Verh. Dtsch. Zool. Ges. 86.2, 131–149 (1993)

Molecular approaches to the phylogenetic study of vertebrates

Molekulare Ansätze zur Verwandtschaftsforschung bei Wirbeltieren

Axel Meyer

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

Summary

Morphology and molecules both should provide congruent information about the evolutionary history of organisms. Molecular data have some advantages for phylogeny reconstruction over morphological data. Here, I am reviewing some of those advantages and potential pitfalls of molecular data for the purpose of molecular systematics. Our understanding of the phylogenetic relationships among vertebrates and the evolution of DNA, particularly mitochondrial DNA, has increased rapidly since the recent invention of the polymerase chain reaction (PCR). This cloning technique together with direct sequencing of PCR-amplified DNA has simplified and dramatically accelerated the accumulation of DNA sequence information for phylogenetic work. Methods of data collection and data analysis for phylogenetic studies on vertebrates with particular emphasis on fish and mitochondrial DNA are outlined. General aspects of the biology of mitochondrial DNA that pertain to phylogenetic reconstruction are reviewed and advantages of DNA sequences over alternative DNA-based genetic markers are highlighted. Examples of phylogenetic work based on mitochondrial and nuclear DNA sequences are used to illustrate the methods, advantages and potential problems with techniques, choice of genes and phylogenetic analyses.

Introduction

Each organism's phenotype and its underlying genotype have experienced the same evolutionary history, hence, both general types of data sets should provide the same estimates of phylogenetic relationships among species (Hillis 1987; Patterson 1987). Data derived from the phenotypes of organisms, which traditionally consist of morphological characters and various kinds of biochemical data reflecting the genotype are expected to share identical evolutionary histories. Molecular data sets are usually easier to obtain than morphological data sets. This is because often only experts of a particular group of organisms are able to identify meaningful morphological characters for a cladistic analysis (see below) which aims to reconstruct the phylogeny of the species under consideration. The creation of molecular data sets does not require such intimate knowledge with the phenotypes of organisms; individual laboratories working on collecting molecular data sets for the purpose of phylogeny reconstruction might therefore often work more readily on a wide phylogenetic range of organisms. Clearly, morphology based approaches have been successfully applied to the reconstruction of phylogenetic relationships among closely as well as distantly related groups of organisms (Patterson 1987). The number of molecular characters that can be found in species is essentially without limits since each species' genome is made up of billions of DNA base pairs each of which potentially contains phylogenetic information. The number of characters that can be identified in the phenotype of organisms is limited by the morphologist's abilities working on the group to identify characters.

Molecular data can have the added advantage over morphological data sets that they can be collected in objective metrics, e.g. DNA sequences of particular genes from several laboratories can be combined and applied to phylogenetic questions that were not intended in the original study. Such a universal metric are e.g. small ribosomal RNA gene sequences that have

been collected for a wide variety of organisms (see below). This potential of some (but not all types of) molecular data sets to be «universal metrics» for the purpose of phylogeny reconstruction is not present in morphological data since each of these data sets must be newly established for every phenotype-based phylogenetic study and are only rarely transferable between studies. Still, one type of data set is not inherently better than another, both exhibit «phylogenetic noise» e.g. homoplasy and provide useful phylogenetic information, the signal-to-noise-ratio is often similar in both kinds of data sets (Hillis 1987; Sanderson and Donoghue 1989; and also behavioral characters: De Queiroz and Wimberger 1993).

Since congruence in the phylogenetic estimates is expected from both kinds of data, it has been argued that the combination of both morphological and molecular data sets should provide «total evidence» (Kluge 1989). There are, however, numerous problems both when both data sets are combined, and when different phylogenetic answers are obtained if these data sets are analyzed separately (reviewed in Swofford 1991; Maddison and Maddison 1992).

Several kinds of biochemical data are typically used to infer phylogenetic relationships among species. Allozyme, immunological and DNA-DNA hybridization data have been widely used but are now increasingly replaced by several types of DNA-based data. Since the advent of the polymerase chain reaction (PCR) in 1985–1986 (Mullis et al 1986; Saiki et al 1985; 1988; Wrishnik et al 1987), our knowledge about DNA and phylogeny of vertebrates has increased dramatically (Fig. 1). The impact of this technological innovation on the understanding of the evolution and the phylogenetic relationships among vertebrates is the focus of this review. My laboratory is mainly concerned with the molecular study of the phylogenetic relationships among fish – the emphasis in this review is going to be on this class of vertebrates. I will attempt to summarize the advantages and pitfalls of currently used DNA-based molecular data to deduce genealogical relationships (see e.g. Hillis and Moritz 1990 for more detail than can be given here). Data collection and analysis will be briefly touched on and recent publications and reviews that provide more detailed information than can be covered in this limited space will be recommended.

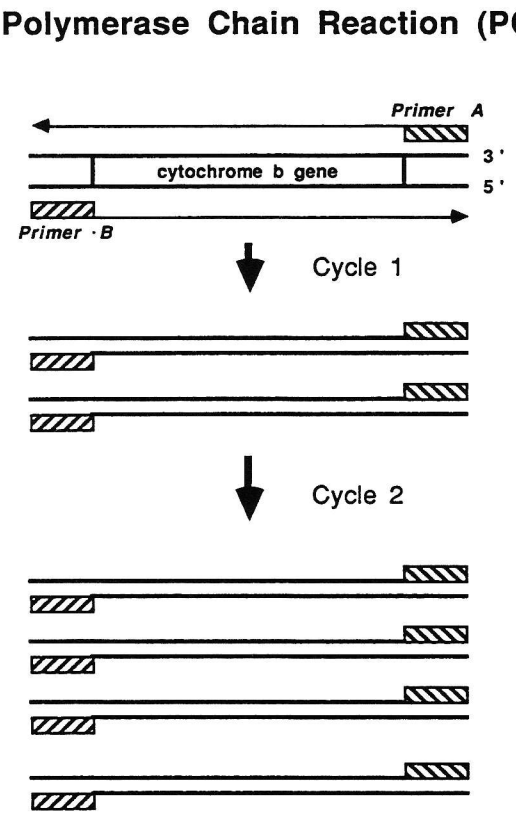

Fig. 1: The principle of the polymerase chain reaction. After Wrishnik et al. 1987. See text for details.

Methods and kinds of DNA data collected for phylogeny reconstruction

Restriction endonuclease methods

DNA (usually mitochondrial DNA, mtDNA) is cut with restriction enzymes and the resulting restriction length polymorphisms (RFLP) are used as binary characters in a phylogenetic analysis (as restriction-fragment data). Often, restriction maps are constructed from the RFLP patterns (after simultaneous digests with two enzymes) and then analyzed in a phylogenetic analysis, coding for absence or presence of particular restriction sites (restriction-site-data). Usually enzymes with six-base recognition sites are used for evolutionary studies among more distantly related species and enzymes with four base pair (bp) recognition sites are used in studies that require more detailed information, e.g. in investigations among more closely related species and population-level questions within species. This method has enjoyed widespread application in the past, but its use has always been limited to rather closely related species since homology between shared bands becomes questionable among more distantly related species (reviews: e.g. Wilson GM et al 1985; Avise 1986; Avise et al 1987; Harrison 1991, but see Wilson et al 1989).

Methodological problems with restriction-fragment data are, that the assumption of independence (see below) is violated, deletions and insertions are problematic and are potential sources of error when trying to establish homology between restriction fragments (for review see e.g. Swofford and Olsen 1990). Therefore, caution should be applied when interpreting RFLP patterns for phylogenetic analysis. Restriction site data, a map derived from RFLP data, also have problems with phylogenetic reconstruction, due to the asymmetry with which restriction sites are gained and lost. The loss of a restriction site is much more likely than a gain, usually RFLP data are therefore analysed using Dollo parsimony (see below) in which a restriction site can be gained only once but lost several times in a phylogenetic tree (Fig. 2). If a particular restriction site is four bp long, and a four bp stretch of DNA is different by only one nucleotide for the recognition sequence only one out of 12 substitutions (substitutions of the same nucleotide [e.g. A for A] at the same site remain undetected) ist going to create this site, i.e. it is unlikely that sites will be gained. However, loosing sites is going to be much more frequent, i.e. any of 12 possible mutations at this restriction site (that will result in a different nucleotide) is going to result in the loss of that site (Templeton 1983 a, b). A special case of parsimony (Dollo parsimony, see below) takes the asymmetry of gains and losses into consideration during the phylogenetic analysis of RFLP data.

An additional disadvantage of restriction data is that results of RFLP and even restriction sites are not immediately transferable between laboratories. That is because the same endonucleases are not always used and the same kinds of gels are not run by all researchers. In this respect, restriction data suffer from the same draw-backs as allozyme data, they tend to be laboratory-specific and often even project-specific results. Although endonuclease data contributed tremendously to our increased understanding of intra- and interspecific genetic variation and phylogenetic relationships, I am biased in favor of actual DNA sequences (see also Wilson et al 1989). For population-level work PCR (see below) and restriction analyses are sometimes combined: known DNA fragments (usually mtDNA) are amplified *via* PCR and then cut with restriction enzymes rather than sequenced (e.g. Orti et al 1993). The advantage of this combined approach being that larger sample sizes can be screened than if every individual is sequenced.

The polymerase chain reaction and direct sequencing

PCR is an enzymatic cloning technique that allows the amplification of any stretch of DNA (within size limits of maximally several thousand base pairs) that is flanked by synthetic oligonucleotide «primers» (Saiki et al 1985; 1988; reviews in e.g. White et al 1989; Arnheim et al 1990; refs. in Erlich 1989; Innis et al 1990) (Fig. 1). The primers are usually around 20 base pairs in length and define the 5' and 3' end of the double-stranded piece of DNA that is going

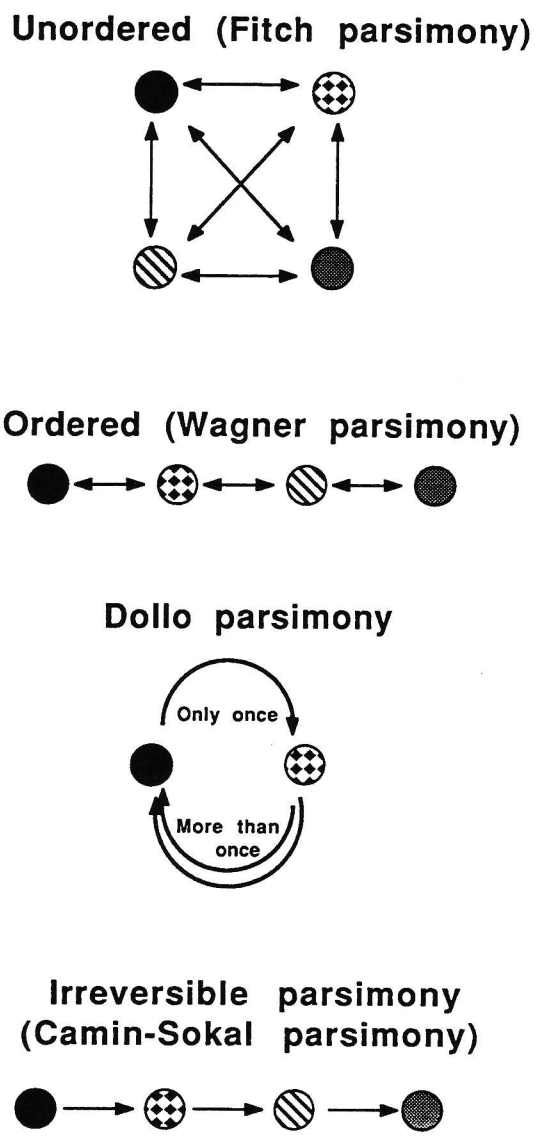

Fig. 2: Different kinds of parsimony. In unordered (or Fitch) parsimony, all character states (e.g. A, G, T or C for DNA sequences) are counted as one step in the analysis. Changes between any two stages involve only one step. Since transversions (see text) often are less likely changes than transitions a step-matrix format is often adopted in the parsimony analysis in which a higher cost is applied for transversion than transitions. In ordered (or Wagner) parsimony, the characters are treated as ordered and the number of steps in the parsimony analysis is counted as the number between their state numbers. This kind of parsimony is often applied to morphological dat sets in which intermediate steps might be assumed between two others. In this example, there is a checkered and striped intermediate between the black and the gray condition. In Dollo parsimony, often applied to RFLP data, a site can only be gained once, but lost several times in the evolutionary history of a lineage. Irreversible (or Camin-Sokal parsimony) applies to characters that are designated as irreversible, multiple gains are allowed, losses are not allowed.

to be amplified. The specificity of the amplification is accomplished through the need for an almost-perfect fit of the primers to the template DNA (Kwok et al 1990). During each cycle of PCR, the number of copies of the DNA-fragment delineated by the primers at either end is doubled. Usually 25–40 cycles are completed in a thermal cycler in about three hours. PCR is much faster and cheaper than conventional cloning techniques. First, a double-stranded PCR product is produced that is then either sequenced (double stranded sequencing, or alternatively «cycle-sequenced»), or subcloned and then sequenced, or cut with restriction enzymes (see above) or used as template DNA for a subsequent asymmetric amplification (Gyllensten and Erlich 1988) or digested with an exonuclease to produce single-stranded DNA for direct sequencing of single-stranded DNA. Sequencing gels of single-stranded DNA often allow one to read more base paris than sequencing gels of double-stranded DNA. Single-

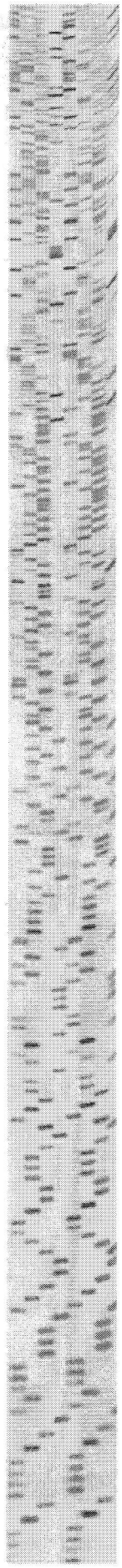

Fig. 3: Example of the quality of direct sequencing of PCR-amplified DNA (from C. Montero). The figure shows partial mitochondrial control region sequences of two individuals of the West African cichlid fish *Astatoreochromis alluaudi*. The loading sequence is C, T, A, G from left to right; the DNA sequence is read from the bottom and determined by the relative position of the bands in the four lanes.

stranded PCR amplified DNA can be as clean as sub-cloned DNA and routinely more than 300–400 bp can be unambiguously determined from a single sequencing reaction (Fig. 3).

The determination of DNA sequences tends to be more time-consuming, costly and technically involved, however DNA sequences of homologous mitochondrial and nuclear genes will allow direct comparisons and study of DNA from different species that have been determined in different laboratories – DNA sequences of the same genes are «universal metrics» that can be transferred between different studies and laboratories. DNA sequences can be stored in data banks (e.g. EMBL, GenBank) and are universally usable, powerful data. The increased costs of DNA sequences compared to RFLP data are far outweighed by their advantage as a universally retrievable, and applicable type of data, since homologous data from independent laboratories can be used in direct comparisons for new studies.

Methods of phylogenetic reconstruction

The history of phylogenetic reconstruction has been turbulent and full of acrimony (reviewed in Hull 1988). There is a large body of literature on phylogenetic reconstruction and several excellent reviews provide an entry to this literature (e.g. Eldredge and Cracraft 1980; Wiley 1981; Felsenstein 1982; 1988; Ax 1984; 1988; Sober 1988; Swofford and Olsen 1990; refs. in Hillis and Moritz 1990). All methods make simplifying assumptions about how DNA sequences evolve (see these reviews for details). No consensus for a single method of phylogenetic reconstruction has been reached among researchers favoring rivaling methodologies. Philosophical as well as practical arguments are used by the proponents of particular methods to argue the superiority of one method over another (see Hull 1988, and below). Simulation studies, intended to determine which method of phylogenetic reconstruction will provide the best estimates have resulted in somewhat conflicting results (e.g. Nei 1991, Kim et al 1993).

Two types of molecular data can be used for phylogenetic analysis. Discrete characters are collected when DNA is sequenced or scored with restriction enzymes. These discrete data provide information about the DNA of a particular individual, often justifiably assumed to be characteristic for the species since most often intraspecific variation will be less than interspecific variation. Further assumptions are the independence and homology of nucleotide positions. If species rather than gene-trees are the purpose of the study (reviewed in Pamilo and Nei 1988), only orthologous rather than paralogous genes should be compared (Patterson 1988). Discrete data can be transformed into similarity or distance data, e.g. by pairwise comparisons of two sequences.

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

The parsimony method (e.g. reviewed in Sober 1988), «the method of minimum net evolution», aims to find the evolutionary tree that requires the fewest changes of nucleotides to explain the evolution of the DNA sequences under consideration. It was pioneered by the German entomologist Willi Hennig. Its philosophy is the hypothetico-deductive approach in which Ockham's principle is invoked, i.e. evolution is believed to proceed by the shortest, simplest pathway. Parsimony only considers so-called «phylogenetically informative sites» in the calculation of the topology of the tree. Sites that do not require different numbers of changes on alternative trees of different topology (e.g. sites that are identical or sites that differ only in one of the species under consideration) are ignored. Different kinds of evolutionary scenarios can be assumed for different kinds of characters and various kinds of parsimony methods are therefore applied to these data sets (see review e.g. Maddison and Maddison 1992) (Fig. 2). For example, Dollo parsimony is applied to RFLP data, most DNA sequence data

are analyzed using Fitch parsimony (Fig. 2) with the underlying assumption that all character state changes among the four possible nucleotides are equally likely. Both distance and parsimony methods only look at part of the information in the data.

Only the maximum-likelihood method attempts to use all the information contained in DNA sequences (Felsenstein 1981) by using statistical criteria to distinguish between alternative trees. It uses a model of likelihoods of substitution changes and attempts to fit the data with a tree. The likelihood of the topology of a tree is the probability of the data given the tree and the model. The maximum-likelihood method chooses a tree with a topology and branch length that has the highest likelihood (reviewed in Swofford and Olson 1990). This method allows for unequal base composition, uneven transition-transversion rates and does not require a molecular clock. The major practical drawback of the maximum-likelihood method is its inherent computational complexity due to exact probability models of sequence change, which in practice limits the number of sequences that can be analyzed in a reasonable amount of time with available computer power.

Unfortunately, studies that test the power of various commonly used methods of phylogenetic reconstruction on known phylogenies are rare; these tests have been weak and did not aid in the decision which method of phylogeny reconstruction best predicts the true phylogeny (Atchley and Fitch 1991; Hillis et al 1992). These studies have been too simple, (all methods extimated the correct phylogeny!) and they have failed to clearly identify a particular methodology as the best. All currently used methods predicted the correct phylogeny, however, parsimony was the best in estimating branch lengths (Hillis et al 1992). However, the strengths and weaknesses of the alternative methods of phylogenetic reconstruction and the kinds of genes to be used and to be avoided have been made more clear in these studies (Felsenstein 1988; Swofford and Olsen 1990; Nei 1991; Atchley and Fitch 1991; Hillis et al 1992).

All methods of phylogeny reconstruction can be led astray and give incorrect estimates of phylogenies. All methods are susceptible to varying degrees to rate variation in the «molecular clock» (reviewed e.g. in Zuckerkandl 1987) in different lineages. Felsenstein (1978) demonstrated that parsimony methods will give an incorrect result if the variation in rates in different lineages under consideration is very significantly different. Two faster evolving lineages will be attracted to each other simply because their faster rate of evolution is going to bring about chance matches that are interpreted as synapomorphies (shared derived characters that are interpreted to indicate common descent). This tendency to produce wrong phylogenetic estimates is enhanced rather than decreased with larger data sets «Felsenstein's paradox» (1978) (Fig. 4). UPGMA (unweighted pair group method using arithmetic averages) however, is the most unreliable method when rates of molecular evolution vary (Swofford and Olson 1990, Nei 1991).

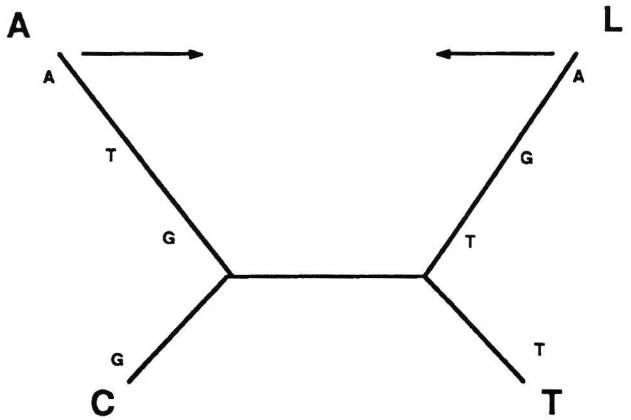

Fig. 4: Felsenstein's paradox (Felsenstein 1978) illustrating how parsimony can be misled by extreme rate variation between lineages included in a phylogenetic analysis. The lineages A and L will be attracted to each add space other (signified by the arrows) because their accelerated rate of DNA substitution will by chance result in equal character states (in this case both having an A) which would be interpreted to be phylogenetically informative. Collecting more sequence information will, paradoxically, exasperate, rather than deminish, the problem and lead to an incorrect phylogenetic estimate.

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

In practice, the data should be subjected to several means of phylogeny reconstruction and differences in results will usually pinpoint areas of weakness in the phylogenetic tree; these branches will also tend to have low bootstrap values (Felsenstein 1985) or be non-significant based on other methods of confidence estimation. Congruent results of different phylogenetic methods will inspire confidence that a phylogenetic estimate has been found that is closer to the true relationships. The identification of «weak branches» through conflicting estimates based on different phylogenetic methods should be pointed out and investigated further by using more or different data. The robustness of a molecular phylogeny is also often judged by whether or not it is congruent with a «well-established» phylogeny based on morphological data (reviewed in Patterson 1987; Swofford 1991); this asymmetry of judgment might change as molecular phylogenetic estimates become more common and available for comparison. Obviously, since the molecules are just a part of the whole organism they are expected to have experienced the same evolutionary history, and therefore report the same evolutionary information as other parts (e.g. morphology) of the species.

The mitochondrial genome

Because of the wide-spread popularity of mitochondrial DNA (mtDNA) for population and phylogenetic work I will first review some of the basic biology of mtDNA. A more detailed review on the evolution of mitochondrial DNA with emphasis on fish is available (Meyer 1993a and references therein).

The mitochondrial (mt) genome of vertebrates is a single, small, double-stranded, circular DNA molecule contained in mitochondria, and up to several thousand copies of the mitochondrial genomes are found per cell. Typically, the size of animal mt genomes is about 16,500 ± 500 basepairs (reviewed by Brown 1981; 1983; 1985; Clark-Walker 1985; Moritz et al 1987) (Fig. 5).

The complete mitochondrial genome has been sequenced or the gene order determined in several invertebrates and vertebrates. Among the vertebrates, complete sequences for human (Anderson et al 1981), cow (Anderson et al 1982), mouse (Bibb et al 1981), rat (Gadaleta et al 1989), fin whale (Arnason et al 1991), harbor seal (Arnason and Johnsson 1992), chicken (Desjardins and Morais 1990) and frog, *Xenopus* (Roe et al 1985) have been published. Additionally, many other partial mitochondrial sequences, too numerous to be listed here, are known.

The mitochondrial gene order of animals is different in every phylum that has been studied. Even within phyla (e.g. echinoderms) differences in gene order exist (Smith et al 1989; 1990; Himeno et al 1990). Mitochondrial gene orders differ slightly among vertebrates (Yoneyama 1987; Desjardins and Morais 1990; 1991; Pääbo et al 1991; reviewed in von Haeseler et al 1993). MtDNA gene order has been used for phylogenetic inferences among classes of echinoderms (Smith et al 1993).

The mitochondrial genome of vertebrates contains 13 genes coding for proteins, two genes coding for ribosomal RNAs (small 12S and large 16S rRNA), 22 genes coding for transfer RNAs (tRNAs) and one major non-coding region (control region) that contains the initiation sites for mtDNA replication and RNA transcription (Fig. 5). The protein-coding genes are seven subunits of NADH dehydrogenase (ND 1, 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) (Fig. 5).

Animal mtDNA is haploid and non-recombining (Hayashi et al 1985; but Horak et al 1974; Olivo et al 1983; see refs. in Hurst 1991) and appears to be almost exclusively maternally inherited. Paternal mitochondria appear to be actively degraded during fertilization or «outre-

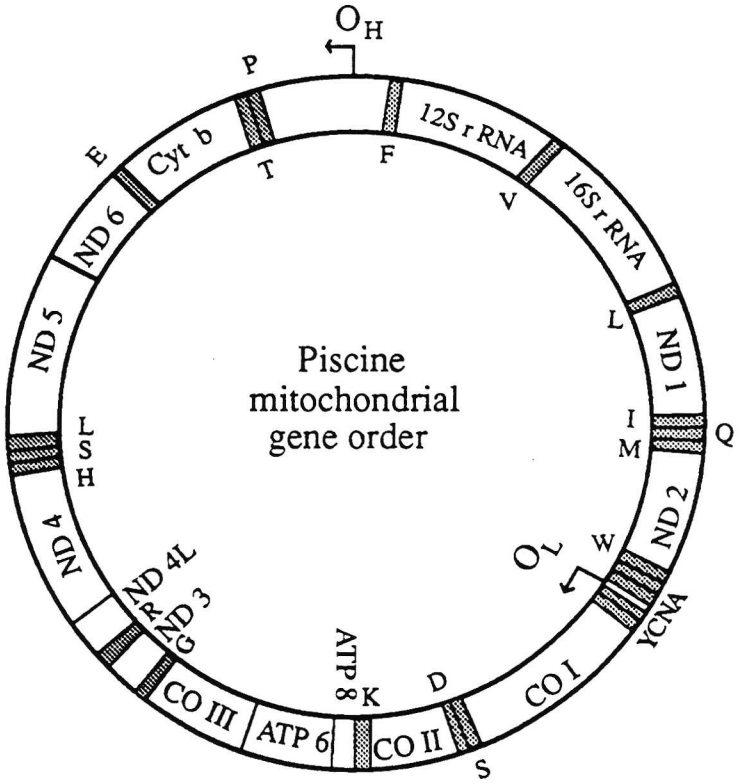

Fig. 5: Vertebrate (here fish as example) mitochondrial gene order. The origins of H- and L-strand replication are indicated in the figure. Transfer RNA genes are shown in shaded boxes. The complete names for the abbreviated names of proteins are given in the text (see also Meyer 1993 a). The tRNA genes coded for by the L-strand are labeled on the outside of the circle, the tRNA genes coded for by the H-strand are labeled on the inside.

plicated» shortly thereafter (Meland et al 1991). Usually, 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 Moritz et al 1987; Moritz 1991). Heteroplasmy has occasionally been found in most major groups of organisms.

In contrast to the nuclear genome, the mitochondrial genome of animals (but not that of fungi and plants) is highly efficient (reviewed in Attardi 1985; Cantatore and Saccone 1987; Gray, 1989), 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) (reviewed in Meyer 1993 a, 1993 b).

Several studies of mitochondrial DNA using restriction enzymes and actual sequences (Brown et al 1982) indicate that mtDNA generally evolves at elevated rates (5–10 times faster) compared to single copy nuclear genes (Brown et al 1979; Perler et al 1980). The faster mtDNA evolution is due to a higher frequency of point and length mutations (Brown et al 1982; Cann et al 1984; Wilson et al 1985). 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 of primates may be 0.5–1% per million years. The rate of silent substitutions (substitutions that do not result in amino acid changes), mainly transitions, is about 4–6 times that of replacement substitutions (Brown et al 1979, 1982). However, among closely related species, if transitions are ignored and only transversions counted, the difference in rates between mitochondrial and nuclear genes would be much less pronounced. Several reasons for the higher rate of mtDNA evolution have been suggested (for references see Meyer 1993 a). The observed differences in rates of evolution between nuclear and mtDNA cannot

be explained exclusively by the increased variation at third positions. Substitution rates at first and second positions and rRNA and tRNA genes are also increased. However, substitution rates at these sites, may not be as different between mitochondrial and nuclear genes as rates at third positions.

Of the three kinds of sequence changes, substitutions are more frequent than indels (additions or deletions), rearrangements are the least common form of DNA change. Additions and deletions are most frequently observed in the control region and intergenic spacers. At a lower frequency, they occur in the tRNAs and rRNAs. Rarely have they been found in protein coding genes.

As sequencing techniques developed it became clear that the basis for the higher rate of mtDNA evolution is due mainly to transitional differences (changes from one purine to another or one pyrimidine to another) (Brown and Simpson 1982; Brown et al 1982; Aquadro and Greenberg 1983; Greenberg et al 1983; Wolstenholme and Clary 1985). Transitions often outnumber transversions by a factor of 10 or 20 in within-species comparisons. Substitutions at third positions quickly accumulate and they become saturated with transitions. Mutations at first and second positions will continue to accumulate despite saturated third positions. For phylogenetic reconstruction purposes (see below) transitions at third positions will therefore not be reliable indicators of evolutionary relationship beyond about 30–40% sequence divergence at these positions (reviewed in Meyer 1993 a, b). 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.

The increase in observed number of transversions with increasing sequence difference is due to several factors (Brown et al 1982; Holmquist 1983; DeSalle et al 1987; Jukes 1987; Meyer 1993 b). Asymmetry in the persistence of transversions and transitions exists, (DeSalle et al 1987) and will produce an observed accumulation of scored transversions with time. Although transitions remain a more common type of mutation event between distantly related species, the percentage of all differences which are scored as transitions will decrease. The observed number of transversions will increase because transversions will become fixed at sites that had already fixed transitions and also at new sites that had not been substituted before.

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

The level at which saturation is reached will depend on several factors e.g. the base compositional bias. The absolute time required to reach the plateau is dependent on the absolute rate with which transversions become fixed. Among congeners, total sequence divergence tends not to exceed about five percent, and scored transversions are usually less than 25% of the total number of substitutions, whereas closely related species exhibit transition differences almost exclusively.

Protein-coding genes

Substitution patterns in protein-coding genes follow some relatively well-understood rules. This regularity in the way in which mutations accumulate make protein-coding genes attractive candidates for phylogenetic studies among vertebrates. 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-synonymus 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. In some codons (leucine), the first posi-

tion 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 constrained second positions of codons. Knowledge of these rules allows one to weight (or exclude, as an extreme form of weighting) kinds of substitutions (transitions and transversions) and positions (first, second and third) differently based on the phylogenetic question that is being addressed. Obviously, among distantly related species transitions in third positions are going to be unreliable tracers of evolutionary descent and represent largely «phylogenetic noise». Transversions are rarer and have a higher chance of being reliable indicators of descent. The most conserved positions (second positions in codons) are going to be the most similar among closely related species but contain phylogenetic information among distantly related species.

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

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. Nuclear rRNA (18S and 28S) and tRNA genes (see below) evolve about 100 times more slowly than their mitochondrial counterparts (Dawid 1972) when the more variable expansion segments of the nuclear rRNA genes are excluded from the analysis (reviewed in Mindell and Honeycutt 1990). Different regions and categories of ribosomal genes have different rates and modes of DNA evolution that need to be considered for phylogenetic work (Vawter and Brown 1993).

The same basic rules of substitutions (see above) apply to rRNA genes (Brown et al 1982). Transitions are more frequent than transversions, this is most apparent among closely related species where the record of transitions has not been overlaid by more slowly accumulating and more persistent transversions (see above). Insertions and deletions are usually small, in the range of 1–5 bp. The 16S gene tends to contain more length variation than the 12S gene. Hence, DNA sequences of the 16S rRNA for distantly related species are more difficult to align than for the 12S rRNA gene. Length mutations are more frequent in rRNA than in protein-coding genes. These length variations make alignment an issue for phylogenetic studies; different alignments, therefore different hypotheses of homology, can result in drastically different phylogenetic inferences. Areas of questionably (criteria for what is well aligned are not always obvious see Swofford and Olsen 1991; Mindell 1991) aligned sequence are best removed from the phylogenetic analysis. Alignment and delineation of which portion of the sequence should be excluded from the analysis and which should be used remain problematic issues.

Secondary structure models are available for both genes, and general agreement about the secondary structure exists (Glotz et al 1981; Dunon-Bluteau and Brun 1986; Hixon and Brown 1986). These secondary structures are conserved across large evolutionary distances.

Among tetrapods, lungfish, the coelacanth, and ray-finned fish, substitutions occur about four times more frequently in proposed loops than in stems; transversions are about nine times more frequent in loops than in stems (Meyer and Wilson 1990; Meyer and Dolven 1992; Meyer, unpublished data). Transversions appear to be indicators of phylogenetic relationships in a slowly evolving portion of the 12S rRNA gene for an evolutionary distance of more than 400 million years (Meyer and Wilson 1990; Mindell and Honeycutt 1990; Meyer and Dolven 1992; see Meyer 1993 a). Among closely related species of primates stems are less likely to contain substitutions than loop-regions (Hixon and Brown 1986). Stems sometimes show «compensatory mutations»: a substitution in one strand of a stem region is compensated for by a change in the other strand of the stem, in order to maintain intra-strand base pairing and a stable secondary structure. The rate of sequence divergene appears higher among closely related species than among distantly related species due to saturation effects. At around 100 to 150 million years of sequence divergence, the rate of substitution appears to decrease (Mindell and Honeycutt 1990; Meyer and Dolven 1992).

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

Control region

The control region, is partially constrained in primary sequence or secondary structure to regulate replication and transcription (Clayton 1991). Of all mitochondrial regions the control region has the highest substitution rate. The rate of evolution of the control region is two to five times higher than those of most mitochondrial protein coding genes (Aquadro and Greenberg 1983). Nucleotide substitutions occur five times more frequently than additions and deletions in the human control region; among closely related species of cichlid fish substitutions outnumber deletions and additions (Meyer et al 1990; Sturmbauer and Meyer 1992). The control region varies tremendously in length, often because of tandem duplications of 200 bp to 4100 bp (Brown 1985), 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 (Densmore et al 1985, Harrison et al 1985).

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

Mitochondrial DNA and phylogenetic analyses

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

reflection of the pre-speciation levels, and furthermore assumes that the rates of mtDNA evolution are the same in all lineages.

MtDNA has found widespread use as a tool for phylogenetic analyses (several reviews are available e.g. Kornfield 1984, 1991; Avise 1986; Avise et al 1987; Moritz et al 1987; Wilson et al 1985). Through the advent of the polymerase chain reaction it has become apparent that mitochondrial DNA sequences may also be a useful tool for the study of evolutionary relationship among more distantly related species (e.g. Meyer and Wilson 1990; Normark et al 1991; Meyer and Dolven 1992).

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

Universal PCR primers for mtDNA (Kocher et al 1989) led to an even further increase in the use of mtDNA for phylogenetic questions. However, the availability of universal PCR primers has led to their sometimes 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 and only contain «phylogenetic noise» rather than information. The 5' end of cytochrome *b* tends to be very conserved in amino acid substitutions and it may not contain enough variation in first and second positions of codons once third positions are saturated. The 3' end of this gene appears to be somewhat less constrained in terms of amino acid substitutions (Irwin et al 1991, reviewed in Esposti et al 1993). Other mitochondrial or nuclear genes may have substitution rates more appropriate for particular questions. On a per base pair basis the control region will provide more information about population level questions and contain more (up to three times more) phylogenetic information than cytochrome *b* (e.g. Meyer et al 1990; Sturmbauer and Meyer 1992; Meyer 1993 a).

Base compositional biases, saturation effects and the low variation in first and second positions of codons all contribute to problems in phylogenetic reconstruction. Similar difficulties in the use of cytochrome *b* were encountered in the reconstruction of the evolutionary relationships among groups of cichlid fish, believed to have diverged more than 80 million years ago (Meyer, unpublished data). Although this gene has been very useful for studies involving populations (see above) and the phylogenetic study of very different groups of animals it may not be the most appropriate gene for all purposes (reviewed in Esposti et al 1993). It would appear that for some phylogenetic problems faster evolving protein-coding genes like ND2, and ATPase 6 and 8 or the ribosomal genes might provide more appropriate data. Primer sequences for many other genes are available (Meyer 1993 b). More «universal» mitochondrial, but also nuclear DNA PCR primers are constantly being published (Palumbi et al 1991; see review in Meyer 1993 b). With these primers, mtDNA can be amplified and sequenced without prior sequence knowledge and they are expected to work for most groups of teleosts.

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

Recently, much progress has been made in many old questions of vertebrate phylogeny (re-

viewed Novacek 1992; Graur 1993). For example, the origin and phylogeny of the mammalian radiation has remained hotly contested and both morphological and molecular analyses often (but not always) result in the same phylogenetic estimate of relationship among orders of mammals (reviewed Novacek 1992; Graur 1993). Sometimes, surprising results are obtained by molecular phylogenetic analyses – e.g. the finding that guinea pigs may not be closely related to rodents (reviewed in Graur 1993). Recently, we hypothesized, based on partial mitochondrial ribosomal sequences that the tooth whales are not a monophyletic group and that baleen whales are the sister group of sperm whales (Fig. 6) (Milinkovitch et al 1993).

The usefulness of mtDNA sequences is determined at one extreme by the stochasticity of the distribution of haplotype polymorphisms within species and at the other extreme by saturation effects due to base compositional biases, mutational bias and selective constraints on the gene product.

Nuclear genes in molecular systematics

The large existing data base and the ease of direct sequencing techniques of nuclear ribosomal genes has lead to their widespread, successful use in molecular systematics (reviewed in Mindell and Honeycutt 1990; Hillis and Dixon 1991). Small (18S) and large (28S) ribosomal genes are arranged in tandem arrays with hundreds or thousands of copies in each genome (Fig. 7). Nuclear ribosomal genes have been widely used for phylogenetic work. Because of the extremely slow rate of evolution of these genes they have mainly been used for phylogenetic work on distantly related species (e.g. origin of tetrapods, Hillis and Dixon 1989; Hillis et al 1989; Meyer, unpublished data; the monophyly of agnathan fish, Stock and Whitt 1992;

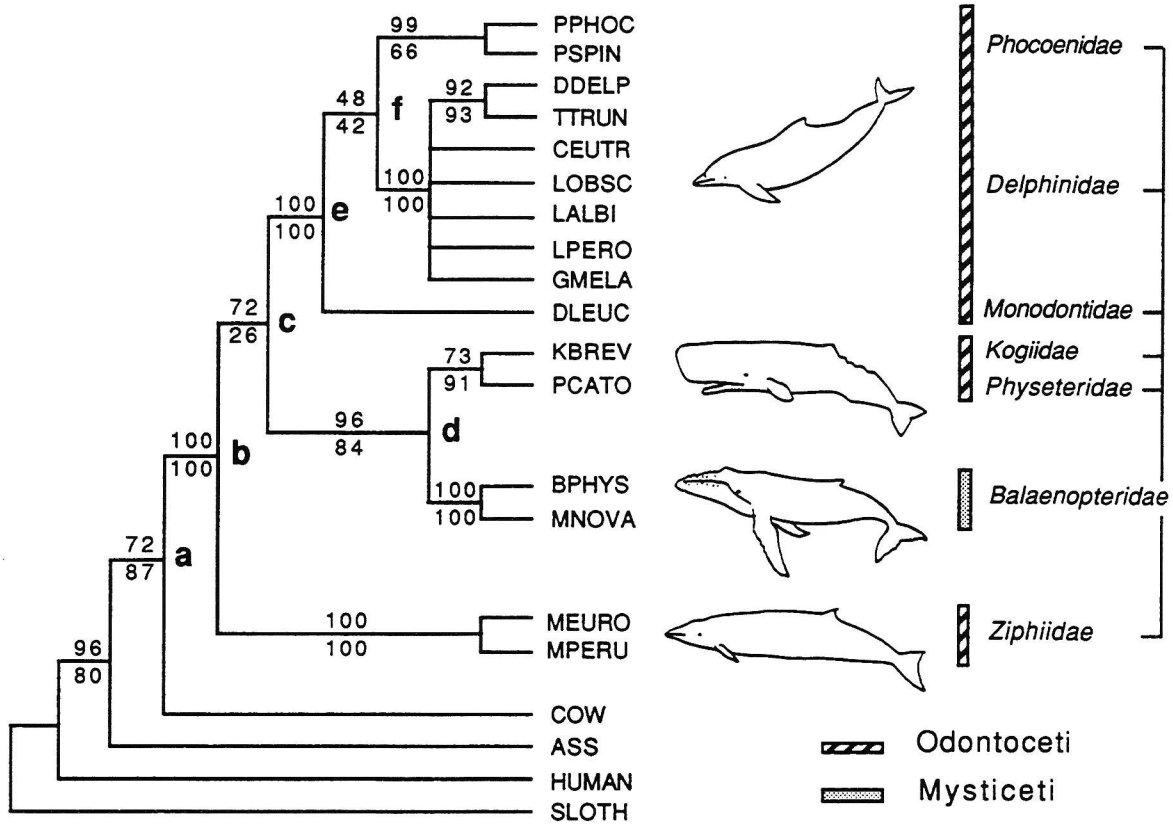

Fig. 6: Phylogenetic hypothesis of the mammalian order Cetacea based on mitochondrial ribosomal sequences (Milinkovitch et al 1993).

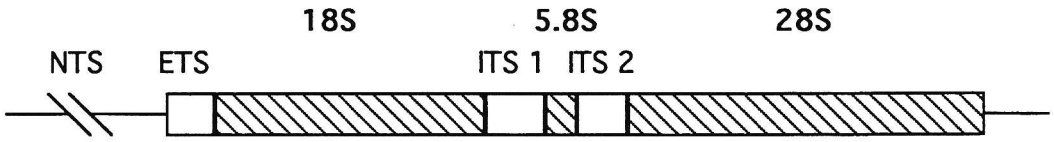

Fig. 7: Arrangements of the tandemly arranged elements of the nuclear ribosomal genes.

Meyer, unpublished data). Many PCR-primers are available (e.g. Hillis and Dixon 1991; Meyer 1993 b) that will ensure the continued use of these genes for phylogenetic work. However, some particular problems, e.g. alignment (see above) plague the use of these genes for phylogenetic work.

Many other nuclear genes (e.g. globins) have been used with varying success for phylogenetic work on vertebrates. This literature is vast and due to space limitations cannot be covered here. DNA sequences are not a panacea for all phylogenetic questions, however, much progress has been made and will continue to be made through the «democratization of the genetic code» by allowing easier access to DNA sequence information through PCR-technology. Molecular approaches should not be viewed as competition to morphological phylogenetic approaches; both aim to elucidate the same evolutionary history. Investigators using either approach need to communicate to increase each others understanding and appreciation for both approaches, each as valid and important as the other.

Acknowledgements

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

References

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Epron IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290: 457–65

Anderson S, de Bruijn MHL, Coulson AR, Epron IC, Sanger F, Young IG (1982) Complete sequence of bovine mitochondrial DNA. J Mol Biol 156: 683–717

Aquadro CF, Greenberg BD (1983) Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. Genetics 103: 287–312

Arnason U, Gullberg A, Widegren B (1991) The complete nucleotide sequence of the mitochondrial DNA of the fin whale, *Balaenoptera physalus*. J Mol Evol 33: 556–568

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

Arnheim N, White T, Rainey WE (1990) Application of PCR: organismal and population biology. BioScience 40: 174–182

Atchley W, Fitch WM (1991) Gene trees and the origins of inbred strains of mice. Science 254: 554–558

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

Avise JC (1986) Mitochondrial DNA and the evolutionary genetics of higher animals. Phil Trans Roy Soc Lond B 312: 325–342

Avise JC (1989) Gene trees and organismal histories: a phylogenetic approach to population biology. Evolution 43: 1192–1208

Avise JC, Helfman GS, Neigel JE, Arnold J (1984) Demographic influences on mitochondrial DNA lineage survivorship in animal populations. J Mol Evol 20: 99–105

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC (1987) Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. Ann Rev Ecol Syst 18: 489–522

Avise JC, Ankney CD, Nelson WS (1990) Mitochondrial gene trees and the evolutionary relationship of mallard and black ducks. Evolution 44: 1109–1119

Ax P (1984) Das Phylogenetische System. Gustav Fischer, Stuttgart

Ax P (1988) Systematik in der Biologie. Gustav Fischer, Stuttgart

Bibb MJ, Van Etten RA, Wright CT, Walber MW, Clayton DA (1981) Sequence and gene organization of mouse mitochondrial DNA. Cell 26: 167–180
Brown GG, Simpson MV (1982) Novel features of animals mtDNA evolution as shown by sequences of two rat cytochrome oxidase subunit II genes. Proc Nat Acad Sci USA 79: 3246–3250
Brown WM (1981) Mechanisms of evolution in animal mitochondrial DNA. Ann N.Y. Acad Sci 361: 119–134
Brown WM (1983) Evolution of mitochondrial DNA. In: Nei N, Koehn RK (eds) Evolution of genes and proteins. Sinauer, Sunderland, 62–88
Brown WM (1985) The mitochondrial genome of animals. In: MacIntyre RJ (ed) Molecular evolutionary genetics. Plenum Press, New York, 95–130
Brown WM, George M Jr., Wilson AC (1979) Rapid evolution of mitochondrial DNA. Proc Natl Acad Sci USA 76: 1967–1971
Brown WM, Prager EM, Wang A, Wilson AC (1982) Mitochondrial DNA sequences of primates: Tempo and mode of evolution. J Mol Evol 18: 225–239
Cann RL, Brown WM, Wilson AC (1984) Polymorphic sites and the mechanism of evolution in human mitochondrial DNA. Genetics 106: 479–499
Cantatore P, Saccone C (1987) Organization, structure, and evolution of mammalian mitochondrial genes. Int Rev Cytol 108: 149–208
Clark-Walker GD (1985) Basis of diversity in mitochondrial DNAs. In: Cavalier-Smith T (ed) The evolution of genome size, Wiley, New York, 277–297
Clayton DA (1991) Replication and transcription of vertebrate mitochondrial DNA. Ann Rev Cell Biol 7: 453–478
Dawid IB (1972) Evolution of mitochondrial DNA sequences in *Xenopus*. Dev Biol 29: 139–151
Densmore LD, Wright JW, Brown WM (1985) Length variation and heteroplasmy are frequent in mitochondrial DNA from parthenogenetic and bisexual lizards (genus *Cnemidophorus*). Genetics 110: 689–707
DeQueiroz A, Wimberger P (1993) The usefulness of behavior for phylogeny estimation: Levels of homoplasmy in behavioral and morphological characters. Evolution 47: 46–60
DeSalle R, Freedman T, Prager EM, Wilson AC (1987) Tempo and mode of sequence evolution in mitochondrial DNA of Hawaiian *Drosophila*. J Mol Evol 26: 157–164
Desjardins P, Morais R (1990) Sequence and gene organization of the chicken mitochondrial genome. J Mol Biol 212: 599–634
Desjardins P, Morais R (1991) Nucleotide sequence and evolution of coding and noncoding regions of a quail mitochondrial genome. J Mol Evol 32: 153–161
Dunon-Bluteau D, Brun G (1986) The secondary structure of the *Xenopus laevis* and human mitochondria small ribosomal subunit RNA are similar. FEBS 198: 333–338
Eldredge N, Cracraft J (1980) Phylogenetic patterns and the evolutionary process. Columbia University Press, New York
Erlich H A (ed) (1989) *PCR technology:* Principle and applications for DNA amplification. Stockton Press, New York
Esposti Degli M, De Vries S, Crimi M, Ghelli A, Patarnello T, Meyer A (1993) Mitochondrial cytochrome *b*: Evolution and structure of the protein. Acta Biophys Biochem Bioenergetics 1143: 243–271
Felsenstein J (1978) Cases in which parsimony and compatibility methods will be positively misleading. Syst Zool 27: 27–33
Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol17: 368–376
Felsenstein J (1982) Numerical methods for inferring evolutionary trees. Quart Rev Biol 57: 379–404
Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39: 783–791
Felsenstein J (1988) Phylogenies from molecular sequences: inference and reliability. Ann Rev Genet 22: 521–565
Gadaleta G, Pepe G, DeCandia G, Quagliariello C, Sbisa E, Saccone C (1989) The complete nucleotide sequence of the *Rattus norvegicus* mitochondrial genome: cryptic signals revealed by comparative analysis between vertebrates. J Mol Evol 28: 497–516
Glotz C, Zwieb C, Brimacombe R, Edwards K, Kossel H (1981) Secondary structure of the large subunit ribosomal RNA from *Echerichia coli, Zea mays* chloroplast, and human and mouse mitochondrial ribosomes. Nucl Acid Res 9: 3287–3306
Graur D (1993) Molecular phylogeny and the higher classification of eutherian mammals. Trends Ecol Evol 8: 141–147
Gray MW (1989) Origin and evolution of mitochondrial DNA. Ann Rev Cell Biol 5: 25–50
Greenberg BD, Newbold JE, Sugino A (1983) Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. Gene 21: 33–49
Gyllensten UB, Ehrlich HA (1988) Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQa locus. Proc Natl Acad Sci USA 85: 7652–7655
Haeseler A v, Janke A, Pääbo S (1993) Molecular phylogenetics. Verh Dtsch Zool Ges 86.2: 119–129
Harrison RG (1991) Molecular changes at speciation. Ann Rev Ecol Syst 22: 281–308

Harrison RG, Rand DM, Wheeler WC (1985) Mitochondrial DNA size variation within individual crickets. Science 228: 1446–1448
Hayashi J-I, Tagashira Y, Yoshida MC (1985) Absence of extensive recombination between inter- and intraspecies mitochondrial DNA in mammalian cells. Exp Cell Res 160: 387–395
Hillis DM (1987) Molecular versus morphological approaches to systematics. Ann Rev Ecol Syst 18: 23–42
Hillis DM, Bull JJ, White ME, Badgett MR, Molineux IJ (1992) Experimental phylogenetics: generation of a known phylogeny. Science 255: 589–592
Hillis DM, Dixon MT (1989) Vertebrate phylogeny: evidence from 28S ribosomal DNA sequences. In: Miyamoto MM, Cracraft J (eds) Phylogenetic analysis of DNA sequences. Oxford Univ. Press, New York, 355–367
Hillis DM, Dixon MT, Ammerman LK (1989) The relationships of the coelacanth *Latimeria chalumnae:* evidence from sequences of vertebrate 28S ribosomal genes: In: Musick JA, Bruton MN, Balon EK (eds) The biology of *Latimeria chalumnae* and evolution of coelacanths. Kluwer Academic Publishers, Dordrecht, 117–130
Hillis DM, Moritz C (eds) (1990) Molecular systematics. Sinauer, Sunderland
Hillis DM, Dixon MT (1991) Ribosomal DNA: Molecular evolution and phylogenetic inference. Quart Rev Biol 66: 411–453
Himeno H, Masaki H, Kawai T, Ohta T, Kumagai I, Miura I, Watanabe K (1987) Unusual genetic codes and a novel gene structure for tRNA-AGY-Ser in starfish mitochondrial DNA. Gene 56: 219–230
Hixon JE, Brown WM (1986) A comparison of the small ribosomal RNA genes from the mitochondrial DNA of the great apes and humans: sequence, structure, evolution and phylogenetic implications. Mol Biol Evol 3: 1–18
Holmquist R (1983) Transitions and transversion in evolutionary descent: an approach to understanding. J Mol Evol 19: 134–144
Horak I, Coon HG, Dawid IB (1974) Interspecific recombination of mitochondrial DNA molecules in hybrid somatic cells. Proc Natl Acad Sci USA 71: 1828–1832
Hull DL (1988) Science as a process. Chicago University Press, Chicago
Hurst LD (1991) Sex, slime and selfish genes. Nature 354: 23–24
Innis MA, Gelfand DH, Sninsky JJ, White TJ (1990) PCR Protocols: a guide to methods and applications. Academic Press, San Diego
Irwin DM, Kocher TD, Wilson AC (1991) Evolution of the cytochrome *b* of mammals. J Mol Evol 32: 128–144
Jacobs HT, Elliott DJ, Math VB, Farharson A (1988) Nucleotide sequence and gene organization of sea urchin mitochondrial DNA. J Mol Biol 202: 185–217
Johansen S, Guddal PH, Johansen T (1990) Organization of the mitochondrial genome of Atlantic cod, *Gadus morhua.* Nucl Acids Res 18: 411–419
Jukes TH (1987) Transitions, transversions, and the molecular evolutionary clock. J Mol Evol 26: 87–98
Kim J, Rohlf FJ, Sokal RR (1993) The accuracy of phylogenetic estimation using the neighbor-joining method. Evolution 47: 471–486
Kluge AG (1989) A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). Syst Zool 38: 7–25
Kocher TD, Thomas WK, Meyer A, Edwards SV, Pääbo S, Villablanca FX, Wilson AC (1989) Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc Natl Acad Sci USA 86: 6196–6200
Kornfield I (1984) Descriptive genetics of cichlid fish. In: Turner Bj (ed) Evolutionary genetics of fishes. Plenum Press, New York, 591–615
Kornfield I (1991) Genetics. In: Cichlid fish: behaviour, ecology and evolution. Keenleyside MHA (ed), Chapman Hall, New York, 103–128
Kwok S, Kellogg DE, McKinney N, Spasic D, Goda L, Levenson C, Sninsky JJ (1990) Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. Nucl Acid Res 18: 999–1005
Li W.-H, Gouy M (1991) Statistical methods for testing molecular phylogenies. In: Miyamoto MM, Cracraft J (eds) Phylogenetic analysis of DNA sequences. Oxford Univ. Press, New York, 249–277
Maddison WP, Maddison DR (1992) MacClade: analysis of phylogeny and character evolution. Version 3.0. Sinauer, Sunderland, Massachusetts
Meland S, Johansen S, Johansen T, Haugli K, Haugli F (1991) Rapid disappearance of one parental mitochondrial genotype after isogamous mating in the myxomycete *Physarum polycephalum.* Curr Genet 19: 55–60
Meyer A (1993 a) Evolution of mitochondrial DNA in fish. In: Hochachka PW, Mommsen P (eds) Biochemistry and molecular biology of fish Vol. 2, Elsevier Press, New York, 1–38
Meyer A (1993 b) Molecular phylogenetic studies of fish. In: Beaumont A (ed) Genetics and evolution of aquatic organisms. Chapman and Hall, London (in press)
Meyer A, Kocher TD, Basasibwaki P, Wilson AC (1990) Monophyletic origin of Lake Victoria cichlid fish suggested by mitochondrial DNA sequences. Nature 347: 550–553
Meyer A, Wilson AC (1990) Origin of tetrapods inferred from their mitochondrial DNA affiliation to lungfish. J Mol Evol 31: 359–364

Meyer A, Dolven SI (1992) Molecules, fossils and the origin of tetrapods. J Mol Evol 35: 102–113
Milinkovitch MC, Orti G, Meyer A (1993) Revised phylogeny of whales suggested by mitochondrial ribosomal DNA sequences. Nature 361:.346–348
Mindell DP (1991) Aligning DNA sequences: homology and phylogenetic weighting. In: Miyamoto MM, Cracraft J (eds) Phylogenetic analysis of DNA sequences. Oxford Univ. Press, New York, 73–89
Mindell DP, Honeycutt RL (1990) Ribosomal RNA in vertebrates: evolution and phylogenetic applications. Ann Rev Ecol Syst 21: 541–66
Moritz C (1991) Evolutionary dynamics of mitochondrial DNA duplications in parthenogenetic geckos, *Heteronotia binoei.* Genetics 129: 221–230
Moritz C, Dowling TE, Brown WM (1987) Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Ann Rev Ecol Syst 18: 269–292
Mullis K, Faloona F, Scharf S, Saiki S, Horn G, Erlich HA (1986) Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction. Cold Spring Harbor Symp Quant Biol 51: 263–273
Nei M (1991) Relative efficiencies of different tree-making methods for molecular data. In: Miyamoto MM, Cracraft J (eds) Phylogenetic analysis of DNA sequences. Oxford Univ. Press, New York, 90–128
Normark BB, McCune AR, Harrison RG (1991) Phylogenetic relationships of Neopterygian fish, inferred from mitochondrial DNA sequences. Mol Biol Evol 8: 819–834
Novacek MJ (1992) Mammalian phylogeny: shaking the tree. Nature 356: 121–125
Ojala D, Montoya J, Attardi G (1981) tRNA punctuation model of RNA processing in human mitochondrial DNA. Nature 290: 470–474
Olivo PD, Van de Walle JJ, Laipis JP, Hauswirth WW (1983) Nucleotide sequence evidence for rapid genotypic shifts in bovine mitochondrial DNA D-loop. Nature 306: 400–402
Orti G, Bell M, Reimchen T, Meyer A (1993) Global survey of mitochondrial DNA sequence variation in the threespine stickleback: evidence for recent colonizations. Evolution (in press)
Pääbo S, Thomas WK, Whitfield KM, Kumazawa Y, Wilson AC (1991) Rearrangements of mitochondrial transfer RNA genes in marsupials. J Mol Evol 33: 426–430
Palumbi SR, Martin A, Romano S (1991). *Simple fool's guide to PCR.* Dept. of Zoology, University of Hawaii, Honolulu
Pamilo P, Nei M (1988) Relationships between gene trees and species trees. Mol Biol Evol 5: 568–583
Patterson C (ed) (1987) Molecules and morphology in evolution: conflict or compromise? Cambridge University Press, Cambridge
Patterson C (1988) Homology in classical and molecular biology. Mol Biol Evol 5: 603–625
Perler F, Efstratiadis A, Lomedico P, Gilbert W, Kollodner R, Dogson J (1980) The evolution of genes: the chicken preproinsulin gene. Cell 20: 555–556
Roe BA, Ma D-P, Wilson RK, Wong JF-H (1985) The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome. J Biol Chem 260: 9759–9774
Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230: 1350–1354
Saiki RK, Gelfand DH, Stoffel S, Scharf S, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239: 487–491.
Saitou N, Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406–425
Sanderson MJ, Donoghue MJ (1989) Patterns of variation and levels of homoplasy. Evolution 43: 1781–1795
Smith MF, Banfield DK, Doteval K, Gorski S, Kowbel DJ (1989) Gene arrangement in sea star mitochondrial DNA demonstrates a major inversion event during echinoderm evolution. Gene 76: 181–185
Smith MF, Banfield DK, Doteval K, Gorski S, Kowbel DJ (1990) Nucleotide sequence of nine protein-coding genes and 22 tRNAs in the mitochondrial DNA of the sea star *Piaster ochraceus.* J Mol Evol 31: 195–204
Smith MJ, Arndt A, Gorski S, Fajber E (1993) The phylogeny of echinoderm classes based on mitochondrial gene arrangements. J Mol Evol 36: 545–554
Sober E (1988) Reconstructing the past: parsimony, evolution and inference. MIT Press, Cambridge
Stock DW, Whitt GS (1992) Evidence from 18S ribosomal RNA sequences that lampreys and hagfish form a natural group. Science 257: 787–789
Sturmbauer C, Meyer A (1992) Genetic divergence, speciation and morphological stasis in a lineage of African cichlid fish. Nature 358: 578–581
Swofford DL (1991) When are phylogeny estimates from molecular and morphological data incongruent? In: Miyamoto MM, Cracraft J (eds) Phylogenetic analysis of DNA sequences Oxford Univ. Press, New York, 295–333
Swofford DL, Olsen GJ (1990) Phylogeny reconstruction. In: Hillis DM, Moritz C (eds) Molecular systematics. Sinauer, Sunderland, 411–501

Templeton AR (1983 a) Convergent evolution and non-parametric inferences from restriction fragment and DNA sequence data. In: Weir B (ed) Statistical analysis of DNA sequence data. Marcel Dekker, New York, 151–179
Templeton AR (1983 b) Phylogenetic inference from restriction endonuclease site maps with particular reference to the humans and apes. Evolution 37: 211–244
Vawter L, Brown WM (1993) Rates and patterns of base change in the small subunit ribosomal RNA gene. Genetics 134: 597–608
White TJ, Arnheim N, Erlich HA (1989) The polymerase chain reaction. Trends Genet 5: 185–189
Wiley E (1981) Phylogenetics. Wiley, New York
Wilson AC, Zimmer EA, Prager EM, Kocher TD (1989) Restriction mapping in the molecular systematics of mammals: a retrospective salute. In: Fernholm B, Bremer K, Jornvall H (eds) The hierarchy of life. Elsevier, Amsterdam, 407–419
Wilson AC, Cann RL, Carr SM, George M, Gyllensten UB, Helm-Bychowski KM, Higuchi RG, Palumbi SR, Prager EM, Sage RD, Stoneking M (1985) Mitochondrial DNA and two perspectives on evolutionary genetics. Biol J Linn Soc 26: 375–400
Wilson GM, Thomas WK, Beckenbach AT (1985) Intra- and interspecific mitochondrial DNA sequence divergence in *Salmo:* rainbow, steelhead, and cutthroat trouts. Can J Zool 63: 2088–2094
Wolstenholme DR, Clary DO (1985) Sequence evolution of *Drosophila* mitochondrial DNA. Genetics 109: 725–744
Wrishnik LA, Higuchi RG, Stoneking M, Wilson AC (1987) Length mutations in human mitochondrial DNA: Direct sequencing of enzymatically amplified DNA. Nucl Acid Res 15: 529–542.
Yoneyama Y (1987) The nucleotide sequences of the heavy and light strand replication origins of the *Rana catesbeiana* mitochondrial genome. Nippon Ika Daioaku Zasshi 54: 429–440 (in Japanese)
Zuckerkandl E (1987) On the molecular evolutionary clock. J Mol Evol 26: 34–46