-A36

5

DNA technology and phylogeny of fish

5.1 MOLECULAR PHYLOGENETIC STUDIES OF FISH

Axel Meyer

Abstract

Knowledge of the evolution of DNA, particularly mitochondrial DNA, of fish and the phylogenetic relationships among fish has increased rapidly since the invention of the polymerase chain reaction (PCR). This enzymatic cloning technique, together with direct sequencing, has simplified and dramatically accelerated the accumulation of DNA sequence information. Methods of data collection and data analysis for phylogenetic studies with particular emphasis on fish are outlined. Aspects of the biology of mitochondrial DNA that pertain to phylogenetic reconstruction are reviewed and advantages of DNA sequences over alternative DNA-based genetic markers are highlighted. Examples of phylogenetic work based on mitochondrial and nuclear DNA sequences are used to illustrate the methods, advantages and potential problems with techniques, choice of genes and analyses.

5.1.1 Introduction

There are several kinds of biochemical data that can be used to infer phylogenetic relationships among species. Allozyme, immunological and DNA–DNA hybridization data have been widely used but are now increasingly being replaced by several types of DNA-based data. Since the advent of the polymerase chain reaction (PCR) in 1985–1986 (Mullis *et al.*, 1986; Saiki *et al.*, 1985, 1988; Wrishnik, Higuchi and Stoneking, 1987), our knowledge about DNA of fish has increased dramatically. The impact of this technological innovation on the understanding of the evolution of DNA of fish and phylogenetic relation-ships among fish is the focus of this section. From my biased perspective, I will attempt to summarize the advantages and disadvantages of currently used DNA-based molecular data to deduce genealogical relationships among fish. Data collection and analysis will be briefly touched on and recent publications and reviews that provide more detailed information will be recommended.

Genetics and Evolution of Aquatic Organisms. Edited by A.R. Beaumont.
Published in 1994 by Chapman & Hall, 2–6 Boundary Row, London SE1 8HN.
ISBN 0-412-49370-5

5.1.2 Methods and kinds of DNA data

(a) Restriction endonuclease methods

With this method, DNA (usually mtDNA) is cut with restriction enzymes (endonucleases) and the resulting restriction fragment length polymorphisms (RFLP) are used as binary characters in a phylogenetic analysis (restriction-fragment data). Often, restriction maps are constructed from the RFLP patterns (after simultaneous digests with two enzymes) and then analysed in a phylogenetic analysis as binary characters, coding for absence or presence of particular restriction sites (restriction-site data). Usually enzymes with six-base recognition sites are used for evolutionary studies among more distantly related species and enzymes with four-base-pair (bp) recognition sites are used in studies that require more detailed information, e.g. in investigations among more closely related species and population-level questions within species. This method has enjoyed widespread application (reviews: Wilson, Thomas and Beckenbach, 1985; Avise, 1986; Avise *et al.*, 1987; Harrison, 1991; but see Wilson *et al.*, 1989) and continues to be applied for a wide range of taxonomic questions in fish.

Methodological problems with restriction-fragment data are that the assumption of independence (see below) is violated, and deletions and insertions are problematic and are potential sources of error when trying to establish homology between restriction fragments (e.g. Swofford and Olsen, 1990). For these reasons caution should be applied when interpreting RFLP patterns for phylogenetic analysis. Restriction-site data, a map derived from RFLP data, also have problems with phylogenetic reconstruction, due to the asymmetry with which restriction sites are gained and loss. The loss of a restriction site is much more likely than a gain. If a particular restriction site is 6 bp long, and a 6-bp stretch of DNA is different by only one nucleotide for the recognition sequence, only 1 out of 18 substitutions (substitutions of the same nucleotide at the same site remain undetected) is going to create this site, i.e. it is unlikely that sites will be gained. However, losing sites is going to be much more frequent, i.e. any of these 18 mutations at this restriction site is going to result in the loss of that site (Templeton, 1983a,b). A special case of parsimony (Dollo parsimony, see below) takes the asymmetry of gains and losses into consideration during the phylogenetic analysis of RFLP data.

An additional disadvantage of restriction data is that results of RFLP and even restriction sites are not immediately transferable between laboratories. That is because the same endonucleases are not always used and the same kinds of gels are not run by all researchers. In this respect, restriction data suffer from the same drawbacks as allozyme data; they tend to be laboratory-specific and often even project-specific results. Although endonuclease data contributed tremendously to our increased understanding of intra- and interspecific genetic variation and phylogenetic relationships among fish and remain a viable technique, I am biased in favour of actual DNA sequences (see also Wilson *et al.*, 1989). For population-level work, PCR (see below) and restriction analyses are sometimes combined: known DNA fragments (usually mtDNA) are amplified via PCR and then cut with restriction enzymes rather than sequenced. The advantage of this combined approach is that larger sample sizes can be screened than if every individual is sequenced.

(b) The polymerase chain reaction and direct sequencing

PCR is an enzymatic cloning technique that allows the amplification of any stretch of DNA (within size limits) that is flanked by synthetic oligonucleotide 'primers' (Saiki *et al.*, 1985, 1988). The primers are usually around 20 bp in length and define the 5′ and 3′ end of the

double-stranded piece of DNA that is going to be amplified. The specificity of the amplification is accomplished through the need for an almost perfect fit of the primers to the template DNA. During each cycle of PCR, the number of copies of the DNA fragment delineated by the primers at either end is doubled. Usually 25–40 cycles are completed in a thermal cycler in about three hours. PCR is much faster and cheaper than conventional cloning techniques. First, a double-stranded PCR product is produced (Figure 5.1.1 (a)) that is then either sequenced (double-stranded sequencing, or, alternatively 'cycle-sequenced') or subcloned and then sequenced or cut with restriction enzymes (see above) or used as template DNA for a subsequent asymmetric amplification (Gyllensten and Erlich, 1988) or digested with an exonuclease to produce single-stranded DNA (Figure 5.1.1 (b)) for direct sequencing of single-stranded DNA. Sequencing gels of single-stranded DNA often allows one to read more base pairs than sequencing gels of double-stranded DNA. Single-stranded PCR-amplified DNA can be as clean as subcloned DNA and routinely more than 300–400 bp can be unambiguously determined from a single sequencing reaction. Figure 5.1.2 shows some partial cytochrome *b* sequences from three-spined sticklebacks that were produced through double- and single-stranded amplification and direct sequencing.

Although only a recent addition to the molecular toolbox, PCR and its application to evolutionary biology has already been reviewed (e.g. White, Arnheim and Erlich, 1989; Arnheim, White and Rainey, 1990; refs in Erlich, 1989; Innis *et al.*, 1990). A new journal entitled *PCR* first appeared in 1991. Details on PCR methodology, technical improvements and modifications and new applications of PCR technology can be found in journals like *PCR* and *BioTechniques*.

The determination of DNA sequences tends to be more time-consuming, costly and technically involved; however, DNA sequences of homologous mitochondrial and nuclear

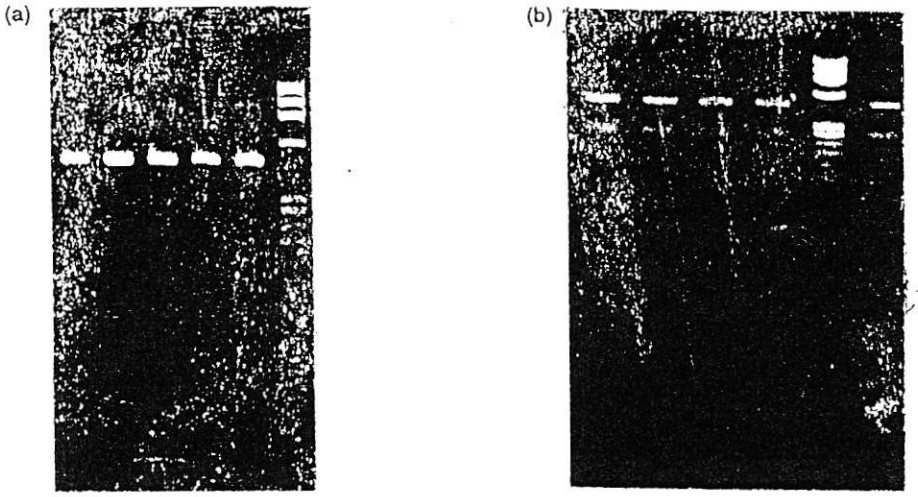

Figure 5.1.1 (a) Double-stranded PCR product of the 5′ end of cytochrome *b* of five species of South American characins (five left lanes, amplified with L14725 and H15148, protocols from Kocher *et al.*, 1989). One major amplification product of the expected size, about 450 bp, is visible. The size standard is Phi-X-174 cut with *Hae* III of 1353, 1078, 872, 602, 310, 281, 271, 234, 194, 118 in lane 6. (b) Single-stranded PCR products produced by asymmetric PCR (Gyllensten and Erlich, 1988). L14725 was diluted 1 : 100. The same five fish as in (a) (size standard in lane 5). The targeted double-stranded product is seen above and the faster migrating single-stranded product is below. Data from G. Orti's unpublished study on the phylogenetic relationships among the major lineages of characin fish.

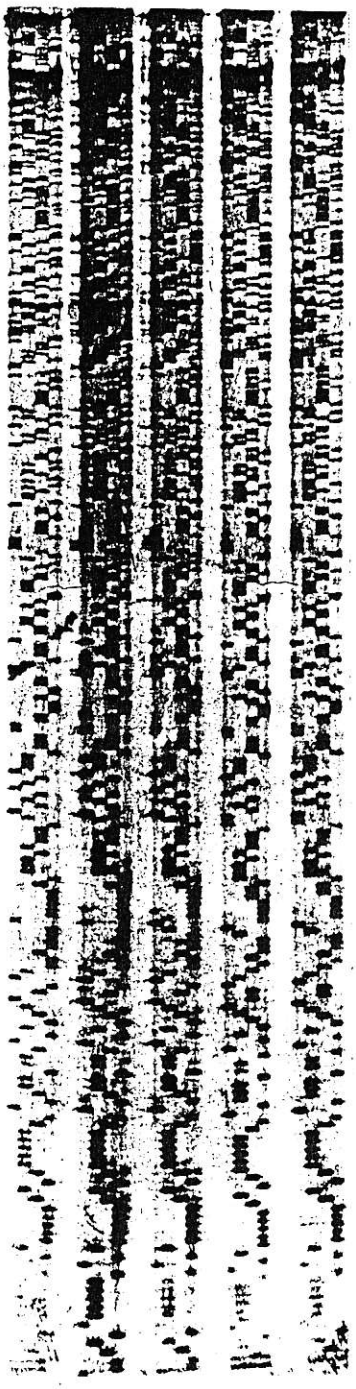

Figure 5.1.2 Example of the quality of direct sequencing of PCR-amplified DNA: 5′ end of the cytochrome *b* gene of five individuals of the three-spined stickleback (*Gasterosteus aculeatus*) from several circumpolar locations (from left to right: Quebec, New York, Japan, Rhode Island, Quebec) (Orti, Bell and Meyer, unpublished results). Up to 350 bp can be read from this wedge gel. The loading sequence is G, A, T, C from left to right.

genes will allow direct comparisons and study of DNA from different species that has been obtained in different laboratories. DNA sequences can be stored in databanks (e.g. EMBL, GenBank) and are universally usable, powerful data. The increased costs of DNA sequences compared to RFLP data are far outweighed by their advantage as a universally retrievable and applicable type of data, since homologous data from independent laboratories can be used in direct comparisons for new studies.

(c) RAPD-PCR

A recent PCR-based technique, random amplified polymorphic DNA (RAPD-PCR) (Welsh and McClelland, 1990; Williams *et al.*, 1990), has stirred up a lot of interest among researchers interested in population-level questions (e.g. stock identity, paternity, mating systems). This technique utilizes only one instead of two different PCR primers. This single primer is small, usually in the range of 10 bp, has a random sequence, and anneals on the genome at priming sites close to each other and in an inverted orientation. During the PCR amplification, at low annealing temperatures and with a relatively large number of cycles, random pieces of DNA are amplified and then visualized usually on ethidium-bromide-stained NuSieve-agarose gels (often 2–4% gels). Based on the number and position of bands on these gels, genetic relatedness between the samples is inferred. RAPD-PCR was greeted with enthusiasm, because it seems to provide very detailed genetic information quickly and relatively inexpensively.

Unfortunately, the large initial interest in RAPD-PCR has not been matched by a flood of successful studies. Several technical and analytical problems have dampened the initial excitement. Part of the appeal of PCR is its technical simplicity. However, in RAPD-PCR many factors (Mg^{2+}, primer, template DNA concentrations and exact thermal cycling conditions) need to be carefully controlled to ensure reproducibility (e.g. Hadrys, Balick and Schierwater, 1992). Analytically, the fact that DNA fragments of uncertain homology between individuals are amplified by RAPD-PCR makes the analysis less tractable. However, if these technical and analytical obstacles are solved in the future, this technique might very well become the panacea for population-level work. Any diagnostic RAPD marker can be used to generate specific probes for Southern analysis.

5.1.3 Methods of phylogenetic data analysis

The history of phylogenetic reconstruction has been turbulent and full of acrimony (review: Hull, 1988). There is a large body of literature on phylogenetic reconstruction and several excellent reviews provide an entry to this literature (Felsenstein, 1982, 1988; Swofford and Olsen, 1990; refs in Hillis and Moritz, 1990). All methods make underlying simplifying assumptions about how DNA sequences evolve. No consensus for a single method of phylogenetic reconstruction has been reached among researchers favouring rivalling methodologies. Philosophical as well as practical arguments are used by the proponents of particular methods to argue the superiority of one method over another. Simulation studies, intended to determine which method of phylogenetic reconstruction will provide the best estimate, tend to give preference to the favoured method of the researcher who conducted the simulation study.

Two types of molecular data can be used for phylogenetic analysis. Discrete characters are collected when DNA is sequenced or scored within restriction enzymes. These discrete data provide information about the DNA of a particular individual, often assumed to be characteristic for the species. Further assumptions are the independence and homology of nucleotide positions. If species rather than gene-trees are the purpose of the study, only orthologous

rather than paralogous genes should be compared. Discrete data can be transformed into similarity or distance data, by pairwise comparisons of two sequences.

Distance methods of phylogenetic reconstruction are based on these pairwise distances (see below for different ways of calculating and correcting for 'multiple hits') of sequences and attempt to fit a tree to a distance matrix. The goodness of fit (e.g. least squares methods) of the observed distances to the expected distances (based on the tree) is measured and the topology that minimizes the discrepancy between expected and observed distances is chosen. Several distance methods exist (reviews: Felsenstein, 1988; Swofford and Olsen, 1990; Nei, 1991).

The parsimony method, 'the method of minimum net evolution', aims to find the evolutionary tree that requires the fewest changes of nucleotides to explain the evolution of the DNA sequences under consideration. Its philosophy is the hypothetico-deductive approach in which Ockham's principle is invoked, i.e. evolution is believed to proceed by the shortest, simplest pathway. Parsimony only considers so-called 'phylogenetically informative sites' in the calculation of the topology of the tree. Sites that do not require different numbers of changes on alternative trees of different topology (e.g. sites that are identical or sites that differ only in one of the species under consideration) are ignored. Both distance and parsimony methods only look at part of the information in the data.

Only the maximum-likelihood method attempts to use all the information contained in DNA sequences (Felsenstein, 1981) by using statistical criteria to distinguish between alternative trees. It uses a model of likelihoods of substitution changes and attempts to fit the data with a tree. The likelihood of the topology of a tree is the probability of the data, given the tree and the model. The maximum-likelihood method chooses a tree with a topology and branch length that has the highest likelihood. This method allows for unequal base composition and uneven transition-transversion rates, and does not require a molecular clock. The major practical drawback of the maximum-likelihood method is its inherent computational complexity due to exact probability models of sequence change, which in practice limits the number of sequences that can be analysed in a reasonable amount of time with available computer power.

Unfortunately, studies that test the power of various commonly used methods of phylogenetic reconstruction on known phylogenies are rare (Atchley and Fitch, 1991; Hillis *et al.*, 1992). They have failed to clearly identify a particular methodology as the best. However, the strengths and weaknesses of the alternative methods of phylogenetic recon-struction and the kinds of genes to be used and to be avoided have been made more clear in these studies (Felsenstein, 1988; Swofford and Olsen, 1990; Nei, 1991; Atchley and Fitch, 1991; Hillis *et al.*, 1992).

In practice, the data should be subjected to several methods of phylogeny reconstruction and differences in results will usually pinpoint areas of weakness in the phylogenetic tree. Congruent results of different phylogenetic methods will inspire confidence that a phylogenetic estimate has been found that is closer to the true relationships. The robustness of a molecular phylogeny is also often judged by whether or not it is congruent with a 'well-established' phylogeny based on morphological data. Obviously, since the molecules are just a part of the whole organism they are expected to have experienced the same evolutionary history, and therefore report the same evolutionary information, as other parts (e.g. morphology) of the species. Generally, a molecular phylogeny is viewed as suspicious if it is in conflict with the traditional phylogeny (reviews: Patterson, 1987; Swofford, 1991).

Several methods testing the confidence in the phylogenetic estimate have been developed; the most commonly used one is the 'bootstrap' (Felsenstein, 1985). Several other methods that evaluate the statistical confidence of molecular phylogenies are available (review: Li and Gouy, 1991).

5.1.4 Biology of the mitochondrial genome

Because of the widespread popularity of mitochondrial DNA (mtDNA) for population and phylogenetic work, I will first review some of the basic biology of mtDNA. A more detailed review on the evolution of mtDNA of fish is available (Meyer, 1993). Fish are probably the least well-studied group of vertebrates in terms of their mtDNA, and to date only a handful of papers with actual mtDNA sequences from fish has been published.

The mitochondrial genome of vertebrates is a single, small, double-stranded, circular DNA molecule contained in mitochondria, and up to several thousand copies of the mitochondrial genomes are found per cell. Typically, the size of animal mitochondrial genomes is about 16 500– 500 bp (reviews: by Brown, 1981, 1983, 1985; Clark-Walker, 1985; Moritz, Dowling and Brown, 1987).

The complete mitochondrial genome has been sequenced or the gene order determined in several invertebrates and vertebrates. The invertebrates include two sea urchins (Jacobs *et al.*, 1988; Cantatore *et al.*, 1989), sea stars (not complete) (Smith *et al.*, 1989, 1990; Jacobs *et al.*, 1989; Himeno *et al.*, 1987), the fruit fly, *Drosophila yakuba* (Clary and Wolstenholme, 1985), *Leishmania tarentolae* (not complete) (de la Cruz, Neckelmann and Simpson, 1984), *Artemia salina* (not complete) (Batuecas *et al.*, 1988), *Daphnia* (not complete) (Stanton, Crease and Herbert, 1991), the blue mussel, *Mytilus edulis* (not complete) (Hoffmann, Boore and Brown, 1992), the nematodes, *Ascaris suum* and *Caenorhabditis elegans* (Okimoto *et al.*, 1992), the platyhelminth Fasciola (not complete) (Garey and Wolstenholme, 1989) and the ciliate, *Paramecium* (Pritchard *et al.*, 1990). Among the vertebrates, complete sequences for human (Anderson *et al.*, 1981), cow (Anderson *et al.*, 1982), mouse (Bibb *et al.*, 1981), rat (Gadaleta *et al.*, 1989;), fin whale (Arnason, Gullberg and Widegren, 1991), harbour seal (Arnason and Johnnson, 1992), chicken (Desjardins and Morais, 1990) and frog *Xenopus* (Roe *et al.*, 1985) have been published. Additionally, many other partial mitochondrial sequences, too numerous to be listed here, are known.

Two complete mitochondrial genomes for two cyprinid fish have been determined (Tzeng *et al.*, 1992) (*Crossostoma lacustre*: GenBank accession no. M91245; the complete sequence for the carp is also available in GenBank accession no. X61010). The next best known fish mtDNA genome is of the Atlantic cod (*Gadus morhua*) (Johansen, Guddal and Johansen, 1990; Johansen, personal communication) (Figure 5.1.3).These studies and restriction and partial sequence studies on carp (Araya *et al.*, 1984) and salmonids (Thomas, Withler and Beckenbach, 1986; Davidson, Birt and Green,1989) and unpublished data from chondrichthyes and agnathans from Wes Brown's laboratory (cf. Moritz, Dowling and Brown, 1987) indicate that the piscine mtDNA gene order complies with the general vertebrate condition. The mitochondrial gene order of animals is different in every phylum that has been studied. Even within phyla (e.g. echinoderms) differences in gene order exist (Smith *et al.*, 1989, 1990; Himeno *et al.*, 1987). Mitochondrial gene orders differ slightly among vertebrates (Yoneyama, 1987; Desjardins and Morais, 1990, 1991; Pääbo *et al.*, 1991; reviewed in Meyer, 1993; Meyer *et al.*, unpublished). The piscine mitochondrial gene order does not differ from vertebrate 'consensus' gene order (Figure 5.1.3)

The mitochondrial genome of vertebrates contains 13 genes coding for proteins, two genes coding for ribosomal RNAs (small 12S and large 16S rRNA), 22 genes coding for transfer RNAs (tRNAs) and one major noncoding region (control region) that contains the initiation sites for mtDNA replication and RNA transcription (Figure 5.1.3). The protein coding genes are seven subunits of NADH dehydrogenase (ND1, 2, 3, 4, 4L, 5, 6), cytochrome *b*, three subunits of cytochrome *c* oxidase (CO I, II, III) and two subunits of ATP synthetase (ATPase 6 and 8).

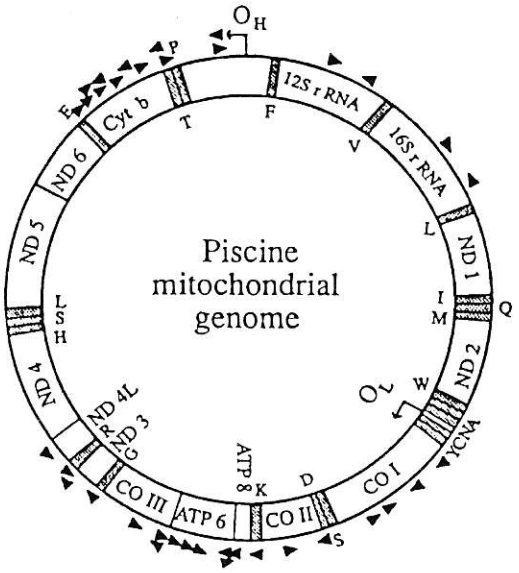

Figure 5.1.3 Piscine mitochondrial gene order. The origins of H- and L-strand replication are indicated in the figure. The origin of the H-strand is in the control region, and the origin of the L-strand replication is in the YCNAW tRNA gene cluster. Transfer RNA genes are shown in shaded boxes. The coding sequences (templates) of all proteins (except ND 6) and the majority of the tRNA genes are on the H-strand. The complete names for the abbreviated names of proteins are given in the text. The tRNA genes coded for by the L-strand are labelled on the outside of the circle, and the tRNA genes coded for by the H-strand are labelled on the inside. A complete mitochondrial fish genome has not been determined, but the arrangement shown here, which is the consensus vertebrate gene order, has been confirmed by sequencing across all gene junctions for the cod mitochondrial genome by Johansen, Guddal and Johansen (1990). The arrows indicate the direction and approximate relative positions of 'universal' PCR primers, the sequence of which is listed in Table 5.1.1.

The two strands of the mitochondrial genome are designated light 'L' and heavy 'H'. These names reflect marked differences in their G + T content in vertebrates resulting in their different behaviour in caesium chloride density gradients. With few exceptions, all genes in the vertebrate mitochondrial genome are encoded by the H-strand; only ND 6 and eight tRNAs are encoded by the L-strand (Figure 5.1.3).

The largest mitochondrial genome (up to 39.3 kb) found so far in animals is from scallops, which also show large intraspecific size variation (Snyder *et al.*, 1987; Gjetvaj Cook and Zouros, 1992). Often, large intraspecific variation is due to tandem duplications involving the control region (see below). In fish, intraspecific size differences can be as large as differences between species (e.g. Bentzen, Leggett and Brown, 1988; Bermingham, Lamb and Avise, 1986; Buroker *et al.*, 1990; reviewed in Brown, 1983, 1985; Moritz, Dowling and Brown, 1987).

Animal mtDNA is haploid and non-recombining (Hayashi, Tagashira and Yoshida, 1985; but Horak, Coon and Dawid, 1974; Olivo *et al.*, 1983; refs in Hurst, 1991) and appears to be almost exclusively maternally inherited. Paternal mitochondria appear to be actively degraded during fertilization (Vaughn, DeBonte and Wilson, 1980) or 'outreplicated' shortly thereafter (Meland *et al.*, 1991). Recent results (Koehler *et al.*, 1991) tend to support the 'bottleneck hypothesis', that maternal inheritance of mtDNA is mediated by the differential amplification of a small number of specific germ-line mtDNA molecules from the

Table 5.1.1 Primers for the polymerase chain reaction

Gene and primer name	*Sequence*	*Source*
Control region		
L15926	TCAAAGCTTACACAGTCTTGTAAACC	Kocher *et al.* (1989)
H16498	CCTGAAGTAGGAACCAGATG	Meyer *et al.* (1990)
L16518	CATCTGGTTCTTTCTTCAGGGCCAT	Meyer (unpublished)
12S rRNA[a]		
H1109	GTGGGGTATCTAATCCCAGTT	Meyer (unpublished)
L1091	AAAAAGCTTCAAACTGGGATTAGATACCCCACTAT	Kocher *et al.* (1989)
H1478	TGACTGCAGAGGGTGACGGGGCGGTGTGT	Kocher *et al.* (1989)
16S rRNA[a]		
L2510	CGCCTGTTTATCAAAAACAT	Palumbi *et al.* (1991), Wilson lab.
H3080	CCGGTCTGAACTCAGATCACGT	Palumbi *et al.* (1991), Wilson lab.
CO I[a]		
L5950	ACAATCACAAAGAYATYGG	Normark, McCune and Harrison (1991)
L6586	OCTGCAGGAGGAGGAGAYCC	Palumbi *et al.* Colf, (1991)
H7086	CCTGAGAATARKGGGAATCAGTG	Palumbi, Cole, (1991)
H7196	AGAAAATGTTGWGGGAARAA	Normark, McCune and Harrison (1991)
CO II		
L7450	AAAGGAAGGAATCGAACCCCC	Normark, McCune and Harrison (1991)
H8055	GCTCATGAGTGGAGGACGTCTT	Normark, McCune and Harrison (1991)
ATPase 8		
L8331	TAAGCRNYAGCCTTTTAAG	Meyer (unpublished)
ATPase 6		
H8517	GGGRACTTTGACTGGTACT	Meyer (unpublished)
L8531	CCCCYTGAAACTGACCATG	Meyer (unpublished)
L8580	AGCCCCACATACCTAGGTATCCC	Meyer (unpublished)
H8674	AARATTTGTTGHGTRAARCGRTT	Meyer (unpublished)
H8907	GGGGTTCCTTCAGGCAATAAA[illegible]	Meyer (unpublished)
H8969	GGGGNCGRATRAANAGRCT	Meyer (unpublished)
H9210	GTAKGCGTGTGCTTGGTGTGCCAT	Meyer (unpublished)

Table 5.1.1 (*contd*)

Gene and primer name	*Sequence*	*Source*
ATPase 6		
H8517	GGGRACTTTGACTGGTACT	Meyer (unpublished)
L8531	CCCCYTGAAACTGACCATG	Meyer (unpublished)
L8580	AGCCCCACATACCTAGGTATCCC	Meyer (unpublished)
H8674	AARATTTGTTGHGTRAARCGRTT	Meyer (unpublished)
H8907	GGGGTTCCTTCAGGCAATAAATG	Meyer (unpublished)
H8969	GGGGNCGRATRAANAGRCT	Meyer (unpublished)
H9210	GTAKGCGTGTGCTTGGTGTGCCAT	Meyer (unpublished)
CO III		
L9225	CACCAAGCACACGCATACCACAT	Meyer (unpublished)
H9407	AAAGTTCCTGTGGTGTGCGGGGG	Meyer (unpublished)
ND3		
L10028	AGTAYANGTRRCTTCCAA	Meyer (unpublished)
H10430	TTGAGCCGAAATCAA	Meyer (unpublished)
ND4L		
L10420	AAYAARAYCNTTGATTTCGRCTCA	Meyer (unpublished)
H10720	TCYGTGCCRTGYGTYCGNGC	Meyer (unpublished)
Cytochrome *b*		
L14725	CGAAGCTTGATATGAAAAACCATCGTTG	Pääbo (1990)
L14841	AAAAAGCTTCCATCCAACATCTCAGCATGATGAAA	Kocher *et al.* (1989)
L15020	GCYAAYGGCGCATCCTTYTTYTT	Meyer (unpublished)
H15149	AAACTGCAGCCCCTCAGAATGATATTTGTCCTCA	Kocher *et al.* (1989)
L15162	GCAAGCTTCTACCATGAGGACAAATATC	Pääbo, *et al.*, 1991; Irwin *et al.* (1991)
L15225	TCCCATACATTGGGACAGACCT	Wilson lab
L15424	ATCCCATTCCACCCATACTACTC	Edwards, Arctander and Wilson (1991)
H15573	AATAGGAAGTATCATTCGGGTTTGATG	Wilson lab, Taberlet, Meyer and Bouvet (1992)
L15774	GTACATGAATTGGAGGACAACCAGT	Wilson lab, Irwin, Kocher and Wilson (1991)
H15915	AACTGCAGTCATCTCCGGTTTACAAGAC	Wilson lab, Irwin, Kocher and Wilson (1991)

Nuclear ribosomal genes:		
18S		
18A	TGGTTGATCCTGCCAGTAG	Wilson lab (unpublished)
18E	TACCATCGAAAGTTGATAGGGCAGA	Hamby *et al.* (1988)
18J	TCTAAGGGCATCACAGACCTGTTATTG	Hamby *et al.* (1988)
18L	CACCTACGGAAACCTTGTTACGACTT	Hamby *et al.* (1988)
		See also several other papers, e.g. Stock and Whitt (1992)
28S		
		See several papers, e.g. Hillis and Dixon (1991), Hillis *et al.* (1992)
ITS 1		
	GTTTCCGTAGCTGAACCTGC	Meyer (unpublished)
	AATTCTCGCAGCTAG	Meyer (unpublished)
IT 2		
	CTCTTAGCGGTGGAT	Meyer (unpublished)
	GTCTGATCTGAGGTC	Meyer (unpublished)

Only PCR primers are listed that have been successfully used with fish DNA. Primer names of mitochondrial PCR primers follow the Kocher *et al.* (1989) convention of naming (i.e. numbering) the primer by the most 3 position of the primer in the human mtDNA sequence (Anderson *et al.*, 1981). L and H refer to the light and heavy strand respectively. If sequences have been published previously the reference is indicated; otherwise, primer sequences have not been published, and were designed and tested in my laboratory and should be referenced as such. Primer sequences are given 5′ to 3′. Y = C or T, R = A or G, K = G or T, H = not G, W = A or T. Several more primer sequences have been published in the *Simple Fool's Guide to PCR*, distributed by S.R. Palumbi's laboratory, University of Hawaii (Palumbi *et al.*, 1991). Primers for nuclear ribosomal PCR primers have been published in several sources, e.g. Hamby *et al.* (1988), Hillis and Dixon and (1991), Stock and Whitt (1992), and are likely to work in fish.

[a]Genes for which more PCR primer sequences have been published.

mitochondrial DNA genotype of the previous generation (Hauswirth and Laipis, 1982; Ashley, Laipis and Hauswirth, 1989; Koehler *et al.*, 1991).

Usually, only one type of mtDNA is found in an organism. However, reports of heteroplasmy, the presence of more than one type of mtDNA in an individual, are accumulating rapidly (reviews: Moritz, Dowling and Brown, 1987; Moritz, 1991). Heteroplasmy has occasionally been found in most major groups of organisms, including several species of fish e.g. shad (Bentzen, Leggett and Brown, 1988), sturgeon (Buroker *et al.*, 1990; Brown, Beckenbach and Smith,1992), cod (Arnason and Rand, 1992) and bowfin (Bermingham, Lamb and Avise, 1986). In the reported cases of heteroplasmy in fish, the proportion of heteroplasmic individuals in the population was 5% in bowfins (Bermingham, Lamb and Avise, 1986), 12% in shad (Bentzen, Leggett and Brown, 1988), and up to 41% in sturgeon (Buroker *et al.*, 1990).

Simulation and modelling of the 'population genetic' behaviour of male mitochondrial genomes suggest that male mitochondria should only rarely be fixed in the population of oocyte mitochondria (Chapman *et al.*, 1982; Birky, Maruyama and Fuerst, 1983; Neigel and Avise, 1986). However, paternal leakage may occur at the low rate of 1 in a 1000 molecules. The universality and rate of parental leakage is still unknown (Gyllensten *et al.*, 1991). Horizontal transfer of mtDNA between species of animals has been observed in rare case – mice (Ferris *et al.*, 1983), voles (Tegelstrom, 1987) – but appears to be a rare phenomenon.

The mitochondrial genetic code is more degenerate and thus less constrained than the 'universal' eukaryotic nuclear genetic code (review: Attardi, 1985). The piscine mitochondrial code appears to be like the vertebrate mitochondrial code in that TGA and ATA align with tryptophan, and methionine respectively (Kocher *et al.*, 1989; Johansen, Guddal and Johansen, 1990; Meyer and Wilson, 1990; Meyer *et al.*, 1990; Tzeng *et al.*, 1992; Meyer, 1993).

In contrast to the nuclear genome, the mitochondrial genome of animals (but not that of fungi and plants) is highly efficient (reviews: Attardi, 1985; Cantatore and Saccone, 1987; Gray, 1989), i.e. it rarely contains duplicate or noncoding sequences. Mitochondrial protein coding genes do not contain introns, and genes are usually separated by less than 10 bp. Two genes may abut directly without intergenic spacers, and in some cases these genes even overlap by several bases. Transfer RNA genes and protein coding genes may overlap considerably (Clary and Wolstenholme, 1985). Some genes that are coded for by different strands overlap as well. The reading frames of the ATPase 6 and the ATPase 8 genes overlap never by less than 6–7 bp in vertebrates. The two reading frames of the ATPase 6 and 8 genes differ by one nucleotide. A single transcript has been found for each of these reading frame pairs (Anderson *et al.*, 1981; Ojala, Montoya and Attardi, 1981); the overlapping reading frames in each pair of genes appear to be translated from a single mRNA.

In fish, several relatively large noncoding regions have been described. Johansen *et al.*, (1990) found a 74-bp insertion in the Atlantic cod (*Gadus morhua*) between the Thr-tRNA and the Pro-tRNA. In *Xenopus* this stretch of noncoding DNA is only 26 bp long (Roe *et al.*, 1985) and in sturgeon only 3 bp long (Gilbert *et al.*, 1988). In several dozen species of fish from several families there are no or less than 5 bp between the Thr-tRNA and Pro-tRNA (Meyer *et al.*, 1990; Sturmbauer and Meyer, 1992; Meyer, unpublished data). In the 15-spine stickleback (*Spinachia spinachia*) we (Meyer, Orti and Bell, unpublished) found a 109-bp insertion between the Glu-tRNA and the cytochrome *b* gene.

5.1.5 Rate of evolution

Several studies of mtDNA using restriction enzymes and actual sequences (Brown *et al.*, 1982) indicate that mtDNA generally evolves at elevated rates (5–10 times faster) compared

to single copy nuclear genes (Brown, George and Wilson, 1979; Perler *et al.*, 1980). The faster mtDNA evolution is due to a higher frequency of point and length mutations (Brown *et al.*, 1982; Cann, Brown and Wilson, 1984; Wilson *et al.*, 1985). The divergence at silent sites of protein coding genes may be about 10% per 10^6 years (generalized from results from primates), ten times the rate found in nuclear protein coding genes. The overall rate of substitution for the complete mitochondrial genome of primates may be 0.5–1% per 10^6 years. The rate of silent substitutions (substitutions that do not result in amino acid changes), mainly transitions, is about 4–6 times that of replacement substitutions (Brown, George and Wilson, 1979; Brown *et al.*, 1982). However, among closely related species, if transitions are ignored and only transversions counted, the difference in rates between mitochondrial and nuclear genes would be much less pronounced. Several reasons for the higher rate of mtDNA evolution have been suggested; for references see Meyer (1993). The observed differences in rates of evolution between nuclear and mtDNA cannot be explained exclusively by the increased variation at third positions. Substitution rates at first and second positions and rRNA and tRNA genes are also increased. However, substitution rates at these sites may not be as different between mitochondrial and nuclear genes as rates at third positions.

(a) Tempo and mode of substitution

Of the three kinds of sequence changes, substitutions are more frequent than indels (additions or deletions), and rearrangements are the least common form of DNA change. Additions and deletions are most frequently observed in the control region and intergenic spacers. At a lower frequency, they occur in the tRNAs and rRNAs. Rarely have they been found in protein coding genes, but they do exist. A single-codon deletion was found among salmonids in the ATPase 6 gene (Thomas and Beckenbach, 1989). A three-codon addition and a one-codon addition in the ATPase 6 gene of percoid fish compared to salmonid fish (Thomas and Beckenbach, 1989) was also discovered (Meyer and Titus, unpublished data). As far as is known, slowly evolving protein coding genes like CO I, II, III and cytochrome *b* do not seem to contain any length variation among neopterygian fish (Normark, McCune and Harrison, 1991; Meyer, unpublished data), and no length variation has been detected among ray-finned or lobe-finned fish for cytochrome *b* and CO III (Meyer and Wilson, 1990; Meyer, unpublished data).

As sequencing techniques developed it became clear that the higher rate of mtDNA evolution is due mainly to transitional differences (changes from one purine to another of one pyrimidine to another) (Brown and Simpson, 1982; Brown *et al.*, 1982; Aquadro and Greenberg, 1983; Greenberg, Newbold and Sugino, 1983; Wolstenholme and Clary, 1985). Transitions often outnumber transversions (changes from a purine to a pyrimidine or vice versa by a factor of 10 or 20 in within-species comparisons. The pronounced predominance of transitions over transversions was found in all positions of codons and all mitochondrial genes. This supports the idea that a mutational bias for transitions over transversions rather than selection and constraints may be largely responsible for the tempo and mode of evolution of the mitochondrial genome. This transition bias appears to decrease with increasing sequence divergence and therefore time since common ancestry.

Substitutions at third positions quickly accumulate and they become saturated with transitions. Mutations at first and second positions will continue to accumulate despite saturated third positions. For phylogenetic reconstruction purposes (see below) transitions at third positions will therefore not be reliable indicators of evolutionary relationships beyond about 30–40% sequence divergence at these positions (Figure 5.1.4).

The increase in observed number of transversions (e.g. Figure 5.1.5) is due to several factors (Brown *et al.*, 1982; Holmquist, 1983; DeSalle *et al.*, 1987; Jukes, 1987; Meyer,

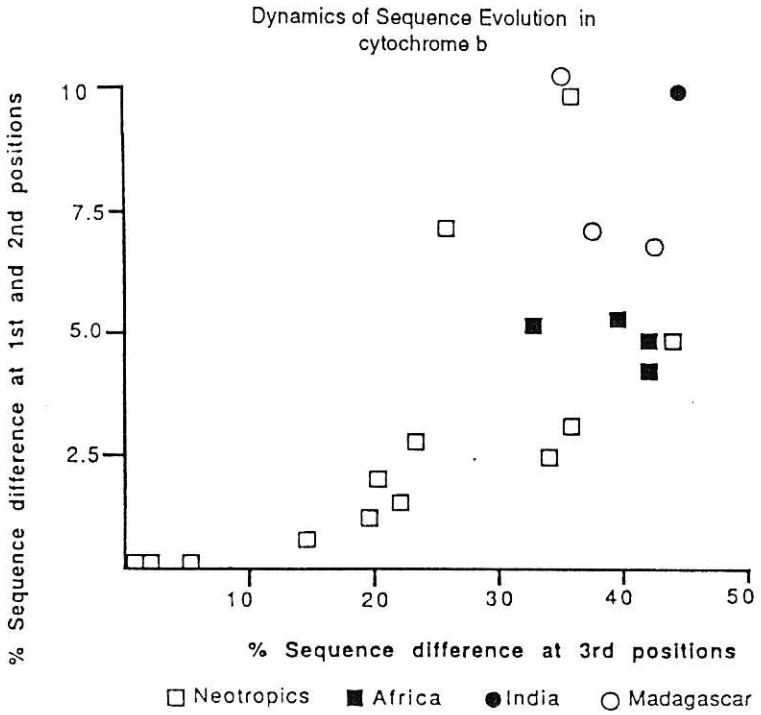

Figure 5.1.4 Mode of sequence evolution for the 5′ end (about 360 bp) of the cytochrome *b* gene for cichlid fish (Meyer, unpublished data). Cichlid fish are a family of fish found in South America, Africa and Madagascar and India. All species were compared to the neotropical species *Cichlasoma citrinellum*. Third position substitutions quickly accumulate and become saturated with substitutions at a maximum level of divergence of around 45%. Hence, for phylogenetic analyses, transitions in the third positions will be reliable indicators of evolutionary relationships for the whole family, although they are informative among closely related groups of species. Mutations at first and second positions continue to accumulate up to a level of about 10% within the family Cichlidae.

1993). Asymmetry in the persistence of transversions and transitions exists (DeSalle *et al.*, 1987) and will produce an observed accumulation of scored transversions with time. Although transitions remain a more common type of mutation event between distantly related species, the percentage of all differences which are scored as transitions will decrease. The observed number of transversions will increase because transversions will become fixed at sites that had already fixed transitions and also at new sites that had not been substituted before.

Transversions, however, accumulate more slowly (Brown and Simpson, 1982; Brown *et al.*, 1982; Aquadro and Greenberg, 1983; Greenberg, Newbold and Sugino, 1983; Wolstenholme and Clary, 1985), eventually outnumbering transitions, i.e. the percentage transversions of the total number of substitutions observed between two species increases with increasing sequence divergence (Figure 5.1.5). The mode of change from transitions predominating to transversions becoming more prevalent differs between genes and may be related to functional constraints and base compositional biases (Holmquist, 1976, 1983; see below).

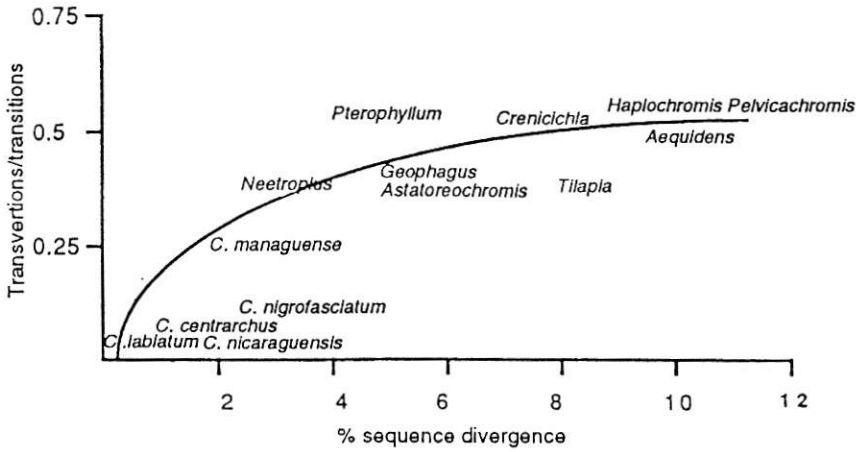

Figure 5.1.5 Dynamics of sequence evolution among cichlid fish for a conservative (about 300 bp) portion of the 12S rRNA gene (Meyer, unpublished data); for methods see Kocher *et al.* (1989) and Meyer *et al.* (1990). Graphed is the mode of sequence divergence and dynamics of the transversion transition ratio from the neotropical species *Cichlasoma citrinellum*. Among closely related species of the genus *Cichlasoma*, transversions were only rarely observed, but the percentage of observed transversions quickly increases to an asymptotic level of about 50%.

These dynamics will have to be accounted for in phylogenetic analysis. Transversions will trace phylogenetic events more reliably because back-mutations will accumulate at a much lower rate. Often, they will be weighted higher in a parsimony analysis or in the calculation of the matrix of genetic divergences for a distance phylogenetic method (e.g. Mindell Honeycutt, 1990; Meyer and Dolven, 1992; Sturmbauer and Meyer, 1992).

Transitions also occur more frequently among closely related species of fish, and transversions become apparent among more distantly related species (Kocher *et al.*, 1989; Meyer *et al.*, 1990; Sturmbauer and Meyer, 1992; Fajen and Breden, 1992). Among conspecific cichlid fish, transitions often outnumber transversions 10 to 1; among distantly related species this ratio changes to about one. Figure 5.1.5 shows this trend quite clearly in a conserved portion of the 12S ribosomal RNA gene for some cichlid fish. With increasing sequence divergence from a neotropical species, *Cichlasoma citrinellum*, the proportion of tranversion substitutions increases rapidly to an upper limit of about 50%. The level at which saturation is reached will depend on several factors, e.g. the base compositional bias (see below). The absolute time required to reach the plateau is dependent on the absolute rate with which transversions become fixed (see below). Among congeners, total sequence divergence tends not to exceed about 5%, and scored transversions are usually less than 25% of the total number of substitutions, whereas closely related species exhibit transition differences almost exclusively. Within the family Cichlidae the sequence divergence (uncollected for multiple substitutions) in this portion of the 12S rRNA gene does not exceed about 15%. This is not the case for species of fish in the family Gasterosteidae (Orti, Bell and Meyer, unpublished data). Patterns like the one seen for the 12S rRNA gene in cichlid fish are typical (Figure 5.1.5). The observed ratio of transversions to the total number of substitutions between species increases relatively rapidly (beyond the level of the species?); it usually quickly reaches a plateau (among congeners?) with some fluctuation (Holmquist, 1983). Even quite distantly related species of fish do not exceed about 20% of total (uncorrected) sequence divergence for this conservative part of the 12S rRNA gene (Meyer and Wilson, 1990; Mindell and Honeycutt, 1990; Meyer and Dolven, 1992).

(b) Protein coding genes

Substitution patterns in protein coding genes follow some relatively well-understood rules. This regularity in the way in which mutations accumulate makes protein coding genes attractive candidates for phylogenetic studies of fish. Mutations in third (and rarely in first) positions of codons that do not result in amino acid (AA) substitutions (silent or synonymous substitutions) accumulate much more rapidly than AA replacement substitutions (non-synonymous substitutions). The most frequently observed substitutions are transitions in third positions of codons; second most frequent are transversions in third positions and silent transitions in some first codon positions (Figure 5.1.6). In some codons (leucine), the first position is degenerate (both TTA/G and CTN codon families code for leucine). Furthermore, since transitions in first positions of codons will usually result in conservative AA substitutions will tend to maintain a functional gene product, mutations will accumulate more quickly in first than in the most constrained second positions of codons (Figure 5.1.6). Knowledge of these rules allows one to weight (or exclude, as an extreme form of weighting) kinds of substitutions (transitions and transversions) and positions (first, second and third) differently based on the phylogenetic question that is being addressed. Obviously, among distantly related species transitions in third positions are going to be unreliable tracers of evolutionary descent and represent largely 'phylogenetic noise'. Transversions are rarer and have a higher chance of being reliable indicators of descent. The most conserved positions (second positions in codons) are going to be the most similar among closely related species but contain phylogenetic information among distantly related species.

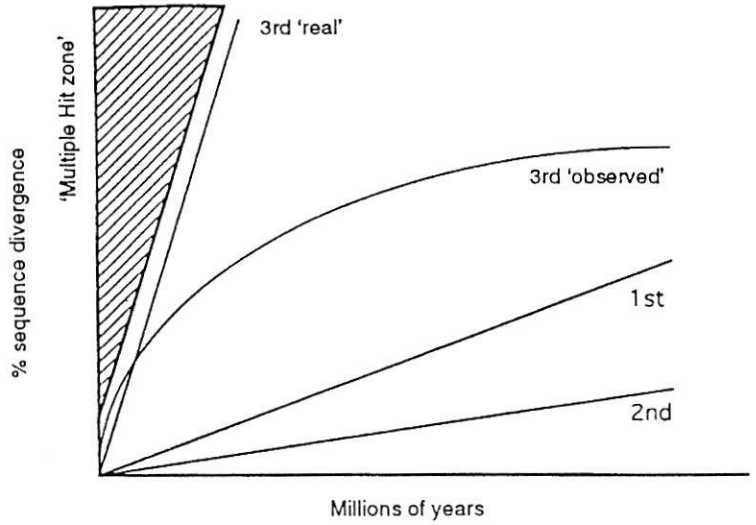

Figure 5.1.6 Hypothetical example of the dynamics of sequence evolution in a mitochondrial protein coding gene. The multiple hit zone is depicted to signify the observation that the likelihood of a multiple substitution at a particular nucleotide increases with sequence divergence. The probability of a multiple hit will not increase linearly but rather exponentially with increasing sequence divergence. Due to multiple substitutions the 'observed' sequence divergence at third positions (also, of course, at first and second positions, but more slowly) will 'decelerate' quickly compared to second and first positions in codons. The 'observed' rate of sequence divergence will increase in a relatively linear fashion for about the first 10–15% sequence divergence. The 'real' rate of evolution can be estimated from a tangent that is laid on the observed mode of divergence. The actual slopes will vary between particular mitochondrial protein coding genes.

Each mitochondrial protein coding gene has its own particular rate of evolution that will depend on factors such as functional constraints on the gene product and base compositional biases (Johansen *et al.*, 1990; Meyer, 1993, Table 1). Despite the fast rate of mtDNA evolution, some genes may be highly conserved; there may be a low ceiling for total divergence, which is partly due to nucleotide base compositional biases (e.g. DeSalle *et al.*, 1987) and strong functional constraints. The genes coding for the subunits of cytochrome oxidase and cytochrome *b* are the most conserved genes, and the most variable ones are some of the ND and the ATPase genes. These slowly evolving protein coding genes have been used to test relationships among distantly related neopterygian fish (Normark, McCune and Harrison, 1991), yet these genes turned out to have a rate of AA substitution that is too slow this phylogenetic question. The early publication and availability of 'universal' PCR primers, particularly for 12S and cytochrome *b* (Kocher *et al.*, 1989), had the effect that cytochrome *b* is being sequenced in many different organisms for many different questions. Often, however, particularly for distantly related species, other genes might have been better choices (see below).

(c) Transfer RNA genes

All vertebrate mitochondrial genomes contain the usual set of 22 tRNAs. They are smaller than their cytoplasmic counterparts, usually 59–75 bp in length. Secondary structures of mitochondrial tRNAs are less conserved but still fold into the cloverleaf secondary structure (review: Sprinzl *et al.*, 1991). Mitochondrial tRNAs are surprisingly variable in their primary sequence (substitutions and, less frequently, small 1–2 bp insertions and deletions are observed) and secondary structure. Lengths of stems and loops of mitochondrial tRNAs are exceedingly variable compared with nuclear tRNA genes. Mitochondrial tRNA genes evolve at least 100 times faster than nuclear tRNA genes (Brown *et al.*, 1982). Despite the observed elevated rates of variation compared to the nuclear tRNAs, the tRNA genes are among the more slowly evolving genes of the mitochondrial genome, but variation exists in their relative rates. Because of their small size they may not be the best choice for genes for phylogenetic inquiries (although they have been used successfully); however, they are often good sites for the design of PCR primers (e.g. Kocher *et al.*, 1989), because of the conserved features of tRNA genes.

(d) Ribosomal RNA genes

Two ribosomal RNA (rRNA) genes are found in animal mitochondrial genomes: the small 12S (About 819–975 bp in vertebrates) rRNA gene and the large 16S (about 1 571–1 640 bp in vertebrates) rRNA gene. Nuclear rRNA (18S and 28S) and tRNA genes (see below) evolve about 100 times more slowly than their mitochondrial counterparts (Dawid, 1972) when the more variable expansion segments of the nuclear rRNA genes are excluded from the analysis (Mindell and Honeycutt, 1990).

The same basic rules of substitutions (see above) apply to rRNA genes (Brown *et al.*, 1982). Transitions are more frequent than transversions; this is most apparent among closely related species where the record of transitions has not been overlaid by more slowly accumulating and more persistent transversions (Figure 5.1.5, see above). Insertions and deletions are usually small, in the range of 1–5 bp. The 16S gene tends to contain more length variation than the 12S gene. Hence, DNA sequences of the 16S rRNA for distantly related species are more difficult to align than for the 12S rRNA gene. Length mutations are more frequent in rRNA than in protein coding genes. These length variations make alignment an issue for phylogenetic studies; different alignments, and therefore different

hypotheses of homology, can result in drastically different phylogenetic inferences. Areas of questionably aligned sequence (criteria for what is well aligned are not always obvious; see Swofford and Olsen (1990) and Mindell (1991)) are best removed from the phylogenetic analysis. Alignment and delineation of which portion of the sequence should be excluded from the analysis and which should be used remain problematic issues.

Secondary structure models are available for both genes, and general agreement about the secondary structure exists (Glotz *et al.*, 1981; Dunon-Bluteau and Brun, 1986; Hixon and Brown, 1986). These secondary structures are conserved across large evolutionary distances. Among tetrapods, lungfish, the coelacanth and ray-finned fish, substitutions occur about four times more frequently in proposed loops than in stems; transversions are about nine times more frequent in loops than in stems (Meyer and Wilson, 1990; Meyer and Dolven, 1992; Meyer, unpublished data). Transversions appear to be indicators of phylogenetic relationships in a slowly evolving portion of the 12S rRNA gene for an evolutionary distance of more than 400 000 000 years (Meyer and Wilson, 1990; Mindell and Honeycutt, 1990; Meyer and Dolven,1992; Meyer, 1993). Among closely related species of primates, stems are less likely to contain substitutions than loop regions (Hixson and Brown, 1986). Stems sometimes show 'compensatory mutations': a substitution in one strand of a stem region is compensated for by a change in the other strand of the stem, in order to maintain intrastrand base pairing and a stable secondary structure. The rate of sequence divergence appears higher among closely related species than among distantly related species due to saturation effects (Figure 5.1.5). At around 100–150 $\times$ 10^6 years of sequence divergence, the rate of substitution appears to decrease (Mindell and Honeycutt, 1990; Meyer and Dolven, 1992).

The overall rRNA and tRNA substitution rates are about half those of the protein coding genes (Brown *et al.*, 1982; Hixson and Brown, 1986; Jacobs *et al.*, 1988; Meyer and Wilson, 1990; Hillis and Dixon, 1991; Mindell and Honeycutt, 1990), making them attractive genes for phylogenetic questions among distantly related species of fish.

(e) Control region

The control region is partially constrained in primary sequence or secondary structure to regulate replication and transcription (Clayton, 1991). The control region is characterized by the displacement loop (D-loop), a stretch of DNA that is complementary to the light L-strand; the D-loop strand displaced the H-strand. The initiation site of D-loop synthesis and the origin of H-strand replication are identical, but, the termination site(s) for the D-loop strand and the H-strand are separate (see above).

Of all mitochondrial genes the control region has the highest substitution rate. The rate of evolution of control region is two to five times higher than that of mitochondrial protein coding genes (Aquadro and Greenberg, 1983). Nucleotide substitutions occur five times more frequently than additions and deletions in the human control region; among closely related species of cichlid fish, substitutions outnumber deletions and additions (Meyer *et al.*, 1990; Sturmbauer and Meyer, 1992). The control region varies tremendously in length, often because of tandem duplications of 200–4 100 bp (Brown, 1985), and is primarily responsible for the observed variation in the total length of the vertebrate mitochondrial genomes. It also contains the highest frequency of length mutations at the population level (Densmore, Wright and Brown, 1985; Harrison, Rand and Wheeler, 1985). In fish (Bermingham, Lamb and Avise, 1986; Bentzen, Leggett and Brown, 1988; Buroker *et al.*, 1990), tandem duplications and various numbers of repeats are usually found here when heteroplasmy is observed.

The characteristically high amount of sequence divergence in the control region is not evenly distributed; there are several regions that exhibit high levels of sequence conservation. Usually A-T-rich regions are found at both ends of the D-loop, with an

evolutionarily conserved central domain in the middle. Although highly variable in sequence and structurally variable in vertebrates, the central domain is probably functionally similar, since conserved sequence blocks (CSBs) can usually be identified in all animal control regions (Brown *et al.*, 1986; Saccone, Attimonelli and Sbiza, 1987).

Because mutations accumulate fastest here, the control region is the molecule of choice for the study of population level phenomena and the study of phylogenetic relationships among closely related species and has been used for both of these purposes in fish (e.g. Meyer *et al.*, 1990; Sturmbauer and Meyer, 1992; Reinthal and Meyer, unpublished).

(f) Slower rate of evolution in ectotherms?

Much variation was expected in terms of amino acid differences in mitochondrial protein coding genes of fish, since up to 8.5% sequence divergence had been found within species of sunfish (Avise *et al.*, 1984a) and also some cichlid fish have high levels of intraspecific variation (Sturmbauer and Meyer, 1992). However, this is not always the case. There is little variation within or between species of salmonids (Thomas and Beckenbach, 1989) and cichlids (Kocher *et al.*, 1989; Meyer *et al.*, 1990; Meyer, Biermann and Pallson, unpublished data) in the more conserved protein coding genes.

Ectothermic animals may be up to five times slower than endothermic animals in the evolution of cytochrome *b* amino acid replacement substitutions (Kocher *et al.*, 1989), and perhaps even 10 times slower (Martin, Naylor and Palumbi, 1992). This may also be true for nucleotide substitutions. Given that synonymous changes in protein coding genes appear to accumulate so rapidly, it was surprising to find that amino acid substitutions in the most conservative mitochondrial genes (cytochrome oxidase and cytochrome *b* genes) accumulate rather slowly in fish (Kocher *et al.*, 1989; Meyer and Wilson, 1990; Normark, McCune and Harrison, 1991). This slower rate allowed mtDNA to be used to investigate phylogenetic relationships among very distantly related species (Meyer and Wilson, 1990; Normark, McCune and Harrison, 1991; Meyer and Dolven 1992). The rate of amino acid substitution in the cytochrome *b* gene proved to be conservative enough to test relationships among groups of lobe-finned fish and ray-finned fish that have not shared a common ancestor for more than 400 000 000 years (Meyer and Wilson, 1990). The rate of amino acid substitution in CO and cytochrome b genes actually was too slow to clearly elucidate the evolutionary relationships among neopterygian fish (Normark, McCune and Harrison, 1991).

Although mtDNA evolution in ectotherms appears to be slower compared to birds and mammals (Kocher *et al.*, 1989; Martin, Naylor and Palumbi, 1992), their nuclear rate is still slower than the mtDNA rate. Comparisons of mtDNA (cytochrome *b*) and nuclear genes (*X-mrk*, *X-src*) between closely related species of poeciliid fish of the genus *Xiphophorus* showed that mtDNA evolves at least 2.5 times faster than these nuclear genes (Meyer, Robertson, Lydeard and Schartl, unpublished data). This estimate for the rate of these nuclear genes includes noncoding regions (introns) and is therefore underestimating the relative mitochondrial rate. Based on earlier findings in primates (e.g. Brown, George and Wilson, 1979) it was expected that the mitochondrial rate generally should be at least five times the nuclear rate.

Whether mitochondrial and nuclear substitution rates are more similar in fish because of slower mtDNA molecular substitution rates or faster nuclear substitution rates is not clear and will require much more data than are currently available. This is important for our understanding of the absolute rates of sequence divergence of both genomes, as well as comparisons between the two. If the mitochondrial rate is generally slower in fish and/or the rate of nuclear DNA evolution increased, we will require a knowledge of divergence dates.

However, too few studies have calibration points from the fossil record to allow us to test these hypotheses.

5.1.6 Mitochondrial DNA and population-level questions

Until recently, mtDNA has been used mainly in population-level work and in studies of molecular relationships among closely related species (reviews: Wilson *et al.*, 1985; Avise, 1986; Avise *et al.*, 1987; Moritz, Dowling and Brown, 1987). The fast rate of evolution of mtDNA (Brown, George and Wilson, 1979) compared to nuclear DNA makes mtDNA useful for high-resolution analyses of recent evolutionary events. This fast rate of mtDNA evolution, coupled with maternal inheritance (but see Zouros *et al.*, (1992) and above), has made mtDNA an extremely popular genetic system with which to study gene flow, hybrid zones, population structure and other population-level questions (reviews: Wilson *et al.*, 1985; Avise, 1986; Moritz, Dowling and Brown, 1987; Harrison, 1991). Mitochondrial DNA lends itself to the study of founder events and female-mediated gene flow, i.e. differences in dispersal between sexes will be apparent through comparisons of the geographic distribution of nuclear DNA and mtDNA (review: Harrison, 1991). Mitochondrial DNA also lends itself to the study of the origin of clonally reproducing species (e.g. Densmore, Wright and Brown, 1985; Quattro, Avise and Vrijenhoek, 1992; reviewed in Moritz, Dowling and Brown, 1987. Mitochondrial DNA, as well as allozyme markers, has been used to identify hybridization events (e.g. Avise and Saunders, 1984; Avise and Van Den Avyle, 1984; Herke *et al.*, 1990). The special mode of inheritance of mtDNA has consequences for its population genetic behaviour (Birky, Maruyama and Fuerst, 1983; Avise *et al.*, 1984b; Avise, Ball and Arnold, 1988; Neigel and Avise, 1986; Pamilo and Nei, 1988; see reviews in Avise, 1986, 1989; Avise *et al.*, 1987; Harrison, 1991).

Until recently most of the work on population structure and phylogenetic relationships of fish based on mtDNA data was based on the analysis of restriction fragment length polymorphisms (RFLPs) (e.g. Graves, Ferris and Dizon, 1984; Wilson *et al.*, 1985; Wilson, Thomas and Beckenbach, 1987; Bermingham, Lamb and Avise, 1986; Gyllensten and Wilson, 1987; Avise *et al.*, 1986; Ovenden, White and Sanger, 1988; Wirgin, Proenca and Crossfield, 1989; reviews in Avise, 1986; Avise *et al.*, 1987; Moritz, Dowling and Brown, 1987). High levels of polymorphism and geographic differentiation within species are sometimes found (see above). Extremely low levels of sequence variation within and between species have been found as well (e.g. Meyer *et al.*, 1990). Restriction analyses are usually unable to provide information about the patterns of substitutions that cause observed differences in RFLPs. Actual DNA sequences do provide this information. The estimates of sequence divergence based on mtDNA RFLPs are not always the same as DNA sequence analyses (e.g. Wilson *et al.*, 1985; Thomas and Beckenbach, 1989). In the case of intra-specific variation in salmonids, the divergence estimate from actual DNA sequences is about half that estimated from RFLP analyses. The opposite trend was observed for interspecific variation, in which sequence data tended to be larger than estimates based on RFLP data (Thomas and Beckenbach, 1989; Beckenbach, Thomas and Sohrabi, 1990).

Differences in the estimates of sequence divergence based on restriction enzymes and DNA sequences can be due to several factors. For example, the selection of restriction enzymes might have been biased to cut preferentially at particular nucleotides. If equal representation of all four nucleotides was expected, but strong base compositional bias existed, the estimates of sequence divergence based on restriction analysis would be inaccurate. Furthermore, restriction enzymes cut indiscriminately throughout the whole mtDNA molecule (including the rapidly evolving control region, as well as more slowly evolving rRNA and tRNA genes), whereas most often only portions of mitochondrial

genome are sequenced, and estimates of sequence difference are influenced by the choice of gene(s) sequenced. Obviously, this has to be taken into consideration. In intraspecific comparisons the rapidly evolving control region might contribute disproportionately to the estimate of sequence divergence based on RFLP analysis of the complete mtDNA genome; in interspecific studies, multiple hits (see below) in the control region might lead to an underestimate of sequence divergence compared to comparisons of actual DNA sequences.

Methods in population genetics and fisheries management change rapidly and mtDNA has found wide application (reviews: Hallermann and Beckmann, 1988; Ovenden, 1990; Ryman and Utter, 1987; Dizon *et al.*, 1992;). RFLP analysis had been the predominant method in population studies (e.g. Wirgin, Proenca and Crossfield, 1989) but PCR has experienced increased application for these questions. Even conservative protein coding genes like those for cytochrome *b* tend to show intraspecific variation, mainly in third positions of codons that can be used to identify fish stocks (Bartlett and Davidson, 1991, 1992; Carr and Marshall, 1991; McVeigh, Bartlett and Davidson, 1991; Finnerty and Block, 1992; Orti, Bell and Meyer, unpublished data) or be used for phylogenetic studies among closely related species (Kocher *et al.*, 1989; Meyer *et al.*, 1990; Meyer, Kocher and Wilson, 1991; Sturmbauer and Meyer, 1992; Reinthal and Meyer, unpublished data).

When using mtDNA to study population differentiation and systematics of closely related species, i.e. recent divergence events, one needs to be aware of some potential difficulties. If the ancestral species contains more than one mtDNA haplotype, lineage sorting at speciation and subsequent random lineage extinctions may cause two species to contain some of the same mtDNA haplotypes, which may not exactly follow species boundaries (e.g. Avise, 1989; Avise *et al.*, 1984b; Avise, Ankney and Nelson, 1990). This is due to random sampling of mtDNA haplotypes at speciation and may introduce errors in phylogenetic reconstruction (see below). The danger of this happening is particularly large if the speciation event is recent and the ancestral species highly polymorphic. If the measured divergence between the mtDNA haplotypes of two young species is used to estimate the time since the origin of these species, this date may be overestimated, since the divergence of haplotypes might have predated the splitting of the species, i.e. the amount of sequence divergence was not zero at the time of speciation and needs to be corrected. This correction can be based on the currently observed level of intraspecific variation (cf. Harrison, 1991). However, current levels of intraspecific variation may not be an accurate reflection of the prespeciation levels, and furthermore assumes that the rates of mtDNA evolution are the same in all lineages.

5.1.7 Mitochondrial DNA and phylogenetic questions

The study of fish mtDNA experienced a boost through the development of so-called 'universal' primers (see above) (Kocher *et al.*, 1989). These primers were designed, based on comparisons of published mtDNA sequences, to anneal to stretches of DNA that are conserved across a wide taxonomic range. Despite the generally fast substitution rate of mtDNA, conserved areas can be identified and primers designed that have applicability for a wide taxonomic range (e.g. a phylum or even beyond) (Kocher *et al.*, 1989). With these 'universal' primers mtDNA from most organisms can be amplified and the DNA sequence determined without prior sequence knowledge from the particular organism studied. The first 'universal' primers (Kocher *et al.*, 1989) and in particular the primers that amplify portions of the mitochondrial 12S ribosomal and cytochrome *b* genes currently enjoy widespread application in many evolutionary and phylogenetic studies from a wide range of taxonomic groups (reviews: Meyer, 1993; Esposti, Degli *et al.*, 1993).

Universal PCR primers for mtDNA (Kocher *et al.*, 1989) led to an even further increase in the use of mtDNA for phylogenetic questions. However, the availability of universal PCR primers has led to their uncritical use for many groups of organisms for many different questions. Cytochrome *b* is a gene that has been targetted for many different inquiries, but it may not be the best gene for the study of evolutionary splits that are ancient enough so that third positions are saturated. The 5′ end of cytochrome *b* tends to be very conserved in amino acid substitutions and it may not contain enough variation in first and second positions of codons once third positions are saturated. The 3′ end of this gene appears to be somewhat less constrained in terms of amino acid substitutions (Irwin, Kocher and Wilson, 1991; reviewed in Esposti Degli, *et al.*, 1992, 1993).

Base compositional biases, saturation effects and the low variation in first and second positions of codons all contribute to problems in phylogenetic reconstruction. Similar difficulties in the use of cytochrome *b* were encountered in the reconstruction of the evolutionary relationships among groups of cichlid fish, believed to have diverged more than 80 000 000 years ago (Meyer, unpublished data). Although this gene has been very useful for studies involving populations (see above) and the phylogenetic study of very different groups of fish, it may not be the most appropriate gene for all purposes (review: Esposti Degli *et al.*, 1993). It would appear that for some phylogenetic problems faster evolving protein coding genes like ND2 and ATPase 6 and 8 or the ribosomal genes might provide more appropriate data. Primer sequences for many other genes have already been published or are available (Table 5.1.1, Figure 5.1.3).

More 'universal' mitochondrial, but also nuclear, DNA PCR primers are constantly being published. Table 5.1.1 represents a compilation of several published (e.g. Kocher *et al.*, 1989; Meyer *et al.*, 1990; Normark, McCune and Harrison, 1991; Palumbi *et al.*, 1991; Grachev *et al.*, 1992) and unpublished PCR primers for several mitochondrial and nuclear genes that are known to work for most species of fish. With these primers, mtDNA can be amplified and sequenced without prior sequence knowledge and they are expected to work for most groups of teleosts.

Mitochondrial DNA has found widespread use as a tool for phylogenetic analyses (reviews: Kornfield, 1984, 1991; Avise, 1986; Avise *et al.*, 1987; Moritz, Dowling and Brown, 1987; Wilson *et al.*, 1985). RFLP analyses were mainly conducted on closely related species since the homology among fragments was not always clear between distantly related species (more than about 10–15% sequence divergence). Through the advent of PCR it has become apparent that mtDNA sequences may also be a useful tool for the study of evolutionary relationships among more distantly related species (e.g. Meyer and Wilson, 1990; Normark, McCune and Harrison, 1991; Meyer and Dolven, 1992).

Multiple mutations at the same nucleotide position (multiple hits) become increasingly more likely and therefore accumulate with increasing sequence divergence (Figure 5.1.6). 'Multiple hits' are problematic phylogenetic studies. They tend to obscure evolutionary relationships, and they also result in an observed upper limit of sequence divergence, a ceiling that is approached asymptotically (Figure 5.1.4 and 5.1.5). Estimates of sequence divergence (important for the calculation of distance matrices for distance methods, see above) between two species, particularly in the 'multiple hit zone', can severely underestimate 'real' divergence and need to be corrected. The more distantly related two species are, the more the simple estimate of sequence divergence (e.g. uncorrected percentage sequence divergence) will differ from an estimate that is derived from some correction that attempts to account for multiple substitutions.

Distance methods are based on a distance matrix, e.g. neighbour-joining (see above, Saitou and Nei, 1987). These distance matrices need to be corrected for multiple hits. Multiple hit corrections can be done in a number of ways; Jukes and Cantor's method (Jukes

and Cantor, 1969) is the simplest but tends to underestimate the evolutionary divergence. More elaborate models are available and provide more reliable estimates of 'real' evolutionary divergence (Figure 5.1.6; Kimura, 1980, 1981; Takahata and Kimura, 1981; Gojobori, Ishii and Nei, 1982; Tajima and Nei, 1984). Multiple hits are also problematic for parsimony analysis and their detrimental effect on phylogenetic analyses needs to be counteracted by downweighting of transitions.

Since the dynamics and rules of substitution are somewhat clearer in protein coding genes than in tRNA and rRNA genes (e.g. the secondary structure and alignments of rRNA are not always clear), they are often a more straightforward choice for a gene used for the study of evolutionary relationships. However, the applicability and usefulness of mitochondrial as well as nuclear rRNA genes is not called into question (Meyer and Dolven, 1992; Meyer and Wilson, 1990; reviewed in Mindell and Honeycutt, 1990; Hillis and Dixon, 1991).

Other mitochondrial or nuclear genes may have substitution rates more appropriate for particular questions. On a per base pair basis the control region will provide more information about population-level questions and contain more (up to three times more) phylogenetic information than cytochrome *b* (e.g. Meyer *et al.*, 1990; Sturmbauer and Meyer, 1992). In a comparison between two closely related species of neotropical cichlid fish (*Cichlasoma citrinellum* and *Neetroplus nematodus*) the variation in a portion of the control region was comparable (18.6%) to the variation in third positions for cytochrome *b* (22.9%) and cytochrome oxidase subunit III (14.3%) (Meyer, unpublished data). The variation in first and second positions for these conservative genes was low (0.7% and 2.8% respectively). The Pro-tRNA was found to vary considerably (8.9%) and a conservative portion of the 12S rRNA gene contained few substitutions (2.4%) (Meyer, unpublished data).

The usefulness of mtDNA sequences is determined at one extreme by the stochasticity of the distribution of haplotype polymorphisms within species and at the other extreme by saturation effects due to base compositional biases, mutational bias and selective constraints on the gene product. Another complication is the potential for rate heterogeneity in the molecular clock even among closely related species, which might complicate phylogenetic analyses. Various phylogenetic methods are more or less sensitive to the effects of rate heterogeneity and the complicating effects they have on the reconstruction of evolutionary relationships (Nei, 1991).

5.1.8 Nuclear genes in molecular systematics of fish

(a) Ribosomal genes

The large existing database and the ease of direct sequencing techniques of nuclear ribosomal genes has led to their widespread, successful use in molecular systematics (reviews: Hamby *et al.*, 1988; Mindell and Honeycutt, 1990; Hillis and Dixon, 1991). Both small (18S) and large (28S) ribosomal genes (Figure 5.1.7) have been used for phylogenetic work. Because of the extremely slow rate of evolution of these genes they have mainly been

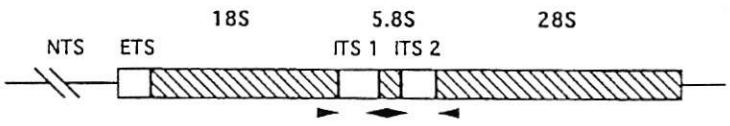

Figure 5.1.7 Arrangements of the elements of the nuclear ribosomal genes. Primer sequences of two primer pairs (Table 5.1.1) that amplify the ITS 1 and ITS 2 regions or both regions plus the 5.8S ribosomal RNA gene.

used for phylogenetic work on distantly related species, e.g. the origin of tetrapods (Hillis and Dixon, 1989; Hillis, Dixon and Ammer, 1989; Stock *et al.*, 1991; Meyer, unpublished data) and the monophyly of agnathan fish (Stock and Whitt, 1992; Meyer, unpublished data). Many PCR primers are available (refs in Table 5.1.1) that will ensure the continued use of these genes for phylogenetic work. Some particular problems, e.g. alignment (see above), plague the use of these genes for phylogenetic work.

(b) Growth hormone and prolactin

Growth hormone (GH), prolactin (PRL) and chorionic somatomammoptropin (CS) are members of a gene family of polypeptide hormones, believed to have evolved from a single ancestral gene. The widespread use of GH in transgenic fish work has led to the accumulation of several (about 19 so far) GH sequences from a variety of groups of fish (e.g. Schneider *et al.*, 1992; Watanabe *et al.*, 1992). The potential for phylogenetic information in GH is largely unexplored.

(c) Protein kinases

A large body of literature exists on several protein (particularly tyrosine) kinase genes in fish. This is largely due to the general interest in these genes as proto-oncogenes, and fish as model systems in cancer research. Several of these genes (e.g. *X-src*) provide useful phylogenetic information at the genus and family level. PCR-based projects are ongoing (Meyer *et al.*, unpublished data) that highlight the universal applicability of nuclear protein coding genes for many phylogenetic questions in fish. These sequences can be used for phylogenetic questions among closely related species (introns contain enough variation) and distantly related species, for which the phylogenetic information contained in exons is utilized.

Acknowledgements

I thank G. Orti and S. Johansen for providing unpublished data and C. Davis for technical assistance and G. Orti for comments on the manuscript. Support from the National Science Foundation for my laboratory (BSR-9107838, BSR-9119867, BSR-9112367, INT-9117104) is gratefully acknowledged.

References

Anderson, S., Bankier, A.T., Barrell, B.G., *et al.*, (1981) Sequence and organization of the human mitochondrial genome. *Nature*, **290**, 457–65.

Anderson, S., de Bruijn, M.H.L., Coulson, A.R., Epron, I.C., Sanger, F. and Young, I.G., (1982) Complete sequence of bovine mitochondrial DNA. *J. Mol. Biol.*, **156**, 683–717.

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

Araya, A., Amthauer, R., Leon, G. *et al.* (1984) Cloning, physical mapping and genome organisation of mitochondrial DNA from *Cyprinus carpio* oocytes. *Mol. Gen. Genet.*, **196**, 43–52.

Arnason, E. and Rand, D.M. (1992) Heteroplasmy of short tandem repeats in mitochondrial DNA of Atlantic cod, *Gadus morhua*. *Genetics*, **132**, 211–20.

Arnason, U., Gullberg, A. and Widegren, B. (1991) The complete nucleotide sequence of the mitochondrial DNA of the fin whale, *Balaenoptera physalus*. *J. Mol. Evol.*, **33**, 556–68.

Arnason, U. and Johnsson, E. (1992) The complete mitochondrial DNA sequence of the harbor seal *Phaca vitulina*. *J. Mol. Evol.*, **34**, 493–505.

Arnheim, N., White, T. and Rainey, W.E. (1990) Application of PCR: organismal and population biology. *BioScience*, **40**, 174–82.

Ashley, M.V., Laipis, P.J. and Hauswirth, W.W. (1989) Rapid segregation of heteroplasmic bovine mitochondria. *Nucleic Acid. Res.*, **17**, 7325–31.

Atchley, W. and Fitch, W.M. (1991) Gene trees and the origins of inbred strains of mice. *Science*, **254**, 554–58.

Attardi, G. (1985) Animal mitochondrial DNA: an extreme example of genetic economy. *Int. Rev. Cyt.*, **93**, 93–145.

Avise, J.C. (1986) Mitochondrial DNA and the evolutionary genetics of higher animals. *Proc. R. Soc. Lond. B.*, **312**, 325–42.

Avise, J.C. (1989) Gene trees and organismal histories: a phylogenetic approach to population biology. *Evolution*, **43**, 1192–208.

Avise, J.C., Ankney, C.D. and Nelson, W.S. (1990) Mitochondrial gene trees and the evolutionary relationship of mallard and black ducks. *Evolution*, **44**, 1109–19.

Avise, J.C. and Saunders, N.C. (1984) Hybridization and introgression among species of sunfish (*Lepomis*): analysis by mitochondrial DNA and allozyme markers. *Genetics*, **108**, 237–55.

Avise, J.C. and Van den Avyle, M.J. (1984) Genetic analysis of reproduction of hybrid White bass × Striped bass in the Savannah River. *Trans. Am. Fish. Soc.*, **113**, 563–70.

Avise, J.C, Bermingham, E., Kessler, L.G. and Saunders, N.C. (1984a) Characterisation of mitochondrial DNA variability in a hybrid swam between subspecies of bluegill sunfish (*Lepomis macrochir*). *Evolution*, **38**, 931–41.

Avise, J.C., Helfman, G.S., Avise, J.C., Neigel, J.E. and Arnold, J. (1984b) Demographic influences on mitochondrial DNA lineage survivorship in animal populations. *J. Mol. Evol.*, **20**, 99–105.

Avise, J.C., Helfman, G.S., Saunders, N.C. *et al.*, (1986) Mitochondrial DNA differentiation in North Atlantic eels: population genetic consequences of an unusual life history pattern. *Proc. Natl. Acad. Sci. USA*, **83**, 4350–54.

Avise, J.C., Arnold J., Ball R.M. *et al.* (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Ann. Rev. Ecol. Syst.*, **18**, 489–522.

Avise, J.C., Ball, R.M. and Arnold, J. (1988). Current versus historical population sizes in vertebrate species with high gene flow: a comparison based on mitochondrial DNA lineages and inbreeding theory for neutral mutations. *Mol. Biol. Evol.*, **5**, 331–44.

Bartlett, S.E. and Davidson, W.S. (1991) Identification of *Thunnus* tuna species by the polymerase chain reaction and direct sequence analysis of their mitochondrial cytochrome *b* genes. *Can. J. Aquat. Sci.*, **48**, 309–17.

Bartlett, S.E. and Davidson, W.S. (1992) FINS (forensically informative nucleotide sequencing): a procedure for identifying the animal origin of biological specimens. *Biotechniques*, **12**, 408–11.

Batuecas, B., Garesse, R., Calleja, M. *et al.* (1988) Genome organization of *Artemia* mitochondrial DNA. *Nucleic Acids, Res.*, **16**, 6515–29.

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

Bentzen, P., Leggett, W.C. and Brown, G.G. (1988) Length and restriction site heteroplasmy in the mitochondrial DNA of American shad (*Alosa sapidissima*). *Genetics*, **118**, 509–18.

Bermingham, E. and Avise, J.C. (1986) Molecular zoogeography of freshwater fish in the southeastern United States. *Genetics*, **113**, 939–65.

Bermingham, E., Lamb, T. and Avise, J.C. (1986) Size polymorphism and heteroplasmy in the mitochondrial DNA of lower vertebrates. *J. Hered.*, **77**, 249–52.

Bibb, M.J., Van Etten, R.A., Wright, C.T. *et al.* (1981) Sequence and gene organization of mouse mitochondrial DNA. *Cell*, **26**, 167–180.

Birky, C.W. Jr, Maruyama, T. and Fuerst, P.A. (1983) An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts and some results. *Genetics*, **103**, 513–27.

Brown, G.G. and Simpson, M.V. (1982) Novel features of animals' mtDNA evolution as shown by sequences of two rat cytochrome oxidase subunit II genes. *Proc. Natl. Acad. Sci. USA*, **79**, 3246–50.

Brown, G.G., Gadaleta, G., Pepe, G. *et al.* (1986) Structural conservation and variation in the D-loop containing region of vertebrate mitochondrial DNA. *J. Mol. Biol.*, **192**, 503–11.

Brown, J.R., Beckenbach, A.T. and Smith, M.J. (1992) Mitochondrial DNA length variation and heteroplasmy in populations of white sturgeon (*Acipenser transmontanus*). *Genetics*, **132**, 221–28.

Brown, W.M. (1981) Mechanisms of evolution in animal mitochondrial DNA. *Ann. N.Y. Acad. Sci.*, **361**, 119–34.

Brown, W.M. (1983) Evolution of mitochondrial DNA, in *Evolution of Genes and Proteins* (eds M. Nei and R.K Koehn), Sinauer, Sunderland, pp. 62–88.

Brown, W.M. (1985) The mitochondrial genome of animals, in *Molecular Evolutionary Genetics* (ed. R.J. MacIntyre), Plenum Press, New York, pp. 95–130.

Brown, W.M., George, M. Jr and Wilson, A.C. (1979) Rapid evolution of mitochondrial DNA. *Proc. Natl. Acad. Sci. USA*, **76**, 1967–71.

Brown, W.M., Prager, E.M., Wang, A. *et al.* (1982) Mitochondrial DNA sequences of primates: tempo and mode of evolution. *J. Mol. Evol.*, **18**, 225–39.

Buroker, N.E., Brown, J.R., Gilbert, T.A. *et al.* (1990) Length heteroplasmy of sturgeon mitochondrial DNA: an illegitimate elongation model. *Genetics*, **124**, 157–63.

Cann, R.L., Brown, W.M. and Wilson, A.C. (1984) Polymorphic sites and the mechanism of evolution and human mitochondrial DNA. *Genetics*, **106**, 479–99.

Cantatore, P., and Saccone, C. (1987) Organization, structure, and evolution of mammalian mitochondrial genes. *Int. Rev. Cytol.*, **108**, 149–208.

Cantatore, P., Roberti, M., Rainaldi, G. *et al.* (1989) The complete nucleotide sequence, gene organization, and genetic code of the mitochondrial genome of *Paracentrotus lividus*. *J. Biol. Chem.*, **264**, 10965–75.

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

Chapman, R.W., Stephens, J.C, Lansman, R.A. *et al.* (1982) Models of mitochondrial DNA transmission genetics and evolution in higher eukaryotes. *Genet. Res.*, **40**, 41–57.

Clark-Walker, G.D. (1985) Basis of diversity in mitochondrial DNAs, in *The Evolution of Genome Size* (ed. T. Cavalier-Smith), Wiley, New York, pp. 277–97.

Clary, D.O. and Wolstenholme, D.R. (1985) The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide, gene organization and genetic code. *J. Mol. Evol.*, **22**, 252–71.

Clayton, D.A. (1991) Replication and transcription of vertebrate mitochondrial DNA. *Annu. Rev. Cell. Biol.*, **7**, 453–78.

Cruz de la, V.F., Neckelmann, C. and Simpson, L. (1984) Sequence of six genes and several open reading frames in the kinetoplast maxicircle DNA of *Leishmania tarentolae*. *J. Biol. Chem.*, **259**, 15136–47.

Davidson, W.S., Birt, T.P. and Green, J.M. (1989) Organization of the mitochondrial genome from Atlantic salmon (*Salmo salar*). *Genome*, **32**, 340–42.

Dawid, I.B. (1972) Evolution of mitochondrial DNA sequences in *Xenopus*. *Dev. Biol.*, **29**, 139–51.

Densmore, L.D., Wright, J.W. and Brown, W.M. (1985) Length variation and heteroplasmy are frequent in mitochondrial DNA from parthenogenetic and bisexual lizards (genus *Cnemidophorus*). *Genetics*, **110**, 689–707.

DeSalle, R., Freedman, T., Prager, E.M. *et al.* (1987) Tempo and mode of sequence evolution in mitochondrial DNA of Hawaiian *Drosophila*. *J. Mol. Evol.*, **26**, 157–64.

Desjardins, P. and Morais, R. (1990) Sequence and gene organization of the chicken mitochondrial genome. *J. Mol. Biol.*, **212**, 599–634.

Desjardins, P. and Morais, R. (1991) Nucleotide sequence and evolution of coding and noncoding regions of a quail mitochondrial genome. *J. Mol. Evol.*, **32**, 153–61.

Dizon, A.E., Lockyer, C., Perrin, W.F. *et al.* (1992) Rethinking the stock concept: a phylogenetic approach. *Conservation Biol.*, **6**, 24–36.

Dunon-Bluteau, D. and Brun, G. (1986) The secondary structure of the *Xenopus laevis* and human mitochondria small ribosomal subunit RNA are similar. *FEBS lett.*, **198**, 333–38.

Edwards, S.V., Arctander, P. and Wilson, A.C., (1991) Mitochondrial resolution of a deep branch in the genealogical tree for perching birds. *Proc. R. Soc. Lond. B*, **243**, 99–107.

Erlich, H.A. (ed.) (1989) *PCR Technology: Principle and Applications for DNA Amplification*, Stockton Press, New York.

Esposti Degli, M., Ghelli, A., Crimi, M. *et al.* (1992) Cytochrome *b* of fish mitochondria is strongly resistant to funiculosin, a powerful inhibitor of respiration. *Arch. Biochem. Biophys.*, **295**, 198–204.

Esposti Degli, M., De Vries, S., Crimi, M. *et al.* (1993) mitochondrial cytochrome *b*: evolution and structure of the protein. *Acta Biophys. Biochem.*, 243–7.

Fajen, A. and Breden, F. (1992) Mitochondrial DNA sequence variation among natural populations of the Trinidad Guppy, *Poecilia reticulata*. *Evolution*, **46**, 1457–65.

Felsenstein, J. (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.*, **17**, 368–76.

Felsenstein, J. (1982) Numerical methods for inferring evolutionary trees. Q. *Rev. Biol.*, **57**, 379–404.

Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**, 783–91.

Felsenstein, J. (1988) Phylogenies from molecular sequences: inference and reliability. *Annu. Rev. Genet.*, **22**, 521–65.

Ferris, S.D., Sage, R.D., Huang, C.M. *et al.* (1983) Flow of mitochondrial DNA across a species boundary. *Proc. Natl. Acad. Sci. USA.*, **80**, 2290–94.

Finnerty, J.R. and Block, B.A. (1992) Direct sequencing of mitochondrial DNA detects highly divergent haplotypes in blue marlin (*Makaira nigricans*). *Mol. Mar. Biol. Biotech.*, **1**, 206–14.

Gadaleta, G., Pepe, G., DeCandia, G. *et al.* (1989) The complete nucleotide sequence of the *Rattus norvegicus* mitochondrial genome: cryptic signals revealed by comparative analysis between vertebrates. *J. Mol. Evol.*, **28**, 497–516.

Garey, J.R. and Wolstenholme, D.R. (1989) Platyhelminth mitochondrial DNA: evidence for early evolutionary origin of a tRNA (Ser) AGN that contains a dihydrouridine arm replacement loop, and of serine-specifying AGA and AGG codons. *J. Mol. Evol.*, **28**, 374–87.

Gilbert, T.L., Brown, J.R., O'Hara, P.J. *et al.* (1988) Sequence of tRNA-Thr and tRNA-Pro from white sturgeon (*Acipenser transmontanus*) mitochondria. *Nucleic. Acids. Res.*, **16**, 11825.

Gjetvaj, B., Cook, D.I. and Zouros, E. (1992) Repeated sequences and large-scale size variation of mitochondrial DNA: a common feature among scallops (Bivalvia: Pectinidae). *Mol. Biol. Evol.*, **9**, 106–24.

Glotz, C., Zwieb, C., Brimacombe, R. *et al.* (1981) Secondary Structure of the large subunit ribosomal RNA from Echerichia coli, Zea mays chloroplast, and human and mouse mitochondrial ribosomes. *Nucleic. Acids Res.*, **9**, 3287–306.

Gojobori, T., Ishii, K. and Nei, M. (1982) Estimation of average number of nucleotide substitutions when the rate of substitution varies with nucleotides. *J. Mol. Evol.*, **18**, 414–23.

Grachev, M.A., Slobodyanyuk, S. Ja., Kholodilov, N.G. *et al.* (1992) Comparative study of two protein coding regions of mitochondrial DNA from three endemic sculpins (Cottoidei) of Lake Baikal. *J. Mol. Evol.*, **34**, 85–90.

Graves, J.E., Ferris, S.D. and Dizon, A.E. (1984) Close genetic similarity of Atlantic and Pacific skipjack tuna (*Katsuwonis pelamis*) demonstrated with restriction endonuclease analysis of mitochondrial DNA. *Mar. Biol.*, **79**, 315–19.

Gray, M.W. (1989) Origin and evolution of mitochondrial DNA. *Annu. Rev. Cell. Biol.*, **5**, 25–50.

Greenberg, B.D., Newbold, J.E. and Sugino, A. (1983) Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. *Gene*, **21**, 33–49.

Gyllensten, U.B. and Ehrlich, H.A. (1988) Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQa locus. *Proc. Natl. Acad. Sci. USA*, **85**, 7652–55.

Gyllensten, U. and Wilson, A.C. (1987) mitochondrial DNA of salmonids: inter- and intraspecific variability detected with restriction enzymes, in *Population Genetics and Fishery Management* (eds N. Ryman and F. Utte), University of Washington Press, Seattle, pp. 301–17.

Gyllensten, U., Wharton, D., Josefsson, A. *et al.* (1991) Paternal inheritance of mitochondrial DNA in mice. *Nature*, **352**, 255–57.

Hadrys, H., Balick, M. and Schierwater, B. (1992) Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. *Mol. Ecol.*, **1**, 55–63.

Hallerman, E.M and Beckmann, J.S. (1988) DNA-level polymorphism as a tool in fisheries science. *Can. J. Fish. Aquat. Sci.*, **45**, 1075–87.

Hamby, R.K., Sims, L., Issel, L. *et al.*, (1988) Direct ribosomal RNA sequencing: optimization of extraction and sequencing methods for work with higher plants. *Plant Mol. Biol. Reporter*, **6**, 175–92.

Harrison, R.G. (1991) Molecular changes at speciation. *Annu. Rev. Ecol. Syst.*, **22**, 281–308.

Harrison, R.G., Rand, D.M. and Wheeler, W.C. (1985) Mitochondrial DNA size variation within individual crickets. *Science*, **228**, 1446–48.

Hauswirth, W.W. and Laipis, P.J. (1982) Mitochondrial DNA polymorphism in a maternal lineage of Holstein cows. *Proc. Natl. Acad. Sci. USA*, **79**, 4686–90.

Hayashi, J-I., Tagashira, Y. and Yoshida, M.C. (1985) Absence of extensive recombination between inter- and intraspecies mitochondrial DNA in mammalian cells. *Exp. Cell. Res.*, **160**, 387–95.

Herke, S.W., Kornfield, I. Moran, P. *et al.* (1990) Molecular confirmation of hybridization between Northern Pike (*Esox lucius*) and Chain Pickerel (*E. niger*). *Copeia*, **1990**, 846–50.

Hillis, D.M. and Dixon, M.T. (1989) Vertebrate phylogeny: evidence from 28S ribosomal DNA sequences, in *Phylogenetic Analysis of DNA Sequences* (eds M.M. Miyamoto and J. Cracraft), Oxford University Press, New York, pp. 355–67.

Hillis, D.M. and Dixon, M.T. (1991) Ribosomal DNA: molecular evolution and phylogenetic inference. *Q. Rev. Biol.*, **66**, 411–53.

Hillis, D.M. and Dixon, M.T. and Ammerman, L.K. (1989) The relationships of the coelacanth *Latimeria chalumnae*: evidence from sequences of vertebrate 28S ribosomal genes, in *The Biology of Latimeria chalumnae and Evolution of Coelacanths* (eds J.A. Musick, M.N. Bruton and E.K. Balon), Kluwer Academic Publishers, Dordrecht, pp. 117–30.

Hillis, D.M. and Moritz, C. (eds) (1990) *Molecular Systematics* Sinauer Associates Inc., Sunderland, MA.

Hillis, D.M., Bull, J.J., White, M.E., Badgett, M.R. and Molineux, I.J. (1992) Experimental phylogenetics: generation of a known phylogeny. *Science*, **255**, 589–592.

Himeno, H., Masaki, H., Kawai, T. *et al.* (1987) Unusual genetic codes and a novel gene structure for tRNA-AGY-Ser in starfish mitochondrial DNA. *Gene.*, **56**, 219–30.

Hixson, J.E. and Brown, W.M. (1986) A comparison of the small ribosomal RNA genes from the mitochondrial DNA of the great apes and humans: sequence, structure, evolution and phylogenetic implications. *Mol. Biol. Evol.*, **3**, 1–18.

Hoffman, R., Boore, J.L. and Brown, W.M. (1992) A novel mitochondrial genome organization in the blue mussel, *Mytilus edulis. Genetics*, **131**, 397–412.

Holmquist, R. (1976) Solution to a gene divergence problem under arbitrary stable nucleotide transition probabilities. *J. Mol. Evol.*, **8**, 337–49.

Holmquist, R. (1983) Transitions and transversion in evolutionary descent: an approach to understanding. *J. Mol. Evol.*, **19**, 134–44.

Horak, I., Coon, H.G. and Dawid, I.B. (1974) Interspecific combination of mitochondrial DNA molecules in hybrid somatic cells. *Proc. Natl. Acad. Sci. USA.*, **71**, 1828–32.

Hull, D.L. (1988) *Science as a Process*, Chicago University Press, Chicago.

Hurst, L.D. (1991) Sex, slime and selfish genes. *Nature*, **354**, 23–4.

Innis, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J. (1990) *PCR Protocols: A Guide to Methods and Applications*, Academic Press, San Diego.

Irwin, D.M., Kocher, T.D. and Wilson A.C. (1991) Evolution of the cytochrome *b* of mammals. *J. Mol. Evol.*, **32**, 128–44.

Jacobs, H.T., Elliott, D.J., Math, V.B. *et al.* (1988) Nucleotide sequence and gene organization of sea urchin mitochondrial DNA. *J. Mol. Biol.*, **202**, 185–217.

Jacobs, H., Asakawa, S., Araki, T. *et al.* (1989) Conserved tRNA gene cluster in starfish mitochondrial DNA. *Curr. Genet.*, **15**, 193–206.

Johansen, S., Guddal, P.H. and Johansen, T. (1990) Organization of the mitochondrial genome of Atlantic cod, *Gardus morhua. Nucleic Acids, Res.*, **18**, 411–19.

Jukes, T.H (1987) Transitions, transversions,and the molecular evolutionary clock. *J. Mol. Evol.*, *26*, 87–98.

Jukes, T.H and Cantor, C.R. (1969) Evolution of protein molecules, in *Mammalian Protein Metabolism* (ed. H.N. Munro), Academic Press, New York, pp. 21–123.

Kimura, M. (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.*, **16**, 111–20.

Kimura, M. (1981) Estimation of evolutionary distances between homologous nucleotide sequences. *Proc. Natl. Acad. Sci. USA*, **78**, 454–58.

Kocher, T.D., Thomas, W.K., Meyer, A. *et al.* (1989). Dynamics of mitochondrial DNA evolution in animals. *Proc. Natl. Acad. Sci. USA*, **86**, 6196–200.

Koehler, G.M., Lindberg, G.L., Brown, D.R. *et al.* (1991) Replacement of bovine mitochondrial DNA by sequence variant within one generation. *Genetics*, **129**, 247–55.

Kornfield, I. (1984) Descriptive genetics of cichlid fish, in *Evolutionary Genetics of Fishes* (ed. B.J. Turner), Plenum Press, New York, pp. 591–615.

Kornfield, I. (1991) Genetics, in *Cichlid Fish: Behaviour, Ecology and Evolution*, (ed. M.H.A. Keenleyside), Chapman and Hall, New York, pp. 103–28.

Li, W.-H. and Gouy, M. (1991) Statistical methods for testing molecular phylogenies, in *Phylogenetic Analysis of DNA Sequences* (eds M.M. Miyamoto and J. Cracraft), Oxford University Press, Oxford, pp. 249–77.

Martin, A.P., Naylor, G.J.P. and Palumbi, S.R. (1992) Rates of mitochondrial DNA evolution in sharks are slow compared with mammals. *Nature*, **357**, 153–5.

McVeigh, H.P., Bartlett, S.E. and Davidson, W.S. (1991) Polymerase chain reaction/direct sequence analysis of the cytochrome *b* gene in *Salmo salar. Aquaculture*, **95**, 225–31.

Meland, S., Johansen, S., Johansen, T. *et al.* (1991) Rapid disappearance of one parental mitochondrial genotype after isogamous mating in the myxomycete *Physarum polycep-halum. Curr. Genet.*, **19**, 55–60.

Meyer, A. (1993) Evolution of mitochondrial DNA in fish, in *Biochemistry and Molecular Biology of Fish*, Vol. 2 (eds P.W Hochachka and P. Mommsen), Elsevier Press (in press).

Meyer, A. and Dolven, S.I. (1992) Molecules, fossils and the origin of tetrapods. *J. Mol. Evol.*, 35, 102–13.

Meyer, A., Kocher, T.D. and Wilson, A.C. (1991) African fish–a reply. *Nature*, **351**, 467–68.

Meyer, A. and Wilson, A.C. (1990) Origin of tetrapods inferred from their mitochondrial DNA affiliation to lungfish. *J. Mol. Evol.*, **31**, 359–64.

Meyer, A., Kocher, T.D., Basasibwaki, P. *et al.* (1990) Monophyletic origin of Victoria cichlid fish suggested by mitochondrial DNA sequences. *Nature*, **347**, 550–53.

Mindell, D.P. (1991) Aligning DNA sequences: homology and phylogenetic weighting, in *Phylogenetic Analysis of DNA Sequences* (eds M.M. Miyamoto and J. Cracraft), Oxford University Press, New York, pp. 73–89.

Mindell, D.P. and Honeycutt, R.L (1990) Ribosomal RNA in vertebrates: evolution and phylogenetic applications. *Annu. Rev. Ecol. Syst.*, **21**, 541–66.

Moritz, C. (1991) Evolutionary dynamics of mitochondrial DNA duplications in parthenogenetic geckos, *Heteronotia binoei*. *Genetics*, **129**, 221–30.

Moritz, C., Dowling, T.E. and Brown, W.M. (1987) Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annu. Rev. Ecol. Syst.*, **18**, 269–92.

Mullis, K., Faloona, F., Scharf, S. *et al.* (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor Symp. Quant. Biol.*, **51**, 263–73.

Nei, M. (1991) Relative efficiencies of different tree-making methods for molecular data, in *Phylogenetic Analysis of DNA Sequences* (eds M.M. Miyamoto and J. Cracraft), Oxford University Press, New York, pp. 90–128.

Neigel, J.E. and Avise, J.C. (1986) Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation, in *Evolutionary Processes and Theory* (eds E. Nevo and S. Karlin), pp. 515–34.

Normark, B.B., McCune, A.R. and Harrison, R.G. (1991 Phylogenetic relationships of Neopterygian fish, inferred from mitochondrial DNA sequences. *Mol. Biol. Evol.*, **8**, 819–34.

Ojala, D., Montoya, J. and Attardi, G. (1981) tRNA punctuation
model of RNA processing in human mitochondrial DNA. *Nature*, **290**, 470–74.

Okimoto, R., Macfarlane, J.L., Clary, D.O. *et al.* (1992) The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. *Genetics*, **130**, 471–98.

Olivo, P.D., Van de Walle, J.J., Laipis, J.P. *et al.* (1983) Nucleotide sequence evidence for rapid genotypic shifts in bovine mitochondrial DNA D-loop. *Nature*, **306**, 400–2.

Ovenden, J.R. (1990) Mitochondrial DNA and marine stock assessment: a review. *Aust. J. Mar. Freshwater Res.*, **41**, 835–53.

Ovenden, J.R., White, R.W.G. and Sanger, A.C. (1988) Evolutionary relationships of *Gadopsis* gspp. inferred from restriction enzyme analysis of their mitochondrial DNA. *J. Fish. Biol.*, **32**, 137–48.

Pääbo, S. (1990). Amplifying ancient DNA, in *PCR Protocols: A guide to Methods and Applications*, (eds M.A., Innes, D.H. Gelfand, J.J. Sninsky and T.J. White), Academic Press, San Diego, pp. 159–66.

Pääbo, S., Thomas, W.K., Whitfield, K.M. *et al.* (1991) Rearrangements of mitochondrial transfer RNA genes in marsupials. *J. Mol. Evol.*, **33**, 426–30.

Palumbi, S.R., Martin, A., Romano, S. *et al.* (1991). *Simple Fool's Guide to PCR*, Department of Zoology, University of Hawaii, Honolulu.

Pamilo, P. and Nei, M. (1988) Relationships between gene trees and species trees. *Mol. Biol. Evol.*, **5**, 568–83.

Patterson, C. (ed.) (1987) *Molecules and Morphology in Evolution: Conflict or Compromise?* Cambridge University Press, Cambridge.

Perler, F., Efstratiadis, A., Lomedico, P. *et al.* (1980) The evolution of genes: the chicken preproinsulin gene. *Cell*, **20**, 555–6.

Pritchard, A.E., Seilhamer, J.J., Mahalingam, R. *et al.* (1990) Nucleotide sequence of the mitochondrial genome of *Paramecium*. *Nucleic. Acids Res.*, **18**, 173–80.

Quattro, J.M., Avise, J.C. and Vrijenhoek, R.C. (1992) Mode and origin and sources of genotypic diversity in triploid gynogenetic fish clones (Poeciliopsis: Poeciliidae). *Genetics*, **130**, 621–218.

Roe, B.A., Ma, D.-P., Wilson, R.K. *et al.* (1985) The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome. *J. Biol. Chem.*, **260**, 9759–74.

Ryman, N. and Utter, F. (ed.) (1987) *Population Genetics and Fishery Management*, University of Washington Press, Seattle.

Saccone, C., Attimonelli, M. and Sbiza, E. (1987) Structural elements highly preserved during the evolution of the D-loop-containing region in vertebrate mitochondrial DNA. *J. Mol. Evol.*, **26**, 205–11.

Saiki, R.K., Scharf, S., Faloona, F. *et al.* (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*, **230**, 1350–54.

Saiki, R.K., Gelfand, D.H., Stoffel, S. *et al.* (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, **239**, 487–91.

Saitou, N. and Nei, M. (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, **4**, 406–525.

Schneider, J.F., Myster, S.H., Hackett, P.B. *et al.* (1992) Molecular cloning and sequence analysis of the cDNA for another pike (Esox lucinus) growth hormone. *Mol. Mar. Biol. Biotechnol.*, **1**, 106–12.

Smith, M.F., Banfield, D.K., Doteval, K. *et al.* (1989) Gene arrangement in sea star mitochondrial DNA demonstrates a major inversion event during echinoderm evolution. *Gene*, **76**, 181–5.

Smith, M.F., Banfield, D.K., Doteval, K. *et al.* (1990) Nucleotide sequence of nine protein coding genes and 22 tRNAs in the mitochondrial DNA of the sea star *Piaster ochraceus*. *J. Mol. Evol.*, **31**, 195–204.

Synder, M., Fraser, A.R, LaRoche, J. *et al.* (1987) A typical mitochondrial DNA from the deep-sea scallop *Placopecten magellanicus*. *Proc. Natl. Acad. Sci. USA*, **84**, 7595–99.

Sprinzl, M., Dank, N., Nock, S. *et al.* (1991) Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic. Acid Res.*, **19**, 2127–71.

Stanton, D.J., Crease, T.J. and Herbert, P.D.N. (1991) Cloning and characterization of *Daphnia* mitochondrial DNA. *J. Mol. Evol.*, **33**, 152–5.

Stock, D.W. and Whitt, G.S. (1992) Evidence from 18S ribosomal RNA sequences that lampreys and hagfish form a natural group. *Science*, **257**, 787–9.

Stock, D.W., Moberg, K.D., Maxson, L.R. *et al.* (1991). A phylogenetic analysis of the 18S ribosomal RNA sequence of the coelacanth *Latimeria chalumnae*, in *the Biology of Latimeria chalumnae and Evolution of Coelacanths*. (eds J.A. Musick, M.N. Bruton and E.K. Balon), Kluwer Academic Publishers, Dordrecht, pp. 99–117.

Sturmbauer, C. and Meyer, A. (1992) Genetic divergence, speciation and morphological stasis in a lineage of African cichlid fish. *Nature*, **358**, 578–81.

Swofford, D.L. (1991) When are phylogeny estimates from molecular and morphological data incongruent? in *Phylogenetic Analysis of DNA Sequences* (eds M.M. Miyamoto and J. Craeraft), Oxford University Press, New York, pp. 295–333.

Swofford, D.L. and Olsen, G.J. (1990) Phylogeny reconstruction, in *Molecular Systematics* (eds D.M. Hillis and C. Moritz), Sinauer, Sunderland, pp. 411–501.

Taberlet, P , Meyer, A. and Bouvet, J. (1992) Unusually large mitochondrial variation in populations of the blue tit, *Parus caeruleus*. *Mol. Ecol.*, **1**, 27–36.

Tajima, F. and Nei, M. (1984) Estimation of evolutionary distance between nucleotide sequences. *Mol. Biol. Evol.*, **1**, 269–85.

Takahata, N. and Kimura, M. (1981) A model of evolutionary base substitutions and its application with special reference to rapid change of pseudogenes. *Genetics*, **98**, 641–57.

Tegelstrom, J.T. (1987) Transfer of mitochondrial DNA from northern red-back vole (*Clethrionomys rutilus*) to the bank vole (*C. glareolus*). *J. Mol. Evol.*, **24**, 218–27.

Templeton, A.R. (1983a) Convergent evolution and non-parametric inference from restriction fragment and DNA sequence data, in *Statistical Analysis of DNA Sequence Data* (ed. B. Weir) Marcel Dekker, New York, pp. 151–79.

Templeton, A.R. (1983b) Phylogenetic inference from restriction endonuclease site maps with particular reference to the humans and apes. *Evolution*, **37**, 221–44.

Thomas, W.K. and Beckenbach, A.T. (1989) Variation in salmonid mitochondrial DNA: evolution constraints and mechanisms of substitution. *J. Mol. Evol.*, **29**, 233–45.

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

Tzeng, C.S., Hui, C.F., Shen, S.C. and Huang, P.C. (1992) The complete nucleotide sequence of the Crossostoma lucustre mitochondrial genome: conservative and variations among vertebrates. *Nucleic Acids Res.*, **20**, 4853–8.

Vaughn, K.C., DeBonte, L.R. and Wilson, K.G. (1980) Organelle alteration as a mechanism for maternal inheritance. *Science*, **208**, 196–7.

Watanabe, K., Igarashi, A., Noso, T. *et al.* (1992) Chemical identification of catfish growth hormone and prolactin. *Mol. Mar. Biol. Biotechnol.*, **1**, 239–49.

Welsh, J. and McClelland, M. (1990) Polymorphisms generated by arbitrarily primed PCR in the mouse: application with arbitrary primers. *Nucleic Acids Res.*, **18**, 7123–18.

White, T.J., Arnheim, N. and Erlich, H.A. (1989) The polymerase-chain reaction. TIG, **5**, 185–9.

Williams, J.G.K., Kubelik, A.R., Livak, K.J. *et al.* (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.*, **18**, 6531–35.

Wilson, A.C., Cann, R.L., Carr, S.M. *et al.* (1985) Mitochondrial DNA and two perspectives on evolutionary genetics. *Biol. J. Linn. Soc.*, **26**, 375–400.

Wilson, A.C., Zimmer, E.A., Prager, E.M. *et al.* (1989). Restriction mapping in the molecular systematics of mammals: a retrospective salute, in The Hierarchy of Life (eds B. Fernholm, K. Bremer and H. Jornvall), Elsevier, Amsterdam, pp. 407–19.

Wilson, G.M., Thomas, W.K. and Beckenbach, A.T. (1985) Intra- and interspecific mitochondrial DNA sequence divergence in *Salmo*: rainbow, steelhead, and cutthroat trouts. *Can. J. Zool.*, **63**, 2088–94.

Wilson, G.M., Thomas, W.K. and Beckenbach, A.T. (1987) Mitochondrial DNA analysis of Pacific Northwest populations of *Oncorhynchus tshawytscha*. *Can. J. Fish. Aquat. Sci.*, **44**, 1301–5.

Wirgin, I.I., Proenca, R. and Grossfield, J. (1989) Mitochondrial DNA diversity among populations of striped bass in the southeastern United States. *Can. J. Zool.*, **67**, 891–907.

Wolstenholme, D.R. and Clary, D.O. (1985) Sequence evolution of Drosophila mitochondrial DNA. *Genetics*, **109**, 725–44.

Wrishnik, L.A., Higuchi, R.G. and Stoneking, M. (1987) Length mutations in human mitochondrial DNA: direct sequencing of enzymatically amplified DNA. *Nucleic Acids Res.*, **15**, 529–42.

Zouros, E., Freeman, K.R., Oberhauser Ball, A. *et al.* (1992) Direct evidence for extensive paternal mitochondrial DNA inheritance in the marine mussel *Mytilus*. *Nature*, **359**, 412–14.

Yoneyama, Y. (1987) The nucleotide sequences of the heavy and light strand replication origins of the *Rana catesbeiana* mitochondrial genome. *Nippon Ika Daioaku Zasshi*, **54**, 429–40 (in Japanese).