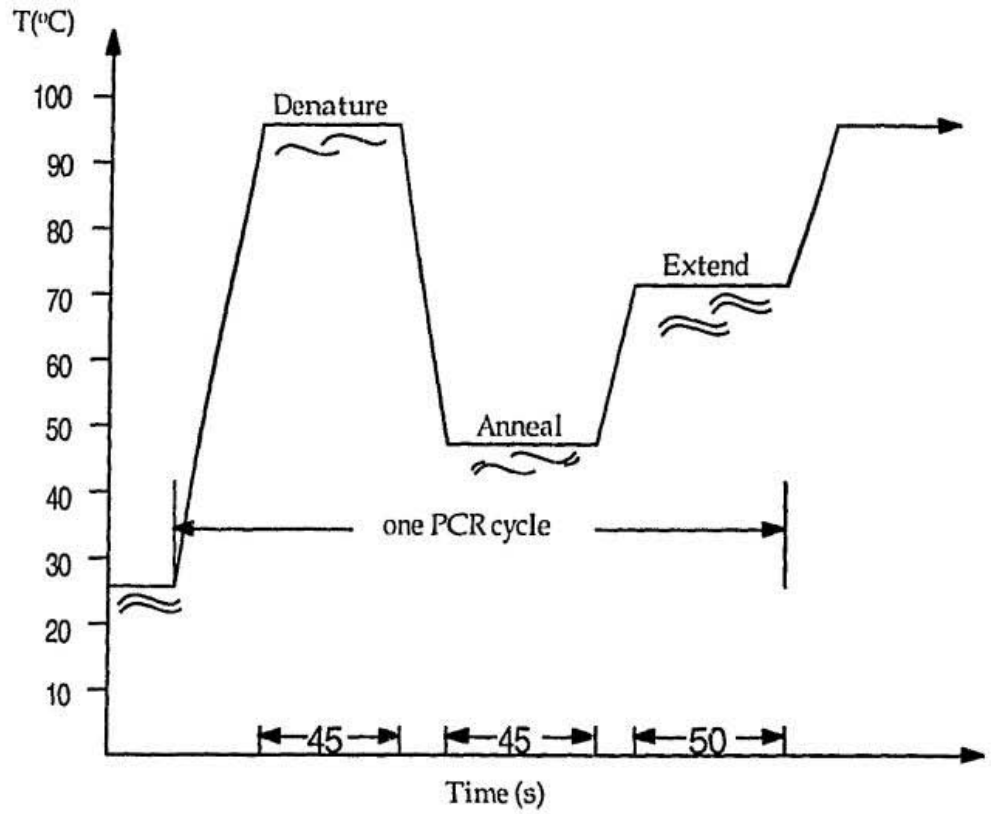

T(°C)
100
90
80
70
60
50
40
30
20
10
Denature
Extend
Anneal
one PCR cycle
45
45
50
Time (s)

Fig. 2.3. Alignment of apoA-I cDNA sequences of brown trout (BT), Atlantic salmon (AS), and rainbow trout (RT) showing the primer designed to sequence over the unknown region of pBTLB27. The sequence obtained revealed an 87bp deletion as shown with a '-'. The nucleotides in the brown trout sequence are numbered from the 5' end of the cDNA sequence.

```
BT   CGTCTGGAGGAGCTGAGGACCCTGGCCGCCCCCTATGCTGAGGAGTACAAGGAGC----------------  662
AS   CGTCTGGAGGAGCTGAGGACCCTGGCAGCCCCCTACGCTGAGGAGTACAAGGAGCAGATGTTCAAGGCTGTT
RT   CGTCTGGAGGAGCTGAGGACCCTGGCCGCCCCCTACGCTGAGGAGTACAAGGAGCAGATGATCAAGGCTGTT

BT   ----------------------------------------------------------------AG  791
AS   GGAGAGGTGCGTGAGAAGGTGGCTCCCCTGTCTGAGGACTTCAAGG-CCAGATGGGCCCCGCCGCCGAGGAG
RT   GGAGAGGTGCGTGAGAAGGTGTCTCCCCTGTCTGAGGACTTCAAGGGCCAGGTGGGCCCCGCAGCCGAACAG

BT   GCCAAGGAAAAGCTCATGGATTTCTACCAGACCATCAGCCAGGCCAT                        734
AS   GCCAAGCAAAAGCTCCTGGCTCTCTACGAGACCATCAGCCAGGCCAT
RT   GCCAAGCAGAAGCTCCTGGCTTTCTACGAGACCATCAGCCAGGCCAT
                                  Primer 65
```

Fig 2.4. Schematic representation of the delipidation of apolipoprotein particles. The procedure outlined is used for each mL of lipoprotein.

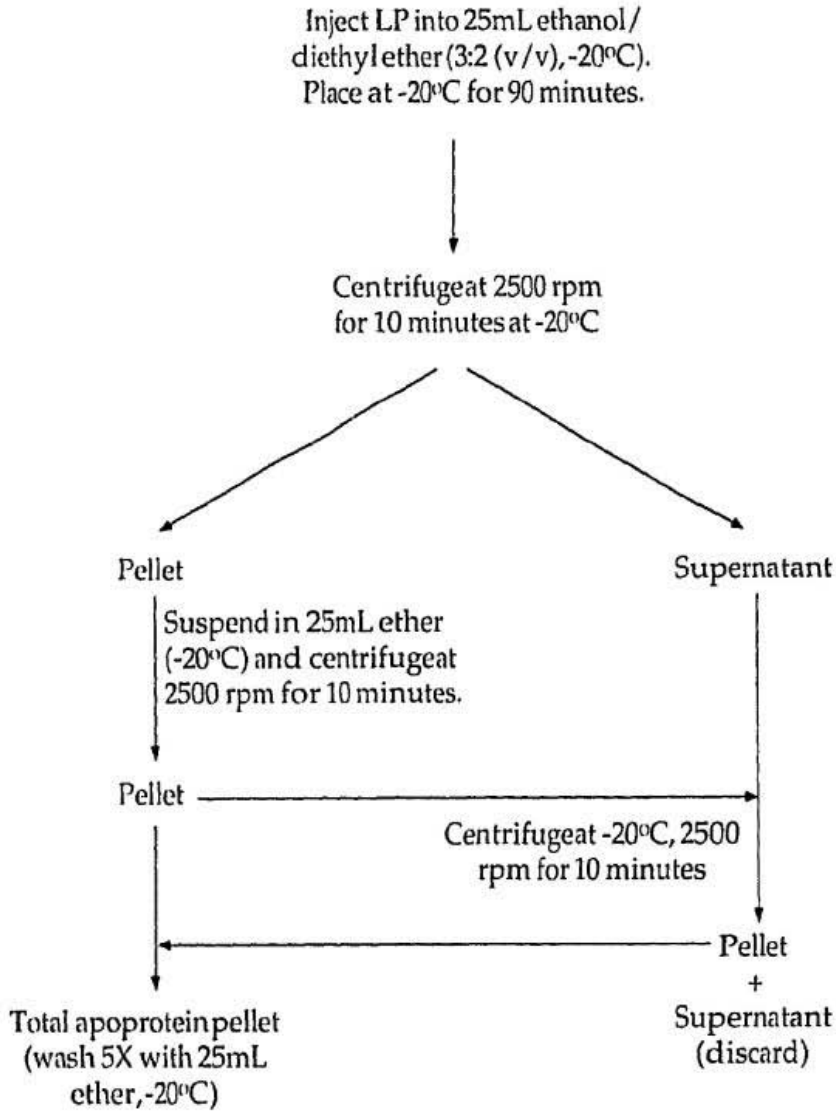

Inject LP into 25mL ethanol/
diethyl ether (3:2 (v/v), -20°C).
Place at -20°C for 90 minutes.
Centrifuge at 2500 rpm
for 10 minutes at -20°C
Pellet
Supernatant
Suspend in 25mL ether
(-20°C) and centrifuge at
2500 rpm for 10 minutes.
Pellet
Centrifuge at -20°C, 2500
rpm for 10 minutes
Pellet
+
Supernatant
(discard)
Total apoprotein pellet
(wash 5X with 25mL
ether, -20°C)

Fig. 2.5. SDS-PAGE analysis of protein isolated from HDL (following delipidation). Duplicate samples from human and brown trout HDL were loaded with human apoA-I (HS; Sigma) as well as a protein marker (M, Electrophoresis Calibration kit, Pharmacia; Phosphorylase b, Bovine Serum Albumin, Ovalbumin, Carbonic Anhydrase, Soybean Trypsin Inhibitor, and μ-Lactalbumin). The molecular weight (Da) of each marker protein is shown.

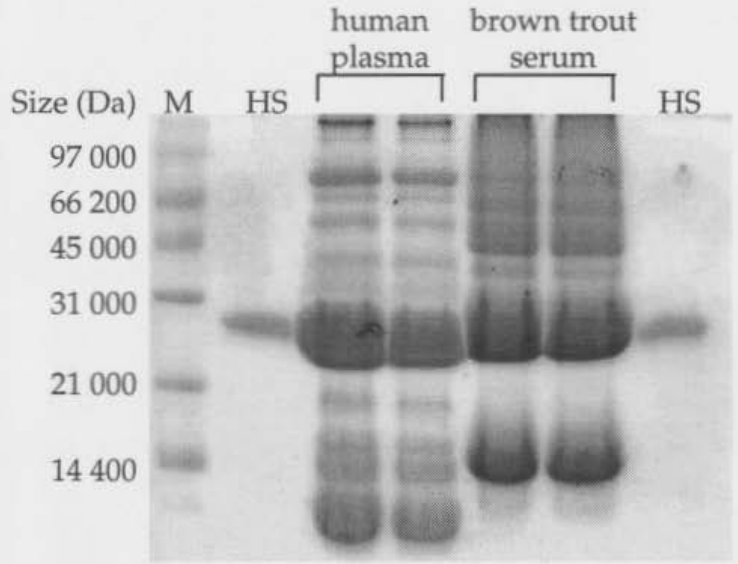

human plasma
brown trout serum
Size (Da)
M
HS
HS
97 000
66 200
45 000
31 000
21 000
14 400

Fig. 2.6. Semi-quantitative ion-exchange of brown trout apoA-I from DEAE-sephadex with NaCl. After incubation of the DEAE-Sephadex wth brown trout HDL protein extract, the resin was pelleted and the supernatant was examined with SDS-PAGE. HS, human apoA-I (Sigma).

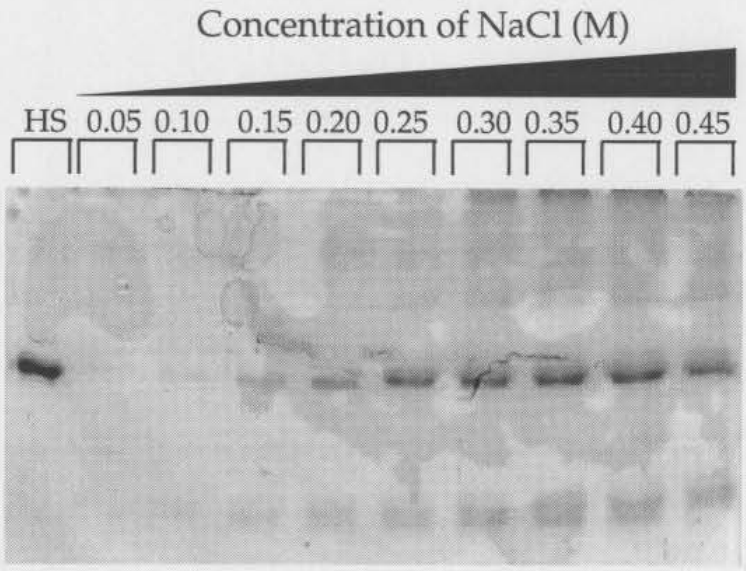

Concentration of NaCl (M)
HS
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45

Fig. 2.7. Amplification of the 87bp deletion region in three salmonids using primers 63 and 65. Three genomic templates, brown trout (BT), rainbow trout (RT), and Atlantic salmon (AS), were amplified. As a positive control, pBTLB27 was also used as a template (+ve). All amplifications and a size marker (M, ØX174 DNA digested with *Hae* III) were then fractionated through Nuseive agarose, stained with ethidium bromide, and photographed under U.V. light. Note: A negative control reaction (lacking template DNA) was performed but no bands were observed.

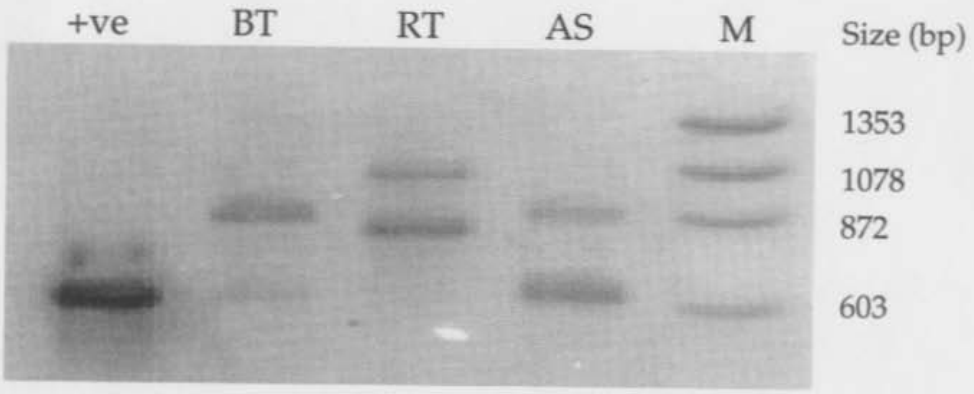
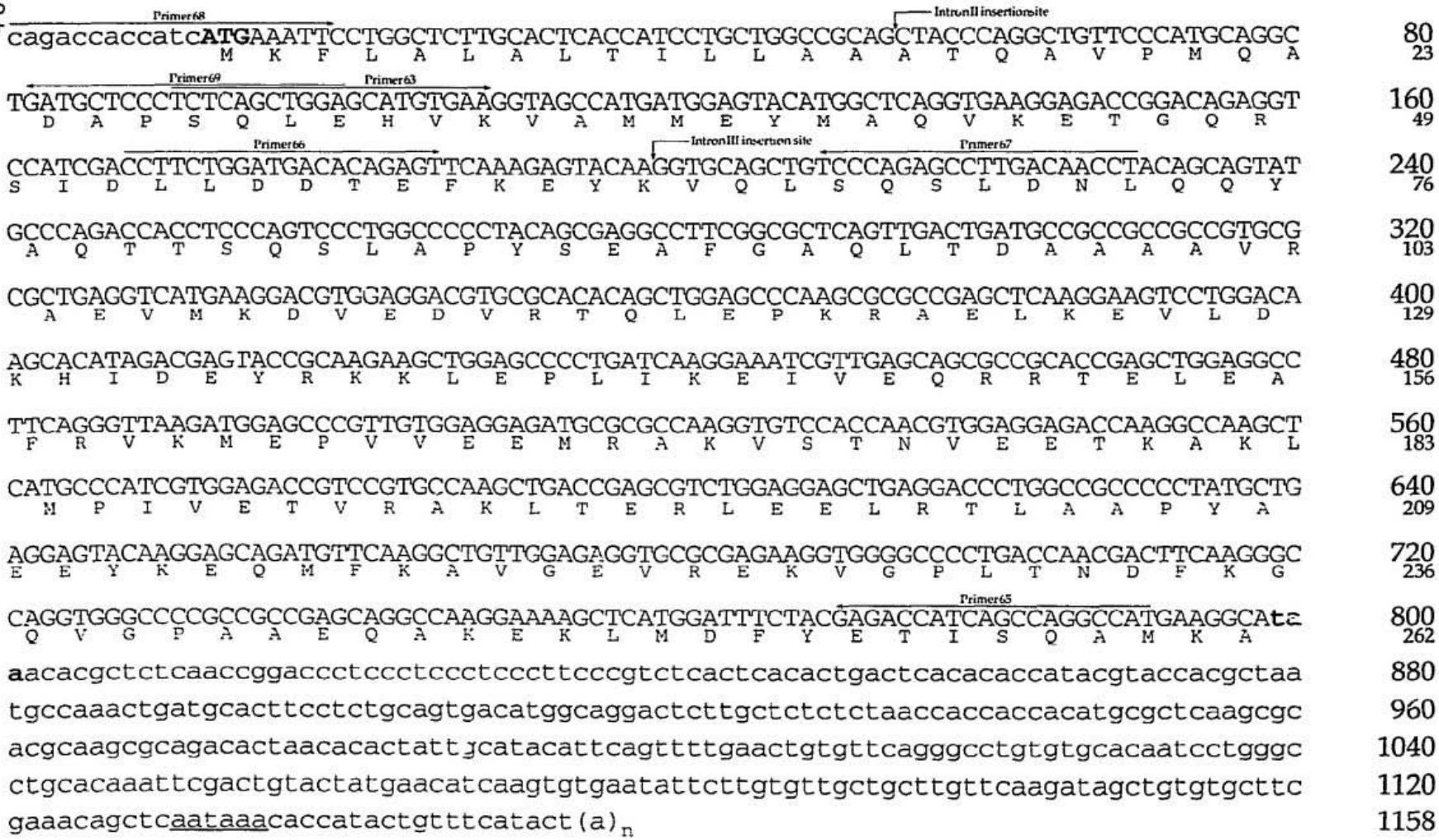

+ve
BT
RT
AS
M
Size (bp)
1353
1078
872
603

Fig. 2.8. The complete cDNA sequence of apoA-I in brown trout, *Salmo trutta*. The cDNA and inferred amino acid sequences are shown. The initiation and stop codons are shown in bold type-face. The polyadenylation signal is underlined. Primer sites are designated with straight arrow and intron/exon boundaries are indicated with a right-angled arrow.




```
cagaccaccatcATGAAATTCCTGGCTCTTGCACTCACCATCCTGCTGGCCGCAGCTACCCAGGCTGTTCCCATGCAGGC      80
          M  K  F  L  A  L  A  L  T  I  L  L  A  A  A  T  Q  A  V  P  M  Q  A          23

TGATGCTCCCTCTCAGCTGGAGCATGTGAAGGTAGCCATGATGGAGTACATGGCTCAGGTGAAGGAGACCGGACAGAGGT     160
  D  A  P  S  Q  L  E  H  V  K  V  A  M  M  E  Y  M  A  Q  V  K  E  T  G  Q  R        49

CCATCGACCTTCTGGATGACACAGAGTTCAAAGAGTACAAGGTGCAGCTGTCCCAGAGCCTTGACAACCTACAGCAGTAT     240
S  I  D  L  L  D  D  T  E  F  K  E  Y  K  V  Q  L  S  Q  S  L  D  N  L  Q  Q  Y      76

GCCCAGACCACCTCCCAGTCCCTGGCCCCCTACAGCGAGGCCTTCGGCGCTCAGTTGACTGATGCCGCCGCCGCCGTGCG     320
  A  Q  T  T  S  Q  S  L  A  P  Y  S  E  A  F  G  A  Q  L  T  D  A  A  A  A  V  R    103

CGCTGAGGTCATGAAGGACGTGGAGGACGTGCGCACACAGCTGGAGCCCAAGCGCGCCGAGCTCAAGGAAGTCCTGGACA     400
  A  E  V  M  K  D  V  E  D  V  R  T  Q  L  E  P  K  R  A  E  L  K  E  V  L  D      129

AGCACATAGACGAGTACCGCAAGAAGCTGGAGCCCCTGATCAAGGAAATCGTTGAGCAGCGCCGCACCGAGCTGGAGGCC     480
K  H  I  D  E  Y  R  K  K  L  E  P  L  I  K  E  I  V  E  Q  R  R  T  E  L  E  A      156

TTCAGGGTTAAGATGGAGCCCGTTGTGGAGGAGATGCGCGCCAAGGTGTCCACCAACGTGGAGGAGACCAAGGCCAAGCT     560
  F  R  V  K  M  E  P  V  V  E  E  M  R  A  K  V  S  T  N  V  E  E  T  K  A  K  L    183

CATGCCCATCGTGGAGACCGTCCGTGCCAAGCTGACCGAGCGTCTGGAGGAGCTGAGGACCCTGGCCGCCCCCTATGCTG     640
  M  P  I  V  E  T  V  R  A  K  L  T  E  R  L  E  E  L  R  T  L  A  A  P  Y  A      209

AGGAGTACAAGGAGCAGATGTTCAAGGCTGTTGGAGAGGTGCCGCGAGAAGGTGGGGCCCCTGACCAACGACTTCAAGGGC     720
E  E  Y  K  E  Q  M  F  K  A  V  G  E  V  R  E  K  V  G  P  L  T  N  D  F  K  G      236

CAGGTGGGCCCCGCCGCCGAGCAGGCCAAGGAAAAGCTCATGGATTTCTACGAGACCATCAGCCAGGCCATGAAGGCAta     800
  Q  V  G  P  A  A  E  Q  A  K  E  K  L  M  D  F  Y  E  T  I  S  Q  A  M  K  A      262

aacacgctctcaaccggaccctccctccctccctttcccgtctcactcacactgactcacacaccatacgtaccacgctaa     880

tgccaaactgatgcacttcctctgcagtgacatggcaggactcttgctctctctaaccaccaccacatgcgctcaagcgc     960

acgcaagcgcagacactaacacactattgcatacattcagttttgaactgtgttcagggcctgtgtgcacaatcctgggc    1040

ctgcacaaattcgactgtactatgaacatcaagtgtgaatattcttgtgttgctgcttgttcaagatagctgtgtgcttc    1120

gaaacagctcaataaacaccatactgtttcatact(a)n                                              1158
```

Fig 2.9. Alignment of human (top sequence) and brown trout (bottom sequence) apoA-I amino acid sequences. The three functional regions are shown: the leader peptide (residue -24 to -1 w.r.t the human sequence), the N-terminal 33 amino acid conserved peptide (residues 11 to 43), and the lipid binding domain (residues 44 to 241). The latter is separated into 22 or 11 amino acid repeats. Amino acids 1-10 have not been assigned a particular function.

```
-24                              +1
MKAAVLTLAVLFLTGSQARHFWQQDEPPQSPWDR          10
MKFLALALTILLAAATQAVPM-QADAPSQ--LEH           8

VKDLATVYVDVLKDSGRDYVSQFEGSALGKQLN           43
VKVAMMEYMAQVKETGQRSIDLLDDTEF-KEYK           40

LKLLDNWDSVTSTFSKLREQLG                      65
VQLSQSLDNLQQYAQTTSQSLA                      62

PVTQEFWDNLEKETEGLRQEMS                      87
PYSEAFGAQLTDAAAAVRAEVM                      84

              KDLEEVKAKVQ                   98
              KDVEDVRTQLE                   95

PYLDDFQKKWQEEMELYRQKVE                     120
PKRAELKEVLDKHIDEYRKKLE                     117

PLRAELQEGARQKLHELQEKLS                     142
PLIKEIVEQRRTELEAFRVKME                     139

PLGEEMRDRARAHVDALRTHLA                     164
PVVEEMRAKVSTNVEETKAKLM                     161

PYSDELRQRLAARLEALKENGG                     186
PIVETVRAKLTERLEELRTLAA                     183

ARLAEYHAKATEHLSTLSEKAK                     208
PYAEEYKEQMFKAVGEVREKVG                     205

PALEDLRQGLL                                219
PLTNDFKGQVG                                216

PVLESFKVSFLSALEEYTKKLNTQ                   243
PAAEQAKEKLMDFYETISQAMKA-                   239
```

Chapter III

Isolation and Characterization of apoA-I Intervening Sequences II and III in Three Salmonids

3.1 INTRODUCTION

3.1.1 The Origins and Evolution of Intervening Sequences.

If the structure of a typical gene is examined among various genomes, there is a clear discrepancy: eubacteria, archaebacteria, and organelles generally lack intron sequences such that the genes consist of continuous stretches of DNA, which are directly related to the amino acid sequence, whereas most eukaryotes exhibit genes interrupted by stretches of non-coding intron DNA (Holland and Blake, 1990; Palmer and Logsdon, 1991). The lack of colinearity between the gene and amino acid sequence in eukaryotes was in itself a startling discovery (Breathnach *et al.*, 1977; Breathnach and Chambon, 1981) but posed an interesting question: did the ancestor to prokaryotes/eukaryotes contain introns which have subsequently been lost in the bacterial lineage or have eukaryotes gained introns since the divergence of the bacterial lineage? The former theory is referred to as the 'introns-early' theory or the exon theory of genes while the latter is called the 'introns-late' theory.

Each of these theories have been very controversial and the issue remains unresolved. However, one major argument, that of exon shuffling, has risen above all others and is recognized here. The concept of exon shuffling was first proposed when it was observed that most proteins containing more than 200 amino acids consisted of two or more structural domains (Holland and Blake, 1990). For instance, one domain might be a binding motif while the other domain might contain a catalytic site. The arrangement of the binding motif and the catalytic site would be such that both would occur at the interface of the two structural domains. This type of

arrangement is seen in the various dehydrogenases which have similar binding properties but different catalytic activities (Holland and Blake, 1990). It thus seemed intuitive that two smaller proteins had been joined to produce one large protein through a mechanism involving gene fusion. Although this type of protein production would be very advantageous from an evolutionary perspective, a reasonable mechanism was not available. Gene duplication followed by recombination seemed reasonable except that there would be a high probability of frame-shift mutations. The discovery of introns in eukaryotes provided this theory with the flexibility it required. If crossing over occurred between the introns of the two genes, then the possibility of frame-shift mutations would be greatly decreased (Holland and Blake, 1990).

Exon shuffling in and of itself does not provide evidence for either the early or late theory of intron origin. However exon shuffling, as evidenced by the placement of introns between structural/functional domains, has been observed in a number of proteins such as pyruvate kinase, phosphoglyceratate kinase, and various dehydrogenases. Given the age of such glycolytic and other enzymes, this provides strong support for the early existence of introns. Although this concept met with resistance from proponents of the intron late theory, it continued to have a strong influence on the issue of intron origin until 1994. Stoltzfus *et al.* (1994) tested the theory that introns divided proteins into domains that were autonomous and could independently fold into domains or smaller units of protein structure such as secondary structures and motifs. They detected no evidence of a correspondence between exons and protein structure in four ancient proteins,

which implied that the exon theory of genes is unfounded. Thus, the role of introns in evolution by promoting exon shuffling seems to be restricted to relatively modern proteins.

3.1.2 Intron Mapping Using PCR.

Once the cDNA sequence of a gene has been determined, the location of the individual exons within the entire gene can also be determined if genomic clones for the gene are available. This PCR-based technique is referred to as exon-mapping (Niu and Crouse, 1993). Subclones of the genomic clones are typically prepared in an appropriate vector and exon-specific primers are used with vector-specific primers to map the position of the exon within the fragment as well as provide sequence information concerning the intron/exon boundary. Inherent in this technique is an *a priori* knowledge of the exon structure of the gene such that appropriate primers can be made complementary to the cDNA sequence in places that will represent each of the exons. One source of such information would be the sequence of the same gene in a closely related species.

If the cDNA sequence is known and additional information about the presence of introns is available, intron mapping can also be performed in a similar fashion. In the case of apoA-I, a number of gene sequences have been published (Table 1.2) and the presence/sequence of the introns determined. A comparison of these sequences revealed that the intron/exon insertion site in each sequence was conserved. Because this was true for the distantly related human and chicken sequence, it was hypothesized that it would hold true in fish as well. To determine the insertion site of intron II and III in the

salmonids, the cDNA sequences of human, chicken, and brown trout were aligned (see section 4.1 for a discussion on sequence alignment). The point of intron insertion in the human and chicken was then noted in the brown trout sequence. Two pairs of primers were designed to flank each of the putative intron insertion sites (see Fig. 2.8). Unlike exon mapping, this technique uses genomic DNA as a template rather than genomic subclones. As a result, only IVS for which both flanking regions are known will be able to be amplified. In the case of apoA-I, IVS I is in the 5'-untranslated region. To determine this sequence, a genomic library would have to be constructed. IVS I could then be isolated by using a vector-specific primer and a cDNA derived primer to amplify the entire 5' region. Alternatively, 5' RACE (Rapid Amplification of cDNA Ends) could be used to determine more of the 5' apoA-I sequence. PCR analysis of genomic DNA could then yield a fragment containing the intron.

In salmonids, intron and exon mapping is an important concern. Following the genome duplication event, many genes became functionally redundant. This allowed mutations to be incorporated at one locus while the other retained the original structure and function. The loci now have two possible fates: duplicate expression will either be lost or retained (Allendorf and Thorgaard, 1984). The former is quite common in salmonids as approximately 50% of the duplicated loci do not produce detectable protein products (Allendorf and Thorgaard, 1984). If the expression of the loci is not affected, additional fates are possible (Fig. 3.1). One way to detect both types of changes would be to examine the intron/exon structure of genes at different

loci. By specifically amplifying fragments from each locus of a gene, it may be possible to detect deletion/insertion mutations.

3.2 EXPERIMENTAL AND DISCUSSION

3.2.1 Isolation and Sequencing of apoA-I: IVS II.

Using primers 68 and 69 (see Fig. 2.8), apoA-I intervening sequence II (IVS II) was amplified from brown trout, rainbow trout, and Atlantic salmon genomic DNA and fractionated through Nuseive agarose as described above (see section 2.2.3; primers were annealed at 56°C). In each case, two to three bands of varying intensities were produced between 200bp and 300bp (data not shown). To obtain better resolution, the reactions were repeated and fractionated through a standard sequencing gel. In addition to the typical PCR reaction components, primer 66 was end-labeled with ^{32}P (as described in the *fmol* DNA Sequencing System, Promega) and mixed with the other reaction components (the volume of water used was adjusted to retain a final volume of 10µL). Thus, a population of the PCR products will be labeled with ^{32}P. Following amplification, each reaction was diluted with 5µL of Sequencing Stop Solution (Promega), denatured at 80°C, and immediately placed on ice. A 3µL aliquot of each reaction was then loaded onto a standard DNA sequencing gel and fractionated for ~3.5 hours at 40 watts. The gel was fixed and dried and exposed to X-ray film for 60 hours. In the reactions which previously produced three products on the agarose gel, only two products corresponding to ~230bp and ~270bp (~230bp and ~285bp in rainbow trout) were produced for each template (data not shown). To sequence the genes, the six products were excised from the gel, purified, and sequenced (cycle sequencing as described in section 2.2.3) using primers 68 and 69. The intron

sequences (apoA-I: IVS II) were identified using the 5'-GT and 3'-AG splice site rules (Lewin, 1994) and are aligned in Fig. 3.2(A)[6].

The most obvious characteristic of these sequences is the 72bp deletion[7] present in the 'lower' set of sequences (nucleotides 13 to 84 with respect to the upper rainbow trout sequence). Analysis of this region reveals some putative transcriptional elements. Although little data are available with respect to transcriptional regulators in teleosts, a number of elements similar to mammalian regulators have been identified. A TATA box, normally found upstream of the transcription start site, is shown. A number of TATA boxes have been identified within introns, but none have exhibited transcriptional activity (Higashimoto and Liddle, 1993; Seib *et al.*, 1994). A short $(dG{\cdot}dT)_n$ repeat is also shown. A region of alternating purines and pyrimidines (which have the potential to form Z-DNA structures) is found in many promoters and is thought to be involved in transcriptional regulation (van Holde and Zlatanova, 1993; Naylor and Clark, 1990; Kuczek and Rogers, 1987). Also, putative binding sites for the *fos-jun* heterodimer (AP-1, Rauscher *et al.*, 1988) and the Myb protein (MRE, Seib *et al.*, 1994) are shown. The possible existence of these types of transcriptional regulators in one set of sequences and not the other provides a possible explanation for the differential expression of two apoA-I cDNAs in rainbow trout. Since AP-1 and MRE are thought to be involved in certain types of cancer, it is possible that regulatory proteins bind to these putative elements to produce the differential expression of the two

---

[6]Genbank Accession numbers for each sequence are L49427 (RT Upper), L49426 (RT Lower), L49425 (BT Upper), L49424 (BT Lower), L49423 (AS Upper), and L49413 (AS Lower).

[7]Deletion is used in reference to other sequences such that there is no ambiguity as to which region of the sequence is being referred to. Sequence information of the region of interest must be available from an ancestor to the salmonids to determine if a DNA fragment is (specifically) an insertion or a deletion.

apoA-I cDNAs. The physiological importance of this phenomenon remains unclear as the inferred amino acid sequences of apoA-I-1 and apoA-I-2 are very similar.

From an evolutionary perspective, IVS II is also very informative. The large deletion in the lower set of sequences is present in all three species. The 7bp deletion in the upper set of sequences (nucleotides 107 to 113 with respect to the lower rainbow trout sequence) is also present in all three species. These deletions most likely occurred prior to the divergence of *Oncorhynchus* from *Salmo*. In contrast, the 18bp deletion in the upper *Salmo* sequences (nucleotides 164 to 181 with respect to the upper rainbow trout sequence) is not present in rainbow trout. This change would have occurred after *Oncorhynchus* diverged from *Salmo* but before the speciation of Atlantic salmon and brown trout. Given the relatively high degree of sequence variation (i.e. insertions and deletions) between the upper and lower sequences, it is likely that they have been amplified from different loci. This is consistent with the tetraploid genome of salmonids and the fact that the primers were designed over coding sequences that are highly conserved (as evidenced by the two sequences reported for rainbow trout apoA-I). However, we expect each locus to have two alleles. The amplification of just two products would suggest that identical alleles are present at each locus. Fig. 3.3 illustrates the possible evolutionary changes that have occurred at both loci of apoA-I: IVS II.

3.2.2 Isolation, Sequencing, and Pedigree analysis of apoA-I: IVS III.

Using primers 66 and 67 (see Fig. 2.8), IVS III was amplified (as described above) from genomic DNA isolated from brown trout, rainbow trout, and Atlantic salmon. For each template, three reactions were prepared and each amplified separately at 48°C, 51°C, and 54°C. As shown in Fig. 3.4, the positive control (pBTLB27 template) shows a single product with a high concentration at both 51°C and 54°C. Although the marker DNA used does not give visible bands in this range, the cDNA sequence predicts a fragment of 101bp. The reactions using genomic DNA produced bands larger than the positive control only when the primers were annealed at 54°C. In the brown trout reaction, one major product larger than the positive control was observed. Rainbow trout and Atlantic salmon genomic templates also produced one major product larger than the positive control, but a few faint secondary products were also observed. When the amplifications were repeated at 56°C, only one major product larger than the positive control was produced for each template (Fig. 3.5). [This demonstrates the importance of optimizing the annealing temperature for a particular set of primers. The higher annealing temperature prevents the primers from annealing to secondary sites thereby preventing the amplification of the secondary bands seen at lower temperatures.]

The major product of each 56°C reaction was excised and purified in 50µL of water as described above. To obtain an adequate amount of product for cycle sequencing, each of these purified products was used as a template for re-amplification (Fig.3.5). As shown, single products were produced in each case. In comparison to the initial amplification, these products fluoresce with a greater intensity. These more concentrated products were then excised,

purified, and sequenced using both primers 66 and 67. The intron sequences (apoA-I: IVS III) are shown in Fig. 3.2(B)[8]. Perhaps the most striking feature of this sequence is the presence of a $(dT{\cdot}dC)_n$ microsatellite. Although the regions flanking the microsatellite were conserved in all three species, the variable nature of this repetitive element is obvious from the three sequences. The rainbow trout intron contained the smallest microsatellite (9 repeats) while the brown trout and Atlantic salmon introns were similar in size to each other but larger than the rainbow trout (30 and 32 repeats, respectively).

Only one product was amplified at IVS III when analyzed on an agarose gel. Given that two loci were amplified at IVS II, we expect to amplify two loci at IVS III as well. Although it is possible that the loci are identical, the presence of the microsatellite (I3M) within IVS III makes this an unlikely possibility. Rather, it is expected that small variations in the number of repeats at each allele of each locus are undectable on an agarose gel. Thus, the exact size of I3M was determined in a small family of individuals.

A female Atlantic salmon was crossed with a male brown trout and the hybrid offspring (fry) sacrificed immediately following yolk sac absorption. Genomic DNA was then prepared from blood samples from both parents and fry tissue (C. McGowan, unpublished results). DNA samples from both parents and six fry were then used as templates to amplify IVS III. Using ^{32}P end-labelled primer 66, the products were fractionated through a standard sequencing gel. The developed gel is shown in Fig. 3.6.

---

[8]Genbank Accession numbers for the sequences are L49430 (RT), L49429 (BT), and L49428 (AS).

With respect to the size of I3M, two products of 217bp and 205 bp are amplified from the Atlantic salmon female. We see that each of the hybrid offspring inherit only one of these two products. This pattern of independent segregation would only occur if the inheritance of these two fragments was disomic (i.e. the two fragments are alleles of the same locus). If these fragments were amplified from separate loci, each offspring could contain both fragments. Thus, the two products formed are from the same locus. As stated above, the primers used are designed over conserved coding sequences and both loci would be expected to amplify. These results suggest that one locus of apoA-I has undergone mutations which prevent the amplification of IVS III.

In brown trout, one product of 213bp is amplified. As in the Atlantic salmon, two alleles of one locus are being amplified. However, the microsatellite at each allele is identical in length and the independent segregation is not possible to discern. However, the intensity of the 213bp product is much more intense (i.e. higher copy number) than the products seen in the offspring, which suggests that the parental band was produced from the combined amplification of two identical alleles.

As described above, two alleles from one locus are amplified by primers 66 and 67. In reference to the Atlantic salmon female, a disomic pattern of inheritance would predict that 50% of the offspring would inherit the 217bp 'slow' allele while the other 50% would inherit the 205bp 'fast' allele. To test this prediction, 47 hybrid individuals were typed with respect to the size of the I3M. Surprisingly, only 17 inherit the slow allele while the remaining 30 inherit the fast allele (data not shown). Although this result is unexpected,

the sample size is too small to make any confident conclusions ($0.95<[\chi^2=3.6]>0.05$) The pattern of inheritance of these alleles was compared to the patterns of a number of variable RAPD (Randomly Amplified Polymorphic DNA) markers in salmon (Colin McGowan, unpublished results). Although the locus was not linked to any of the markers examined, I3M is an important marker for gene mapping in salmonids. The importance of I3M is that it is contained within an expressed gene and the primers are complementary to the coding region. Unlike many of the random markers, I3M can be used in comparitive gene mapping. For example, the inheritance pattern can be examined in a number of species. Because the same locus is being typed in each of the species, it will be possible to determine whether the genomic structure has been conserved.

One problem with this technique is illustrated by the number of bands produced for each allele. In each reaction shown in Fig. 3.6, 3-4 faint bands (ghost bands) can be seen below each of the highest, most intense bands. This phenomenon has been well documented when amplifying microsatellites (Weber and May, 1989; Estoup *et al.*, 1993; Taylor *et al.*, 1994). It is thought that *Taq* polymerase has difficulties amplifying the repetitive elements *in vitro* and may actually 'slip' along the microsatellite region of the DNA. The fact that each ghost band is 2bp apart from any other ghost band and that the microsatellite consists of a dinucleotide repeat implies that 3-4 repeats are being slipped over. In any case, the highest, most intense band is typed as the true allele size.

3.2.3 Detection of Insertion and Deletion Mutations in apoA-I loci.

To examine the structure of the apoA-I loci more closely, two primers were designed that were specific for either the upper or lower locus (see Fig. 3.2(A)). Primer 68Up was designed in the region corresponding to the 72bp deletion of the lower locus and was thus specific for the upper locus. Primer 68Lo was designed to flank the deletion site on both ends and was specific for the lower locus. PCR reactions were then prepared and amplified as above using brown trout, Atlantic salmon, and rainbow trout genomic templates (primers were annealed at 56°C). Each reaction was then analyzed by gel electrophoresis through Nuseive agarose.

To show that these primers were specific for the separate loci, Primers 68Up and 68Lo were used with primer 69 to amplify IVS II. As shown in Fig. 3.7, primer 68Lo specifically amplified one product which migrated just below the 194bp marker. The expected size of this fragment is 172bp. Each template produced a fragment of the same size which is also expected from the sequence. When primer 68Up was used, the primary product of the brown trout and Atlantic salmon templates migrated very close to the 234bp marker. The expected size of this fragment was 231bp. The rainbow trout template produced a fragment which was a little larger than the brown trout and salmon fragments. This is due to the 18bp deletion in the upper brown trout and Atlantic salmon sequence. Clearly, these primers are specifically amplifying one locus.

Given the specificity of the primers, it was possible to test for the presence of insertions and deletions at each locus separately. Evidence for such a scenario has been shown above. When primers 63 and 65 were used to determine the 87bp absent in the cDNA sequence, both brown trout and

Atlantic salmon templates produced fragments which were the same size as the pBTLB27 positive control template (Fig. 2.7). Also, when IVS III was amplified at different annealing temperatures, products similar and/or smaller in size than the positive control were observed (Fig. 3.4). Finally, only one locus is amplified using primers 66 and 67. These results indicate that the introns and perhaps some of the coding region have been deleted in at least one allele of one locus. Thus, each of the locus-specific primers was used with primer 65 to amplify through/across both introns and the region which contained the 87bp DNA deletion. The expected sizes of the products if no deletions are present are shown in Table 3.1.

When primer 68Lo and primer 65 were used, a product of 954-996bp was expected. As shown in Fig. 3.7, each template produced fragments which migrated just below the 1078bp marker. The product from the rainbow trout template migrates, as expected, a little faster than the products from the other two templates. Thus, it would seem that the lower apoA-I locus contains both introns and does not contain any cDNA deletions. This has been confirmed in brown trout and rainbow trout by DNA sequencing. These results are in sharp contrast to those obtained when primer 68Up and primer 65 was used. We would expect the products from the upper locus to be slightly larger than those produced from the lower locus. However, the most intense product produced from the Atlantic salmon and brown trout templates is approximately 400bp. This would imply that a large fragment of the coding region and perhaps IVS III have been deleted. When this experiment was repeated at various annealing temperatures, similar results were obtained. Sequence analysis of this fragment did not reveal a significant similarity to

Table 3.1. Expected size (in base pairs) of products produced from the amplification of an apoA-I fragment from using primer 65 and primer 68Up/68Lo.

Species	Primers	cDNA	IVS II	IVS III	Total(bp)
Atlantic salmon, brown trout	65/68Lo	716	126	154	996
Rainbow trout	65/68Lo	716	126	112	954
Atlantic salmon, brown trout	65/68Up	716	167	154	1037
Rainbow trout	65/68Up	716	185	112	1013

any part of the apoA-I cDNA or IVS sequence. The rainbow trout template produces different amplification products at the upper locus. When the primers were annealed at 56°C, no distinct bands were visible. When the primers were annealed at 60°C, a fragment of approximately 1000bp was observed (data not shown). Thus, the upper locus of brown trout and Atlantic salmon appears to have undergone some deletion mutations while a normal gene is present at the same locus in rainbow trout.

Although the observations indicate that the upper apoA-I locus has undergone major changes in gene structure, they are not conclusive. This is due primarily to the presence of secondary bands amplified by both locus-specific primers. This is clearly seen when the primers are used with primer 65. As shown in Fig 3.5, products of ~300bp are amplified with primer 68Lo while multiple products both larger and smaller than the primary product are amplified with primer 68Up. Thus, there may be mutations at both loci with a full-length, structural allele at the lower locus. Characterization of these secondary products as well as designing primers with a higher specificity would give a more detailed and accurate description of the apoA-I loci.

Fig. 3.1. Schematic diagram of the possible fates of duplicate genetic loci following a genome duplication. A representative allele is shown by lines (non-coding DNA) and open boxes (exons, E). Individual alleles at a single locus are designated by letters (i.e. A^1 and A^2 are alleles of locus A).

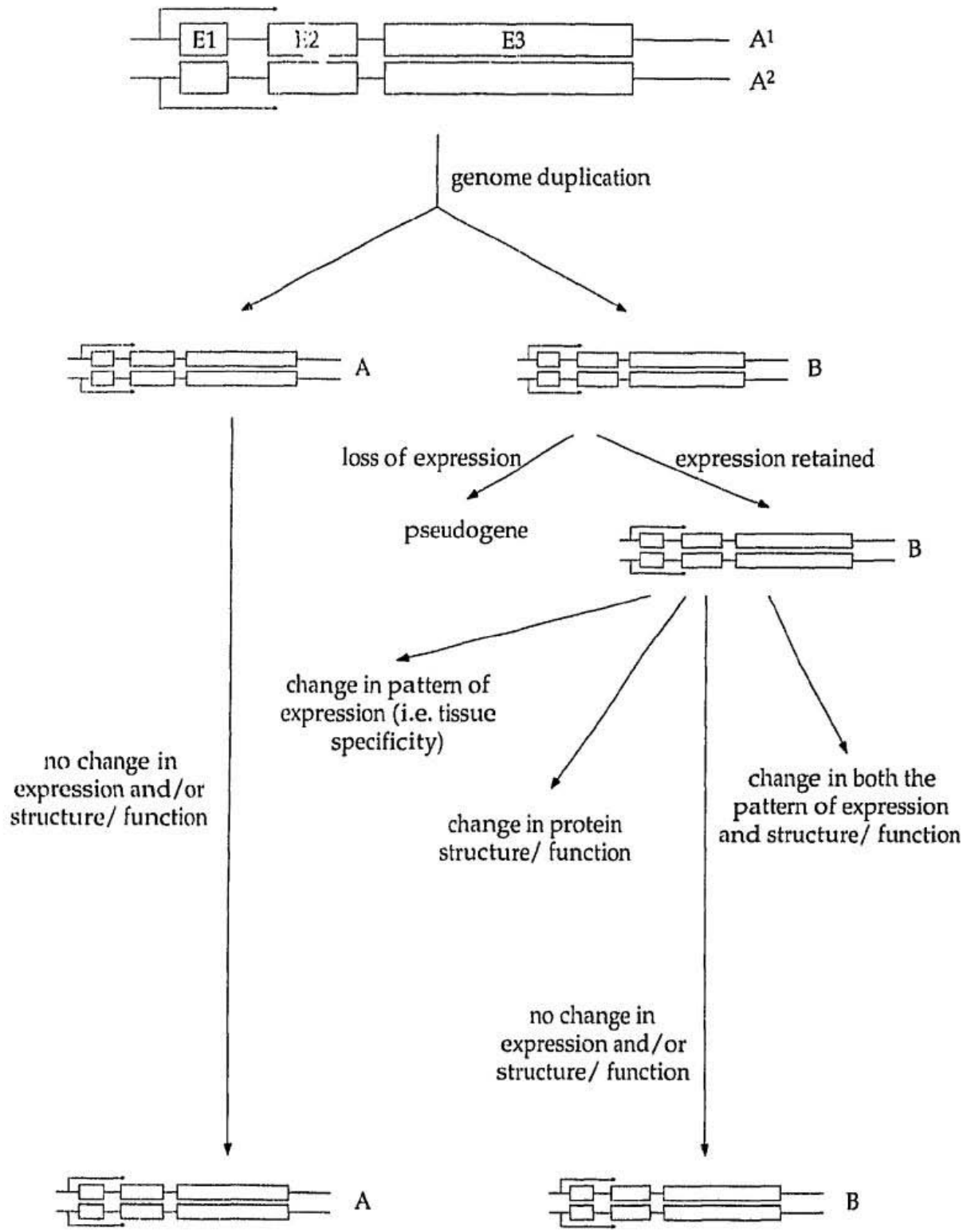

E1
E2
E3
A1
A2
genome duplication
A
B
loss of expression
expression retained
pseudogene
B
change in pattern of expression (i.e. tissue specificity)
no change in expression and/or structure/ function
change in protein structure/ function
change in both the pattern of expression and structure/ function
no change in expression and/or structure/ function
A
B

Fig. 3.2. Alignment of IVS II and III sequences from three salmonid species. Rainbow trout (RT), brown trout (BT), and Atlantic salmon (AS) sequences are shown. The 5'-GT and 3'-AG intron splice sites are shown in bold type-face and alignment gaps are designated with '-'. A, two IVS II sequences are shown for each species. The larger sequence is called 'upper' while the lower sequence (containing the large deletion) is called 'lower'. Putative transcriptional elements are also indicated; MRE; myb-responsive element; AP-1/AP-2, activator protein binding sites. The positions of locus specific primers (68Up, 68Lo) are indicated with arrows; B, the sequence of IVS III is shown. The microsatellite sequence is underlined. The other allele for this IVS differs only in the number of (TC) repeats.

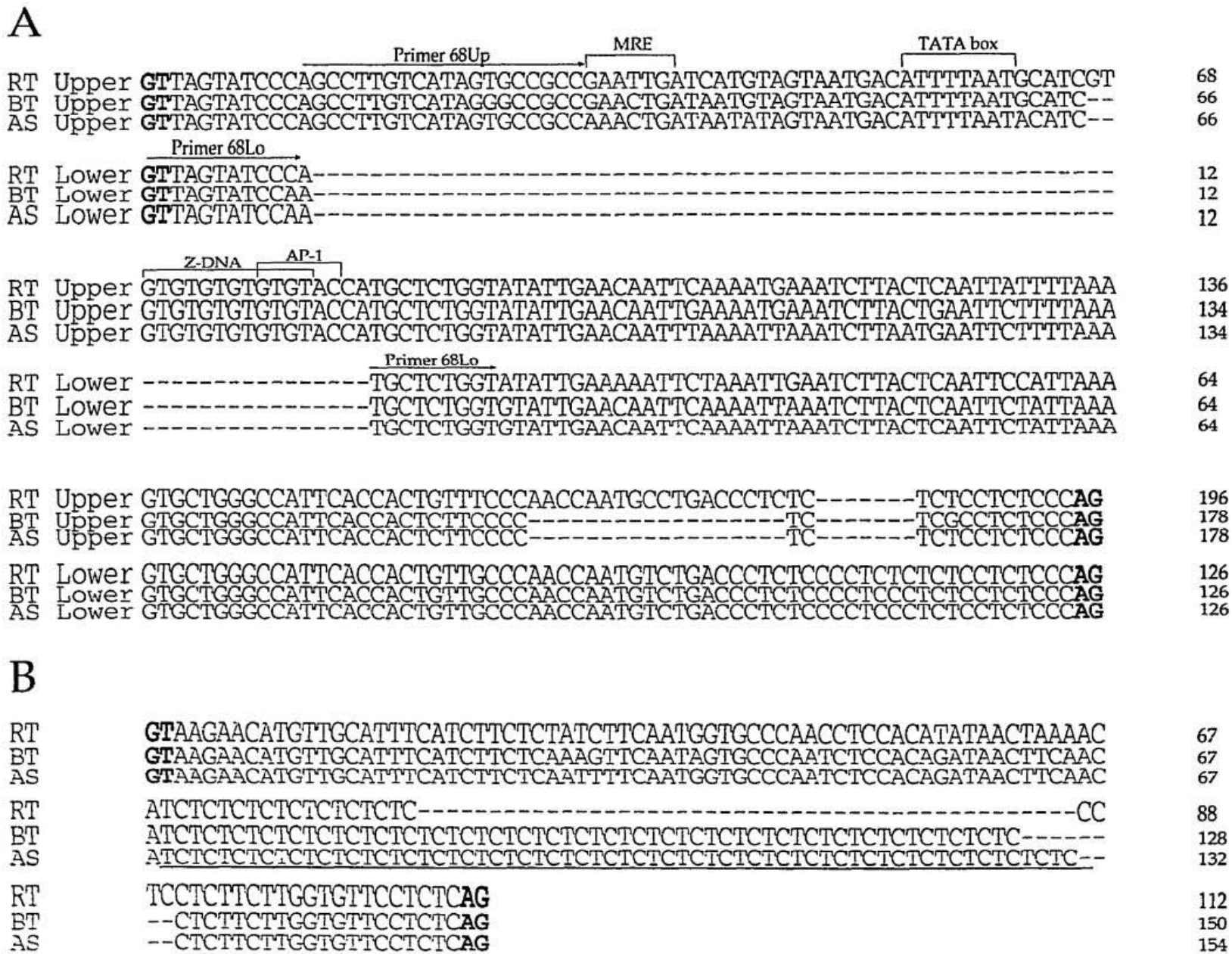
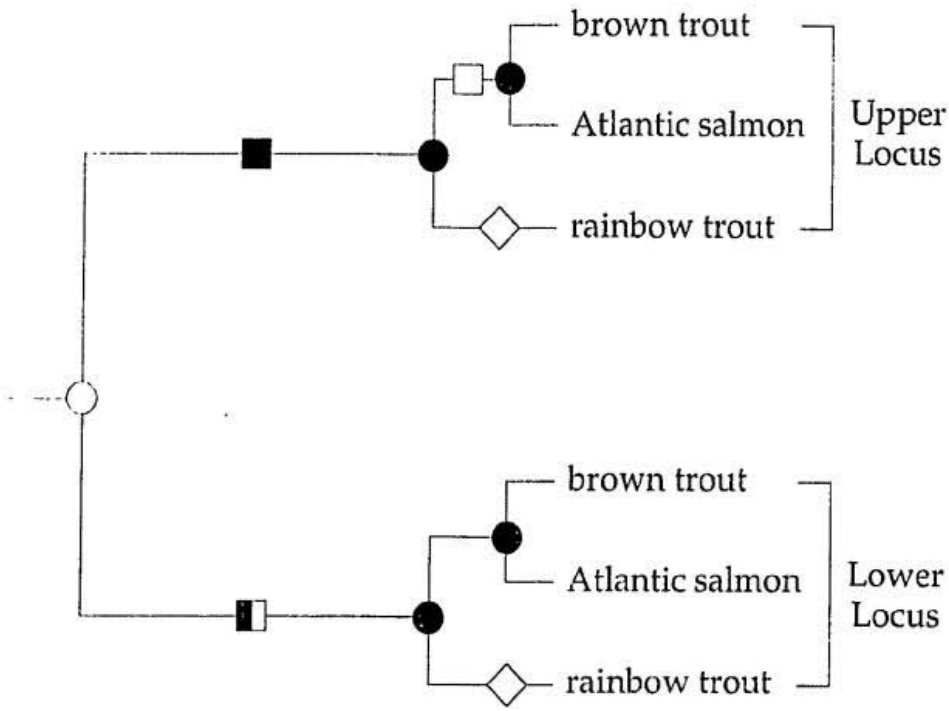

A
Primer 68Up
MRE
TATA box
RT Upper GTTAGTATCCCAGCCTTGTCATAGTGCCGCCGAATTGATCATGTAGTAATGACATTTTAATGCATCGT 68
BT Upper GTTAGTATCCCAGCCTTGTCATAGGGCCGCCGAACTGATAATGTAGTAATGACATTTTAATGCATC-- 66
AS Upper GTTAGTATCCCAGCCTTGTCATAGTGCCGCCAAACTGATAATATAGTAATGACATTTTAATACATC-- 66
Primer 68Lo
RT Lower GTTAGTATCCCA---------- 12
BT Lower GTTAGTATCCAA---------- 12
AS Lower GTTAGTATCCAA---------- 12
Z-DNA
AP-1
RT Upper GTGTGTGTGTGTACCATGCTCTGGTATATTGAACAATTCAAAATGAAATCTTACTCAATTATTTTAAA 136
BT Upper GTGTGTGTGTGTACCATGCTCTGGTATATTGAACAATTGAAAATGAAATCTTACTGAATTCTTTTAAA 134
AS Upper GTGTGTGTGTGTACCATGCTCTGGTATATTGAACAATTTAAAATTAAATCTTAATGAATTCTTTTAAA 134
Primer 68Lo
RT Lower ----------TGCTCTGGTATATTGAAAAATTCTAAATTGAATCTTACTCAATTCCATTAAA 64
BT Lower ----------TGCTCTGGTGTATTGAACAATTCAAAATTAAATCTTACTCAATTCTATTAAA 64
AS Lower ----------TGCTCTGGTGTATTGAACAATTCAAAATTAAATCTTACTCAATTCTATTAAA 64
RT Upper GTGCTGGGCCATTCACCACTGTTTCCCAACCAATGCCTGACCCTCTC-------TCTCCTCTCCCAG 196
BT Upper GTGCTGGGCCATTCACCACTCTTCCCC----------TC-------TCGCCTCTCCCAG 178
AS Upper GTGCTGGGCCATTCACCACTCTTCCCC----------TC-------TCTCCTCTCCCAG 178
RT Lower GTGCTGGGCCATTCACCACTGTTGCCCAACCAATGTCTGACCCTCTCCCCTCTCTCTCCTCTCCCAG 126
BT Lower GTGCTGGGCCATTCACCACTGTTGCCCAACCAATGTCTGACCCTCTCCCCTCCCTCTCCTCTCCCAG 126
AS Lower GTGCTGGGCCATTCACCACTGTTGCCCAACCAATGTCTGACCCTCTCCCCTCCCTCTCCTCTCCCAG 126
B
RT GTAAGAACATGTTGCATTTCATCTTCTCTATCTTCAATGGTGCCCAACCTCCACATATAACTAAAAC 67
BT GTAAGAACATGTTGCATTTCATCTTCTCAAAGTTCAATAGTGCCCAATCTCCACAGATAACTTCAAC 67
AS GTAAGAACATGTTGCATTTCATCTTCTCAATTTTCAATGGTGCCCAATCTCCACAGATAACTTCAAC 67
RT ATCTCTCTCTCTCTCTCTC----------CC 88
BT ATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC------ 128
AS ATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC-- 132
RT TCCTCTTCTTGGTGTTCCTCTCAG 112
BT --CTCTTCTTGGTGTTCCTCTCAG 150
AS --CTCTTCTTGGTGTTCCTCTCAG 154

Fig. 3.3. Proposed evolutionary changes in two apoA-I loci based on sequence data (Fig. 3.2).


brown trout
Atlantic salmon
rainbow trout
Upper Locus
brown trout
Atlantic salmon
rainbow trout
Lower Locus

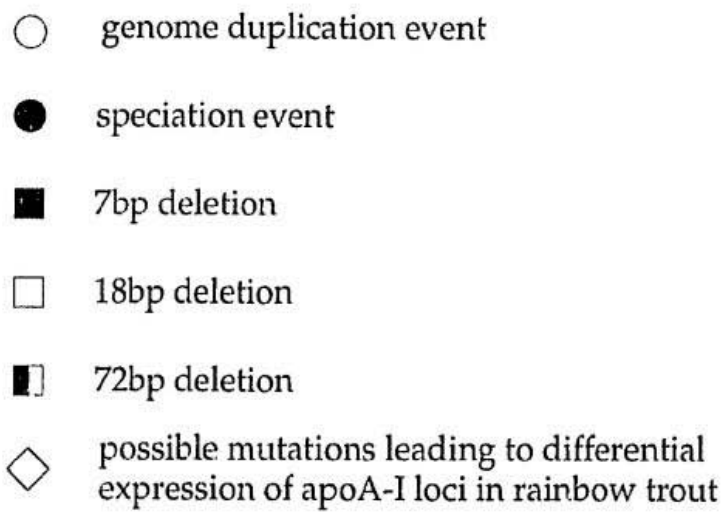

genome duplication event
speciation event
7bp deletion
18bp deletion
72bp deletion
possible mutations leading to differential expression of apoA-I loci in rainbow trout

Fig. 3.4. Amplification of the IVS III using primers 66 and 67 at various annealing temperatures (T_A). Three genomic templates, brown trout (BT), rainbow trout (RT), and Atlantic salmon (AS), were amplified. As a positive control, pBTLB27 was also used as a template (+ve). All amplifications were then fractionated through Nuseive agarose, stained with ethidium bromide, and photographed under U.V. light. Note: A negative control reaction (lacking template DNA) was performed but no bands were observed.

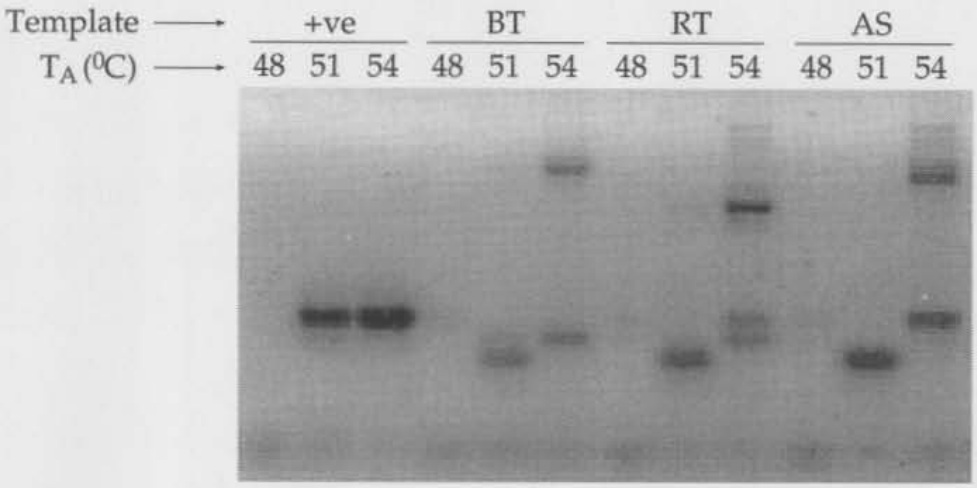

Template ⟶
+ve
BT
RT
AS
T_A (^{0}C) ⟶
48 51 54 48 51 54 48 51 54 48 51 54

Fig. 3.5. Amplification and reamplification of IVS III using primers 66 and 67. Three genomic templates, brown trout (BT), rainbow trout (RT), and Atlantic salmon (AS), were amplified at 56°C4. As a positive control and negative control, pBTLB27 (+ve) and water (-ve) were also used as templates. Each of the genomic products were isolated and used as templates for a reamplification. All amplifications were fractionated through Nuseive agarose, stained with ethidium bromide, and photographed under U.V. light. The size (bp) of the two major product are indicated at the left.

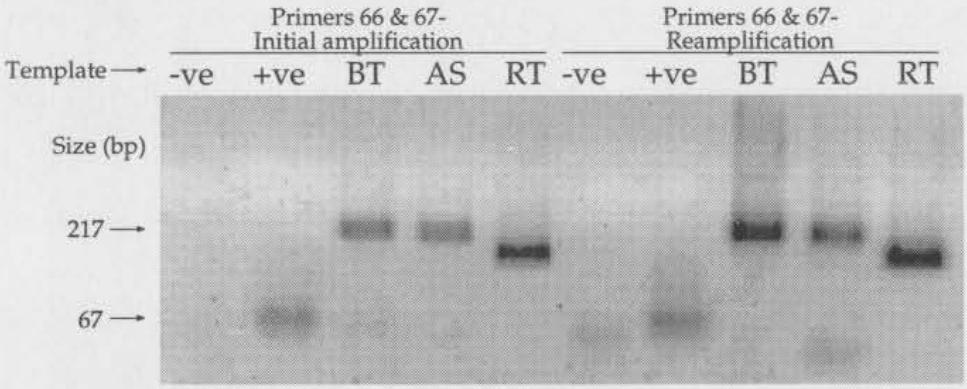

Primers 66 & 67-
Initial amplification
Primers 66 & 67-
Reamplification
Template
-ve
+ve
BT
AS
RT
-ve
+ve
BT
AS
RT
Size (bp)
217
67

Fig. 3.6. Amplification of the variable microsatellite within IVS III in an Atlantic salmon/brown trout hybrid family. Genomic DNA from both parents, Atlantic salmon (AS) and brown trout (BT), and six of the hybrid offspring were used as templates. Primer 67 and ^{32}P end-labelled primer 66 were used in the reactions. All amplifications were fractionated through a standard sequencing gel which was then dried, and exposed to autoradiography film. The developed film is shown. Note: A negative control reaction (lacking template DNA) was performed but no bands were observed.

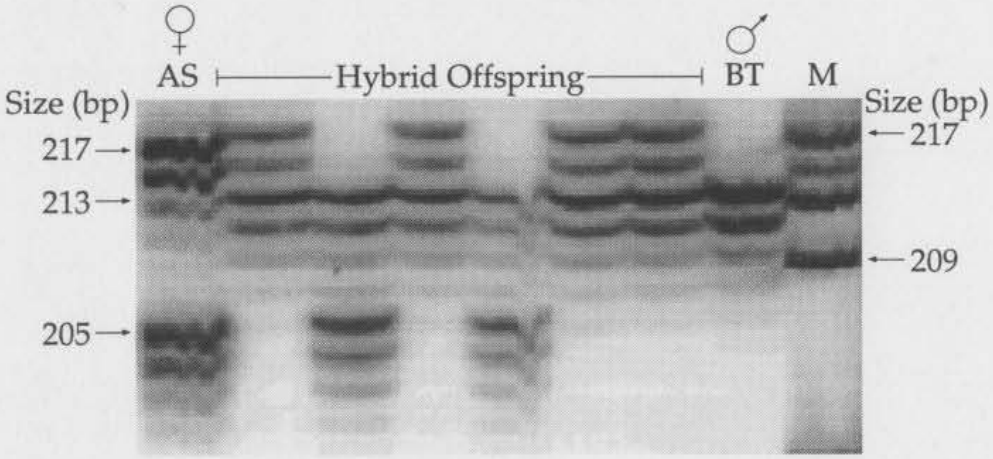
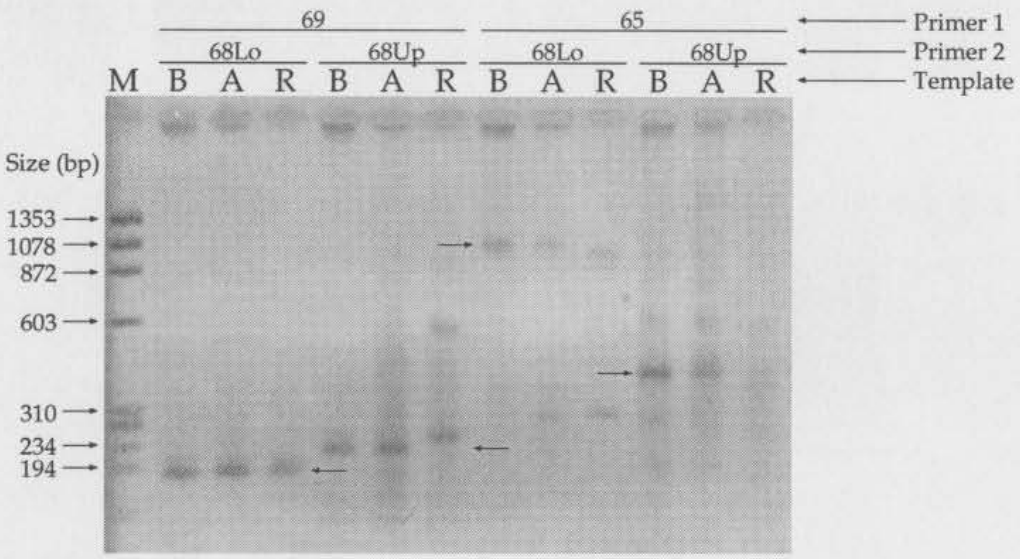

♀
AS
Hybrid Offspring
♂
BT
M
Size (bp)
217
213
205
Size (bp)
217
209

Fig. 3.7. Amplification of two regions of apoA-I using locus specific primers. Three genomic templates, brown trout (B), Atlantic salmon (A), and rainbow trout (R), were used in each set of reactions. The combination of primers used in each set is shown. All reactions and a marker (M) were fractionated through Nuseive agarose, stained with ethidium bromide, and photographed under U.V. light. The size of the marker fragments is shown on the left and the major product of each set is indicated with an arrow. Note: A negative control reaction (lacking template DNA) was performed but no bands were observed.


69
65
Primer 1
68Lo
68Up
68Lo
68Up
Primer 2
M B A R B A R B A R B A R
Template
Size (bp)
1353
1078
872
603
310
234
194

Chapter IV

Evolutionary Analyses Using ApoA-I cDNA and Intervening Sequences;

4.1 INTRODUCTION

4.1.1 Phylogenetic Reconstruction using Molecular Sequences.

The first step in any phylogenetic study using genetic sequences is an alignment of the sequences from each taxon (for a discussion on methods of phylogenetic reconstruction, see Li and Graur, 1991). If we consider just two nucleotide sequences, the alignment consists of a series of paired bases (one from each sequence) of which there are three types. Matched bases are identical and thus presumably represent a conservation in the sequence at that point. Mismatched bases are not identical and indicate that a substitution has occurred in one (or both) of the sequences. The third type of paired base within a sequence alignment consists of a base from one sequence that is absent from the other sequence. A null base or gap is usually represented by a dash ('-') and indicates that an insertion (or deletion) has occurred at that site. The goal of an alignment is to minimize both the number of mismatched base pairs and the number of null bases or gaps. However, these two factors are intimately related; as the number of mismatches decrease, the number of gaps increase and vice versa. If it is assumed that substitutions are more likely to occur than insertions and deletions, it is possible to produce a reasonable alignment by introducing a gap penalty. The alignment is thus biased against insertions/deletions and they will only be introduced if it leads to a decrease in the overall number of mismatches (i.e. more matched bases).

Sequence alignments are typically performed on computers due to the number of calculations required to produce an alignment with a minimum number of mismatches and gaps. This is particularly true when more than two sequences are being considered as the number of possible alignments

increases exponentially. Nonetheless, present day computer software and the ever-increasing processing speed of personal computers offer a high degree of flexibility concerning sequence alignment. For instance, if the sequences of interest are highly conserved, a high gap penalty would be chosen to produce a very stringent alignment with few gaps. On the other hand, if the sequences are not highly conserved and/or deep evolutionary relationships are being examined, a lower gap penalty would be more appropriate.

Once an acceptable sequence alignment has been produced, the phylogeny can be inferred in a number of ways. However, all methods for constructing phylogenetic trees fall into one of two categories; maximum parsimony methods and distance estimation methods. The principle of maximum parsimony (Eck and Dayhoff, 1966; Fitch, 1977) is a minimum evolution method. In other words, it identifies the smallest number of evolutionary changes to account for the differences within a group of sequences or taxa. To do this, the number of informative sites must first be tabulated. In an aligned group of sequences, a site is informative only when a substitution which favors one tree over another occurs. In other words, at least two nucleotides must occur at a given site in different taxa, and each nucleotide must be represented at least twice. Once the number of informative sites has been determined, all the possible tree topologies containing the given taxa are produced and the minimum number of substitutions at each informative site is tallied for each tree. The total number of substitutions for each tree is then determined and the one with the least number of substitutions is chosen as the likeliest approximation of the actual tree. However, there a number of problems with this method. From a

practical perspective, maximum parsimony methods can be very time consuming if every possible tree is produced and tested. Again, the time required to produce a maximum parsimony tree is related hyperexponentially to the number of taxa being examined. This type of exhaustive algorithm can be replaced by branch-and-bound or heuristic algorithms, which are faster but only approximate exhaustive methods. Also, more than one minimum evolution tree is often produced such that a true tree cannot be determined. Finally, the assumption that a minimum evolution tree is the true tree is not necessarily valid. This may be the case if the rate of substitution among different lineages is not approximately constant.

The other major class of phylogenetic inference methods are called distance estimation methods. Each of these methods is based on a distance matrix which compares the distance between each pair of taxa (i.e. three taxa (A, B, and C) can be separated into three pairs (AB, AC, and BC) and the distance matrix would thus contain three distance values). Distance is usually measured as the proportion of nucleotide or amino acid differences (p-distance) between the two sequences. However, most current software packages provide options to account for variations in the rate of synonymous versus nonsynonymous and transition versus transversion substitutions, as well as variations in the frequency of occurrence of the four nucleotides (nucleotide sequence data only).

Once all the pairwise distance estimations have been tabulated, one of a number of algorithms may be used to produce the phylogenetic tree. One of the simplest and most common methods for producing trees from distance matrices is the Unweighted Pair Group Method using arithmetic Averages

(UPGMA). This was originally developed for constructing taxonomic phenograms (Sokal and Michener, 1958) but has been shown to produce a reliable tree with a reasonably high probability if the distance among taxa are relatively high and the molecular clock is valid (Tateno *et al.*, 1982; Sourdis and Krimbas, 1987; see section 4.1.2).

The UPGMA method begins by identifying the pair of taxa or operational taxanomic units (OTUs) which have the smallest p-distance. These simple OTUs (i.e. OTU_A and OTU_B) are the most closely related and are grouped together as a composite OTU (OTU_{AB}). A new matrix is then calculated to determine the pairwise distances between the new set of OTUs (OTU_{AB} and the remaining simple OTUs). If the distance between OTU_A and OTU_B is given by d_{AB}, then distance of OTU_{AB} to a simple OTU, OTU_C ($d_{(AB)C}$) is the average distance between OTU_A and OTU_C and between OTU_B and OTU_C ($d_{AC} + d_{BC}/2$). If $d_{(AB)C}$ is the smallest distance in the matrix, OTU_C will be added to the composite OTU_{AB} at a node which is the midpoint between $OTU_{(AB)C}$ ($d_{(AB)C}/2$). On the other hand, if another pair of OTUs (OTU_D and OTU_E) have the smallest p-distance, then another composite OTU will be formed and the distance matrix recalculated. The distances of OTU_{DE} to another simple OTU will be calculated as above. However, the distance between OTU_{AB} and OTU_{DE} ($d_{(AB)(DE)}$) is the average distance between all the simple OTUs in each composite OTU ($d_{AD} + d_{AE} + d_{BD} + d_{BE}$)/4).

Through this sequential clustering algorithm, the UPGMA method produces one composite OTU from all the simple OTUs. The root for the tree would be placed at the average distance between the two final OTUs (simple and/or composite) as described above. The root represents the common

ancestor of all the taxa being examined and is equidistant from each simple OTU. The presence of a root also adds directionality to the tree such that an increase in distance represents evolutionary time. These trees can thus be used to estimate divergence times of the taxa. Unrooted trees, such as those produced by other distance estimation methods and maximum parsimony methods, only specify relationships among OTUs and do not define evolutionary pathways with respect to time.

When the topology of the tree has been determined, the reliability of each branch must then be tested. A common method for such analysis is called bootstrapping (see Swofford and Olsen, 1990). This algorithm first generates a new set of OTUs by randomly selecting an identical number of nucleotides from the original data set with replacement. In other words, each nucleotide for the new set is randomly selected from the complete original set. In this way, any randomly generated set will contain some nucleotides two or more times while other nucleotides will not be represented at all. This process is typically repeated a few hundred times and each set of nucleotides is used to produce another phylogenetic tree. The topology of each tree is individually compared to the original tree and a bootstrap value or bootstrap confidence level (*BCL*) is obtained for each of the internal branches. To determine the *BCL*, clusters are identified in the original tree. A cluster is a group that contains all of the descendants of the most recent common ancestor of the constituent species. The presence or absence of each original cluster is tallied for each of the replicate trees. The percentage of times that a cluster is present is the *BCL* for that specific cluster. The significance of these values clearly depends upon the number replicate trees produced. When a

bootstrap test is complete, each cluster or node on a tree has a *BCL*. For example, if a *BCL* for a given cluster is 95, then 95% of the randomly sampled trees contain that cluster and that group is thus supported at the 95% level.

4.1.2 Using Conserved Sequences as a 'Molecular Clock'.

Although sequence variation within conserved macromolecules (such as nucleic acids and proteins) is useful for inferring evolutionary relationships, they can also serve as a type of chronometer or molecular clock (for a discussion, see Woese, 1991). In the case of a DNA sequence, for example, the 'state' of the chronometer would be read as the variation in the sequence between two points in time. However, any sequence within one species can only be obtained from one state, the present state[9]. An alternative to this method is to compare two versions of the sequence from two different organisms which have, at some point, shared a common ancestor. The variation between these two sequences thus approximates the variation between the present and ancestral states of either of the original sequences. It is this variation which we designate as a relative measure of evolutionary time. To quantify this relative scale, we can plot the extent of variation or distance against the known fossil record. If the rates of evolution (i.e. incorporation and fixation of a genetic mutation within a species) is approximately constant among all lineages, a linear relationship is observed (Romero-Herrera *et al.*, 1973). This 'standard curve' can then be used extrapolate the divergence times for species which are not represented in the

---

[9]Some exceptions to this generality are *Equus quagga* (an extinct member of the horse family, Higuchi et al., 1984), *Hymenaea protera* (an extinct legume, Poinar et al., 1993), *Smilodon fatalis* (the extinct saber-toothed cat, Janczewski et al., 1992), and *Anthonomus grandis* (the weevil, Cano et al., 1993).

fossil record. In theory, the molecular clock concept is quite sound. In practice, however, a number of problems arise. For instance, the assumption of a constant rate of evolution between lineages is not always valid. Li *et al.* (1987) showed that the substitution rate in rodent lineages could be four to six times higher than in primates. Also, the fossil record may contain a number of biases (for a discussion see Carroll, 1988). Nonetheless, the molecular clock concept continues to provide important insight into evolutionary history.

4.2 EXPERIMENTAL and DISCUSSION

4.2.1 Inferring Phylogeny Using apoA-I cDNA and Intervening Sequences.

When the cDNA sequence of apoA-I in brown trout had been determined, computerized searches of the both the GenBank nucleotide sequence database and the Science Citation Index journal publication database were performed to obtain all known cDNA or gene sequences for apoA-I (see Table 1.2). The coding region from each of the fourteen sequences was used to obtain the inferred amino acid sequences which were then aligned using the program Clustal V (Higgins *et al.*, 1992). This sequence alignment was used to align the cDNA sequence (Fig. 4.1). Using the program MEGA (Kumar *et al.*, 1993) a distance matrix was produced using the p-distance of all nonsynonymous codons (this decreased errors due to multiple substitutions at synonymous codons). The phylogeny was then reconstructed from this matrix using the UPGMA method and the reliability of the internal branches was tested with 1000 bootstrap replications. Similarly, all known sequences for apoA-I:IVS II were aligned (Fig. 4.2) and the phylogeny reconstructed as

described above. The nucleotide sequences for the introns were aligned directly as they do not encode proteins.

When using molecular sequence data for any comparative study, there is always the question of the reliability of the sequences. Thus, one must assume that sequences published and available through databases are accurate. Although this is not always the case, common sequence errors such as a single base substitutions often go unnoticed. As such, all of the sequences obtained for use in this study, with the exception of the Atlantic salmon sequence, were taken at face value. When the salmon sequence was obtained from GenBank, translated, and compared to other inferred amino acid sequences, two regions of 13 and 23 residues were identified in Atlantic salmon which had a very low sequence identity/similarity to corresponding regions in both rainbow trout and brown trout (data not shown). This was particularly striking given that the sequence identity between the three species outside these two regions was greater than 90%. A closer examination of the nucleotide sequence revealed four putative frame-shift mutations (deletions) in the Atlantic salmon sequence, one of which produced a premature stop codon resulting in the apparent loss of four carboxy-terminal residues. Given that the probability of such mutations is quite low in such a highly conserved molecule, four nucleotides were inserted into the salmon cDNA sequence to keep the inferred primary structure in frame. Since the nucleotides at the four positions were conserved in each of the other three salmonid sequences, it was assumed that these positions were conserved in the Atlantic salmon sequence as well.

The cDNA-derived tree is shown in Fig. 4.3(A). The topology of the mammalian cluster agrees with the accepted phylogeny (Miller and Harley, 1994). If these species are grouped by order, this tree groups all of the mammals in their respective orders with the chicken as sister group. Similarly, all the Salmoniformes group together on a completely separate branch from the birds and mammals. This topology corresponds to the pattern of divergence of the mammals and birds from the osteichthyes. At the level of species, we can see that the brown trout sequence is most closely related to apoA-I-2, the rainbow trout apoA-I mRNA that is only expressed in hepatocellular carcinomas (Delcuve *et al.*, 1992). The Atlantic salmon sequence is most closely related to apoA-I-1, the gene expressed in normal cells. These results would seem to conflict given that Atlantic salmon and brown trout are in the same genus (*Salmo*) and we would thus expect them to group together. However, the sequence information in this cluster is incomplete as it lacks the second gene from both Atlantic salmon and brown trout. This observation actually raises a paradox: the brown trout cDNA sequence was derived from (apparently) healthy liver tissue but is more similar to the rainbow trout sequence specific to the cancerous state. Thus, the differential expression of apoA-I may not be present in brown trout and Atlantic salmon. This theory is supported by the results described in section 3.2.3. Both brown trout and Atlantic salmon do not exhibit full length genes at the upper locus whereas rainbow trout has a true gene at this locus. Another possibility is that the Atlantic salmon sequence used is not accurate. Thus, the branching arrangement in the salmonid monophyletic may not be reliable.

The IVS II-derived tree is shown in Fig. 4.3(B). Although not all the taxa in the cDNA tree are represented, the overall topology of the mammalian and avian lineages has been retained. The bootstrap values for the branches within these lineages are comparable to those obtained with the coding sequences. The topology of the salmoniformes branch appears quite different from that seen in the cDNA tree. This is due to the presence of two sequences for each salmonid species. As shown, intron sequences of similar loci (i.e. upper and lower) are more closely related than are sequences within the same species. This would imply that the genome duplication which produced the two apoA-I loci occurred prior to the speciation of the *Salmo* and *Oncorhynchus* genera. The common ancestor of these genes thus represents the genome duplication which occurred in salmonid evolution (Allendorf and Thorgaard, 1984). However, some of the bootstrap values for this monophyletic are low. This may be due to the relatively short length of the IVS region. To get a more reliable estimate of the evolution of these genes in salmonids, the coding regions for the missing Atlantic salmon and brown trout genes are required.

Although the radiation of Mammalia, Aves, and Osteicthyes is generally accepted, the relationships of the orders within these groups are not as easily defined. This is particularly true in Mammalia. One highly controversial issue is the placement of Rodentia in relation to Primates, Artiodactyla, and Carnivora. Morphological and paleontological evidence suggests that the two latter orders diverged from the mammalian lineage prior to the divergence of Rodentia and Primates (Romer, 1968; Kielan-Jaworowska *et al.*, 1979; Novacek, M.J., 1982). On the other hand, some

research suggests that Rodentia is actually a sister group to the other three orders (Szalay, 1977; Li *et al.*, 1990). This has important implications with respect to rates of substitution. For instance, relative rates of nucleotide substitution between human and mouse, assuming that Artiodactyla and Carnivora diverged before Rodentia and Primates, would be an overestimate if Rodentia was indeed an outgroup to the other orders. Both the cDNA and IVS trees strongly support the concept of Rodentia as the outgroup and are in agreement with Li *et al.* (1990) as well as Sparrow *et al.* (1995) who analyzed apoA-I evolution using mature amino acid sequences.

Another order with a controversial evolutionary position is Lagomorpha. Although paleontologists would group it with Rodentia (Szalay, 1977), it has been suggested that Lagomorpha is an outgroup to Rodentia and Primates (Shoshani, 1986). Protein sequence data have also been used to imply that Lagomorpha and Primates are in the same group (Li *et al.*, 1990; Sparrow *et al.*, 1995). The data presented above would suggest that neither of these schemes is correct. Rather, it agrees with Li *et al.* (1990) that within Infraclass Eutheria (the placental mammals) Lagomorpha diverged after Rodentia but before Carnivora, Artiodactyla, and Primates. Unlike the position of Rodentia, this placement of Lagomorpha is in direct contrast with Sparrow *et al.* (1995). This unexpected result is not due to the fact that amino acid sequences were used in the latter and nucleotide sequences were used in the present study. When the phylogeny was reconstructed using mature inferred amino acid sequences, the topology of the resultant tree was unchanged (data not shown). Rather, the differences may be due to the use of different methods to construct sequence alignments and inferring

phylogenies. For example, the sequence alignment in Sparrow *et al.* (1995) suggests that the Atlantic salmon sequence underwent two major changes; at the amino-terminus, a 20 residue insertion and at the carboxy-terminus, a 26 residue deletion. This gives an alignment of 264 residues whereas the Clustal V alignment is only 243 residues (data not shown).

4.2.2 Constructing Molecular Clocks Using the Fossil Record.

When the phylogenetic trees were produced, distance estimations for each branch point of the trees could be determined by simply summating the individual branch lengths that constitute the total branch of interest. The divergence time for each branch point is then derived from the known fossil record. The accepted divergence times (from the fossil records) for a number of branch points are shown in Table 4.1. Using the origin as a constant, divergence time versus distance was plotted and a best-fit, linear regression line calculated (Fig. 4.4). Because the rate of evolution of the rodent lineage is controversial (Li *et al.*, 1987; O'hUigin and Li, 1992), regression analysis was performed with (Fig. 4.4(A)) and without (Fig. 4.4(B)) the rodent data. To calculate average divergence times of points under-represented in the fossil record, the distance of the point was taken from each of the UPGMA trees and the time extrapolated from the best-fit line (Fig. 4.4(B). If more than one estimation was calculated (i.e. from different trees) the values were averaged.

Table 4.1. A listing of evolutionary branch points and divergence times based on the fossil record. Distance estimations from the cDNA and IVS trees as well as references are shown.

Branch Point	cDNA Distance	IVS II Distance	Time (MYA)	Reference
Mouse/Rat	0.079	0.073	12	Flynn *et al.*, 1985; Bulmer *et al.*, 1991
Human/Baboon	0.011	0.017	30	Romero-Herrera and Lehmann, 1972
Cow/Pig	0.029	-	60	Savage and Russell, 1983; Romer, 1968
Mammalia/Lagomorpha	0.058	0.146	80	Kimura, 1987; Li *et al.*, 1990
Mammalia/Rodentia	0.100	0.169	100	Li *et al.*, 1990
Mammalia/Aves	0.162	0.256	280	Carrol, 1988
Mammalia/Teleost	0.256	0.315	380	Carrol, 1988

Fig. 4.4(A) and 4.4(B) shows the molecular clocks produced using distance estimations from both the cDNA and IVS trees. The former considers all the data points while the latter ignores the data from the rodent lineage. As shown, the slope of both best-fit lines produced from IVS data is approximately 1.3X steeper than those produced from the cDNA data. The lack of a functional/structural constraint on the intron sequences has resulted in a higher rate of evolution in comparison to the cDNA sequences (Miyata *et al.*, 1980; Kimura, 1983). However, the slope of the IVS and cDNA lines are not greatly affected by the presence of the rodent data points. When these points are omitted from the regression analysis, the slope is only slightly altered. In contrast, the r^2 value is greatly affected by the removal of the rodent data points. This is particularly evident in the case of the cDNA data where the r^2 is very close to one (1.0), indicating that the non-rodent lineages examined are evolving at an approximately constant rate.

As described above, the mouse/rat and Mammalia/Rodentia divergence points deviate significantly from both best-fit lines (this is especially pronounced by the mouse/rat data points. One explanation for this increase in the rodent lineage specifically is the generation-time effect (Wu and Li, 1985). This states that the short generation time in rodents allows for an increase in substitution rate. Li *et al.* (1990) argued that the rabbit also has a shorter generation time yet it does not exhibit an increased rate of evolution. Our data would suggest that the rabbit does indeed have an increased rate and thus agrees with the generation-time concept. This is evident in the IVS data where the Mammalia/Lagomorpha divergence point deviates from the best-fit line as much as either of the rodent points. Although various alternative

explanations for the increased rate among rodents have been proposed (including differences in the number of DNA replications per unit time in the germ line and the efficiency of DNA repair; Li *et al.*, 1990), the answer may simply be that rodents (and perhaps lagomorphs) diverged from other eutherians earlier than is presently accepted (~100 MYA). This concept is supported by the fact that the fossil record concerning rodents is based primarily on cheek teeth (for a discussion see Wilson *et al.*, 1987). The classification of rodent fossils is further called into question by the placement of the spiny mouse. Fossil evidence would group this species with the family Muridae (which includes true mice, rats, and hamsters) whereas molecular data strongly indicates that the spiny mouse is an out-group to this family. Further, the divergence time of the rat/mouse has been estimated from a large number of genes to be 29MYA (O'hUigin and Li, 1992), more than twice the value suggested from fossil evidence. Thus, the divergence time of the rodents remains uncertain. If the rate of evolution of apoA-I is in fact constant among all lineages, the estimated divergence time for Rodentia would be 177 million years ago (148-206 MYA) while the mouse and rat would have diverged 106 million years ago (89-123 MYA).

As noted above, fossil evidence may not be reliable and is often unavailable for the majority of species studied. Thus, one of the main uses of a molecular clock is to estimate divergence times of these monophyletics which are not represented in the fossil record. Although the phylogeny of the salmonid family is generally accepted, the divergence times of the genera are not well established. The four genera comprising the Salmonid family are *Hucho, Salvelinus, Salmo,* and *Oncorhynchus*. As shown in Figure 1.3,

Hucho diverged first followed by Salvelinus, *Salmo* and *Oncorhynchus* thus share the most recent common ancestor (Ferguson and Flemming, 1983; Phillips and Pleyte, 1991). Recent estimates of divergence times within the Salmonids estimate that *Salmo* and *Oncorhynchus* diverged between 10-16 MYA (Andersson and Matsunaga, 1993) while *Salvelinus* and *Oncorhynchus* diverged about 18 MYA (Andersson *et al.*, 1995). The apoA-I data suggest that *Salmo* and *Oncorhynchus* diverged between 31-40 MYA. Even the lowest estimate (from the IVS data) is significantly higher than those stated above.

The time of the genome duplication event which predated salmonid evolution can also be estimated from the phylogenetic trees by comparing the distances between duplicate loci. This method has recently been used to estimate the time of the tetraploidization in *Cyprinus carpio* (the common carp; Larhammar and Risinger, 1994). The apoA-1 data predicts that this event occurred between 48-67 MYA in salminids. Again, the earliest estimate (from the IVS data) would predate the *Salvelinus/Oncorhynchus* divergence estimates by approximately 30 MYA. Intuitively, this is an unusually long time for a speciation event given that an extra genome was produced and could evolve relatively free selective pressures. It would thus appear that our estimates are consistently greater than those described by Andersson above. If only one set of results can represent the true divergence times, these values must be compared to an independent time line (i.e. the fossil record). Evidence from North America has suggested that three species similar to species of *Oncorhynchus* diverged earlier than 5.5-7.6 MYA (Smith *et al.*, 1982). Assuming that the *Salmo* species were diverging at about the same time as the *Oncorhynchus* (given that these two genera share the most recent

common ancestor in salmonid evolution), the estimated divergence time for brown trout/Atlantic salmon of 15 MYA (3-27 MYA based on the upper and lower locus of the IVS data) appears to be an over-estimate as well. The most probable cause of these over-estimates is the rate of evolution of the apoA-I gene. As stated previously, apoA-I has a 25% higher rate of non-synonymous substitutions than many mammalian genes (O'hUigin *et al.*, 1990). This is also compounded by the fact that IVS evolve at a higher rate than coding sequences. Although apoA-I cDNA and IVS accurately reconstruct the phylogeny, distance/divergence estimations based on the rate of evolution of this gene are likely to be consistently higher than the true value.

Although apoA-I evolves faster than average mammalian genes, its rate appears to be constant among many lineages (with the exception of Rodentia). This is demonstrated by the low variability of the cDNA regression line (Fig. 4.4(B), R^2=0.985). This result is in contrast to studies of rates of evolution in mitochondrial DNA (for a review see Rand, 1994) where a distinct trend is observed: the rate of evolution is proportional to the thermal habitat of the species. For instance, cold-blooded animals generally have slower rates than warm-blooded animals. This variation with habitat temperature has not been documented in nuclear genes and appears to be absent from this research. However, the data presented here have not been corrected for multiple substitutions at individual nucleotide positions. This error is proportional to the length of time which has transpired since the species in question diverged. For instance, the substitutions incorporated into a sequence at time=zero (i.e. the present) would be zero. If we compare the rate of substitutions in primates and Osteichthyes to time=zero, we expect

that the latter will be an under-estimate. Because the chance that a mutation will occur at a site that has already undergone a mutation (i.e. a revertant mutation) increases with time, the brown trout should contain more of these revertant mutants resulting in an under-estimation[10]. Thus, the true slope of our cDNA regression line may be greater than that shown. Further analysis of our data in comparison to a corrected slope is necessary before definitive conclusions concerning the rate of substitution among a particular lineage are drawn.

The results of this section are based entirely upon the molecular clock hypothesis. Assuming that this is valid, it may be an inappropriate method to calculate evolutionary distances where duplicate loci are being examined (for a discussion see Allendorf and Thorgaard, 1984). With respect to polyploids, we must assume that genes in separate species evolve at the same rate as duplicate genes in a single species. However, we would expect the duplicate genes to evolve more rapidly as they would be relatively free of selective pressures as long as the structure/function of the other locus was retained (Allendorf and Thorgaard, 1984).

Another potential problem of using duplicate loci comparisons relates to the pattern of inheritance following the duplication event. In a new tetraploid, we would expect multivalent formation and tetrasomic inheritance to occur. If we consider just one chromosome, all four homologs can associate as a unit during gamete formation. Because recombination can occur among all the homologs, no divergence can take place. Divergence will commence when the four homologs begin to pair as two pairs of

---

[10]Note that although the apparent rate of substitution among old lineages will be less than expected, the true rate is constant and thus the molecular clock remains valid.

chromosomes (i.e. disomic inheritance; this situation is more favorable as it increases zygote survival and allows divergence at one of the loci). This scenario is supported by the finding that some pairs of duplicate loci have not yet returned to complete disomic inheritance (Wright *et al.*, 1980; May *et al.*, 1982). Thus, estimates of divergence times obtained as described above will be minimum estimates of the time of the genome duplication and will more accurately reflect the time of the return to disomic inheritance of the loci in question (Allendorf and Thorgaard, 1984).

Fig. 4.1. Clustal V alignment of apoA-I cDNA sequences. Alignment gaps are given by a '-'. The nucleotides are numbered with respect to the baboon sequence. The four nucleotides inserted into the salmon sequence (nucleotides 262, 298, 299, and 719) are in lower-case typeface and underlined.

```
                                                                                                           400
Baboon    AAGGAGACAGAGGGCCTGAGGCAGGAGATGAGCAAGGATCTGGAGGAGGTGAAGGCCAAGGTGCAGCCCTACCTGGACGACTTCCAGAAGAAGTGGCAGG
Monkey    AAGGAGACAGAGGGCCTGAGGCAGGAGATGAGCAAGGATCTGGAGGAGGTGAAGGCCAAGGTGCAGCCCTACCTGGACGACTTCCAGAAGAAGTGGCAGG
Human     AAGGAGACAGAGGGCCTGAGGCAGGAGATGAGCAAGGATCTGGAGGAGGTGAAGGCCAAGGTGCAGCCCTACCTGGACGACTTCCAGAAGAAGTGGCAGG
Dog       AAGGAGACGGAGGTGCTGCGGCAGGAGATGAGCAAGGACCTGGAGGAGGTGAAGCAGAAGGTGCAGCCCTACCTGGACGACTTCCAGAAGAAGTGGCAGG
Pig       AAGGAGACCGAGGCGCTGCGGCAGGAGATGAGCAAGGACCTGGAGGAGGTGAAGAAGAAGGTGCAGCCCTACCTGGACGACTTCCAGAACAAGTGGCAGG
Cow       AAGGAGACCGCGTCGCTGAGGCAGGAGATGCACAAGGACCTGGAGGAGGTGAAGCAGAAGGTGCAGCCCTACCTGGACGAGTTCCAGAAGAAGTGGCACG
Rabbit    AAGGAGACAGAGGGGCTGAGGGAGGAGATGAACAAGGATCTGCAGGAGGTGAGGCAGAAAGTGCAGCCCTTCCTGGACGAATTCCAGAAGAAGTGGCAGG
Mouse     AAAGAAACAGATTGGGTGAGACAGGAGATGAACAAGGACCTAGAGGAAGTGAAACAGAAGGTGCAGCCCTACCTGGACGAATTCCAGAAGAAATGGAAAG
Rat       AAAGAAACAGATTGGCTGAGAAACGAGATGAACAAGGACCTGGAGAATGTGAAACAGAAGATGCAGCCCCACCTAGATGAATTCCAGGAGAAGTGGAACG
Chicken   AAGGACACCGAGGCTCTGCGTGCTGAGCTCACCAAGGACCTGGAGGAGGTGAAGGAGAAGATCCGGCCCTTCCTGGACCAGTTCTCCGCCAAGTGGACGG
Btrout    GATGCCGCCGCCGCCGTGCGCGCTGAGGTCATGAAGGACGTGGAGGACGTGCGCACACAGCTGGAGCCCAAGCGCGCCGAGCTCAAGGAAGTCCTGGACA
Asalmon   GACGCCACCGCCACCGTGCGCGCTGAGGTCATGAAGGACGTGGAGGACGTGCGCACCCAGCTGGAGCCCAAGCGCGCCGAGCTCACGGAAGTCCTGAACA
Rtrout1   GACGCCACCGCCGCTGTGCGCGCTGAGGTCATGAAGGACGTGGAGGAGCTGCGCTCCCAGCTGGAGCCCAAGCGCGCTGAACTCAAGGAAGTCCTGGACA
Rtrout2   GACGCCACCGCCGCCGTGCGCGCTGAGGTCATGAAGGACGTGGAGGAGCTGCGCTCCCAGCTGGAGCCCAAGCGCGCCGAGCTCAAGGAAGTCCTGGACA

                                                                                                           500
Baboon    AGGAGATGGAGCTCTACCGCCAGAAGGTGGAGCCGCTGCGCGCGGAGCTCCACGAGGGCACGCGCCAGAAGCTGCACGAGCTGCACGAGAAACTGAGCCC
Monkey    AGGAGATGGAGCTCTACCGCCAGAAGGTGGAGCCGCTGCGCGCGGAGCTCCACGAGGGCACGCGCCAGAAGCTGCACGAGCTGCACGAGAAACTGAGCCC
Human     AGGAGATGGAGCTCTACCGCCAGAAGGTGGAGCCGCTGCGCGCAGAGCTCCAAGAGGGCGCGCGCCAGAAGCTGCACGAGCTGCAAGAGAAGCTGAGCCC
Dog       AGGAGGTGGAGCTGTACCGCCAGAAGGTGGCGCCGCTGGGCTCGGAGCTGCGCGAGGGCGCGCGCCAGAAGCTGCAGGAGCTGCAGGAGAAGCTGAGCCC
Pig       AGGAGATGGAGACGTACCGCCAGAAGATGGCGCCGCTGGGCGCCGAGTTCCGCGAGGGCGCGCGCCAGAAGGTGCAGGAGCTGCAGGAGAAGCTGAGCCC
Cow       AGGAGGTGGAGATCTACCGCCAGAAGGTGGCGCCGCTGGGCGAGGAGTTTCGCGAGGGCGCGCGCCAGAAGGTGCAGGAGCTGCAGGACAAGCTGAGCCC
Rabbit    AGGAGGTGGAGCGCTACCGCCAGAAAGTGGAGCCCCTGGGCGCGGAGCTGGGCGAGAGCGCGCGCCAGAAGCTGACCGAGCTGCAGGAGAAGCTGAGCCC
Mouse     AGGATGTGGAGCTCTACCGCCAGAAGGTGGCGCCTCTGGGCGCCGAGCTGCAGGAGAGCGCGCGCCAGAAGCTGCAGGAGCTGCAAGGGAGACTGTCCCC
Rat       AGGAGGTCGAGGCCTACCGCCAGAAGCTGGAGCCTCTGGGCACCGAGCTTCACAAAAACGCGAAG-----------GAGATGCAAAGGCATCTAAAGGT
Chicken   AGGAGCTGGAGCAGTACCGCCAGCGCCTGACGCCCGTGGCTCAGGAGCTGAAGGAGCTCACCAAGCAGAAGGTGGAGCTGATGCAGGCCAAGCTGACCCC
Btrout    AGCACATAGACGAGTACCGCAAGAAGCTGGAGCCCCTGATCAAGGAAATCGTTGAGCAGCGCCGCACCGAGCTGGAGGCCTTCAGGGTTAAGATGGAGCC
Asalmon   AGCACATAGATGAGTACCGCAAGAAGCTGGAGCCCCTGATCAAGCAGCACATCGAACTGCGCCGCACCGAGATGGATGCCTTCAGGGCCAAGATTGACCC
Rtrout1   AGCACATAGATGAGTACCGCAAGAAGCTGGAGCCCCTGATCAAGGAGCACATCGAACTGCGCCGCACCGAGATGGAGGCCTTCAGGGCTAAGATGGAGCC
Rtrout2   AGCACATAGACGAGTACAGCAGGAAGCTGGAGCCCCTGATCAAGGACATCGTCGAGCAGCGCCGCACCGAGCTGGAGGCCTTCAGGGTTAAGATAGAGCC

                                                                                                           600
Baboon    GCTGGGGGAGGAGGTGCGCGACCGCGCGCGCGCCCACGTGGACGCGCTGCGCACGCATCTGGCCCCCTACAGCGATGAGCTGCGCCAGCGCCTGGCCGCG
Monkey    GCTGGGCGAGGAGGTGCGCGACCGCGCGCGCGCCCACGTGGACGCGCTGCGCACGCATTTGGCCCCCTACAGCGATGAGCTGCGCCAGCGCTTGGCCGCG
Human     ACTGGGCGAGGAGATGCGCGACCGCGCGCGCGCCCATGTGGACGCGCTGCGCACGCATCTGGCCCCCTACAGCGACGAGCTGCGCCAGCGCTTGGCCGCG
Dog       GCTGGCGGAGGAGCTGCGGGACCGCGCGCGCACCCACGTGGACGCGCTGCGCGCCCAGCTGGCCCCCTACAGCGACGACCTGCGCGAGCGCCTGGCCGCG
Pig       GCTGGCCGAGGAGCTCCGCGACCGCCTGCGCGCCCACGTGGAGGCGCTGCGCCAGCAGCTGGCGCCCTACAGCGACGACCTGCGCCAGCGCATGGCCGCC
Cow       TCTGGCCCAGGAGCTCCGCGACCGCGCGCGCGCCCACGTGGAGACGCTGCGGCAGCACGTGGCGCCCTACAGCGACGACCTGCGCCAGCGGCTGACCGCG
Rabbit    GCTGGCCGAGGAGCTGCGCGACAGCGCGCGCACCCACGTGGATACACTGCGCACGAAGCTGGCGCCCTACAGCAACGAGCTG---CAGAGGCTGGCCGCG
Mouse     TGTGGCTGAGGAATTTCGCGACCGCATGCGCACACACGTAGACTCTCTGCGCACACAGCTAGCGCCCCACAGCGAACAGATGCGCGAGAGCCTGGCCCAG
Rat       TGTGGCCGAGGAGTTTCGAGACCGCATGCGTGTGAATGCAGACGCGCTGCGCGCAAAGTTTGGGCTCTACAGCGATCAGATGCGCGAGAACCTGGCCCAG
Chicken   GGTGGCTGAGGAGGCGCGGGATCGTCTGCGTGGGCACGTGGAGGAGCTGCGCAAGAACCTGGCGCCATACAGCGATGAGCTGCGGCAGAAGCTGAGCCAG
Btrout    CGTTGTGGAGGAGATGCGCGCCAAGGTGTCCACCAACGTGGAGGAGACCAAGGCCAAGCTCATGCCCATCGTGGAGACCGTCCGTGCCAAGCTGACCGAG
Asalmon   CGTTGTGGAGGAGATGCGCGCCAAGGTGGCCGTCAATGTGGAGGAGACCAAGACCAAGCTCATGCCGATTGTGGAGATTGTCCGTGCCAAGCTGACCGAG
Rtrout1   CATTGTGGAGGAGTTGCGCGCCAAGGTGGCCAGTAAGGTGGAGGAGACCAAGACCAAGCTCATGCCGATTGTGGAGATTGTCCGTGCCAAGCTGACCGAG
Rtrout2   CGTGGTGGAGGAGATGCGTGCCAAGGTGTCCGCTAACGTGGAGGAGACCAAGACCAAGCTCATGCCGATTGTGGAGACCGTCCGTGCCAAGCTGACCGAG
```

```
                                                                                                      700
Baboon   CGCCTTGAGGCGCTCAAGGAGAACGGCGGCGCCCGCCTGGCGGAGTACCACGCCAAGGCCAGCGAGCATCTGAGCACGCTCAGCGAGAAGGCCAAGCCCG
Monkey   CGCCTTGAGGCGCTCAAGGAGAACGGCGGCGCCCGCCTGGCCGAGTACCACGCCAAGGCCAGCGAGCATCTGAGCACGCTCAGCGAGAAGGCCAAGCCCG
Human    CGCCTTGAGGCTCTCAAGGAGAACGGCGGCGCCAGACTGGCCGAGTACCACGCCAAGGCCACCGAGCATCTGAGCACGCTCAGCGAGAAGGCCAAGCCCG
Dog      CGGCTGGAGGCGCTCAAGGAGGGCGGCGGCGCCAGCCTGGCCGAGTACCACGCCAGGGCCAGCGAGCAGCTGAGCGCGCTCGGCGAGAAGGCCAGGCCCG
Pig      CGCTTCGAGGCGCTCAAGGAGGGCGGCGGC---AGCCTGGCCGAGTACCAGGCCAAGGCCCAGGAGCAGCTGAAAGCGCTGGGCGAGAAGGCCAAGCCGG
Cow      CGCCTGGAGGCGCTCAAGGAGGGCGGCGGC---AGCCTGGCCGAGTACCACGCCAAGGCCAGCGAGCAGCTGAAGGCGCTGGGCGAGAAGGCCAAGCCGG
Rabbit   CGCCTCGAGTCCATCAAGGAGGGCGGCGGCGCCAAGCTGGCCGAGTACCAGGCCAAGGCCAGGGAGCACCTGAGCGTGCTCAGTGAGAAGGCCAGGCCTG
Mouse    CGCCTGGCTGAGCTCAAGAGCAAC------CCTACCTTGAACGAGTACCACACCAGGGCCAAAACCCACCTGAAGACACTTGGCGAGAAAGCCAGACCTG
Rat      CGCCTGACCGAGATCAGGAACCAC------CCTACCCTGATCGAGTATCATACCAAGGCCGGCGACCACCTGAGGACACTTGGTGAGAAGGCCAAACCCG
Chicken  AAGCTGGAGGAGATCCGTGAGAAGGGCATCCCCCAGGCTTCCGAGTACCAGGCCAAGGTGATGGAGCAGCTCAGCAACCTGCGTGAGAAGATGACGCCTC
Btrout   CGTCTGGAGGAGCTGAGGACCCTGGCCGCCCCCTATGCTGAGGAGTACAAGGAGCAGATGTTCAAGGCTGTTGGAGAGGTGCGCGAGAAGGTGGGGCCCC
Asalmon  CGTCTGGAGGAGCTGAGGACCCTGGCAGCCCCCTACGCTGAGGAGTACAAGGAGCAGATGTTCAAGGCTGTTGGAGAGGTGCGTGAGAAGGTGGCTCCCC
Rtrout1  CGTCTGGAGGAGCTGAGGACCCTGGCCGCCCCCTACGCTGAGGAGTACAAGGAGCAGATGATCAAGGCTGTTGGAGAGGTGCGTGAGAAGGTGTCTCCCC
Rtrout2  CGTCTGGAGGAGCTGAGGACCCTGGCCTCCCCCTACGCTGAGGAGTACAAGGAGCAGATGGTCAAGGCTGTTGGAGAGGTGCGTGAGAAGGTGGCTCCCC

                                                                                                       801
Baboon   CGCTCGAGGACCTCCGCCAGGGCCTGCTGCCCGTGCTGGAGAGCTTCAAGGTCAGCTTCCTGAGCGCTCTCGAGGAGTACACTAAGAAGCTCAGCACCCAG
Monkey   CGCTCGAGGACCTCCGCCAGGGCCTGCTGCCCGTGCTGGAGAGCTTCAAGGTCAGCTTCCTGAGCGCTCTCGAGGAGTACACTAAGAAGCTCAGCACCCAG
Human    CGCTCGAGGACCTCCGCCAAGGCCTGCTGCCCGTGCTGGAGAGCTTCAAGGTCAGCTTCCTGAGCGCTCTCGAGGAGTACACTAAGAAGCTCAACACCCAG
Dog      CGCTGGAGGACCTGCGCCAGGGCCTGCTGCCCGTGCTGGAGAGCTTCAAGGTCAGCCTGCTGGCTGCCATCGACGAGGCCACCAAGAAGCTGAACGCGCAG
Pig      CGCTGGAGGACCTCCGCCAGGGCCTGCTGCCCGTGCTGGAGAACCTCAAGGTCAGCATCCTGGCCGCCATCGACGAGGCCTCCAAGAAGCTGAACGCCCAG
Cow      TGCTGGAGGACCTCCGCCAGGGCCTGCTGCCCGTGCTGGAGAGCCTCAAGGTCAGCATCCTGGCCGCCATAGACGAGGCCTCCAAGAAGCTGAACGCCCAG
Rabbit   CGCTGGAGGACCTGCGCCAGGGCCTGCTGCCCGTGCTGGAGAGCTTCAAGGCCAGCGTGCAGAACGTCGTGGACGAGGCCACCAAGAAGCTGAACACCCAG
Mouse    CGCTGGAGGACCTGCGCCATAGTCTGATGCCCATGCTGGAGACGCTTAAGACCAAAGCCCAGAGTGTGATCGACAAGGCCAGCGAGACTCTGACTGCCCAG
Rat      CGCTGGATGACCTAGGCCAGGGCCTGATGCCGGTGCTGGAAGCCTGGAAGGCCAAAATCATGAGTATGATCGATGAGGCCAAAAAGAAGCTGAACGCT---
Chicken  TGGTGCAGGAATTCAGGGAGCGCCTCACCCCCTATGCTGAGAACCTCAAGAACCGCTTGATCTCCTTCCTGGACGAACTCCAGAAGTCCGTGGCC------
Btrout   TGACCAACGACTTCAAGGGCCAGGTGGGCCCCGCCGCCGAGCAGGCCAAGGAAAAGCTCATGGATTTCTACGAGACCATCAGCCAGGCCATGAAGGCA---
Asalmon  TGTCTGAGGACTTCAAGGgCCAGATGGGCCCCGCCGCCGAGGAGGCCAAGCAAAAGCTCCTGGCTCTCTACGAGACCATCAGCCAGGCCATGAAGGCA---
Rtrout1  TGTCTGAGGACTTCAAGGGCCAGGTGGGCCCCGCAGCCGAACAGGCCAAGCAGAAGCTCCTGGCTTTCTACGAGACCATCAGCCAGGCCATGAAGGCA---
Rtrout2  TGTCTGAGGACTTCAAGGGCCAGTTGGGCCCCGCCGCCGAGCAGGCCAAGGAAAAGCTCATGGCTTTGTACGAGACCATCAGCCAGGCCATGAAGGCA---
```

Fig. 4.2. Clustal V alignment of apoA-I:IVS II sequences. Alignment gaps are given by a '-'. The nucleotides are numbered with respect to the human sequence.

```
                                                                                                     95
Human     GTAGGTGTCCCCTAACC--TAGG-A-----GCCAACCATCGGGGG-GCTTTCTCCCTAAATCCCCGTGGCCCACCCTCC--TGGGCAGAGGCAGCAGG-TTTCT----CAC
Monkey    GTAGGTGTCCCCCAGCC--TAGGGA-----GCCAACCATCGGGGG-GCTTTCTCCCCAAACCCCTGTGGCCCACCCTCC--TGGGCAGAGGCAGCG---------------
Pig       GTAGGTG-CCCCCAACC--TAGGGA-----GCCAACCTTCGGGGGTGTCTTCTGCCTGAACCCCTGCAGTCCACCCTCC--TGAGCATGGGGAGCAGAATTTCT----CCC
Rabbit    GTCGGTG-CCCCC-ACC--CGGTG--------------TAGGGAGC---------CCCAAAACCCCAGGGCTCACCC-------------AAGGAGTAGAATTTCT----CAC
Rat       GTGGGTGCCTCTTGGTC--TC------------------CATGGA--CTAACTCCCT----CCCTG-GGC---------------ACAGAGAGAACAGAATTCC-----CAC
Mouse     GTAGGTGCCTCTTGACC--TG------------------CGTGGG-----ACTACCT----TCCTG-GGC---------------ACAGAGA--ACAGAATTCC-----CAC
Chicken   GTACGTCCGAGCAGGGCAGCGGGGATGGGGGTGAGCCCGTGCGGGCGCT-GCAGGGAGCGCTCCCGAAGCGCTCCGTGCATTAAACCGCAGCCGAGGGCTCCCCGCAGCCC
BT-Upper  GTTAGTATCCC--AGCC--TTGTCATAGGG-----CCGCCGAACTGATAATGTAGTAATGACATTTTAATGCATCGTGTGT-GTGTGT--ACCATG----CTCTGGTATAT
AS-Upper  GTTAGTATCCC--AGCC--TTGTCATAGTG-----CCGCCAAACTGATAATATAGTAATGACATTTTAATACATCGTGTGT-GTGTGT--ACCATG----CTCTGGTATAT
RT-Upper  GTTAGTATCCC--AGCC--TTGTCATAGTG-----CCGCCGAATTGATCATGTAGTAATGACATTTTAATGCATCGTGTGT-GTGTGTGTACCATG----CTCTGGTATAT
BT-Lower  GTTAGTATCCA-----------------------------------------------------------------------------------ATG----CTCTGGTGTAT
AS-Lower  GTTAGTATCCA-----------------------------------------------------------------------------------ATG----CTCTGGTGTAT
RT-Lower  GTTAGTATCCC-----------------------------------------------------------------------------------ATG----CTCTGGTATAT

                                                                                                     186
Human     TGGCCCCCTC--TCCCCCA---------CCTCCAAGCT-TGGCCTTTCGGCTCAGATCTCAG-------CCCACAGCTGGCCTGATCTGGGTCTCCCCTCCCACCCT-CAG
Monkey    -----CCCTC--TCCCCCA---------CCTCCAAGCT-TGGCCTTTCGGCTCAGATCTCAG-------CCCACAGCTGGCCTGATCTGGGTCTCCCCTCCCACCCT-CAG
Pig       T--CCTGCCC--TTCTCCG---------CT-----GCC-TGGCATTTCAGCTCAGATCCTGA-------CCCGGAGCTGGCCTGATCTGGGTCTTCCCTCCCACCCC-CAG
Rabbit    TAGGATCCCC--TCCCACA---------TCCCAACCCC-TGGCACTCTGGCTCAG------------------AGCTGGCCTAACCCGGGGCTGCCTTCTCACGCC-CAG
Rat       TG---CTCTC--CTCCAAG---------GCTCCAAGTC-TGACATGTCAGCTCAGGCCCCAG-------CCCAGAGTTGACCTAACATGGGTCTCCTCTCC-ATCCT-CAG
Mouse     TG---TTCTC--TTCCCTG---------ACTCCGAGTC-TAAC---------------------------------------CTAACATGGGTCTCCCCTCC-ATCCC-CAG
Chicken   TGGAGCCCGCAGCTCCCCAGCG-GCGGGGCTCTGAGCCACGGCTCCTGGGCACGGCGGTGGGGTGGGGTCCTCCCGTCTGCCCGTCCCCAGCGTGCCGTCCGTCCCCGCAG
BT-Upper  TGAACAATTGAAAATGAAATCTTACTGAATTCTT--TTAAAG-TGCTGGGC-CATTCACCAC-TC-TTCCC----------------CTC-------TCGCCTCTCCCAG
AS-Upper  TGAACAATTTAAAATTAAATCTTAATGAATTCTT--TTAAAG-TGCTGGGC-CATTCACCAC-TC-TTCCC----------------CTC-------TCTCCTCTCCCAG
RT-Upper  TGAACAATTCAAAATGAAATCTTACTCAATTATT--TTAAAG-TGCTGGGC-CATTCACCAC-TGTTTCCCAACCAATGCCTGACCCTCTC-------TCTCCTCTCCCAG
BT-Lower  TGAACAATTCAAAATTAAATCTTACTCAATTCTA--TTAAAG-TGCTGGGC-CATTCACCAC-TGTTGCCCAACCAATGTCTGACCCTCTCCCCTCCCTCTCCTCTCCCAG
AS-Lower  TGAACAATTCAAAATTAAATCTTACTCAATTCTA--TTAAAG-TTCTGGGC-CATTCACCAC-TGTTGCCCAACCAATGTCTGACCCTCTCCCCTCCCTCTCCTCTCCCAG
RT-Lower  TGAAAAATTCTAAATTGAATCTTACTCAATTCCA--TTAAAG-TGCTGGGC-CATTCACCAC-TGTTGCCCAACCAATGTCTGACCCTCTCCCCTCTCTCTCCTCTCCCAG
```

Fig. 4.3. UPGMA trees produced from cDNA (A) and IVS-II (B) data using the program MEGA. Distance measurements were calculated from the number of pairwise nucleotide differences (p-distance) of nonsynonymous codons. Bootstrapping was performed with 1000 replications. *BCL* values for each node are shown in a large font size while branch lengths are shown below each branch in a small font size. The genome duplication event is denoted by a ●.

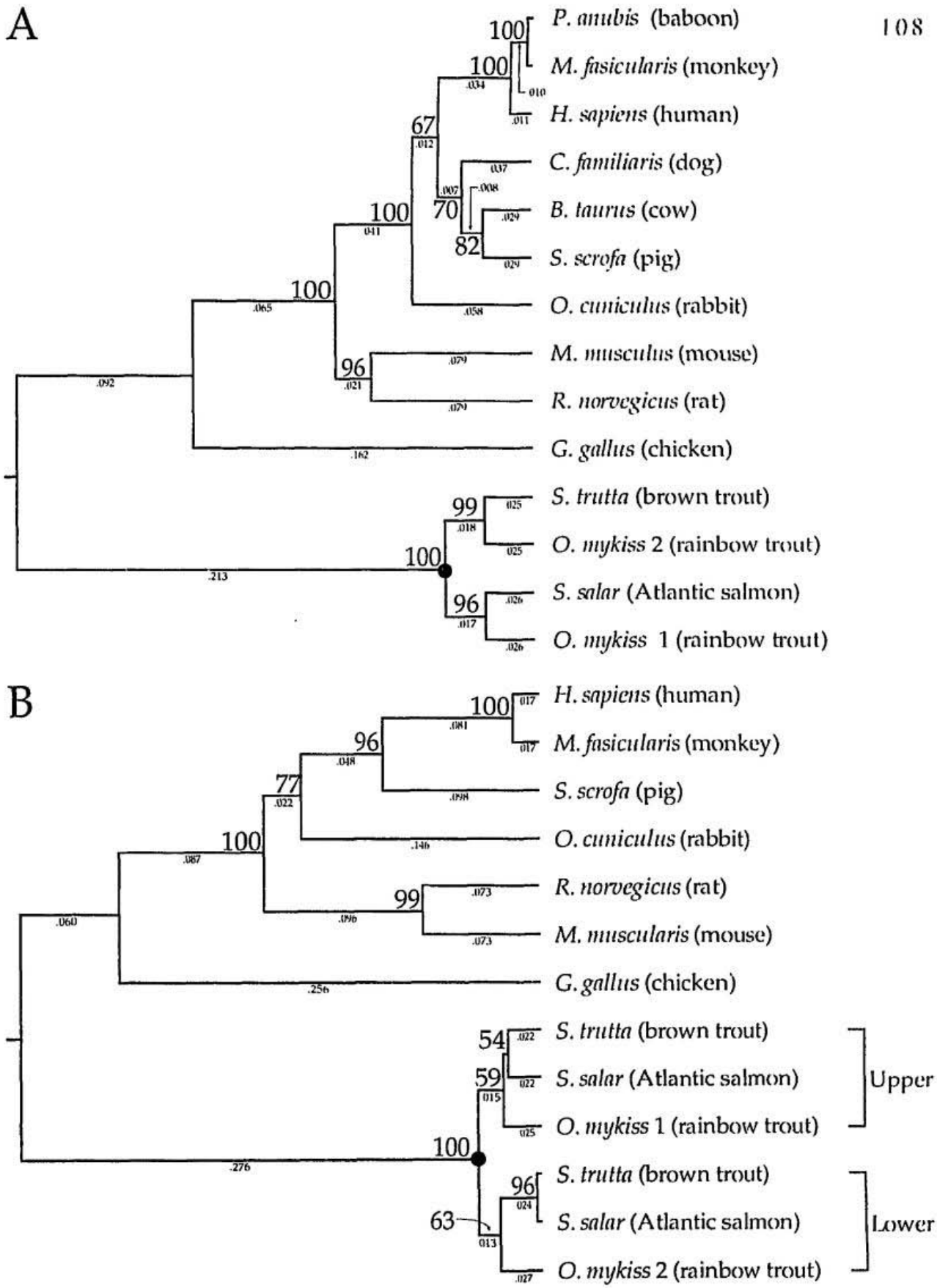
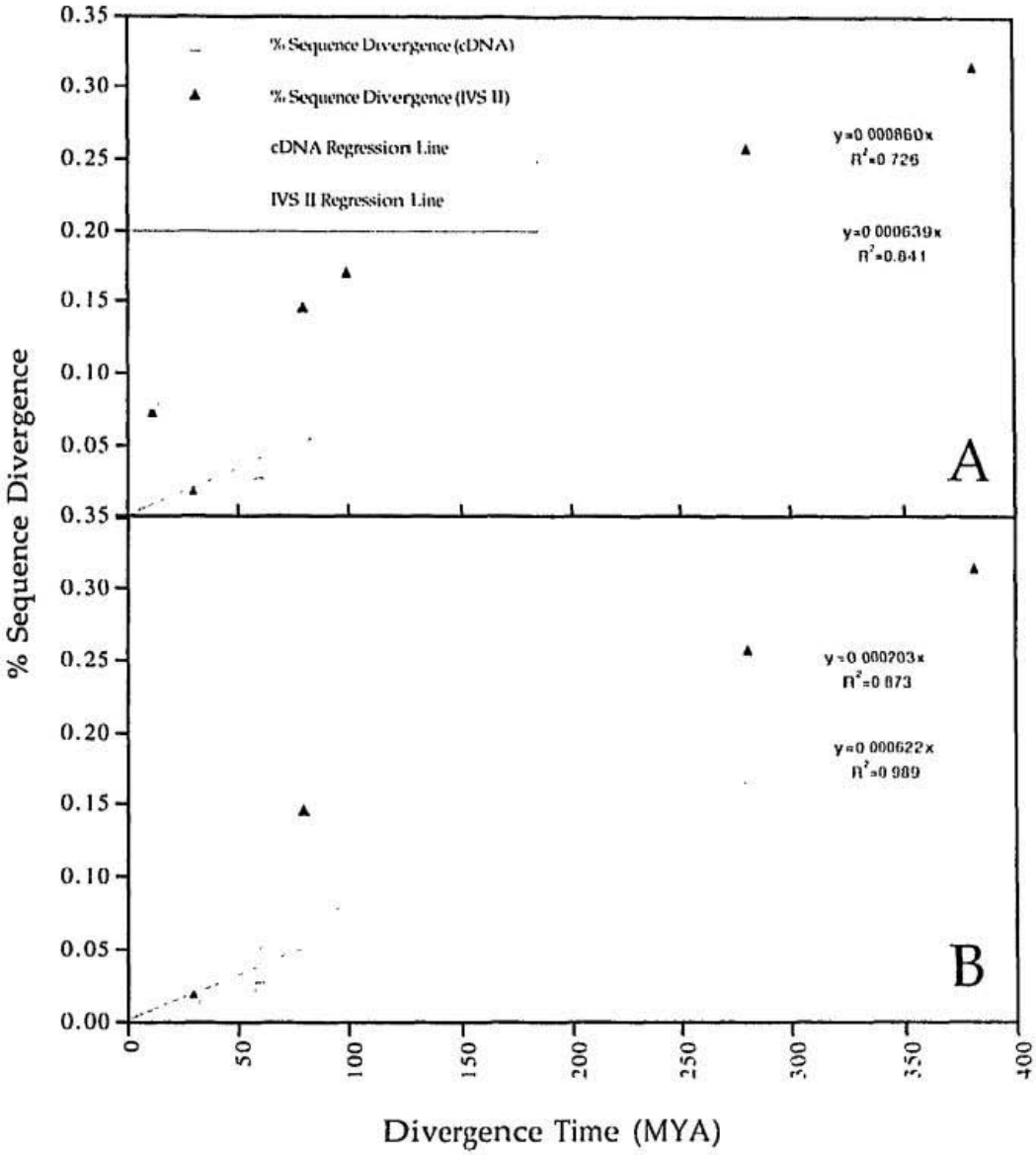

A
P. anubis (baboon)
M. fasicularis (monkey)
H. sapiens (human)
C. familiaris (dog)
B. taurus (cow)
S. scrofa (pig)
O. cuniculus (rabbit)
M. musculus (mouse)
R. norvegicus (rat)
G. gallus (chicken)
S. trutta (brown trout)
O. mykiss 2 (rainbow trout)
S. salar (Atlantic salmon)
O. mykiss 1 (rainbow trout)
B
H. sapiens (human)
M. fasicularis (monkey)
S. scrofa (pig)
O. cuniculus (rabbit)
R. norvegicus (rat)
M. muscularis (mouse)
G. gallus (chicken)
S. trutta (brown trout)
S. salar (Atlantic salmon)
O. mykiss 1 (rainbow trout)
Upper
S. trutta (brown trout)
S. salar (Atlantic salmon)
O. mykiss 2 (rainbow trout)
Lower

Fig. 4.4. Molecular clocks produced from cDNA and IVS II data. In each case, the evolutionary distance (i.e. branch length) of a particular divergence point was taken from Fig. 4.3 and plotted against the divergence time (MYA) as determined from the fossil record. □ represents data points from the cDNA tree. ▲ represents data points from rodent lineages. The regression lines pass through the origin and the equations/R^2 values are shown. A, molecular clock based on all distance data. B, molecular clock based on non-rodent distance data.


% Sequence Divergence (cDNA)
% Sequence Divergence (IVS II)
cDNA Regression Line
IVS II Regression Line
y=0.000860x
R²=0.726
y=0.000639x
R²=0.841
A
y=0.000703x
R²=0.873
y=0.000622x
R²=0.989
B
% Sequence Divergence
Divergence Time (MYA)
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
50
100
150
200
250
300
350
400

References

Allendorf, F.W. and G.H. Thorgaard. Tetraploidy and the evolution of salmonid fishes. In *Evolutionary Genetics of Fishes*, B.J. Turner, Ed., Plenum Press, New York, New York, 1984, pp. 1-53.

Andersson, E., and T. Matsunaga. 1993. Complete cDNA sequence of a rainbow trout *IgM* gene and evolution of vertebrate *IgM* constant domains. Immunogenetics **38**:243-250.

Andersson, E., B. Peixoto, V. Törmänen, and T. Matsunaga. 1995. Evolution of the immunogobulin M constant region genes of salmonid fish, rainbow trout (*Oncorhynchus mykiss*) and Arctic charr (*Salvelinus alpinus*): implications concerning divergence time of species.Immunogenetics **41**:312-315.

Aron, L., S. Jones, and C.J. Fielding. 1978. Human plasma lecithin-cholesterol acyltransferase. J. Biol. Chem. **253**:7220-7226.

Assmann, G., A. von Eckardstein, and H. Funke. Mutations in apolipoprotein genes and HDL metabolism. In *Structure and Function of Apolipoproteins*, Rosseneu, M., Ed., CRC Press, Boca Raton, 1992, pp. 85-122.

Babin, P.J. 1987. Plasma lipoprotein and apolipoprotein distribution as a function of density in the rainbow trout (*Salmo gairdneri*). Biochem. J. **246**:425-429.

Babin, P.J. and Vernier, J-M. 1989. Plasma lipoproteins in fish. J. Lipid Res. **30**:467-489.

Baker, M.E. 1988. Is vitellogenin an ancestor of apolipoprotein B-100 of human low-density lipoprotein and human lipoprotein lipase? Biochem. J. **255**:1057-1060.

Barber, D.L., E.J. Sanders, R. Aebersold, and W.J. Schneider. 1991. The receptor for yolk lipoprotein deposition in the chicken oocyte. J. Biol. Chem. **266**:18761-18770.

Bhattacharyya. N., R. Chattapadhyay, A. Hirsch, and D. Banerjee. 1991. Isolation, sequencing, and characterization of the chicken apolipoprotein-A-I-encoding gene. Gene **104**:163-168.

Birchbauer, A., G. Knipping, B. Juritsch, H. Aschauer, and R. Zechner. 1993. Characterization of the apolipoprotein AI and CIII genes in the domestic pig. Genomics **15**:643-652.

Breathnach, R., and P. Chambon. 1981. Organization and expression of eukaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349-383.

Breathnach, R., J.L. Mandel, and P. Chambon. 1977. Ovalbumin gene is split in chicken DNA. Nature **270**:314-319.

Britten, R.J., and D.E. Kohne. 1968. Repeated sequences in DNA. Science **161**:529-540.

Bulmer, M., K.H. Wolfe, and P.M. Sharp. 1991. Synonymous substitution rates in mammalian genes: Implications for the molecular clock and the relationship of mammalian orders. Proc. Natl. Acad. Sci. USA **88**:5974-5978.

Cano, R.J., H.N. Poinar, N.J. Pieniazek, A. Acra, and G.O. Poinar, Jr.. 1993. Amplification and sequencing of DNA from a 120-135-million-year-old weevil. Nature **363**:536-538.

Carroll, R.L. Vertebrate Paleontology and Evolution. W.H. Freeman and Company, New York, 1988.

Chan, L., and W. Li. Apolipoprotein gene expression, structure, and evolution. In *Structure and Function of Apolipoproteins*, Rosseneu, M., Ed., CRC Press, Boca Raton, 1992, pp. 33-61.

Delcuve, G.P., J.M. Sun, and J.R. Davie. 1992. Expression of rainbow trout apolipoprotein A-I genes in liver and hepatocellular carcinoma. J. Lipid Res. **33**:251-262.

Eck, R.V. and M.O. Dayhoff. Atlas of Protein Sequence and Structure 1966. National Biomedical Research Foundation, Silver Spring, MD, 1966.

Eisenberg, S. 1984. High density lipoprotein metabolism. J. Lipid Res. **25**:1017-1058.

Estoup, A., P. Presa, F. Kreig, D. Vaiman, and R. Guyomard. 1993. $(CT)_n$ and $(GT)_n$ microsatellites: a new class of genetic markers for *Salmo trutta L.* (brown trout). Heredity **71**:488-496.

Ferguson, A., and C.C. Flemming. Evolutionary and taxonomic signifigance of protein variation in the brown trout (*Salmon trutta L.*) and other salmonid fishes. In *Protein Polymorphism: Adaptive and Taxonomic Signifigance*, Oxford, G.S., and D. Rollinson, Eds., Academic Press, London, 1983.

Fitch, W.M. 1977. On the problem of discovering the most parsimonious tree. Am. Natur. **111**:223-257.

Flynn, L.J., L.L. Jacobs, and E.H. Lindsay. Problems in muroid phylogeny: Relationships to other rodents and origin of major groups. In *Evolutionary Relationships Among Rodents*, W.P. Luckett and J-L. Hartenberger, Eds., Plenum Press, New York, 1985, pp. 589-616.

Francone, O.L., A. Guraker, and C.J. Fielding. 1989. Distribution and functions of lecithin:cholesterol acyltransferase and cholesteryl ester transfer protein in plasma lipoproteins. Evidence for a functional unit containing these activities together with apolipoproteins A-I and D that catalyzes the esterification and transfer of cell derived cholesterol. J. Biol. Chem. **264**:7066-7072.

Funk, V.A. and Brooks, D.R. Phylogenetic Systematics as the Basis of Comparative Biology. Smithsonian Institution Press, Washington, 1990.

Haddad, I.A., J.M. Ordovas, T. Fitzpatrick, and S.K. Karathanasis. 1986. Linkage, evolution, and expression of the rat apolipoprotein A-I, C-III, and A-IV genes. J. Biol. Chem. **261**:13268-13277.

Higashimoto, Y., and R. Liddle. 1993. Isolation and characterization of the gene encoding rat glucose-dependent insulinotropic peptide. Biochem. Biophys. Res. Commun. **193**:182-190.

Higgins, D.G., A.J. Bleasby, and R. Fuchs. 1992. CLUSTAL V: improved software for multiple sequence alignment. Comput. Appl. Biosci. **8**:189-191.

Higuchi, R., B. Bowman, M. Freiberger, O.A. Ryder, and A.C. Wilson. 1984. DNA sequences from the quagga, an extinct member of the horse family. Nature **312**:282-284.

Hixson, J.E., S. Borenstein, I.A. Cox, D.I. Rainwater, and J.L. VandeBerg. 1988. The baboon gene for apolipoprotein A-I: characterization of a cDNA clone and identification of DNA polymorphisms for genetic studies of cholesterol metabolism. Gene **74**:483-490.

Holland, S.K., and C.C.F. Blake. Proteins, exons, and molecular evolution. In *Intervening Sequences in Evolution and Development*, E.M. Stone and R.J. Schwartz, Eds., Oxford University Press, Inc., New York, New York, 1990, pp. 10-42.

Janczewski, D.N., N. Yuhki, D.A. Gilbert, G.T. Jefferson, and S.J. O'Brien. 1992. Molecular phylogenetic inference from saber-toothed cat fossils of Rancho La Brea. Proc. Natl. Acad. Sci. USA **89**:9769-9773.

Karathanasis, S.K., V.I. Zannis, and J.L. Breslow. 1993. Isolation and characterization of the human apolipoprotein A-I gene. Proc. Natl. Acad. Sci. USA **80**:6147-6151.

Kido, Y., M. Himberg, N. Takasaki, and N. Okada. 1994. Amplification of distinct subfamilies of short interspersed elements during evolution of the Salmonidae. J. Mol. Biol. **241**:633-644.

Kielan-Jaworowska, Z., T.M. Brown, and J.A. Lillegraven. Eutheria. In *Mesozoic Mammals: The First Two-Thirds of Mammalian History*, Lillegraven, J.A., Z. Kielan-Jaworowska, and W.A. Clemens, Eds., Univ. of California Press, Berkley, 1979, pp. 221-258.

Kimura, M. 1987. J. Mol. Evol. **26**:24-33.

Kimura, M. The Neutral Theory of Molecular Evolution. Cambridge: The Press Syndicate of the University of Cambridge, 1983.

Kuczek, E.S., and G.E. Rogers. 1987. Sheep wool (glycine + tyrosine)-rich keratin genes. A family of low sequence homology. Eur. J. Biochem. **166**:79-85.

Kumar, S., K. Tamura, and M. Nei. 1993. MEGA: Molecular Evolutionary Genetics Analysis, version 1.0. The Pensylvania State University, University Park, PA.

Larhammar, D., and C. Risinger. 1994. Molecular genetic aspects of tetraploidy in the common carp *Cyprinus carpio*. Mol. Phylogenet. Evol. **3**:59-68.

Law, S.W., G. Gray, and H.B. Brewer, Jr.. 1983. cDNA cloning of human apoA-I: amino acid sequence of preproapoA-I. Biochem. Biophys. Res. Commun. **112**:257-264.

Lewin, B.L. Genes V, Oxford University Press Inc., New York, New York, 1994.

Li, W-H. and D. Graur. Fundamentals of Molecular Evolution, Sinauer Associates, Inc., Sunderland, Mass, 1991.

Li, W-H., C-I. Wu, and C-C. Luo. 1985. A new method for estimating synonymous and non-synonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol. Biol. Evol. **2**:150-174.

Li, W-H., M. Gouy, P.M. Sharp, C. OhUigin, and Y-W Yang. 1990. Molecular phylogeny of Rodentia, Lagomorpha, Primates, Artiodactyla, and Carnivora and molecular clocks. Proc. Natl. Acad. Sci. USA **87**:6703-6707.

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

Lim, S.T. and G.S. Bailey. 1977. Gene duplication in Salmonid fishes: Evidence for duplicated but catalytically equivalent A_4 lactate dehydrogenases. Biochem. Genet. **15**:707-721.

Lipman, D.J., and W.R. Pearson. 1985. Rapid and sensitive protein similarity searches. Science **227**:1435-1441.

Litt, M., and J.A. Luty. 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am. J. Hum. Genet. **44**:397-401.

Luo, C-C., W-H. Li, and L. Chan. 1989. Structure and expresion of dog apolipoprotein A-I, E, and C-I mRNAs: implications for the evolution and functional constraints of apolipoprotein structure. J. Lipid Res. **30**: 1735-1746.

May, B., M. Stoneking, and J. E. Wright. 1982. Joint segregation of biochemical loci in Salmonidae. III. Linkage associations in Salmonidae including data from rainbow trout (*Salmo gairdneri*). Biochem. Gene. **20**:29-40.

McLaughlin, P.J., and M.O. Dayhoff. Evolution of species and proteins: A time scale. In *Atlas of Protein Sequence and Structure*, M.O. Dayhoff, Ed., National Biomedical Research Foundation, Washington, 1972. pp. 47-52.

Miller, G.J. and N.E. Miller. 1975. Plasma high-density-lipoprotein concentration and development of ischaemic heart-disease. Lancet **1**:16-19.

Miller, S.A., and J.P. Harley. Zoology, second edition, Wm. C. Brown Communications Inc., Dubuque, Iowa, 1994.

Miyata, T., T. Yasunaga, and T. Nishida. 1980. Nucleotide sequence divergence and functional constraint in mRNA evolution. Proc. Natl. Acad. Sci. USA **77**:7328-7332.

Morrisett, J.D., R.L. Jackson, and A.M. Gotto, Jr. 1977. Lipid-protein interactions in the plasma lipoproteins. Biochim. Biophys. Acta **472**:93-133.

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

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

Murray, R.W., and K.R. Marotti. 1992. Nucleotide sequence of the cynomolgus monkey apolipoprotein A-I gene and corresponding flanking regions. Biochim. Biophys. Acta **1131**:207-210.

Naylor, L.H., and E.M. Clark. 1990. $d(TG)_n \cdot d(CA)_n$ sequences upstream of the rat prolactin gene form Z-DNA and inhibit gene transcription. Nucl. Acids Res. **18**:1595-1601.

Niu, L., and G.F. Crouse. 1993. Exon mapping by PCR. Nucl. Acids Res. **21**:769-770.

Novacek, M.J. Information for molecular studies from anatomical and fossil evidence on higher eutherian phylogeny. In *Macromolecular Sequences in Systematic and Evolutionary Biology*, Goodman, M., Ed., Plenum Press, New York, 1982, pp. 3-41.

O'hUigin, C., and W-H. Li. 1992. The molecular clock ticks regularly in muroid rodents and hamsters. J. Mol. Evol. **35**:377-384.

O'hUigin, C., L. Chan, and W-H. Li. 1990. Cloning and sequencing of bovine apolipoprotein A-I cDNA and molecular evolution of apolipoproteins A-I and B-100. Mol. Biol. Evol. **7**:327-339.

Ohno, S., U. Wolf, and N.B. Atkin. 1968. Evolution from fish to mammals by gene duplication. Hereditas (Lund) **59**:169-187.

Palmer, J., and J. Logsdon. 1991. The recent origins of introns. Curr. Opin. Genet. Dev. **1**:470-477.

Pan, T.C., Q.L. Hao, T.T. Yamin, P.H. Dai, B.S. Chen, S.L. Chen, P.A. Kroon, and Y.S Chao. 1987. Rabbit apolipoprotein A-I mRNA and gene. Evidence that rabbit apolipoprotein A-I is synthesized in the intestine but not in the liver. Eur. J. Biochem. **170**:99-104.

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

Poinar, H.N., G.O. Poinar, Jr., and R.J. Cano. 1993 (unpublished). Genbank Accession #L080474

Powell, R., D.G. Higgins, J. Wolff, L. Byrnes, M. Stack, P.M. Sharp, and F. Gannon. 1991. The salmon gene encoding apolipoprotein A-I: cDNA sequence, tissue expression and evolution. Gene **104**:155-161.

Pownall, H.J., A.M. Gotto, Jr., R.D. Knapp, and J.B. Massey. 1986. The helical hydrophobic momment avoids prolines in phospholipid binding proteins. Biochem. Biophys. Res. Commun. **139**:202-208.

Pownall, H.J., and A.M. Gotto, Jr. Human plasma apolipoproteins in biology and medicine. In *Structure and Function of Apolipoproteins*, Rosseneu, M., Ed., CRC Press, Boca Raton, 1992, pp. 1-32.

Pownall, H.J., R.D. Knapp, A.M. Gotto, Jr., and J.B. Massey. 1983. Helical amphipathic momment: application to plasma lipoproteins. FEBS Lett. **159**:17-23.

Rand, D.M. 1994. Thermal habit, metabolic rate, and the evolution of mitochondrial DNA. Trends Ecol. Evol. **9**:125-131.

Rassmann, K., C. Schlötterer, and D. Tautz. 1991. Isolation of simple sequence loci for use in polymerase chain reaction-based DNA fingerprinting. Electrophoresis **12**:113-118.

Rauscher F., L. Sambucetti, T. Curran, R. Distel, and B. Spiegelman. 1988. Common DNA binding site for Fos protein complexes and transcription factor AP-1. Cell **52**:471-480.

Romer, A.S. Vertebrate Paleontology. Univ. of Chicago Press, Chicago, 1968.

Romero-Herrera, A.E., and H. Lehmann. 1972. The myoglobin of primates. Biochim. Biophys. Acta **278**:62-67.

Romero-Herrera, A.E., H. Lehmann, K.A. Joysey, and A.E. Friday. 1973. Molecular evolution of myoglobin and the fossil record: a phylogenetic synthesis. Nature **246**:389-395.

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

Savage, D.E., and D.E. Russell. Mammalian Paleofaunas of the World. Addison-Wesley, 1983.

Schmitz, G., A. Roebenek, U. Lohmann, and G. Assmann. 1985. Interaction of high density lipoproteins with cholesteryl ester-laden macrophages: biochemical and morphological characterization of cell surface receptor binding, endocytosis and resecretion of high density lipoproteins by macrophages. EMBO J. **4**:613-672.

Seib, T., C. Welter, M. Engel, B. Theisinger, and S. Dooley. 1994. Presence of regulatory sequences within intron 4 of human and murine c-myb genes. Biochim. Biophys. Acta **1219**:285-292.

Shoshani, J. 1986. Mammalian phylogeny-comparison of morphological and molecular results. Mol. Biol. Evol. **3**:222-224.

Slotte, J.P., J.F. Oram, and E.L. Bierman. 1987. Binding of high density lipoproteins to cell receptors promotes translocation of cholesterol from intracellular membranes to the cell surface. J. Biol. Chem. **262**:12904-12907.

Smith, G.R., K. Swirydczuk, P.G. Kimmel, and B.H. Wilkinson. 1982. First biostratiography of late Miocene to Pleistocene sediments of the western Snake River plain, Idaho. Idaho Bureau of Mines and Geology Bull. **26**:519-541.

Sokal, R.R., and C.D. Michener. 1958. A statistical method for evaluating systematic relationships. Univ. Kansas Sci. Bull. **28**: 1409-1438.

Sourdis, J., and C. Krimbas. 1987. Accuracy of phylogenetic trees estimated from DNA sequence data. Mol. Bio. Evol. **4**:159-166.

Sparrow, D.A., P.M. Laplaud, M. Saboureau, G. Zhou, P.J. Dolphin, A.M. Gotto, Jr., and J.T. Sparrow. 1995. Plasma lipid transport in the hedgehog: partial characterization of structure and function of apolipoprotein A-I. J. Lipid Res. **36**:485-495.

Steyrer, E., D.L. Barber, and and W.J. Schneider. 1990. Evolution of lipoprotein receptors. The chicken oocyte receptor for very low density lipoprotein and vitellogenin binds the mammalian ligand apolipoprotein E. J. Biol. Chem. **265**: 19575-19581.

Stoffel, W., R. Muller, E. Binczek, and K. Hoffman. 1992. Mouse apolipoprotein AI. cDNA-derived primary structure, gene organisation and complete nucleotide sequence. Biol. Chem. Hoppe-Seyler **373**:187-193.

Stoltzfus, A., D.F. Spencer, M. Zuker, J.M. Logsdon, Jr., and W.F. Doolittle. 1994. Testing the exon theory of genes: the evidence from protein structure. Science **265**:202-207.

Svardson, G. 1945. Chromosome studies of salmonidae. Rep. Swed. State Inst. Freshwater Fish. Res. **23**:1-151.

Swofford, D.L. and G.J. Olsen. Phylogeny reconstruction. In *Molecular Systematics*, Hills, D.M. and C. Moritz, Eds., Sinauer Associates, Inc., Sunderland, Mass., 1990, pp. 411-501.

Szalay, F.S. Phylogenetic relationships and a classification of the eutherian mammalia. In *Major Patterns in Vertebrate Evolution*, Hecht, M.K., P.C. Goody, and B.M. Hecht, Eds., Plenum Press, New York, 1977, pp. 15-374.

Tabas, I. and A.R. Tall. 1984. Mechanism of the association of HDL_3 with endothelial cells, smooth muscle cells, and fibroblasts. Evidence against the role of specific ligand and receptor proteins. J. Biol. Chem. **259**:13897-13905.

Tateno , Y., M. Nei, and F. Tajima. 1982. Accuracy of estimated phylogenetic trees from molecular data. I. Distatly related species. J. Mol. Evol. **18**:387-404.

Tautz, D. 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res. **17**:6463-6471.

Taylor, A.C., W.G Sherwin, and R.K. Wayne. 1994. Genetic variation of microsatellite loci in a bottlenecked species: the northern hairy-nosed wombat *Lasiorhinus krefftii*. Mol. Ecol. **3**:277-290.

van Holde, K., and J. Zlatanova. 1993. Unusual DNA structures, chromatin, and transcription. BioEssays **16**:59-68.

Weber, J.L., and P.E. May. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am. J. Hum. Genet. **44**:388-396.

Wilson, A.C., H. Ochman, and E.M. Prager. 1987. Molecular time scale for evolution. Trends in Genet. **3**:241-247.

Winterø, A.K., M. Fredholm, and P.D. Thomsen. 1992. Variable $(dG\text{-}dT)_n\cdot(dC\text{-}dA)_n$ sequences in the porcine genome. Genomics **12**:281-288.

Woese, C.R. The use of ribosomal RNA in reconstructing evolutionary relationships among bacteria. In *Evolution at the Molecular Level*, Selander, R.K., A.G. Clark, and T.S. Whittam, Eds., Sinauer Associates Inc., Sunderland, Massachusetts, 1991, pp. 1-24.

Wright, J. E., B. May, M. Stoneking, and G.M. Lee. 1980. J. Pseudolinkage of the duplicate loci for supernatant aspartate aminotransferase in brook trout, *Salvelinus fontinalis*. Hered. **71**:223-228.

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