Hochachka and Mommsen (eds.), Biochemistry and molecular biology of fishes, vol. 2

CHAPTER 1

Evolution of mitochondrial DNA in fishes

AXEL MEYER

Department of Ecology and Evolution, State University of New York, Stony Brook, NY 11794-5245, U.S.A.

I. Introduction

Mitochondria originated from free-living purple bacteria that became part of the eukaryotic organism through an 'endosymbiotic' event[142,231]. Whether mitochondria are mono- or polyphyletic however, is not clear[98] (references in ref. 97). The evolution of mitochondrial DNA (mtDNA) has been reviewed[9,40,41,42,57,95–97,154,187,219]. Several reviews that deal with the molecular biology and biogenesis of mitochondria have also been published[10,51,56,61,215]. This review attempts to summarize, in the limited space provided, the current knowledge of mtDNA evolution with particular emphasis on piscine systems. Fishes are probably among the least well-studied groups of vertebrates in terms of mtDNA however, and to date, only a handful of papers with actual mtDNA sequences from fishes have been published. Therefore, I will use unpublished data mainly from my laboratory in an attempt to identify and

highlight patterns and possible underlying processes of the evolution of mtDNA of fishes.

The complete mitochondrial genome has been sequenced or the gene order determined in several invertebrates and vertebrates. The invertebrates include two sea urchins[50,127], sea stars (not complete)[116,128,192,193], the fruit fly, *Drosophila yakuba*[58], *Leishmania tarentolae* (not complete)[63], *Artemia salina* (not complete)[25], *Daphnia* (not complete)[199], the blue mussel, *Mytilus edulis* (not complete)[120], the nematodes, *Ascaris suum* and *Caenorhabditis elegans*[164], the platyhelminth, *Fasciola* (not complete)[87] and the ciliate, *Paramecium*[175]. Among the vertebrates, complete sequences for human[1], cow[2], mouse[34], rat[88], fin whale[6], chicken and quail (not complete)[68,69] and frog, *Xenopus laevis*[180] have been published. Additionally, many other partial mitochondrial sequences, too numerous to be listed here, are known. A complete mitochondrial genome for a fish has not been determined, but a research program to sequence the complete mitochondrial genome of the Atlantic cod (*Gadus morhua*) is in progress[129] (Johansen, pers. comm.).

II. The mitochondrial genome

1. Size

The mitochondrial genome of animals is a single, small, double-stranded, circular (it is linear in *Hydra*[221], and in *Paramecium*[175]) DNA molecule contained in multiple copies in mitochondria. Depending on the cell type, up to several thousand mitochondrial genomes are found per cell. Typically, the size of animal mt genomes is around 16,500 ± 500 basepairs (bp) (reviewed in refs. 40–42, 57, 154). In animals, the size distribution of the mt genome does not follow a particular phylogenetic plan. However, vertebrates generally show a smaller size range. As yet, complete mt genomes have only been sequenced in a handful of animals, and surprises may lie ahead as more species are studied. The largest mt genome (up to 39.3 kb) found so far in animals is in scallops which also show large intraspecific size variation[90,194]. Here, as found elsewhere, large intraspecific variation is often due to tandem duplications that involve the control region (see below). Fungi and plants have hugely different mtDNA sizes and patterns of evolution but will not be considered here (see reviews in refs. 94–96). In fishes, intraspecific differences in size can be as large as differences between species[27,29,46] (reviewed in refs. 41, 42, 154).

2. The mitochondrial genetic system

Mitochondrial DNA is a genetic system that is unlike the nuclear genetic system in a number of ways. Animal mtDNA is haploid and non-recombining[113] (but see refs. 123, 163, and references in ref. 124). Plant mitochondrial DNA appears to be a mosaic (a combination of mitochondrial, chloroplast and nuclear genes), assembled through horizontal transfer (cf. ref. 97), and is known to be recombining (references in ref. 97).

The animal mitochondrial genome is almost exclusively maternally inherited. Paternal mitochondria appear to be actively degraded during fertilization[217] or 'out-replicated' shortly thereafter[145]. Recent results[137] 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 mitochondrial DNA genotype of the previous generation[8,111,137]. This proposed bottleneck seems to be a plausible explanation for the observation that a complete replacement of one mtDNA haplotype by another within a single generation can occur in cows[137]. In *Drosophila*, experimental introduction of foreign mtDNA showed that the donor's mtDNA persisted together with the host's mtDNA for several generations. In some cases the donor's mtDNA even replaced the host's mtDNA and was propagated[143,160]. Other studies indicate that the correct nuclear environment is necessary for mitochondria to function (cf. ref. 154).

Generally, only one type of mitochondrial DNA is found in an organism. However reports of heteroplasmy, the presence of more than one type of mtDNA in an individual, are accumulating rapidly (for reviews of heteroplasmy see refs. 153, 154). Heteroplasmy has occasionally been found in most major groups of organisms including several species of fishes, e.g. shad (*Alosa sapidissima*)[27], sturgeon (*Acipenser spp.*)[46] and bowfin (*Amia calva*)[29]. In the reported cases of heteroplasmy in fishes, the proportion of heteroplasmic individuals in the population was 5% in bowfins[29], 12% in shad[27] and up to 41% in sturgeon[46]. Other reported incidents of heteroplasmy include humans[99,173], cows[112], lizards[66,153], newts[220], *Drosophila*[103] and *Mytilus*[118].

Most cases of heteroplasmy involve tandem duplications of some portion of the mt genome, largely in the D-loop region (see below). Length mutations are not the only form of heteroplasmy; simple substitutions are found as well[27]. In cows the heteroplasmic condition (point mutations) can be reversed to a homoplasmic situation within very few generations[111,137], but heteroplasmy due to length variation persists for generations in insects[106,197].

The widespread occurrence of heteroplasmy due to length variation suggests that it might arise easily. Proposed mechanisms for the origin and maintenance of heteroplasmy in copy numbers of tandem repeats include intermolecular recombination[178], polymerase stalling in homopolymer regions[112], replicative misalignment[74], slipped mispairing during replication[66], and misalignment prior to replication through illegitimate elongation[46]. In the heteroplasmy condition, a smaller mtDNA may be more likely to be passed on to the next generation or out-replicate the larger mt genome after fertilization[177,197].

Modelling of the 'population genetical' behavior of mitochondrial genomes suggests that male mitochondria should only rarely be fixed in the large population of oocyte mitochondria[35,55]. However, exceptions to the dogma of exclusive maternal inheritance are beginning to appear in the literature. In the blue mussel, *Mytilus*, experimental production of hybrids demonstrated an extensive paternal contribution[234], which had been suggested before[118]. Similarly, in *Drosophila*[138] and mice, paternal leakage has been documented[102]. Paternal leakage may occur at a low rate of one in a 1000 molecules; the universality and rate of parental leakage

is still unknown[102]. Transfer of genes between mitochondria, chloroplasts and the nucleus in plants and fungi (references in 97) is common, but in animals only a few cases of transfer of mitochondrial genes into the nucleus, and never the other way around, are known[133] (see references in ref. 191). Horizontal transfer of mtDNA between species of animals has been observed in some cases (mice[83], voles[207]) but is probably a rare phenomenon.

3. *Mitochondrial genes*

The mitochondrial genome of animals 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 non-coding region (AT-rich region in invertebrates, control region in vertebrates) that contains the initiation sites for mtDNA replication and RNA transcription (Fig. 1).

Mitochondrial protein coding genes code for enzyme subunits involved in the electron transport chain[10,110,215]. These 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). When only 12 of 13 genes are found, it is the ATPase 8 gene that is missing[120]. Obviously, many more gene products are required for the functioning of the mitochondrial biochemical machinery; the majority are imported into the mitochondria from the cytosol (e.g.

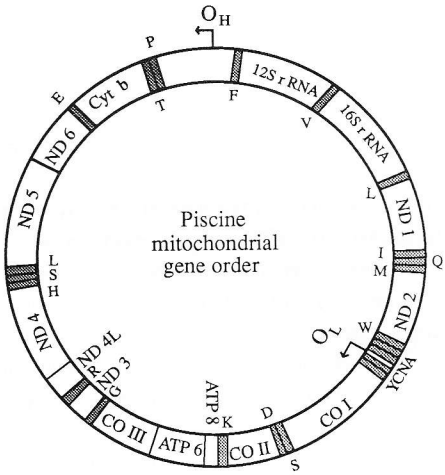


Fig. 1. Piscine mitochondrial gene order. The origins of H- and L-strand replication are indicated in the figure. The origin of the H-strand is in the control region, the origin of the L-strand replication is in the YCNAW tRNA gene cluster. Transfer RNA genes are shown in shaded boxes. The coding sequences (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, 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 by 129.

refs. 107, 215). Apparently, the 'streamlined' nucleus-dependent, mitochondrial genome of today is the result of the progressive transfer of genetic control to the nucleus[219] from a previously more complex and larger mt genome.

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 behavior in $CsCl_2$ density gradients. Endotherm vertebrates show more strand bias than ectotherms, which is caused by a bias against G in the light strand and C in the heavy strand (reviewed in refs. 40–42). With few exceptions, all genes in the vertebrate mitochondrial genome are encoded by the 'H' strand. Of the 37 structural genes in the vertebrate mitochondrial genome, only ND 6 and eight tRNAs are encoded by the L-strand (Fig. 1). This bias towards the 'H' strand as coding strand is generally not as pronounced in invertebrates (e.g. *Drosophila*). The number of coding bases on the L-strand in vertebrates is less than 10% of the total number of coding bases in the mt genome, whereas it is about 50% in *Drosophila*. The coding strand in vertebrates has a strong compositional bias that is correlated with the inequality of 'gene density'.

4. *Gene order*

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[116,192,193]. The piscine mitochondrial gene order does not differ from the vertebrate 'consensus' gene order (Fig. 1). The most complete piscine mtDNA sequence data are from the Atlantic cod (*Gadus morhua)*[129] (Johansen, pers. comm.) and conform to the 'consensus' vertebrate gene order (Fig. 1). Restriction and partial sequence studies on carp (*Cyprinus carpio*)[5] and salmonids (*Salmo salar*)[64](*Oncorhynchus* spp.)[208] and unpublished data from chondrichthyes and agnathans from Wes Brown's laboratory (cf. ref. 154) suggest that the piscine mtDNA gene order complies with the general vertebrate condition.

Mitochondrial gene order differs slightly among vertebrates[68,69,169,232,233]. Marsupial and bird mtDNA gene orders differ somewhat from the 'consensus' vertebrate condition. The most common differences observed are due to tRNA gene transpositions. More exceptions from the 'universal' vertebrate gene order are likely to be found once more groups, e.g. reptiles, are studied in this regard.

The regular interspersion of tRNA genes among protein-coding genes (Fig. 1) is remarkable. It has been suggested[9,24,45,162] that the secondary structure of the tRNA genes serves as a punctuational signal during transcription (see below). Interestingly, when two protein-coding genes are not separated by a tRNA gene (e.g. ATPase 6–COIII), a stable hairpin loop structure can sometimes be inferred from the primary sequence at the 3′ end of the gene (mouse[34]; *Drosophila*[45]). At TA–ATG junctions of genes these secondary structures may be required for proper transcription.

This secondary structure may also exist in some fishes. Figure 2 shows such a suggested hairpin structure based on the primary sequence of rainbow trout (*Oncorhynchus mykiss*)[208]. The structure involves a seven bp stem and a 29 bp loop.

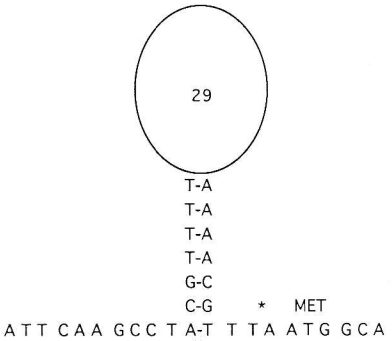


Fig. 2. Suggested hairpin-structure for rainbow trout at the ATPase 6–CO III junction. Sequence data from ref. 208. It is not known whether this structure exists and whether it serves as punctuation signal as had been suggested for *Drosophila* and several other species (see text). Unlike the other species for which this secondary structure had been suggested, the stem in rainbow trout (*Oncorhynchus mykiss*) does not end exactly at the stop codon of the ATPase 6 gene.

The stem, however, does not stop exactly at the TA end of the ATPase 6 gene. A strong secondary structure cannot be inferred from the primary sequence of other species of fishes from four families (Cichlidae, Pomacentridae, Embiotocidae and Labridae) (Meyer and Titus, unpublished data).

5. *Genetic code*

The mitochondrial genetic code is more degenerate and thus less constrained than the 'universal' eukaryotic nuclear genetic code (reviewed in ref. 9). Not all animal mitochondrial genetic codes are identical[1,21,162]. The codes of vertebrates[1], echinoderms[50,127] and *Drosophila* (e.g. ref. 58) differ slightly. The most remarkable differences in all vertebrate mitochondrial codes compared to the nuclear code is the use of TGA as a tryptophan codon rather than as a stop codon and the use of ATA as methionine and not isoleucine[165,166]. The simplified mitochondrial code requires fewer tRNA species. Twenty-four tRNAs would be required to translate the degenerate mt code; however, only 22 tRNAs are usually found in the mt genome. This is possible, since the codons AGA/G are not used as arginine but are stop codons in mammalian and amphibian systems (but not in *Drosophila yakuba*), and tRNA-Met and tRNA-F-Met seem to be the product of the same tRNA gene (reviewed in ref. 9).

Various start codons aside from AUG are allowed as initiator codons (e.g. AUA, AUU, AUC, reviewed in ref. 9). GUG is the translation initiation codon of the COI gene in cod[129], the ND1 gene of rat[88], the ND5 gene of *Drosophila yakuba*[58] and the ATPase 8 gene of sea urchins[50,127]. Some mitochondrial genes lack proper stop codons, ending in T or TA; the complete stop codon (ochre, TAA) is created during RNA processing through polyadenylation of the mRNAs[1,162].

The piscine mitochondrial code is similar to the vertebrate mitochondrial code in that TGA and ATA align with tryptophan and methionine, respectively[129,136,146,148,161].

6. *Efficiency and compactness*

In contrast to the nuclear genome, the mitochondrial genome of animals (but not that of fungi and plants) is highly efficient (reviewed in refs. 9, 97), i.e. it rarely contains duplicate or non-coding sequences. Mitochondrial protein-coding genes do not contain introns, and genes are usually separated by less than 10 base pairs (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, as in *Drosophila yakuba* where the tRNA-Ser and ND1 genes overlap by 18 bp[58]. Some genes that are coded for by different strands overlap as well (e.g. the ND5 and ND6 genes by 14 bp in the mouse[34] and by 17 bp in the cow[2]). 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[1,162]; the overlapping reading frames in each pair of genes appear to be translated from a single mRNA. The ND4 and ND4L genes also overlap. This ND4–ND4L gene overlap is shorter, never more than seven base pairs. In *Drosophila* these genes abut without overlap, and in echinoderms the ND4 and the ND4L genes are separated.

Interestingly, overlap in ATPase 6 genes is shorter in ectotherms (*Xenopus* 10 bp[180]; *Gadus* 10 bp[129]; sculpins 10 bp[93]; cichlid fishes 10 bp [Meyer and Titus unpublished]) compared to endotherms (human 46 bp[1]; cow 40 bp[2]; mouse 43 bp[34]; rat 43 bp[88]; fin whale 31 bp[6]). However, the chicken[68] has the ectothermic condition of a 10 bp overlap. It is not known what the situation in reptiles and marsupials is. It would seem worthwhile to further investigate this phylogenetic trend and its potential molecular causes, if they exist. Given that these genes overlap by so many bases, it may not be surprising that ATPase 6 and 8 genes have not been reported apart in any animal mitochondrial DNA. However, the gene for ATPase 8 appears to be absent in the blue mussel mt genome[120].

It has been suggested that the compactness of the mt genome is partly responsible for the stability of gene order in vertebrates (see above), since the lack of introns and large intergenic spacers makes the mt genome less likely to rearrange. The intergenic spacers appear to be non-functional, because substitutions occur at these sites with about the same frequency as in third positions of protein-coding genes, and additions and deletions are as frequent as in the control region[42].

All non-coding nucleotides of a typical vertebrate mt genome outside the control region sum up to less than 100: 64 in mice [34]; 57 in cows[2]; 87 in humans[1]; 183 in *Drosophila yakuba*[58]. Five large stretches of apparently non-coding DNA have been found in the *Mytilus* mt genome[120]. In vertebrates, the only large non-coding region other than the control region is a 31–34 bp region between the tRNA-Asp and the tRNA-Cys. This is the region at which the origin of the L-strand replication is located (OL in Fig. 1, see below).

In fishes, several sizable non-coding regions have been discovered. In the Atlantic cod (*Gadus morhua*) a 74 bp insertion was found between the tRNA-Thr and the tRNA-Pro[129] in *Xenopus* this stretch of non-coding DNA is only 26 bp long[180] and

in sturgeon only 3 bp long[89]. In several dozen species of fishes from several families there are none or less than 5 bp between the Thr-tRNA and Pro-tRNA[146,200] (Meyer, unpublished data). In the fifteen-spine stickleback (*Spinachia spinachia*) we[150] found a 109 bp insertion between the Glu-tRNA and the cytochrome *b* gene. This insertion is identical in all tested individuals of *Spinachia* and is absent in all closely related species of sticklebacks.

7. *Replication and transcription*

Comprehensive reviews of the transcription and replication of mtDNA have been published[9,59–61]. Replication of mtDNA has chiefly been studied in the mouse[71] (reviewed in refs. 59, 61). Replication of both strands is initiated at two fixed points and proceeds unidirectionally, without interruption. The daughter H-strand synthesis (origin is in the control region, see below) is initiated first and extends towards the cytochrome *b* gene. Once the replicated daughter H-strand has reached the light strand origin of replication (in a tRNA cluster, O_L Fig. 1) the replication of the L-strand begins; it proceeds in the opposite direction and uses the displaced parental H-strand as template strand. This region contains features that are conserved across large evolutionary distances. It forms a large hairpin-loop in which the stem is about 11 bp long and the loop, less conserved in sequence and length, varies between 10 and 19 bp (in *Xenopus*[229]). Another noteworthy feature of the loop is the fact that it is usually high in T content (over 50%) (cf. ref. 42). The only known fish sequence of this region however, has 50% C rather than T[129]. This light-strand replication origin appears to be missing in chicken[68]. Daughter strands separate once the H-strand synthesis is complete and before the L-strand synthesis is finished. Replication appears to be quite different in sea urchins than in vertebrates (review and references in ref. 42).

Replication of mtDNA is continuous, asymmetric and unidirectional, hence it requires fewer enzymes to function than the nuclear machinery[54] (reviewed in refs. 59, 61). A single DNA polymerase (DNA polymerase gamma) has been found in mitochondria[61] as opposed to the typical three found in the eukaryotic nucleus. Repair and proof-reading enzymes, long believed absent, have recently been found in animal mitochondria (cf. ref. 61).

Again, transcription is only well studied in mammalian (mouse)[9,53,60,61] and amphibian (*Xenopus*)[36,47] systems. Transcription of the H-strand is initiated at a promoter site in the control region and a second site in proximity to the tRNA-Phe. The initiation of L-strand transcription occurs at only a single site in the control region. The transcript elongation produces polycistronic RNAs that are later processed into tRNAs, rRNAs and mRNAs.

The origin of replication has not been studied in fishes, but as in all other vertebrates the points of origin for replication appear to be about 5000 bp apart; the origin of L-strand replication has been found in cod in the usual tRNA cluster (Fig. 1[129]). In the control region, conserved sequence blocks (e.g. CSB-2), the transcription initiation region, termination associated sequence (5′–TACAT–3′, TAS[85]) and pyrimidine stretches, implicated to be important in replication, have

been identified in cod[129] based on sequence comparisons with mouse and *Xenopus*. CSB 1 of mammals is found in *Xenopus* and the sturgeon[46] but is missing in the cod control region[129]. Transcription and replication in fish mtDNA may proceed similarly to the known models. An open reading frame (ORF) of considerable size in the control region has been proposed to exist in the C domain in mammals[182] (e.g. ref. 119), yet its reality is unclear and disputed[129]. It is not present in the cod D-loop[129].

III. Rate of evolution

Studies using restriction enzymes indicate that mtDNA generally evolves at elevated rates (5–10 times faster) compared to single copy nuclear genes[44,172]. These findings were confirmed by actual sequences, after the development of DNA sequencing techniques[44]. The faster mtDNA evolution is due to a higher frequency of point and length mutations[44,48,223]. It came as a surprise that mtDNA was found to vary so much in DNA sequence. Mitochondrial DNA was originally assumed to be conservative, because it codes for proteins that play crucial roles in cellular metabolism. The divergence at silent sites of protein-coding genes may be about 10% per million years (generalized from results from primates), ten times the rate found in nuclear protein-coding genes. The overall rate of substitution for the complete mt genome in primates may be 0.5–1% per million years. The rate of silent substitutions, mainly transitions (changes from one purine to another or one pyrimidine to another), is about 4–6 times that of replacement substitutions[43,44]. However, among closely related species, if transitions are ignored and only transversions (changes from a purine to a pyrimidine or *vice versa*) 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 e.g.: mt DNA polymerase might lack proofreading capacity, mtDNA repair enzymes might not be present, mitochondria might lack an efficient system for the removal of pyrimidine dimers and mtDNA. Furthermore, greater exposure of mtDNA to oxidizing agents like radicals and superoxide might cause a higher mutation rate. A more error-prone DNA replication system (through poorer fidelity of dNTP selection, or cytosine deamination)[37], lack of editing (but see below), the high turnover of mtDNA have all been implicated as partly responsible for the high rate of evolution of mtDNA. Another suggested reason for the higher rate of evolution of mtDNA is that inaccuracies might be allowed, since mtDNA does not code for proteins involved in translation, replication, or transcription[109] and mtDNA is therefore free from the 'error-catastrophe loop'[223].

Mouse mitochondria seemed to lack excision and recombination repair capability[62]. The presumed absence of proof-reading and repair enzymes had been cited as a potential reason for the observed higher rate of accumulation of mutations in mtDNA compared to nuclear DNA. Recently, a mtDNA polymerase (gamma) was described from mitochondria (cf. ref. 61) which shows 3′ to 5′ exonuclease activity. Furthermore, the mitochondrial DNA polymerase gamma has a high

fidelity that is mediated by 3′ to 5′ exonuclease (proofreading) activity[141]. Its fidelity and proofreading capability seems to contradict the notion that many errors are introduced during replication or that repair is inefficient. DNA repair enzymes (e.g. uracil-DNA glycosylase and AP endonuclease) have been identified in mammalian mitochondrial systems (cf. refs. 212, 213). Currently, it is therefore not clear that the rapid evolution of mtDNA can be explained by lack of repair.

A case can be made for eased selection due to relaxed functional constraints on mitochondrial gene products as cause for the observed higher rate in mtDNA evolution. Relaxed codon recognition, the fact that four nucleotides are often tolerated at the third position of codons and recognized by the same species of tRNA, and that mutations in these third positions are therefore effectively silent might be one such relaxed constraint against mutations. The observed differences in rates of evolution between nuclear and mitochondrial DNA 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.

1. Tempo and mode of substitution

Of the three kinds of sequence changes, substitutions are more frequent than indels (additions or deletions), rearrangements are least common. Rearrangements have not been reported in the mitochondrial genome of fishes; but reports of length variation and heteroplasmy exist for a few species (see above). Several approaches can provide information about the dynamics, patterns, and processes of nucleotide substitutions. One such approach is intraspecific or intrapopulational DNA sequence comparison. It has the advantage that multiple mutational events at the same nucleotide position, that might obscure a correct scoring (interpretation) of sequence differences, are unlikely among conspecifics or among closely related species. Intraspecific comparisons may provide the best estimate of the pattern of nucleotide substitution and might even help elucidate the underlying processes that cause the observed patterns of evolution.

Additions and deletions are most frequently observed in the control region and intergenic spacers. They occur in the tRNAs and rRNAs as well, but at a lower frequency. Rarely have they been found in protein-coding genes, but they do exist. Irwin *et al.*[126] found single codon additions and deletions in the cytochrome *b* gene among mammals. A single codon deletion was found among salmonids in the ATPase 6 gene[208]. A three-codon addition and a one-codon addition in the ATPase 6 gene of percoid fishes compared to salmonid fishes[208] was also discovered (Meyer and Titus, unpublished data). Slowly evolving protein-coding genes like CO I, II, III and cytochrome *b* do not seem to contain any length variation among Neopterygian fishes[161], and no length variation has been detected among ray-finned or lobe-finned fishes for cytochrome *b* and CO III[148] (Meyer, unpublished data).

As sequencing techniques were developed it became clear that the basis for the higher rate of the mtDNA evolution is due mainly to transitional differ-

ences[4,37,44,99,230]. Transitions often outnumber transversions 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. Any DNA sequence is a snap-shot representation of the substitutional (evolutionary) history that occurred in an evolutionary lineage. Obviously, this snap-shot will be an incomplete representation of history, blurred by multiple substitutions that may have occurred at each site. Studies of closely related species or conspecifics will not be as affected by multiple substitutions and should provide a better representation of evolutionary events and modes and dynamics of substitutions.

Figure 3 shows the dynamics of sequence evolution for the 5′ end of the cytochrome *b* gene for cichlid fishes (Meyer, unpublished data). Substitutions at third positions quickly accumulate and they become saturated with transversions at a maximum level of about 50%. 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

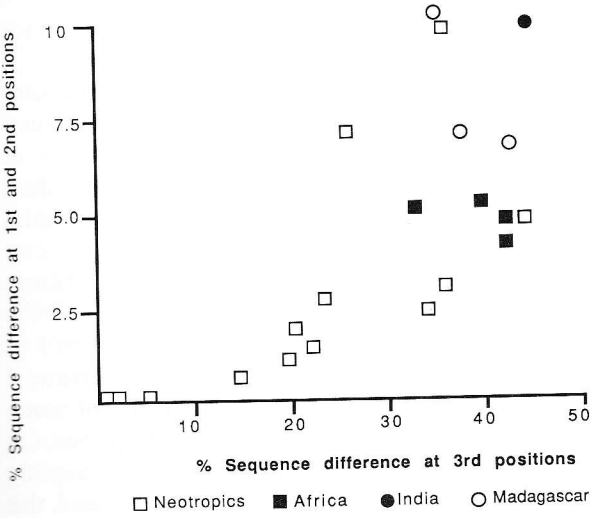


Fig. 3. Mode of sequence evolution for the 5′ end (about 360 bp) of the cytochrome *b* gene for cichlid fishes (Meyer, unpublished data). Cichlid fishes are a family of fishes found in South America, Africa and Madagascar and India. All species were compared to the Neotropical species *Cichlasoma citrinellum*. Third positions substitutions quickly accumulate and become saturated with substitutions at a maximum level of divergence of around 45% (see Fig. 7 and text for explanation). Hence, for phylogenetic analyses, transitions in third positions will not 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 fishes tested.

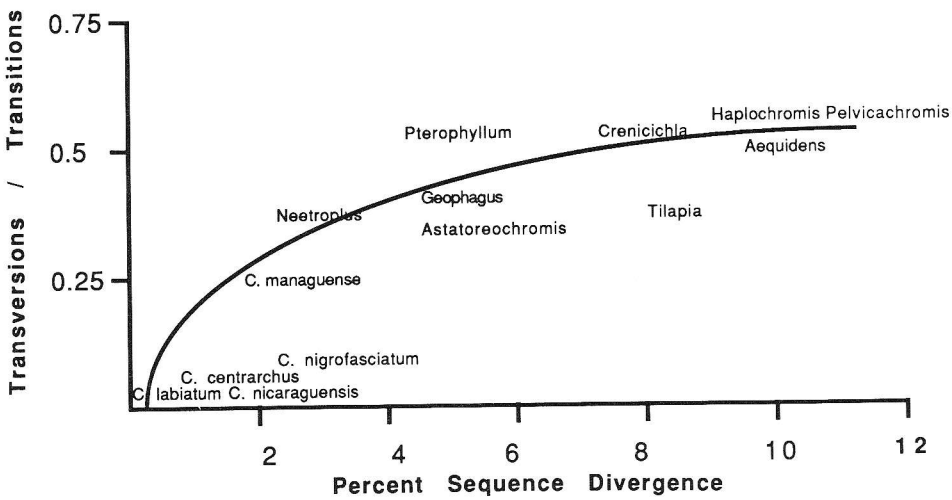


Fig. 4. Dynamics of sequence evolution among cichlid fishes for a conservative (about 300 bp) portion of the 12S rRNA gene (Meyer, unpublished data). Graphed is the mode of sequence divergence and dynamics of the transversion–transition ratio from the Neotropical species *Cichlasoma citrinellum*. Among 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%.

indicators of evolutionary relationships beyond about 30–40% sequence divergence at these positions.

Transversions accumulate slowly[4,37,44,99,230], eventually outnumbering transitions, i.e. the percent transversions of the total number of substitutions observed between two species increases with increasing sequence divergence (Fig. 4). The mode of change from transitions predominating to transversions becoming more prevalent differs between genes and may be related to constraints and base compositional biases[121,122] (see below).

Transitions also occur more frequently among closely related species of fishes, and transversions become apparent among more distantly related species[136,146]. Among conspecific cichlid fishes, transitions often outnumber transversions 5 : 1 to 10 : 1; among distantly related species this ratio changes to about one. Figure 4 shows this trend in a conserved portion of the 12S ribosomal RNA gene for some cichlid fishes. With increasing sequence divergence from the Neotropical species, *Cichlasoma citrinellum*, the proportion of transversions substitutions increases rapidly to an upper limit of about 50 percent. The absolute time required to reach the plateau is inversely related to the absolute rate with which transversions become fixed (see below). Among congeners, total sequence divergence does not exceed about five percent, and scored transversions are 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 (uncorrected for multiple substitutions) in this portion of the 12S rRNA gene does not exceed about 15%. This is also the case for species of fishes in the family Gasterosteidae (Orti,

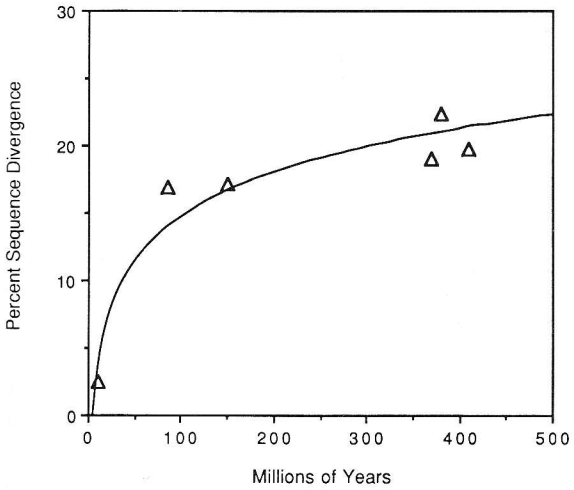


Fig. 5. Dynamics of sequence evolution among lobe-finned and ray-finned fishes. Depicted is the rate of sequence divergence in a conservative 245 bp portion of the 12S rRNA gene over a period of about 400 million years[149]. Compared are two species of the African lungfish *Protopterus* (*annectens* and *aethiopicus*) assumed to have separated recently, these two species are compared with the Southamerican lungfish *Lepidosiren paradoxa* (more than 85 my of separation), the Australian lungfish, *Neoceratodus fosteri* (more than 135–150 my of separation), the coelacanth, *Latimeria chalumnae* (more than 380 my of separation), the frog, *Xenopus laevis* (about 380 my of separation) and a ray-finned fish, the Midas cichlid, *Cichlasoma citrinellum* (more than 400 my of separation). For this conservative portion of the 12S rRNA gene sequences do not seem to diverge more than about 20%. The plateau is reached at around 100–150 million years of divergence. Also around this divergence time has the ratio of transitions to transversions reached a 1:1 plateau that is maintained for more than 400 million years of divergence (data not shown).

Bell and Meyer, unpublished data). Patterns like the one seen in Fig. 4 for the 12S rRNA gene in cichlid fishes are typical. 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[122].

Even quite distantly related species of fishes do not exceed about 20% of total (uncorrected) sequence divergence for this conservative part of the 12S rRNA gene[148,151] (Fig. 5). Figure 5 shows the dynamics of sequence divergence until this plateau is reached among lobe-finned fishes, tetrapods and ray-finned fishes. This plateau is approached at 100–150 million years since the common ancestry between the African and Australian lungfishes (about 17%), and does not exceed about 20% after more than 400 million years of separation. This may be the upper level of sequence divergence possible for this portion of the 12S rRNA gene, and may be determined by requirements of its secondary structure. Transversions outnumber transitions (56%:44%) at about 150 million years, and the ratio of about one is maintained (with some fluctuation) after this date of divergence[149].

The increase in observed number of transversions (e.g. Fig. 4) is due to several factors[44,67,122,130]. There are 12 possible types of substitutions among the four types

a) C → T → C

b) C → G → C or T

c) C → T → G or A

Fig. 6. Relationship of transitions and transversions. See text for details.

of nucleotides: four are transitions, eight are transversions. Although the transition bias is strong, transversions appear to increase with increasing sequence divergence because transversions erase the record of previous transitions[44]. Figure 6 (from ref. 67) attempts to explain three types of interactions that can occur at one nucleotide site. Figure 6a shows that if two transitions occur at the same site in an evolutionary lineage the original base will always be restored, erasing the record of the first transition. Two transversions at one site will also destroy their record (Fig. 6b). However, two possible mutational pathways exist: (a) the second transversion restores the original base, or (b) the new base will be another base, which would be scored as a transition. If one of the two substitution is a transition and the other a transversion the net effect will always be the scoring of a transversion (Fig. 6c). This is true whether the transition occurred first or second. DeSalle *et al.*[67] summarize these rules of substitution: (1) Transversions will erase the record of all previous transitions and transversions at a site. (2) Only transversions can introduce transversions at a given site. (3) Transitions cannot erase the record of a transversion (or odd number of transversions) at a nucleotide position (transitions can erase the record of a particular transversion, but not the record that a transversion difference occurred). This asymmetry[67] between the effect of transitions and transversions will produce an observed accumulation of scored transversions with time, and the percentage of all differences which are scored as transitions will decrease. Although transitions remain a more common type of mutation event between distantly related species, 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.

2. *Protein-coding genes*

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 amino acid 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 (Fig. 7). 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 that will tend to maintain a functional gene product, mutations will accumulate more quickly in first than in the most

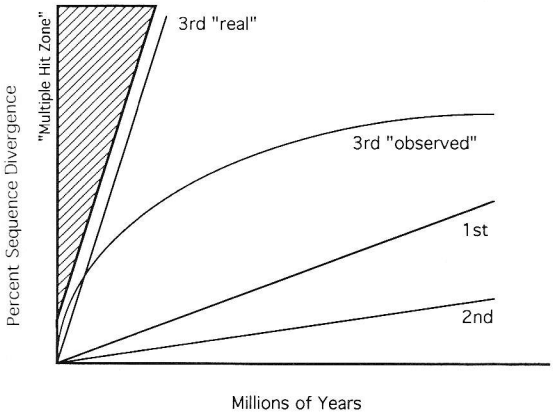


Fig. 7. Hypothetical example of the dynamics of sequence evolution in a mt protein-coding gene. The multiple hit zone is depicted to signify the observation that the likelihood of a multiple substitution at a particular nucleotide increases with sequence divergence. The probability of a multiple hit 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 mt protein-coding genes.

constrained second positions of codons (Fig. 7). 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. Table 1 (from[129]) lists the 13 protein-coding genes from less to more conserved, based on a comparisons of cod (*Gadus morhua*[129]) and frog (*Xenopus*[180]) sequences. The genes coding for the subunits of cytochrome oxidase and cytochrome *b* are the more conserved genes, and the most variable ones are the ND and ATPase genes. 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. ref. 67) and strong functional constraints.

3. *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, and in all cases the 3′ terminal-CCA is not encoded, but added post-transcriptionally. Secondary structures of mitochondrial tRNAs are less conserved but still fold into the cloverleaf secondary structure (but see below). With few exceptions the acceptor stem is 7 bp long, the anticodon stem is 5 bp and the anticodon loop contains 7 nucleotides. A 'U' residue in the position preceding (5') the anticodon is usually preserved; a purine is usually in the position immediately 3′ to the anticodon. Mitochondrial tRNA genes are not only smaller and lower in G + C contents,

TABLE 1

Rate of evolution of mitochondrial genes from cod, *Gadus morhua*[129] [a]

Gene	No. of nucleotides	%DNA similarity	No. of AA positions	%AA similarity
ND4L	294	54	98	46
ND2	413	58	138	56
ND4	482	61	160	57
ND6	226	62	75	64
ATPase 6	211	62	71	49
ATPase 8	165	67	55	55
ND3	348	68	116	74
ND5	1352	70	450	72
ND1	432	70	144	76
Cyt *b*	255	72	85	65
COIII	547	73	182	85
COII	672	74	224	80
COI	1386	79	462	93
12S rRNA	751	72		
16S rRNA	207	80		

[a] Nucleotide and amino acid (AA) sequences were compared to frog sequences (*Xenopus laevis*[180]). Genes are presented in order from fastest (in terms of nucleotide similarity) to most conserved. The complete sequences have not been determined for all genes; the number of nucleotides upon which the comparison is based, is presented. Brown *et al.*[39] determined the complete cytochrome *b* sequence for the sturgeon (*Aciperser transmontanus*) and reported an AA and nucleotide similarity with *Xenopus* of 73.0% and 79.2% respectively.

but also have non-standard basepairing (mainly G:U) in their stems compared to their nuclear counterparts. 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 (reviewed in ref. 198). Mitochondrial tRNA genes evolve at least 100 times faster than nuclear tRNA genes[44]. Compensatory substitutions in stems on the complementary strand have been observed in tRNAs as well as the rRNA genes (see below). 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. Some tRNA genes are more conserved[2] (e.g., the Met tRNA) while others are quite variable, even to the point of missing entire loops (e.g. Ser[AGY] tRNA, e.g. ref. 58).

4. *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 1571–1640 bp in vertebrates) rRNA gene. Based on DNA–DNA hybridization work, Dawid[65] noted that mitochondrial rRNA genes evolve much more rapidly than their nuclear counterpart. The nuclear rRNA (18S and 28S) and tRNA genes (see above) evolve about 100 times more slowly than their mitochondrial

counterparts[65] when the more variable expansion segments of the nuclear rRNA genes are excluded[151].

The same basic rules of substitutions (see above) apply to rRNA genes[44]. 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 (Fig. 4, see above). Insertions and deletions are usually small, in the range of 1–5 bp. The 16S gene contains 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.

Secondary structure models exist for both genes, and general agreement about the secondary structure exists[72,91,117]. These secondary structures are conserved across large evolutionary distances. Although regions of secondary structure are more conserved than other regions, this conservation is not always related to the inferred secondary structure of the genes in an obvious way (e.g. ref. 190). Among tetrapods, lungfishes, the coelacanth, and ray-finned fishes, 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[148] (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 million years[148,149,151] (Figs. 5, 8, see below). Among more closely related species of primates stems are less likely to contain substitutions than loop-regions[117]. 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 (Fig. 4). At around 100 to 150 million years of sequence divergence, the rate of substitution appears to decrease[148,151] (Fig. 5).

The overall rRNA and tRNA substitution rates are about half those of the protein-coding genes[44,115,117,127,148,151], although variation in the rate among protein-coding genes exists (Table 1).

5. *Control region*

The control region contains several regions that are constrained in primary sequence or secondary structure to regulate replication and transcription[61]. 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 displaces 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 the control region is two to five times higher than mitochondrial protein-coding genes[4]. Extreme variability in the control region (in terms

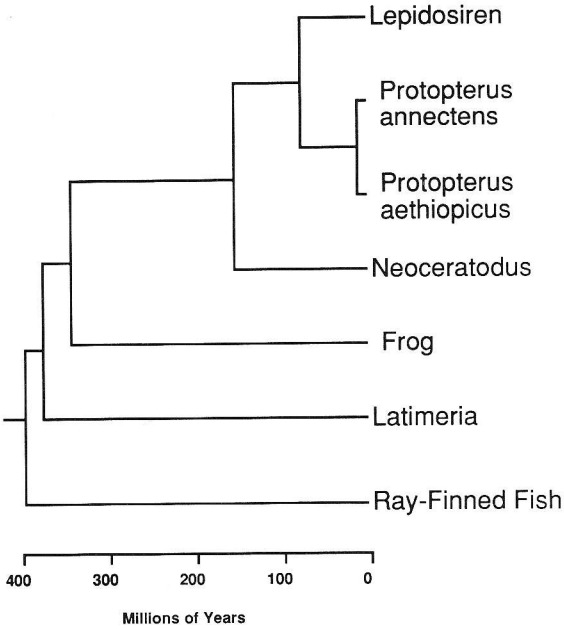


Fig. 8. Phylogenetic tree of lungfishes, based on about 245 bp of a conservative portion of the 12S rRNA gene[149] constructed based on the parsimony principle (computer program PAUP[202]). Based on this example it appears that mtDNA can be used to reconstruct the phylogenetic relationships among distantly related species that diverged more than 400 million years ago.

of substitution and length mutations) was discovered as early as the 1970's[80,81,216]. Nucleotide substitutions occur five times more frequently than additions and deletions in the human control region[4]; among closely related species of cichlid fishes substitutions far outnumber deletions and additions[146]. The control region varies tremendously in length, often because of tandem duplications of 200 bp to 4100 bp[42], and is primarily responsible for the observed variation in the total length of the vertebrate mt genome. It also contains the highest frequency of length mutations at the population level[66,106]. In fishes[27,29,46], it is usually here where tandem duplications and various numbers of repeats are found when heteroplasmy is observed. Direct sequence comparisons of individuals with mtDNAs of different lengths may provide the clues for the actual mechanism that might be responsible for the introduction of length variation. Slippage and mispairing might be the mechanism by which additions and deletions occur[46].

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[38,182].

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 fishes[79,146,179,200].

6. *Slowed rate of evolution in ectotherms?*

Much variation was expected in terms of amino acid differences in mitochondrial protein-coding genes of fishes, since up to 8.5% sequence divergence had been found within species of sunfishes[13]. This is not always the case, however. There is little variation within or between species of salmonids[208] and cichlids[136,146] in the more conserved protein-coding genes.

Cold-blooded animals may be up to five times slower than warm-blooded animals in the evolution of cytochrome *b* amino acid replacement substitutions[136] (Fig. 9). This may be true as well 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 fishes[136,148,161]. This slowed rate allowed mtDNA to be used to investigate phylogenetic relationships among very distantly related species[148,149,161]. The rate of amino acid substitutions in the cytochrome *b* gene proved to be conservative enough to test relationships among groups of lobe-finned fishes and ray-finned fishes that have not shared a common ancestor for more than 400 million years[148]. The number of amino acid substitutions in CO and cytochrome *b* genes actually was too small to clearly elucidate the evolutionary relationships among Neopterygian fishes[161].

Although the mtDNA evolution in ectotherms appears to be slower compared to birds and mammals (Fig. 9), 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 fishes of the genus *Xiphophorus* showed

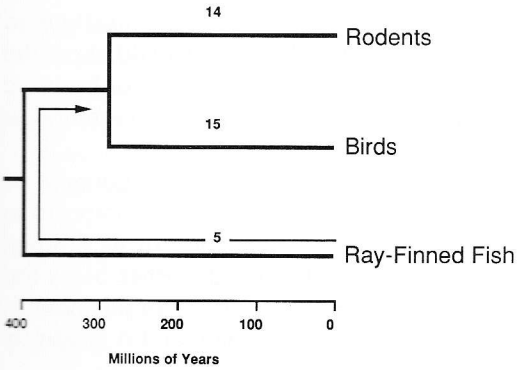


Fig. 9. Slowed rate of amino acid substitutions in ectotherms (from 136). Number of estimated number of amino acid substitutions (in the 5′ end of cytochrome *b*) that occurred along these lineages. Fishes appear to have an about five times slower rate of amino acids substitutions. See text for details.

that mtDNA evolves at least 2.5 times faster than these nuclear genes (Meyer *et al.*, unpublished data). This estimate for the rate of these nuclear genes included introns and therefore overestimates the real nuclear rate. Based on earlier findings on primates (e.g. ref. 43) 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 fishes (compared to mammals) because of slowed mtDNA molecular substitution rates or faster nuclear substitution rates will require much more data than are currently available. If the mitochondrial rate is generally slowed in fishes and/or the rate of nuclear DNA evolution increased will require the knowledge of divergence dates. However, too few studies have calibration points from the fossil record that allow to test these hypotheses. This is important for our understanding of the absolute rates of sequence divergence of both genomes, as well as comparisons between the two.

Slowed rates of mtDNA evolution had previously been suggested for sea urchins[170,218] and flies[189,195]. However, in some cases the alternative explanation, an increase in single copy nuclear gene substitution rate, rather than a slowed mitochondrial substitution rate could not be excluded[209].

IV. Potential mechanisms of mitochondrial DNA evolution

1. Base compositional and codon biases

Base compositional biases in nuclear and mitochondrial genes have been well documented[30–33]. Usually ectotherms have lower levels of G and C than warm-blooded animals. This holds true for the whole genome, and nuclear protein-coding genes in particular. This anti G–C bias in ectotherm vertebrates has been found in all codon positions of nuclear protein-coding genes. Codon bias and compositional biases have been observed in all animal mitochondrial genomes, the heavy strand is rich in G and T (reviewed in refs. 41, 42).

Figure 10 shows the base composition of several species of fishes for the cytochrome *b* gene and the D-loop. Because the D-loop and third positions in codons may be functionally less constrained, these two data sets should show the clearest demonstration of base compositional biases. D-loop sequences of both the cod[129], and the frog[180] show the typical anti G bias found in other vertebrates. Furthermore, both seem to favor A and T over C.

In the cytochrome *b* gene, this bias detected from the D-loop sequences of these two species is exacerbated in third positions. The overall base composition of cytochrome *b* in the cod is very similar to the base composition of its D-loop, but only in the third positions is the anti-G bias pronounced. The other bases are almost equally represented. The frog, which has a relatively even use of nucleotides (except G) in the cytochrome *b* gene favors A and T in the D-loop and A relatively strongly in the third positions of cytochrome *b*.

All species of fishes investigated show a strong bias against G in the cytochrome *b* gene and particularly in third positions (Fig. 10). Biases against T in third positions

exist in some, but not all species of fishes (Fig. 10). Both birds[73] and mammals[126] have an extremely strong bias toward A and C in the third position of the cytochrome *b* gene. Whether these differences are phylogenetically determined or functionally constrained is not known.

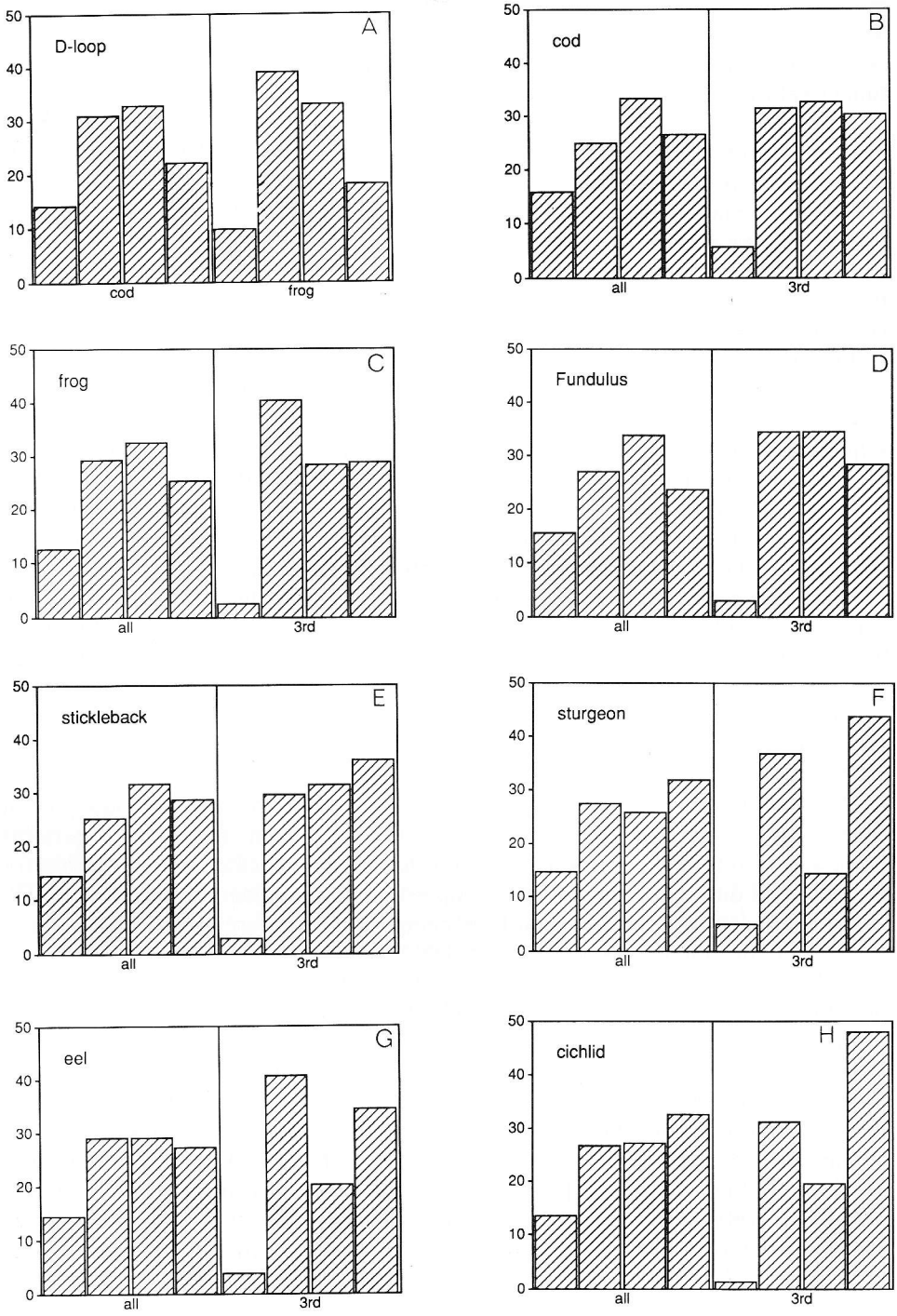
More than 80% of all codons end in A or C, about 14% in T and less than 6% in G in chicken[68]. Codon bias is the effect of, not the cause of, base compositional differences between the two strands. The main reason for the very infrequent use of G is that the main coding strand is low in G (around 12–14% in most vertebrates). Chicken shows a rearrangement of the mitochondrial gene order, the ND6 gene is coded for by the opposite strand[68]. The codons of this gene have chiefly G or T in third positions, in keeping with the predominant strand bias. The light strand, which codes for a majority of the mitochondrial genes, is C and A rich; therefore, most codons in fishes end in A and C and rarely in G.

A base compositional bias, e.g. the often observed bias against guanines in third positions of protein-coding genes, has the effect that the ceiling for saturation of sequence divergence is lower (e.g. Fig. 7) than if all bases had the same probability of occurring at a particular site. For example, in *Drosophila* a strong A + T bias is observed[67] and limits the magnitude of (uncorrected) sequence divergence obtainable. Caution must be taken when phylogenies are reconstructed, since base compositional biases will obscure phylogenetic relationships.

In nuclear genes, the genes most often used have a more biased base composition compared to other less frequently translated genes[157]. It has been argued that the base compositional bias will make translation faster by skewing the codon usage towards the most common tRNAs. In the mitochondrial genome two arguments would seem to speak against this mechanism: most amino acids are coded for by only one tRNA species, and all mitochondrial genes are part of a polycistronic mRNA. Directional selection and mutational bias may lead to the same outcome.

Whether mutational pressure or selection are responsible for any biases is not entirely clear; selection has often been invoked to explain the observed patterns (reviewed in ref. 78). Directional selection has been implicated as responsible for the observed differences in the base composition between man, cow and mouse[108]. It has also been hypothesized that base compositional differences within genomes are caused by differences in between DNA synthesis, repair and replication[84]. Since base compositional bias is not restricted to third positions in codons, mutational pressure rather than selection might be largely responsible for the observed compositional biases[132].

Because of the base compositional bias, choice among synonymous codons is clearly non-random[3]. There is a correlation between the frequency of nucleotides (base composition) and the synonymous substitution rate in nuclear genes of rat, human and *Drosophila*[155,211]. Rates are usually higher for A (and T) biases and lower for C (and G) biased genes. The quickly evolving control region (AT-rich region in *Drosophila*) is AT-rich. Quickly evolving pseudogenes also tend to have elevated levels of A and low levels of C contents compared to functional genes (e.g. ref. 155).


A
D-loop
cod
frog
B
cod
all
3rd
C
frog
all
3rd
D
Fundulus
all
3rd
E
stickleback
all
3rd
F
sturgeon
all
3rd
G
eel
all
3rd
H
cichlid
all
3rd
0
10
20
30
40
50

2. *Potential mechanisms driving mitochondrial DNA evolution*

Hasegawa *et al.*[109] noted (see above) that since mtDNA does not code for enzymes involved in replication and translation, the gene products may be relatively free from stringent constraints involving the proper function of the gene product. MtDNA may be free from the error catastrophe loop[223], i.e. the catastrophic accumulation of errors due to improperly functioning of replication and translation enzymes.

The observation that transitions occur at all positions at a high frequency, rather than only third positions in codons, makes it likely that a bias in creating, rather than in selecting against, transversions is responsible for the high occurrence of transitions. Several molecular mechanisms during replication (e.g. cytosine deamination[37]) may be responsible for the preponderance of transitions, base compositional biases and the generally elevated mutation rate in mtDNA compared to nuclear genes. Cytosine deamination, which results in a uracil instead of a cytosine in one strand, leads to a concomitant increase of adenosine residues in the opposite strand. It has been suggested[228] that the evolution towards A + T rich genomes may reflect the absence of a mechanism for repairing C to T deaminations (e.g. absence of a mitochondrial uracil N-glycosylase) or a tendency for the more efficient use of A and T by DNA polymerase[132,230].

W. Kelly Thomas (pers. comm.) developed an hypothesis that centers around the argument that cytosine deamination occurs frequently in the first strand of replication (the H-strand in most animals). Replication of the two strands is asymmetric (see above) and one strand will stay single-stranded for some time (cf. ref. 61). Thomas argues that the longer the time of single-strandedness, the higher the chance for transitional mutations due to cytosine deamination or directional mutation. A strong anti-C bias in the H-strand will be the result, hence an anti-G and pro A bias will result in the L-strand. This results in a strong base compositional

---

Fig. 10. Base compositional biases in control region (D-loop) and cytochrome *b* genes. All graphs are shown in order G, A, T, C.

A. Comparison of the base composition of the control regions in cod, *Gadus morhua* (1069 bp)[129] and frog, *Xenopus laevis* (2204 bp)[180]. Both show a pronounced anti-G bias, T is most common in the cod, the frog has a pro-A bias. B–H: base compositions of all nucleotides of the cytochrome *b* gene (in the left panels) and only the third positions in codons (in the right panels) for several species of fishes.

B. the cod (*Gadus morhua*) cytochrome *b* sequence (1142 bp: Johansen pers. comm.) shows a quite even base composition (except for the usual anti-G bias), as in the control region (Fig. 10A) T is the most common base.

C. The frog (*Xenopus laevis*) cytochrome *b* (1137 bp[180]) sequence shows an extreme anti-G bias and, as in the control region (Fig. 10A) favors A, which is most clearly seen in third positions.

D. *Fundulus heteroclitus* (411 bp: Meyer, unpublished data), like the cod (Fig. 10B) shows an even distribution of nucleotides at third positions.

E. The three-spined stickleback, *Gasterosteus aculeatus* (794 bp: Orti, Bell and Meyer, unpublished data) shows the usual anti-G bias in third positions and favors C slightly in third positions.

F. The sturgeon, *Acipenser transmontanus* (1140 bp[39]) shows an pronounced bias towards A and C in the third positions.

G. The American eel, *Anguilla rostrata* (708 bp: Meyer, unpublished data, amino acid sequence reported in 77) also shows a strong pro A and C bias.

H. The cichlid fish, *Hemichromis bimaculatus* (684 bp: Meyer, unpublished data) as the sturgeon and eel, shows a strong A and C bias. The base compositional bias is very strong anti-G and pro-C.

bias in both strands and would explain the observed transition bias in mtDNA. This model is supported by an astonishingly high difference in the cytosine contents in the H-strand in relation to the distance (therefore the time it is single-stranded) from the replication origin of the L-strand. The cytosine content decreases with decreasing distance from the origin of replication of the L-strand. Thomas' model is testable, since in echinoderms several different methods of replication have been found in this phylum.

V. Mitochondrial DNA and the polymerase chain reaction

Since the advent of the polymerase chain reaction in 1985–1986, our knowledge about mitochondrial DNA has increased dramatically. The polymerase chain reaction (PCR) is an enzymatic cloning technique that allows the amplification of any piece of DNA (within size limits) that is flanked by synthetic oligonucleotide 'primers'[156,183,184]. PCR is much faster and cheaper than conventional cloning techniques. The polymerase chain reaction and its application to evolutionary biology has been reviewed[7] (references in refs. 75, 125).

The study of mitochondrial DNA experienced a boost through the development of so-called 'universal' primers and direct sequencing of DNA without subcloning[100,136]. 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, these conserved areas can be identified and primers designed that have applicability for a wide taxonomic range (e.g. a phylum or even beyond)[136]. With these primers, mtDNA from fishes can be amplified and sequenced without prior sequence knowledge from fishes. Table 2 lists several published[93,136,146,161] and previously unpublished primer sequences for a number of mitochondrial genes that have been successfully used in several groups of fishes and are expected to work for most groups of teleosts.

Both restriction fragment data and sequence data provide enough resolution for most population level work. However, another effect of the 'universal' primers is that a universal currency, actual DNA sequences rather than restriction fragment length polymorphisms (RFLPs), has been created. Although RFLP analyses have contributed tremendously to our increased understanding of intra- and interspecific genetic variation and the evolution of mtDNA in fishes and remain a viable technique, results were not immediately transferable between laboratories. DNA sequences of the same mitochondrial gene(s) will allow the direct comparison and study of mtDNA from different species that have been determined in different laboratories.

1. 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 in refs.

TABLE 2

Primers for the polymerase chain reaction that have been successfully used in fishes [a]

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		
H1109	GTGGGGTATCTAATCCCAGTT	(Meyer, unpublished)
L1091	AAAAAGCTTCAAACTGGGATTAGATACCCCACTAT	(Kocher *et al.*, 1989)
H1478	TGACTGCAGAGGGTGACGGGGCGGTGTGT	(Kocher *et al.*,1989)
16S rRNA		
L2510	CGCCTGTTTATCAAAAACAT	(Wilson lab)
H3080	CCGGTCTGAACTCAGATCACGT	(Wilson lab)
CO I		
L5950	ACAATCACAAAGAYATYGG	(Normark *et al.*, 1991)
L6586	CCTGCAGGAGGAGGAGAYCC	(Palumbi, COIf, SFG)
H7086	CCTGAGAATARKGGGAATCAGTG	(Palumbi,COIe, SFG, modified)
H7196	AGAAAATGTTGWGGGAARAA	(Normark *et al.*, 1991)
CO II		
L7450	AAAGGAAGGAATCGAACCCCC	(Normark *et al.*, 1991)
H8055	GCTCATGAGTGGAGGACGTCTT	(Normark *et al.*, 1991)
ATPase 8		
L8331	TAAGCRNYAGCCTTTTAAG	(Meyer, unpublished)
ATPase 6		
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)
Cytochrome *b*		
L14725	CGAAGCTTGATATGAAAAACCATCGTTG	(Pääbo *et al.*, 1991)
L14841	AAAAAGCTTCCATCCAACATCTCAGCATGATGAAA	(Kocher *et al.*, 1989)
L15020	GCYAAYGGCGCATCCTTYTTYTT	(Meyer lab, unpublished)
H15149	AAACTGCAGCCCCTCAGAATGATATTTGTCCTCA	(Kocher *et al.*, 1989)
L15162	GCAAGCTTCTACCATGAGGACAAATATC	(Pääbo *et al.*, 1991; Irwin *et al.*, 1991)
L15424	ATCCCATTCCACCCATACTACTC	(Edwards *et al.*, 1991)
H15573	AATAGGAAGTATCATTCGGGTTTGATG	(Wilson lab)
L15774	GTACATGAATTGGAGGACAACCAGT	(Irwin *et al.*, 1991)
H15915	AACTGCAGTCATCTCCGGTTTACAAGAC	(Irwin *et al.*, 1991)

[a] Primer names follow the Kocher *et al.*[136] convention of naming the primer by the most 3′ position of the primer in the human mtDNA sequence[1]. 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. 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. SFG = Simple fool's guide to PCR, distributed by S.R. Palumbi's laboratory, University of Hawaii.

11, 18, 154, 223). The fast rate of evolution of mtDNA[43] 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 have made mtDNA an extremely popular genetic system with which to study gene flow, hybrid zones, population structure and other population level questions (reviewed in, e.g., refs. 11, 18, 105, 154, 223). MtDNA 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 and mtDNA (reviewed in ref. 105). MtDNA also lends itself well for the study of the origin of clonally reproducing species[66,176] (reviewed in ref. 154). Mitochondrial DNA as well as allozyme markers have been used to identify hybridization events[13,14,114]. The special inheritance of mtDNA has consequences for its population genetical behavior[15,35,159,171] (see reviews in refs. 11, 18, 105) but that topic will not be covered here.

Until recently most of the work on population structure and phylogenetic relationships of fishes based on mitochondrial DNA data was based on the analysis of restriction fragment length polymorphisms (RFLPs)[17,101,224,225] (reviews in refs. 11, 18, 154). Within species, high levels of polymorphism and geographic differentiation are sometimes found. In some sunfishes (*Lepomis spp.*) up to 8.5% intraspecific variation has been described[16]. Extremely low levels of sequence variation within and between species have been found as well[94,146,168,226]. 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[208,223]. In the case of intraspecific 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[26,208].

Differences in the estimates of sequence divergence based on restriction enzymes and DNA sequences, can be due to several factors, e.g. 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 will be inaccurate. Furthermore, since 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 disproportionally to the estimate of sequence divergence based on RFLP analysis of the complete mtDNA genome; in interspecific studies, multiple hits 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 in refs. 70, 104, 167, 181). RFLP analysis had been the predominant method in population studies[226] but PCR has experienced increased application for these questions. Even conservative protein-coding genes like cytochrome *b* tend to show intraspecific variation, mainly in third positions of codons that can be used to identify fish stocks[22,23,52,144] or be used for phylogenetic studies among closely related species[136,146,147,179,299] (see below).

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[20]. 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. ref. 105). However, current levels of intraspecific variation may not be an accurate reflection of the pre-speciation levels, and this furthermore assumes that the rates of mtDNA evolution are the same in all lineages.

2. *Mitochondrial DNA and phylogenetic questions*

MtDNA has found widespread use as a tool for phylogenetic analyses, several reviews are available[11,18,140,141,154,223]. 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 the polymerase chain reaction it has become apparent that mitochondrial DNA sequences may be also a useful tool for the study of evolutionary relationships among more distantly related species[136,146,148]. The relative ease with which DNA sequences of mitochondrial genes can be determined through PCR, universal primers, and direct sequencing has led to a virtual explosion of new mtDNA sequence information. Mitochondrial DNA data can be phylogenetically analyzed in several ways, e.g. converted to distances, or as multistate characters (for review of phylogenetic methods using DNA sequences see ref. 203).

Multiple mutations at the same nucleotide position (multiple hits) become increasingly more likely and therefore accumulate with increasing sequence divergence (Fig. 7). They tend to obscure evolutionary relationships through homoplasy (for discussion of homoplasy see ref. 186), and they also result in an observed upper limit of sequence divergence, a ceiling that is approached asymptotically (Figs. 4 and 7). Estimates of sequence divergence 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 percent sequence divergence) will differ from an estimate that is derived from some correction that attempts to account for multiple substitutions.

Distance methods for phylogenetic reconstruction are based on a distance matrix (e.g. neighbor-joining[185]). If they are used to reconstruct phylogenetic relationships (see below) sequence distances between two species need to be corrected for multiple hits. Multiple hit corrections can be done in a number of ways; Jukes and Cantor's method[131] is the simplest method but tends to underestimate the evolutionary divergence. More elaborate models are available and provide more reliable estimates of 'real' evolutionary divergence[92,134,135,205,206] (Fig. 7).

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 often are 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[148,149,152] (reviewed in refs. 115, 151).

Universal PCR primers for mtDNA[136] 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 targeted 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 substitution 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 [126](reviewed in ref. 76). Even using complete cytochrome *b* sequences could not aid in the reconstruction of some phylogenetic relationships (e.g. among mammals[126]). This may be due to several factors, e.g. short branch length, due to rapid speciation during the radiation of mammals. Furthermore, 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 fishes, believed to have diverged more than 80 million years ago (Meyer, unpublished data). Although this gene has been very useful for studies involving populations (e.g. refs. 22, 23) and the phylogenetic study of very different groups of fishes[148,161] it may not be the most appropriate gene for all purposes (reviewed in ref. 76). 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 (Table 1). Primer sequences for many other genes are available. Table 2 lists several published and previously unpublished PCR primers that have proven to work reliably in many groups of fishes.

Other mitochondrial or nuclear genes may have substitution rates more appro-

priate for particular questions. For example, the control region, since it is the fastest evolving mtDNA region, may be more useful for population level questions than the cytochrome *b* gene. 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. refs. 148, 200). In a comparison between two closely related species of Neotropical cichlid fishes (*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 CO 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 tRNA-Pro 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 for phylogenetic purposes is determined at one extreme by the stochasticity of the distribution of haplotype polymorphisms within species or among closely related 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[158].

3. *Prospectus*

For almost every topic covered in this paper more questions about process, pattern and mode of substitution have been raised. Much remains to be studied. Several developments in molecular biology, namely the polymerase chain reaction and direct sequencing of DNA, will profoundly affect the rate at which knowledge about piscine mtDNA evolution will accumulate, and it is hoped that in a few years our understanding of mtDNA evolution in fishes will have been amplified.

Acknowledgements. I thank S. Dolven, G. Orti, T. Städler, C. Sturmbauer and T. Titus for discussion and comments on the manuscript. R. Hoffmann, J. Boer, F. Breden, W. Brown, A. Fajen, W.K. Thomas and S. Johansen kindly provided unpublished data, or manuscripts. Allan C. Wilson has been an example, mentor and inspiration to me, as Mensch as well as scientist. Support from the Systematic Biology Panel of the National Science Foundation (BSR-9107838 and BSR-9119867) is gratefully acknowledged.

VI. References

1. Anderson S., A.T. Bankier, B.G. Barrell, M.H.L. de Bruijn, A.R. Coulson, J. Drouin, I.C. Eperon, D.P. Nierlich, B.A. Roe, F. Sanger, P.H. Schreier, A.J.H. Smith, R. Staden and I.G. Young. Sequence and organization of the human mitochondrial genome. *Nature* 290: 457–465, 1981.

2. Anderson, S., M.H.L. de Bruijn, A.R. Coulson, I.C. Eperon, F. Sanger and I.G. Young. Complete sequence of bovine mitochondrial DNA. *J. Mol. Biol.* 156: 683–717, 1982.
3. Aota, S. and T. Ikemura. Diversity in G + C content at the third position of codons in vertebrate genes and its cause. *Nucl. Acids Res.* 14: 6345–6355, 1986.
4. Aquadro, C.F. and B.D. Greenberg. Human mitochondrial DNA variation and evolution: Analysis of nucleotide sequences from seven individuals. *Genetics* 103: 287–312, 1983.
5. Araya, A., R. Amthauer, G. Leon and M. Krauskopf. Cloning, physical mapping and genome organization of mitochondrial DNA from *Cyprinus carpio* oocytes. *Mol. Gen. Genet.* 196: 43–52, 1984.
6. Arnason, U.A. Gullberg and B. Widegren. The complete nucleotide sequence of the mitochondrial DNA of the fin whale, *Balaenoptera physalus*. *J. Mol. Evol.* 33: 556–568, 1991.
7. Arnheim, N., T. White and W.E. Rainey. Application of PCR: Organismal and population biology. *BioScience* 40: 174–182, 1990.
8. Ashley, M.V., P.J. Laipis and W.W. Hauswirth. Rapid segregation of heteroplasmic bovine mitochondria. *Nucl. Acids Res.* 17: 7325–7331, 1989.
9. Attardi, G. Animal mitochondrial DNA: An extreme example of genetic economy. *Int. Rev. Cyt.* 93: 93–145, 1985.
10. Attardi, G. and G. Schatz. Biogenesis of mitochondria. *Ann. Rev. Cell. Biol.* 4: 289–333, 1988.
11. Avise, J.C. Mitochondrial DNA and the evolutionary genetics of higher animals. *Phil. Trans. Soc. Lond. B* 312: 325–342, 1986.
12. Avise, J.C. Gene trees and organismal histories: a phylogenetic approach to population biology. *Evolution* 43: 1192–1208, 1989.
13. Avise, J.C. and N.C. Saunders. Hybridization and introgression among species of sunfish (*Lepomis*): Analysis by mitochondrial DNA and allozyme markers. *Genetics* 108: 237–255, 1984.
14. Avise, J.C. and M.J. Van den Avyle. Genetic analysis of reproduction of hybrid White bass x Striped bass in the Savannah River. *Trans. Am. Fish. Soc.* 113: 563–570, 1984.
15. Avise, J.C., Neigel, J.E. and J. Arnold. Demographic influences on mitochondrial DNA lineage survivorship in animal populations. *J. Mol. Evol.* 20: 99–105, 1984.
16. Avise, J.C, Bermingham, E., Kessler, L.G., and N.C. Saunders. Characterization of mitochondrial DNA variability in a hybrid swarm between subspecies of bluegill sunfish (*Lepomis macrochir*). *Evolution* 38: 931–941, 1984.
17. Avise, J.C., Helfman, G.S., Saunders, N.C. and L. Stanton Hales. Mitochondrial DNA differentiation in North Atlantic eels: Population genetic consequences of an unusual life history pattern. *Proc. Nat. Acad. Sci. USA* 83: 4350–4354, 1986.
18. Avise, J.C., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E., Reeb, C.A. and N.C. Saunders. Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. *Ann. Rev. Ecol. Syst.* 18: 489–522, 1987.
19. Avise, J.C., Ball, R.M. and J. Arnold. 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: 331–344, 1988.
20. Avise, J.C., C.D. Akney and W.S. Nelson. Mitochondrial gene trees and the evolutionary relationship of mallard and black ducks. *Evolution* 44: 1109–1119, 1990.
21. Barrell, B.G., Bankier, A.T. and J. Drouin. A different genetic code in human mitochondria. *Nature* 282: 189–194, 1979.
22. Bartlett, S.E. and Davidson, W.S. Identification of *Thunnus* tuna species by the polymerase chain reaction and direct sequence analysis of their mitochondrial cytochrome *b* genes. *Can. J. Fish. Aquat. Sci.* 48: 309–317, 1991.
23. Bartlett, S.E. and W.S. Davidson. FINS (Forensically informative nucleotide sequencing): A procedure for identifying the animal origin of biological specimens. *Biotechniques* 12: 408–411, 1992.
24. Battey, J. and D.A. Clayton. The transcription map of human mitochondrial DNA implicates transfer RNA excision as a major processing event. *J. Biol. Chem.* 255: 11599–11606, 1980.
25. Batuecas B., Garesse, R., Calleja, M., Valverde, J.R. and R. Marco. Genome organization of *Artemia* mitochondrial DNA. *Nucl. Acids Res.* 16: 6515–6529, 1988.
26. Beckenbach, A.T., W.K. Thomas and H. Sohrabi. Intraspecific sequence variation in the mitochondrial genome of rainbow trout (*Oncorhynchus mykiss*). *Genome* 33: 13–15, 1990.
27. Bentzen, P., Leggett, W.C. and G.G. Brown. Length and restriction site heteroplasmy in the mitochondrial DNA of American shad (*Alosa sapidissima*). *Genetics* 118: 509–518, 1988.
28. Bermingham, E. and J.C. Avise. Molecular zoogeography of freshwater fishes in the southeastern

United States. *Genetics* 113: 939–965, 1986.
29. Bermingham, E., Lamb, T. and J.C. Avise. Size polymorphism and heteroplasmy in the mitochondrial DNA of lower vertebrates. *J. Hered.* 77: 249–252, 1986.
30. Bernardi, G. and G. Bernardi. Compositional constraints and genome evolution. *J. Mol. Evol.* 24: 1–11, 1986.
31. Bernardi, G. and G. Bernardi. Compositional patterns in the nuclear genomes of cold-blooded vertebrates. *J. Mol. Evol.* 31: 265–281, 1990.
32. Bernardi, G. and G. Bernardi. Compositional transition in the nuclear genomes of cold-blooded vertebrates. *J. Mol. Evol.* 31: 282–293, 1990.
33. Bernardi, G. and G. Bernardi. Compositional properties of nuclear genes from cold-blooded vertebrates. *J. Mol. Evol.* 33: 57–67, 1991.
34. Bibb, M.J. R.A. Van Etten, C.T. Wright, M.W. Walberg and D.A. Clayton. Sequence and gene organization of mouse mitochondrial DNA. *Cell* 26: 167–180, 1981.
35. Birky, C.W. Jr., T. Maruyama and P.A. Fuerst. 1983. An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. *Genetics* 103: 513–527, 1983.
36. Bogenhagen, D.F., B.K. Yoza and S.S. Cairns. Identification of the initiation sites for transcription of *Xenopus laevis* mitochondrial DNA. *J. Biol. Chem.* 261: 8488–8494, 1986.
37. Brown, G.G. and M.V. Simpson. Novel features of animal mtDNA evolution as shown by sequences of two rat cytochrome oxidase subunit II genes. *Proc. Nat. Acad. Sci. USA* 79: 3246–3250, 1982.
38. Brown, G.G., G. Gadaleta, G. Pepe, C. Saccone and E. Sbiza. Structural conservation and variation in the D-loop containing region of vertebrate mitochondrial DNA. *J. Mol. Biol.* 192: 503–511, 1986.
39. Brown, J.R., T.L. Gilbert, D.J. Kowbel, P.J. O'Hara, N.E. Buroker, A.T. Beckenbach and M.J. Smith. Nucleotide sequence of the apocytochrome B gene in white sturgeon mitochondrial DNA. *Nucl. Acids Res.* 17: 4389, 1989.
40. Brown, W.M. Mechanisms of evolution in animal mitochondrial DNA. *Ann. N.Y. Acad. Sci.* 361: 119–134, 1981.
41. Brown, W.M. Evolution of mitochondrial DNA. In: *Evolution Of Genes And Proteins*, edited by M. Nei and R.K Koehn, Sunderland, MA: Sinauer, 1983, pp. 62–88.
42. Brown, W.M. The mitochondrial genome of animals. In: *Molecular Evolutionary Genetics*, edited by R.J. MacIntyre, New York: Plenum Press, 1985, pp. 95–130.
43. Brown, W.M., M. George, Jr. and A.C. Wilson. Rapid evolution of mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* 76: 1967–1971, 1979.
44. Brown, W.M., E.M. Prager, A. Wang and A.C. Wilson. Mitochondrial DNA sequences of primates: Tempo and mode of evolution. *J. Mol. Evol.* 18: 225–239, 1982.
45. Bruijn de, M.H.L. *Drosophila melanogaster* mitochondrial DNA, a novel organization and genetic code. *Nature* 304: 234–241, 1983.
46. Buroker, N.E., J.R. Brown, T.A. Gilbert, P.J. O'Hara, A.T. Beckenback, W.K. Thomas and M.J. Smith. Length heteroplasmy of sturgeon mitochondrial DNA: An illegitimate elongation model. *Genetics* 124: 157–163, 1990.
47. Cairns, S.S. and D.F. Bogenhagen. Mapping of the displacement loop within the nucleotide sequence of *Xenopus laevis* mitochondrial DNA. *J. Biol. Chem.* 261: 8481–8487, 1986.
48. Cann, R.L., Brown, W.M. and A.C. Wilson. Polymorphic sites and the mechanism of evolution in human mitochondrial DNA. *Genetics* 106: 479–499, 1984.
49. Cantatore P., M.N. Gadaleta, M. Roberti, C. Saccone and A.C. Wilson. Duplication and remoulding of tRNA genes during the evolutionary rearrangement of mitochondrial genomes. *Nature* 329: 853–855, 1987.
50. Cantatore P., M. Roberti, G. Rainaldi, M.N. Gadaleta and C. Saccone. The complete nucleotide sequence, gene organization, and genetic code of the mitochondrial genome of *Paracentrotus lividus*. *J. Biol. Chem.* 264: 10965–10975, 1989.
51. Cantatore P. and C. Saccone. Organization, structure, and evolution of mammalian mitochondrial genes. *Int. Rev. Cytol.* 108: 149–208, 1987.
52. Carr S.M. and D. Marshall. 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, 1991.
53. Chang, D. and D.A. Clayton. Precise identification of individual promoters for transcription of each strand promoter. *Cell* 36: 635–643, 1984.
54. Chang, D. and D.A. Clayton. Priming of human mitochondrial DNA replication occurs at the light strand promoter. *Proc. Natl. Acad. Sci. USA* 82: 351–355, 1985.

55. Chapman, R.W., J.C. Stephens, R.A. Lansman and J.C. Avise. Models of mitochondrial DNA transmission genetics and evolution in higher eukaryotes. *Genet. Res.* 40: 41–57, 1982.
56. Chomyn, A. and G. Attardi. Mitochondrial gene products. *Curr. Top. Bioener.* 15: 295–329, 1987.
57. Clark-Walker, G.D. Basis of diversity in mitochondrial DNAs. In: *The Evolution Of Genome Size*, edited by T. Cavalier-Smith. New York: Wiley, 1985, pp. 277–297.
58. Clary, D.O. and D.R. Wolstenholme. The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization and genetic code. *J. Mol. Evol.* 22: 252–271, 1985.
59. Clayton, D.A. Replication of animal mitochondrial DNA. *Cell* 28: 693–705, 1982.
60. Clayton, D.A. Transcription of the mammalian mitochondrial genome. *Ann. Rev. Biochem.* 53: 573–594, 1984.
61. Clayton, D.A. Replication and transcription of vertebrate mitochondrial DNA. *Ann. Rev. Cell. Biol.* 7: 453–478, 1991.
62. Clayton, D.A., J.N. Doda and E.C. Freidberg. The absence of a pyrimidine dimer repair mechanism in mammalian mitochondria. *Proc. Natl. Acad. Sci. USA* 71: 2777–2781, 1974.
63. Cruz de la, V.F., N. Neckelmann and L. Simpson. Sequence of six genes and several open reading frames in the kinetoplast maxicircle DNA of *Leishmania tarentolae*. *J. Biol. Chem.* 259: 15136–15147, 1984.
64. Davidson, W.S., T.P. Birt and J.M. Green. Organization of the mitochondrial genome from Atlantic salmon (*Salmo salar*). *Genome* 32: 340–342, 1989.
65. Dawid, I.B. Evolution of mitochondrial DNA sequences in *Xenopus*. Dev. Biol. 29: 139–151, 1972.
66. Densmore, L.D. Wright, J.W. and W.M. Brown. Length variation and heteroplasmy are frequent in mitochondrial DNA from parthenogenetic and bisexual lizards (genus *Cnemidophorus*). *Genetics* 110: 689–707, 1985.
67. DeSalle, R., T. Freedman, E.M. Prager and A.C. Wilson. Tempo and mode of sequence evolution in mitochondrial DNA of Hawaiian *Drosophila*. *J. Mol. Evol.* 26: 157–164, 1987.
68. Desjardins, P. and R. Morais. Sequence and gene organization of the chicken mitochondrial genome. *J. Mol. Biol.* 212: 599–634, 1990.
69. Desjardins, P. and R. Morais. Nucleotide sequence and evolution of coding and noncoding regions of a quail mitochondrial genome. *J. Mol. Evol.* 32: 153–161, 1991.
70. Dizon, A.E., C. Lockyer, W.F. Perrin, D.P. Demaster and J. Sisson. Rethinking the stock concept: A phylogenetic approach. *Conservation Biol.* 6: 24–36, 1992.
71. Doda, J.N., C.T. Wright and D.A. Clayton. Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. *Proc. Natl. Acad. Sci. USA* 78: 6116–6120, 1981.
72. Dunon-Bluteau, D. and G. Brun. The secondary structure of the *Xenopus laevis* and human mitochondria small ribosomal subunit RNA are similar. *FEBS Letts.* 198: 333–338, 1986.
73. Edwards, S.V., P. Arctander and A.C. Wilson. Mitochondrial resolution of a deep branch in the genealogical tree for perching birds. *Proc. R. Soc. Lond. B* 243: 99–107, 1991.
74. Efstratiadis, A.J., J.W. Posakony, T. Maniatis, R.W. Lawn, C. O.Connell, R.A. Spritz, J.K. Deriel, B.G. Forget, S.M. Weissman, J.L. Slightom, A.E. Blechi. O. Smithies, F.E. Baralle, C.C. Shoulders and N.J. Proudfoot. The structure and evolution of the human beta-globin gene family. *Cell* 21: 653–668, 1980.
75. Erlich, H.A. (ed.) *PCR Technology: Principle and applications for DNA amplification*. New York: Stockton Press, 1989.
76. Esposti Degli, M., S. De Vries, M. Crimi, A. Ghelli, T. Patarnello, T. and A. Meyer. Mitochondrial cytochrome *b*: Evolution and structure of the protein. *Biochem. Biophys. Acta (Bioenergetics)* 1143: 243–271, 1993.
77. Esposti Degli, M., A. Ghelli, M. Crimi, A. Baracca, G. Solaini, T. Tron and A. Meyer. Cytochrome *b* of fish mitochondria is strongly resistant to funiculosin, a powerful inhibitor of respiration. *Arch. Biochem. Biophys.* 295: 198–204, 1992.
78. Eyre-Walker, A.C. An analysis of codon usage in mammals: selection or mutation bias? *J. Mol. Evol.* 33: 442–449, 1991.
79. Fajen, A. and F. Breden. Mitochondrial DNA sequence variation among natural populations of the Trinidad Guppy, *Poecilia reticulata*. *Evolution* 46: 1457–1465, 1992.
80. Fauron, C.M.-R. and D.R. Wolstenholme. Structural heterogeneity of mitochondrial DNA molecules within the genus *Drosophila*. *Proc. Natl. Acad. Sci. USA* 73: 3623–3627, 1976.
81. Fauron, C.M.-R. and D.R. Wolstenholme. Extensive diversity among *Drosophila* species with respect to nucleotide sequences within the adenine + thymine-rich region of mitochondrial DNA molecules. *Nucl. Acids Res.* 8: 2439–2452, 1980.

82. Fauron, C.M.-R. and D.R. Wolstenholme. Intraspecific diversity of nucleotide sequences within the adenine + thymine rich region of mitochondrial DNA molecules of *Drosophila mauritiana, Drosphila melanogaster* and *Drosophila simulans*. *Nucl. Acids Res.* 8: 5391–5410, 1980.
83. Ferris, S.D., R.D. Sage, C.M. Huang, J.N. Nielsen, U. Ritte and A.C. Wilson. Flow of mitochondrial DNA across a species boundary. *Proc. Natl. Acad. Sci. USA* 80: 2290–2294, 1983.
84. Filipski, J. Why the rate of silent codon substitutions is variable within a vertebrate's genome. *J. theor. Biol.* 134: 159–164, 1988.
85. Foran, L.A., J.E. Hixon and W.M. Brown. Similarities of ape and human sequences that regulate mitochondrial DNA transcription and their DNA synthesis. *Nucl. Acids Res.* 17: 5841–5861, 1989.
86. Edwards, S.V., P. Arctander and A.C. Wilson. Mitochondrial resolution of a deep branch in the genealogical tree for perching birds. *Proc. Roy. Soc. B* 243: 99–107, 1991.
87. Garey, J.R. and D.R. Wolstenholme. Platyhelminth mitochondrial DNA: Evidence for early evolutionary origin of a tRNA (Ser) AGN that contains a dihydrourindine arm replacement loop, and of serine-specifying AGA and AGG codons. *J. Mol. Evol.* 28: 374–387, 1989.
88. Gadaleta, G., G. Pepe, G. DeCandia, C. Quagliariello, E. Sbisa and C. Saccone. The complete nucleotide sequence of the *Rattus norvegicus* mitochondrial genome: cryptic signals revealed by comparative analysis between vertebrates. *J. Mol. Evol.* 28: 497–516, 1989.
89. Gilbert, T.L., J.R. Brown, P.J. O'Hara, N.E. Buroker, A.T. Beckenbach and M.J. Smith. Sequence of tRNA-Thr and tRNA-Pro from white sturgeon (*Acipenser transmontanus*) mitochondria. *Nucl. Acids Res.* 16: 11825, 1988.
90. Gjetvaj, B., D.I. Cook and E. Zouros. Repeated sequences and large-scale size variation of mitochondrial DNA: A common feature among scallops (Bivalvia: Pectinidae). *Mol. Biol. Evol.* 9: 106–124, 1992.
91. Glotz, C., C. Zwieb, R. Brimacombe, K. Edwards and H. Kossel. Secondary structure of the large subunit ribosomal RNA from *Escherichia coli, Zea mays* chloroplast, and human and mouse mitochondrial ribosomes. *Nucl. Acids Res.* 9: 3287–3306, 1981.
92. Gojobori, T., K. Ishii and M. Nei. Estimation of average number of nucleotide substitutions when the rate of substitution varies with nucleotides. *J. Mol. Evol.* 18: 414–423, 1982.
93. Grachev, M.A., S. Ja. Slobodyanyuk, N.G. Kholodilov, S.P. Fyodorov, S.I. Belikov, D. Yu. Sherbakov, V.G. Sideleva, A.A. Zubin and V.V. Kharchenko. Comparative study of two protein-coding regions of mitochondrial DNA from three endemic sculpins (Cottoidei) of Lake Baikal. *J. Mol. Evol.* 34: 85–90, 1992.
94. Graves, J.E., S.D. Ferris and A.E. Dizon. Close genetic similarity of Atlantic and Pacific skipjack tuna (*Katsuwonus pelamis*) demonstrated with restriction endonuclease analysis of mitochondrial DNA. *Mar. Biol.* 79: 315–319, 1984.
95. Gray, M.W. Mitochondrial genome diversity and the evolution of mitochondrial DNA. *Can. J. Zool.* 60: 157–171, 1982.
96. Gray, M.W. The bacterial ancestry of plastids and mitochondria. *BioScience* 33: 693–699, 1983.
97. Gray, M.W. Origin and evolution of mitochondrial DNA. *Ann. Rev. Cell. Biol.* 5: 25–50, 1989.
98. Gray, M.W. and W.F. Doolittle. Has the endosymbiont hypothesis been proven? *Microbiol. Rev.* 46: 1–42, 1982.
99. Greenberg, B.D., J.E. Newbold and A. Sugino. Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. *Gene* 21: 33–49, 1983.
100. Gyllensten, U.B. and H.A. Erlich. 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: 7652–7655, 1988.
101. Gyllensten, U. and A.C. Wilson. Mitochondrial DNA of salmonids: Inter- and Intraspecific variability detected with restriction enzymes. In: *Population Genetics And Fishery Management*. edited by N. Ryman and F. Utter, Seattle: Univ. Washington Press, 1987, pp. 301–317.
102. Gyllensten, U., D. Wharton, A. Josefsson and A.C. Wilson. Paternal inheritance of mitochondrial DNA in mice. *Nature* 352: 255–257, 1991.
103. Hale, L.A. and R.S. Singh. Extensive variation and heteroplasmy in size of mitochondrial DNA among geographical populations of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* 83: 8813–8817, 1986.
104. Hallerman, E.M. and J.S. Beckmann. DNA-level polymorphism as a tool in fisheries science. *Can. J. Fish. Aquat. Sci.* 45: 1075–1087, 1988.
105. Harrison, R.G. Molecular changes at speciation. *Ann. Rev. Ecol. Syst.* 22: 281–308, 1991.
106. Harrison, R.G., D.M. Rand and W.C. Wheeler. Mitochondrial DNA size variation within individual crickets. *Science* 228: 1446–1448, 1985.

107. Hartl, F.-U. and W. Neupert. Protein sorting to mitochondria: Evolutionary conservation of folding and assembly. *Science* 247: 930–938, 1990.
108. Hasegawa, M. and H. Kishino. Heterogeneity of tempo and mode of mitochondrial DNA evolution among mammalian orders. *Jap. J. Genet.* 64: 243–258, 1989.
109. Hasegawa, M., T. Yano and T. Miyata. Evolutionary implications of error amplification in the self-replicating and protein-synthesizing machinery. *J. Mol. Evol.* 20: 77–85, 1984.
110. Hatefi, Y. The mitochondrial electron transport and oxidative phosphorylation system. *Ann. Rev. Biochem.* 54: 1015–1069, 1985.
111. Hauswirth, W.W. and P.J. Laipis. Mitochondrial DNA polymorphism in a maternal lineage of Holstein cows. *Proc. Natl. Acad. Sci. USA* 79: 4686–4690, 1982.
112. Hauswirth, W.W., M.J. Van De Walle, P.J. Laipis and P.D. Olivo. Heterogeneous mitochondrial DNA D-loop sequences in bovine tissue. *Cell* 37: 1001–1007, 1984.
113. Hayashi, J-I., Y. Tagashira and M.C. Yoshida. Absence of extensive recombination between inter- and intraspecies mitochondrial DNA in mammalian cells. *Exp. Cell. Res.* 160: 387–395, 1985.
114. Herke, S.W., I. Kornfield, P. Moran and J.R. Moring. Molecular confirmation of hybridization between Northern Pike (*Esox lucius*) and Chain Pickerel (*E. niger*). *Copeia* 1990: 846–850, 1990.
115. Hillis, D.M. and M.T. Dixon. Ribosomal DNA: Molecular evolution and phylogenetic inference. *Quart. Rev. Biol.* 66: 411–453, 1991.
116. Himeno, H., H. Masaki, T. Kawai, T. Ohta, I. Kumagai, I. Miura and K. Watanabe. Unusual genetic codes and a novel gene structure for tRNA-AGY-Ser in starfish mitochondrial DNA. *Gene* 56: 219–230, 1987.
117. Hixon, J.E. and W.M. Brown. 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, 1986.
118. Hoeh, W.R., K.H. Blakley and W.M. Brown. Heteroplasmy suggests limited biparental inheritance of *Mytilus* mitochondrial DNA. *Science* 251: 1488–1490, 1991.
119. Hoelzel, A.R., J.M. Hancock and G.A. Dover. Evolution of the cetacean mitochondrial D-loop region. *Mol. Biol. Evol.* 8: 475–493, 1991.
120. Hoffmann, R., J.L. Boore and W.M. Brown. A novel mitochondrial genome organization in the blue mussel, *Mytilus edulis*. *Genetics* 131: 397–412, 1992.
121. Holmquist, R. Solution to a gene divergence problem under arbitrary stable nucleotide transition probabilities. *J. Mol. Evol.* 8: 337–349, 1976.
122. Holmquist, R. 1983. Transitions and transversion in evolutionary descent: an approach to understanding. *J. Mol. Evol.* 19: 134–144, 1983.
123. Horak, I., H.G. Coon and I.B. Dawid. Interspecific recombination of mitochondrial DNA molecules in hybrid somatic cells. *Proc. Natl. Acad. Sci. USA* 71: 1828–1832, 1974.
124. Hurst, L.D. Sex, slime and selfish genes. *Nature* 354: 23–24, 1991.
125. Innis, M.A., D.H. Gelfand, J.J. Sninsky and T.J. White. *PCR Protocols: A Guide To Methods And Applications*. Academic Press, San Diego, 1990.
126. Irwin, D.M., T.D. Kocher and A.C. Wilson. Evolution of the cytochrome *b* of mammals. *J. Mol. Evol.* 32: 128–144, 1991.
127. Jacobs, H.T., D.J. Elliott, V.B. Math and A. Farquharson. Nucleotide sequence and gene organization of sea urchin mitochondrial DNA. *J. Mol. Biol.* 202: 185–217, 1988.
128. Jacobs, H., S. Asakawa, T. Araki, K. Miura, M.J. Smith and K. Watanabe. Conserved tRNA gene cluster in starfish mitochondrial DNA. *Curr. Genet.* 15: 193–206, 1989.
129. Johansen, S., P.H. Guddal and T. Johansen. Organization of the mitochondrial genome of Atlantic cod, *Gardus morhua*. *Nucl. Acids Res.* 18: 411–419, 1990.
130. Jukes, T.H. Transitions, transversions, and the molecular evolutionary clock. *J. Mol. Evol.* 26: 87–98, 1987.
131. Jukes, T.H. and C.R. Cantor. Evolution of protein molecules. In: *Mammalian Protein Metabolism*, edited by H.N. Munro, New York: Academic Press, 1969, pp. 21–123.
132. Jukes, T.H. and V. Bhushan. Silent nucleotide substitutions and G + C content of some mitochondrial and nuclear genes. *J. Mol. Evol.* 24: 39–44, 1986.
133. Kamimura, N.S., M. Ishii, M. Liandon and J.W. Shay. Three mitochondrial DNA sequences are contiguous in human genomic DNA. *J. Mol. Biol.* 210: 703–707, 1989.
134. Kimura, M. A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16: 111–120, 1980.
135. Kimura, M. Estimation of evolutionary distances between homologous nucleotide sequences. *Proc. Natl. Acad. Sci. USA* 78: 454–458, 1981.

136. Kocher. T.D., W.K. Thomas, A. Meyer, S.V. Edwards, S. Pääbo, F.X. Villablanca and A.C. Wilson. Dynamics of mitochondrial DNA evolution in animals: Amplification and sequencing with conserved primers. *Proc. Natl. Acad. Sci. USA* 86: 6196–6200, 1989.
137. Koehler, C.M., G.L. Lindberg, D.R. Brown, D.C. Beitz, A.E. Freeman, J.E. Mayfield and A.M. Myers. Replacement of bovine mitochondrial DNA by sequence variant within one generation. *Genetics* 129: 247–255, 1991.
138. Kondo, R.Y., E.T. Satta, H. Matsuura, N. Ishiwa, N. Takahata and S.I. Chigusa. Incomplete maternal transmission of mitochondrial DNA in *Drosophila*. *Genetics* 126: 657–663, 1990.
139. Kornfield, I. Descriptive genetics of cichlid fishes. In: *Evolutionary Genetics Of Fishes*, edited by B.J. Turner, New York: Plenum Press, 1984, pp. 591–615.
140. Kornfield, I. Genetics. In: *Cichlid Fishes: Behaviour, Ecology And Evolution*, edited by M.H.A. Keenleyside, New York: Chapman Hall, 1991, pp. 103–128.
141. Kunkel, T.A. and A. Soni. Exonucleolytic proofreading enhances the fidelity of DNA synthesis by chicken embryo DNA polymerase-gamma. *J. Biol. Chem.* 263: 4450–4459, 1988.
142. Margulis, L. *Origin Of Eukaryotic Cells*. New Haven, CT: Yale Univ. Press, 1970.
143. Matsuura, E.T., S.I. Chigusa and Y. Niki. Induction of mitochondrial DNA heteroplasmy by intra- and interspecific transplantation of germ plasm in *Drosophila*. *Genetics* 122: 663–667, 1989.
144. McVeigh, H.P., S.E. Bartlett and W.S. Davidson. Polymerase chain reaction/direct sequence analysis of the cytochrome *b* gene in *Salmo salar*. *Aquaculture* 95: 225–231, 1991.
145. Meland, S., S. Johansen, T. Johansen, K. Haugli and F. Haugli. Rapid disappearance of one parental mitochondrial genotype after isogamous mating in the myxomycete *Physarum polycephalum*. *Curr. Genet.* 19: 55–60, 1991.
146. Meyer A., T.D. Kocher, P. Basasibwaki and A.C. Wilson. Monophyletic origin of Victoria cichlid fishes suggested by mitochondrial DNA sequences. *Nature* 347: 550–553, 1990.
147. Meyer A., T.D. Kocher and A.C. Wilson. 1991. African fishes — a reply. *Nature* 351: 467–468, 1991.
148. Meyer, A. and A.C. Wilson. Origin of tetrapods inferred from their mitochondrial DNA affiliation to lungfish. *J. Mol. Evol.* 31: 359–364, 1990.
149. Meyer. A. and S.I. Dolven. Molecules, fossils and the origin of tetrapods. *J. Mol. Evol.* 35: 102–113, 1992.
150. Meyer. A., G. Orti, and M.A. Bell. Unusually large insertion in the mitochondrial genome of the fifteen-spine stickleback (*Spinachia spinachia*). manuscript, 1992.
151. Mindell, D.P. and R.L. Honeycutt. Ribosomal RNA in vertebrates: Evolution and phylogenetic applications. *Ann. Rev. Ecol. Syst.* 21: 541–566, 1990.
152. Miyamoto, M.M. and S.M. Boyle. The potential importance of mitochondrial DNA sequence data to eutherian mammal phylogeny. In: *The Hierarchy of Life*, edited by B. Fernholm, K. Bremer and H. Jornvall, Amsterdam: Elsevier, 1989, pp. 437–450.
153. Moritz, C. Evolutionary dynamics of mitochondrial DNA duplications in parthenogenetic geckos, *Heteronotia binoei*. *Genetics* 129: 221–230, 1991.
154. Moritz, C., T.E. Dowling and W.M. Brown. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Ann. Rev. Ecol. Syst.* 18: 269–292, 1987.
155. Moriyama, E.N. and T. Gojobori. Rates of synonymous substitution and base composition of nuclear genes in *Drosophila*. *Genetics* 130: 855–864, 1992.
156. Mullis, K., F. Faloona, S. Scharf, S. Saiki, G. Horn and H.A. Erlich. Specific enzymatic amplification of DNA *in vitro*: The polymerase chain reaction. *Cold Spring Harbor Symp. Quant. Biol.* 51: 263–273, 1986.
157. Muto, A. and S. Osawa. The guanine and cytosine content of genomic DNA and bacterial evolution. *Proc. Natl. Acad. Sci. USA* 84: 166–169, 1987.
158. Nei, M. Relative efficiencies of different tree-making methods for molecular data. In: *Phylogenetic Analysis Of DNA Sequences*. edited by M.M. Miyamoto and J. Cracraft, New York: Oxford Univ. Press, 1991, pp. 90–128.
159. Neigel, J.E. and J.C. Avise. Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation. In: *Evolutionary Processes And Theory*, edited by E. Nevo and and S. Karlin, 1986, pp. 515–534.
160. Niki, Y., S.I. Chigusa and E.T. Matsuura. Complete replacement of mitochondrial DNA in *Drosophila*. *Nature* 341: 551–552, 1989.
161. Normark, B.B., A.R. McCune and R.G. Harrison. Phylogenetic relationships of Neopterygian fishes, inferred from mitochondrial DNA sequences. *Mol. Biol. Evol.* 8: 819–834, 1991.
162. Ojala, D., J. Montoya and G. Attardi. tRNA punctuation model of RNA processing in human

mitochondrial DNA. *Nature* 290: 470–474, 1981.
163. Olivo, P.D., J.J. Van de Walle, J.P. Laipis and W.W. Hauswirth. Nucleotide sequence evidence for rapid genotypic shifts in bovine mitochondrial DNA D-loop. *Nature* 306: 400–402, 1983.
164. Okimoto, R., J.L. Macfarlane, D.O. Clary and D.R. Wolstenholme. The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. *Genetics* 130: 471–498, 1992.
165. Osawa, S., T. Ohama, T.H. Jukes and K. Watanabe. Evolution of the mitochondrial genetic code I. Origin of AGR serine and stop codons in metazoan mitochondria. *J. Mol. Evol.* 29: 202–207, 1989.
166. Osawa, S, A. Muto, T.H. Jukes and T. Ohama. Evolutionary changes in the genetic code. *Proc. Roy. Soc. Lond. B.* 241: 19–28, 1990.
167. Ovenden, J.R. Mitochondrial DNA and marine stock assessment: A review. *Aust. J. Mar. Freshwater Res.* 41: 835–853, 1990.
168. Ovenden, J.R., R.W.G. White and A.C. Sanger. Evolutionary relationships of *Gadopsis spp.* inferred from restriction enzyme analysis of their mitochondrial DNA. *J. Fish Biol.* 32: 137–148, 1988.
169. Pääbo, S., W.K. Thomas, K.M. Whitfield, Y. Kumazawa and A.C. Wilson. Rearrangements of mitochondrial transfer RNA genes in marsupials *J. Mol. Evol.* 33: 426–430, 1991.
170. Palumbi, S.R. and A.C. Wilson. Mitochondrial DNA diversity in the sea urchins *Stronglyocentrotus purpuratus* and *S. droenachiensis*. *Evolution* 44: 403–415, 1990.
171. Pamilo, P. and M. Nei. Relationships between gene trees and species trees. *Mol. Biol. Evol.* 5: 568–583, 1988.
172. Perler, F., A. Efstratiadis, P. Lomedico, W. Gilbert, R. Kolodner and J. Dodgson. The evolution of genes: the chicken preproinsulin gene. *Cell* 20: 555–556, 1980.
173. Poulton, J., M.E. Deadman and R.M. Gardiner. Duplications of mitochondrial DNA in mitochondrial myopathy. *Lancet* 236–240, 1989.
174. Prager, E.M. and A.C. Wilson. Ancient origin of Lactalbumin from lysozyme: Analysis of DNA and amino acid sequences. *J. Mol. Evol.* 27: 326–335, 1988.
175. Prichard, A.E., J.J. Seilhamer, R. Mahalingam, C.L. Sable, S.E. Venuti and D.J. Cummings. Nucleotide sequence of the mitochondrial genome of *Paramecium*. *Nucl. Acids Res.* 18: 173–180, 1990.
176. Quattro, J.M., J.C. Avise, and R.C. Vrijenhoek. Mode and origin and sources of genotypic diversity in triploid gynogenetic fish clones (*Poeciliopsis*: Poeciliidae). *Genetics* 130: 621–628, 1992.
177. Rand, D.M. and R.G. Harrison. Mitochondrial DNA transmission genetics in crickets. *Genetics* 114: 955–970, 1986.
178. Rand, D.M. and R.G. Harrison. Molecular population genetics of mtDNA size variation in crickets. *Genetics* 121: 351–369, 1989.
179. Reinthal, P.N. and A. Meyer. Genetic tests of speciation models in cichlid fishes from Lake Malawi, Africa. manuscript, 1992.
180. Roe, B.A., D.-P. Ma, R.K. Wilson and J.F.-H. Wong. The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome. *J. Biol. Chem.* 260: 9759–9774, 1985.
181. Ryman, N. and F. Utter (ed.). *Population Genetics And Fishery Management*. Seattle: Univ. of Washington Press, 1987.
182. Saccone, C., M. Attimonelli and E. Sbiza. Structural elements highly preserved during the evolution of the D-loop-containing region in vertebrate mitochondrial DNA. *J. Mol. Evol.* 26: 205–211, 1987.
183. Saiki, R.K., S. Scharf, F. Faloona, K.B. Mullis, G.T. Horn, H.A. Erlich and N. Arnheim. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230: 1350–1354, 1985.
184. Saiki, R.K., D.H. Gelfand, S. Stoffel, S. Scharf, R. Higuchi, G.T. Horn, K.B. Mullis and H.A. Erlich. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239: 487–491, 1988.
185. Saitou, N. and M. Nei. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4: 406–425, 1987.
186. Sanderson, M.J. and M.J. Donoghue. Patterns of variation in levels of homoplasy. *Evolution* 43: 1781–1795, 1989.
187. Sederoff, R.R. Structural variation in mitochondrial DNA. Adv. Genet. 22: 1–108, 1984.
188. Sederoff, R.R. Molecular mechanisms of mitochondrial-genome evolution in higher plants. *Am. Nat.* 130: S30–S45 (suppl.), 1987.
189. Sharp, P.M. and W.-H. Li. On the rate of DNA sequence evolution in *Drosophila*. *J. Mol. Evol.* 28: 398–402, 1989.

190. Simon, C. Molecular systematics at the species boundary: Exploiting conserved and variable regions of the mitochondrial genome of animals via direct sequencing from amplified DNA. In: *Molecular Taxonomy*, NATO Advanced Studies Institute, edited by G.M. Hewitt, A.W.B. Johnston and J.P.W. Young, Berlin: Springer Verlag, H57: 33–71, 1991.
191. Smith, M.F., W.K. Thomas and J.L. Patton. Mitochondrial DNA-like sequence in the nuclear genome of an akodontine rodent. *Mol. Biol. Evol.* 9: 204–215, 1992.
192. Smith, M.F., D.K. Banfield, K. Doteval, S. Gorski and D. J. Kowbel. Gene arrangement in sea star mitochondrial DNA demonstrates a major inversion event during echinoderm evolution. *Gene* 76: 181–185, 1989.
193. Smith, M.F., D.K. Banfield, K. Doteval, S. Gorski and D. J. Kowbel.Nucleotide sequence of nine protein-coding genes and 22 tRNAs in the mitochondrial DNA of the sea star *Pisaster ochraceus J. Mol. Evol.* 31: 195–204, 1990.
194. Snyder, M., A.R. Fraser, J. LaRoche, K.E. Gartner-Kepkay and E. Zouros. Atypical mitochondrial DNA from the deep-sea scallop *Placopecten magellanicus*. *Proc. Natl. Acad. Sci. USA* 84: 7595–7599, 1987.
195. Solignac, M., M. Monnerot and J.-C. Mounolou. Mitochondrial DNA heteroplasmy in *Drosophila mauritiana*. *Proc. Natl. Acad. Sci. USA* 80: 6942–6946, 1983.
196. Solignac, M., M. Monnerot and J.-C. Mounolou. Mitochondrial DNA evolution in the *melanogaster* species subgroup of *Drosophila*. *J. Mol. Evol.* 23: 31–40, 1986.
197. Solignac, M., J. Genermont, M. Monnerot, and J.-C. Mounolou. *Drosophila* mitochondrial genetics: evolution of heteroplasmy through germ line cell divisions. *Genetics* 117: 687–696, 1987.
198. Sprinzl, M., N. Dank, S., Nock and A. Schon. Compilation of tRNA sequences and sequences of tRNA genes. *Nucl. Acids Res.* 19: 2127–2171, 1991.
199. Stanton, D.J., T.J. Crease and P.D.N. Herbert. Cloning and characterization of *Daphnia* mitochondrial DNA. *J. Mol. Evol.* 33: 152–155, 1991.
200. Sturmbauer, C. and A. Meyer. Genetic divergence, speciation and morphological stasis in a lineage of African cichlid fishes. *Nature* 358: 578–581. 1992.
201. Sueoka, N. Directional mutation pressure and neutral molecular evolution. *Proc. Natl. Acad. Sci. USA* 85: 2653–2657, 1988.
202. Swofford, D.L. *PAUP: Phylogenetic Analysis Using Parsimony*, Version 3.0rβ1. Illinois Natural History Survey, Champaign, IL, 1991.
203. Swofford, D.L. and G.J. Olsen. Phylogeny reconstruction. In: *Molecular Systematics*. edited by D.M Hillis and C. Moritz, Sunderland MA: Sinauer. 1990, pp. 411–501.
204. Taberlet, P., A. Meyer and J. Bouvet. Unusual mitochondrial DNA polymorphism in two local populations of Blue Tit (*Parus caeruleus*). *Mol. Ecol.* 1: 27–36, 1992.
205. Tajima, F. and M. Nei. Estimation of evolutionary distance between nucleotide sequences. *Mol. Biol. Evol.* 1: 269–285, 1984.
206. Takahata, N. and K. Kimura. A model of evolutionary base substitutions and its application with special reference to rapid change of pseudogenes. *Genetics* 98: 641–657, 1981.
207. Tegelström, H. Transfer of mitochondrial DNA from the northern red-back vole (*Clethrionomys rutilus*) to the bank vole (*C. glareolus*). *J. Mol. Evol.* 24: 218–227, 1987.
208. Thomas, W.K. and A.T. Beckenbach. Variation in salmonid mitochondrial DNA: Evolutionary constraints and mechanisms of substitution. *J. Mol. Evol.* 29: 233–245, 1989.
209. Thomas, W.K. and A.C. Wilson. Mode and tempo of molecular evolution in the nematode *Ceanorhabditis*: Cytochrome oxidase II and calmodulin sequences. *Genetics* 128: 269–279, 1991.
210. Thomas, W.K., R.E. Withler and A.T. Beckenbach. Mitochondrial DNA analysis of Pacific salmonid evolution. *Can. J. Zool.* 64: 1058–1064, 1986.
211. Ticher, A. and D. Graur. Nucleic acid composition, codon usage, and the rate of synonymous substitution in protein-coding genes. *J. Mol. Evol.* 28: 286–298, 1989.
212. Tomkinson, A.E., R.T. Bonk and S. Linn. Mitochondrial endonuclease activities specific for apurinic/apyrimidinic sites in DNA from mouse cells. *J. Biol. Chem.* 263: 12532–12537, 1988.
213. Tomkinson, A.E., R.T. Bonk, J. Kim, N. Bartfeld and S. Linn. 1990. Mammalian mitochondrial endonuclease activities specific for ultraviolet-irradiated DNA. *Nucl. Acids Res.* 18: 929–935, 1990.
214. Topal, M.D. and J.R. Fresco. Complementary base pairing and the origin of substitution mutations. *Nature* 263: 285–289, 1976.
215. Tzagoloff, A. and A.M. Myers. Genetics of mitochondrial biogenesis. *Ann. Rev. Biochem.* 55: 249–285, 1986.
216. Upholt, W.B. and I.B. Dawid. Mapping of mitochondrial DNA of individual sheep and goats: Rapid evolution in the D-loop region. *Cell* 11: 571–583, 1977.

217. Vaughn, K.C., L.R. DeBonte and K.G. Wilson. Organelle alteration as a mechanism for maternal inheritance. *Science* 208: 196–197, 1980.
218. Vawter, L. and W.M. Brown. Nuclear and mitochondrial DNA comparisons reveal extreme rate variation in the molecular clock. *Science* 234: 194–196, 1986.
219. Wallace, D.C. Structure and evolution of organelle genomes. *Microbiol. Rev.* 46: 208–240, 1982.
220. Wallis, G.P. Mitochondrial DNA insertion polymorphism and germ line heteroplasmy in the *Triturus cristatus* complex. *Heredity* 58: 229–238, 1987.
221. Warrior, R. and J. Gall. The mitochondrial DNA of *Hydra attenuata* and *Hydra littoralis* consists of two linear molecules. *Arch. Sc. Geneve* 3: 439–445, 1985.
222. White, T.J., N. Arnheim and H.A. Erlich. The polymerase chain reaction. *Trends Genet.* 5: 185–189, 1989.
223. Wilson, A.C., R.L. Cann, S.M. Carr, M. George, U.B. Gyllensten, K.M. Helm-Bychowski, R.G. Higuchi, S.R. Palumbi, E.M. Prager, R.D. Sage and M. Stoneking. Mitochondrial DNA and two perspectives on evolutionary genetics. *Biol. J. Linn. Soc.* 26: 375–400, 1985.
224. Wilson, G.M., W.K. Thomas and A.T. Beckenbach. Intra- and interspecific mitochondrial DNA sequence divergence in *Salmo*: rainbow, steelhead, and cutthroat trouts. *Can. J. Zool.* 63: 2088–2094, 1985.
225. Wilson, G.M., W.K. Thomas and A.T. Beckenbach. Mitochondrial DNA analysis of Pacific Northwest populations of *Oncorhynchus tshawytscha*. *Can. J. Fish. Aquat. Sci.* 44: 1301–1305, 1987.
226. Wirgin, I.I., R. Proenca and J. Grossfield. Mitochondrial DNA diversity among populations of striped bass in the southeastern United States. *Can. J. Zool.* 67: 891–907, 1989.
227. Woese, C.R. Bacterial evolution. *Microbiol. Rev.* 51: 221–271, 1987.
228. Wolf, K. and L. Del Giudice. The variable mitochondrial genome of ascomycetes: organization, mutational alterations, and expression. *Adv. Genet.* 25: 185–308, 1988.
229. Wong, J.F.H., D.P. Ma, R.K. Wilson and B.A. Roe. DNA sequence of the *Xenopus laevis* mitochondrial heavy and light strand replication origin and flanking tRNA genes. *Nucl. Acids Res.* 11: 4977–4995, 1983.
230. Wolstenholme, D.R. and D.O. Clary. Sequence evolution of *Drosophila* mitochondrial DNA. *Genetics* 109: 725–744, 1985.
231. Yang. D., Y. Oyaizu, H. Oyaizu, G.J. Olson and C.R. Woese. Mitochondrial origins. *Proc. Natl. Acad. Sci. USA* 82: 4443–4447, 1985.
232. Yang, W. and X. Zhou. rRNA genes are located far away from the D-loop region in Peking duck mitochondrial DNA. *Curr. Genet.* 13: 351–355, 1988.
233. Yoneyama, Y. The nucleotide sequences of the heavy and light strand replication origins of the *Rana catesbeiana* mitochondrial genome. *Nippon Ika Daioaku Zasshi* 54: 429–440, 1987 (in Japanese).
234. Zouros, E., K.R. Freeman, A. Oberhauser Ball and G.H. Pogson. Direct evidence for extensive paternal mitochondrial DNA inheritance in the marine mussel *Mytilus*. *Nature* 359: 412–414, 1992.

Note added in proof

Since this chapter was completed (April 1992), two complete fish mt genomes have been published. Huang, F.L. (1992) reports the complete mtDNA sequence of carp, *Cyprinus carpio* (accession no. X61010). The complete mtDNA sequence of *Crossostoma lacustre* (GenBank accession no. M91245) was determined by Tzeng, C.-S., C.-F. Hui, S.-C. Shen and P.C. Huang. 'The complete nucleotide sequence of the *Crossostoma lacustre* mitochondrial genome: conservation and variations among vertebrates.' *Nucl. Acids Res.* 20: 4853–4858, 1992.