INVITED ARTICLE

Evolution, Weighting, and Phylogenetic Utility of Mitochondrial Gene Sequences and a Compilation of Conserved Polymerase Chain Reaction Primers

CHRIS SIMON,[1] FRANCESCO FRATI,[1,2] ANDREW BECKENBACH,[3] BERNIE CRESPI,[3] HONG LIU,[1,4] AND PAUL FLOOK[5]

Ann. Entomol. Soc. Am. 87(6): 651–701 (1994)

ABSTRACT DNA-sequence data from the mitochondrial genome are being used with increasing frequency to estimate phylogenetic relationships among animal taxa. The advantage to using DNA-sequence data is that many of the processes governing the evolution and inheritance of DNA are already understood. DNA data, however, do not guarantee the correct phylogenetic tree because of problems associated with shared ancestral polymorphisms and multiple substitutions at single nucleotide sites. Knowledge of evolutionary processes can be used to improve estimates of patterns of relationships and can help to assess the phylogenetic usefulness of individual genes and nucleotides. This article reviews molecular processes, discusses the correction of genetic distances and the weighting of DNA data, and provides an assessment of the phylogenetic usefulness of specific mitochondrial genes. The *Appendix* presents a compilation of conserved polymerase chain reaction primers that can be used to amplify virtually any gene in the mitochondrial genome. DNA data sets vary tremendously in degree of phylogenetic usefulness. Correction or weighting (or both) of DNA-sequence data based on level of variability can improve results in some cases. Gene choice is of critical importance. For studies of relationships among closely related species, the use of ribosomal genes can be problematic, whereas unconstrained sites in protein coding genes appear to have fewer problems. In addition, information from studies of amino acid substitutions in rapidly evolving genes may help to decipher close relationships. For intermediate levels of divergence where silent sites contain many multiple hits, amino acid changes can be useful for construction phylogenetic relationships. For deep levels of divergence, protein coding genes may be saturated at the amino acid level and highly conserved regions of ribosomal RNA and transfer RNA genes may be useful. Because of the arbitrariness of taxonomic categories, no sweeping generalizations can be made about the taxonomic rank at which particular genes are useful. As more DNA-sequence data accumulate, we will be able to gain an even better understanding of the way in which genes and species evolve.

KEY WORDS mitochondrial DNA, phylogeny, polymerase chain reaction

THE ADVENT OF the polymerase chain reaction (PCR) (Saiki et al. 1985, Mullis et al. 1986) marked the beginning of a revolution in molecular biology (Guyer & Koshland 1990) and a synthesis of molecular, evolutionary, and systematic thinking. PCR speeds the DNA sequencing process and facilitates the selection of specific gene regions (Innes et al. 1990). The use of *universal* PCR primers (Kocher et al. 1989, Simon et al. 1991; *Appendix*) permits the sequencing of the DNA of species for which no previous sequence information exists. Another great advantage of PCR is that only tiny amounts of starting material are necessary, making it possible to sequence small individuals and degraded samples (Pääbo 1990). Studies of insects can now include museum specimens (e.g., Simon 1988, Fang et al. 1993), fossils (e.g., DeSalle et al. 1992a, Cano et al. 1993), and individual specimens of minute species (e.g., ants, Jermiin & Crozier 1994; collembolans, F.F. & C.S., unpublished data; thrips, B.C., unpublished data). In the past several years, sequence data have accumulated rapidly. Systematic studies have demonstrated that some genes are clearly better than others for reconstructing evolutionary relationships among taxa at particular levels of divergence. The majority of genes chosen for study are mitochondrial because they are easier to manipulate; are clonally inherited, single-copy, and nonrecombining; and are abundant (Brown 1985, Moritz et

[1] Department of Ecology and Evolutionary Biology, University of Connecticut, U-43, Storrs, CT 06269.
[2] Current address: Department of Evolutionary Biology, University of Siena, Via P.A. Mattioli 4, 53100 Siena, Italy.
[3] Department of Biological Sciences, Simon Fraser University, Burnaby, British Columbia, V5A 1S6 Canada.
[4] Current address: Mitsubishi Kasei Institute of Life Sciences, 11, Minamiooya, Machida-shi, Tokyo 194, Japan.
[5] Zoologisches Institut, Rheinsprung 9, CH-4051, Basel, Switzerland.

0013-8746/94/0651–0701$02.00/0

al. 1987, Attardi & Schatz 1988, Wolstenholme 1992).

Recent reviews provide excellent summaries of phylogenetic techniques and their assumptions (Felsenstein 1988; Swofford & Olsen 1990, 1995; Maddison & Maddison 1992; Penny et al. 1992; Swofford 1993); thus we need not review these procedures here. Progress has also been made in the assessment of the significance of phylogenetic conclusions (Kishino & Hasegawa 1989, Archie 1989, Faith 1991, Hillis & Huelsenbeck 1992, Felsenstein & Kishino 1993, Hillis & Bull 1993, Swofford 1993). Other recent reviews discuss the application of molecular techniques to systematic and evolutionary problems (Hillis & Mortiz 1990, 1995; Simon 1991; Cameron et al. 1992; Miyamoto & Cracraft 1992; Crozier 1993; Hillis et al. 1994) and protocols for direct DNA sequencing (Hillis & Moritz 1990, 1995; Innes et al. 1990; Simon et al. 1991; DeSalle et al. 1992). In this article we discuss issues relevant to the assessment of the phylogenetic usefulness of individual genes and illustrate these with examples from insect mitochondrial DNA (mtDNA).

This article is part review and part an original synthesis of recently published data. The first third of the article discusses general molecular processes and patterns of DNA evolution in various gene regions. The second third of the article is a synthesis of the assumptions and practical aspects of correcting of genetic distances and the weighting of DNA data for phylogenetic analysis. The final section assesses the phylogenetic usefulness of individual mitochondrial genes. The *Appendix* is a compilation of conserved PCR primers. The text is written in modular form to allow those familiar with molecular processes and the general characteristics of mitochondrial genes to skip the first part and those familiar with correcting genetic distances and weighting DNA data to skip the second part. The final section is of general interest; its purpose is to summarize and synthesize initial DNA-sequence studies and assess the phylogenetic information content of specific mitochondrial genes at various levels of divergence. Some of the conclusions we present are not apparent in the original articles. Because comparative DNA-sequence data are just beginning to accumulate, it will be interesting to see if the generalizations presented here stand the test of time. The *Appendix* will be useful for continued reference; it presents conserved PCR primers useful for all mitochondrial genes, with comments on their effectiveness. Although most of these primers were designed with insects in mind, the high level of conservation of these primers and the comparative context in which they are presented make them easily adaptable for other animal groups. Brower & DeSalle (1994), in this issue (pp. 702–716), discuss the application of nuclear gene sequences to insect systematics.

General Properties of Molecular Data

The use of molecular data is appealing because we understand the genetic basis and have a working knowledge of many of the molecular processes involved in nucleotide substitution. The use of molecular data, however, does not guarantee correct phylogenetic trees. Two major problems exist. The first of these—retained similarity resulting from shared ancestral polymorphisms—has the potential to affect all genes that exhibit polymorphisms. If shared ancestral polymorphism is a problem, it should be recognized, but little can be done to sort out relationships among species other than to examine within-species variation and to sequence different, unlinked genes. Thus, ancestral polymorphism will be discussed only briefly here. The second problem—the masking of information caused by multiple substitutions at a single site—can be circumvented or corrected to a degree. It affects some genes or gene segments more severely than others with the result that they vary in their phylogenetic usefulness. A discussion of the problem of multiple substitutions and solutions for it will serve to introduce our assessment of individual genes.

Shared Ancestral Polymorphisms. Problems in reconstructing species histories from molecular data arise when variants fixed in different species originated from polymorphic variation in the common ancestor. Thus, substitutional changes are older than the species that carry them and have become fixed or lost randomly among species. In such cases, the phylogenetic relationships of the genes are unlikely to reflect the phylogenetic relationships of the species that possess them. This problem has been discussed at length (Neigel & Avise 1986, Pamilo & Nei 1988, Avise 1989, Crozier 1990, Doyle 1992, Crozier 1993). The severity of the problem increases as time between speciation events decreases, for example, in species radiations. In such cases, the problem of randomly shared ancestral polymorphisms will be compounded by a scarcity of informative, shared-derived character-state changes on the short internal branches of the tree that describe its evolutionary history. This problem exists regardless of whether the radiation took place recently or in the distant past and so can be a problem at higher taxonomic levels as well. Given that bursts of speciation appear to be common in the history of organisms, this problem cannot be ignored.

The result of shared ancestral polymorphisms is that reliable phylogenetic relationships will be difficult to derive. Phylogenetic groupings will be supported poorly as indicated by measures of fit of the data to the tree and measures of support for individual nodes. Comparisons among unlinked nuclear genes and between mitochondrial

and nuclear genes can reveal discrepancies between gene and species phylogenies.

Information content of all genes will be negatively affected by shared ancestral polymorphisms. Genes that tend to be polymorphic within species will be most problematic. Information on within species nucleotide polymorphisms is beginning to accumulate for insect mitochondrial genes (Martin & Simon 1990, Simon et al. 1993, Xiong & Kocher 1993a, Vogler & DeSalle 1993, Vogler et al. 1993a).

Multiple Substitutions at a Site. Other problems in the assessment of relationships based on molecular data derive from the fact that there are only four nucleotide character states—G, A, C, and T—and that base substitutions at single nucleotide sites can be obscured by later substitutions at the same site (also called multiple hits). Amino acids as well as nucleotides can experience multiple substitutions. Thus, like morphological character states, molecular character states that are shared among taxa may actually be the result of chance convergence rather than common ancestry. With molecular characters, however, we have the advantage that our knowledge of molecular processes, structure, and function can help us determine statistically which sites are most likely to have experienced multiple substitutions.

In molecular systematics, as in any systematic study, choice and weighting of characters to be used will make as much or more of an effect on the final outcome as choice of analytical method. Choosing and weighting, however, are not easy tasks. Genes vary in their phylogenetic usefulness and nucleotide sites within genes similarly vary. Rapidly evolving genes and nucleotide sites are useful for comparisons of closely related taxa; slowly evolving genes and nucleotide sites are useful for comparisons of distantly related taxa (Kocher et al. 1989). This is because in comparisons of recently diverged taxa only the most rapidly evolving nucleotide sites will have accumulated substitutions. Caution must be exercised, however, in studies of groups whose relationships are completely unknown. Species that appear similar morphologically mistakenly could be assumed to be recently diverged when they are not (C.S., L. Nigro, J. Sullivan, A. Martin, A. Grapputo, A. Franke & C. McIntosh, unpublished data; Xiong & Kocher 1991). If a substantial number of sites is fixed (i.e., not free to vary as a result of structural or functional constraints), the few sites that do vary will eventually accumulate multiple hits (Fitch 1986). With increasing time after divergence, genetic distances are progressively underestimated. Thus, as pointed out by Shoemaker & Fitch (1989) a low percentage of sequence divergence does not guarantee that genetic distances are estimated correctly.

Differences in average evolutionary rate among gene regions or among taxa are expressed as observed number of nucleotide substitutions per unit time. Rate differences can be caused by differences in the number of sites free to vary or by actual differences in the number of mutations per unit time. Here, we distinguish mutation (actual changes introduced at the molecular level) from substitution (changes persisting in viable adults).

In comparisons of distantly related taxa, only the most slowly evolving nucleotide sites will have escaped the masking effects of multiple substitutions at the same site. The trick is to find enough of these slowly evolving sites to establish a robust tree of relationships. For assessing relationships among these very deep branches of phylogenetic trees, major molecular, cellular, and morphological events may supplement or be more useful than sparse sequence data (e.g., Donoghue & Sanderson 1992, Downie & Palmer 1992). Examples of major molecular events that could provide phylogenetic information include insertions and deletions of introns (in protein genes) or short sequences (in ribosomal RNA genes), shifts in secondary and higher molecular structures or gene order rearrangements by inversions, or both, and transpositions or translocations (e.g., Downie & Palmer 1992, Sankoff et al. 1992, Sankoff 1993, Smith et al. 1993).

Multiple substitutions are a problem for all phylogenetic algorithms. For distance-based phylogenetic analysis, multiple substitutions at a site will cause under- or overestimation of the true genetic distances among taxa. For character-based analysis, multiple substitutions will obscure shared character information (synapomorphies) and create false character state matches (high homoplasy–low character consistency). To solve the problem of multiple substitutions for distance-based analysis, where genetic distances are used to link related taxa, several distance correction methods have been suggested. The difficulty lies in designing a realistic correction that incorporates all important aspects of molecular substitution. To solve the problem of multiple substitutions for character-based analysis where shared derived characters are used to link related taxa, classes of nucleotide sites least likely to have experienced multiple substitutions can be weighted more heavily in tree construction. Before discussing methods for correcting and weighting data containing multiple substitutions, it is necessary to review briefly the general structural and functional properties of different genes and noncoding DNA regions.

General Properties of Mitochondrial Ribosomal RNA, Transfer RNA, and Protein Coding Genes

Wolstenholme (1992) provides an excellent general introduction to the molecular biology and evolution of the animal mitochondrial genome which in the majority of taxa has 2 ribo-

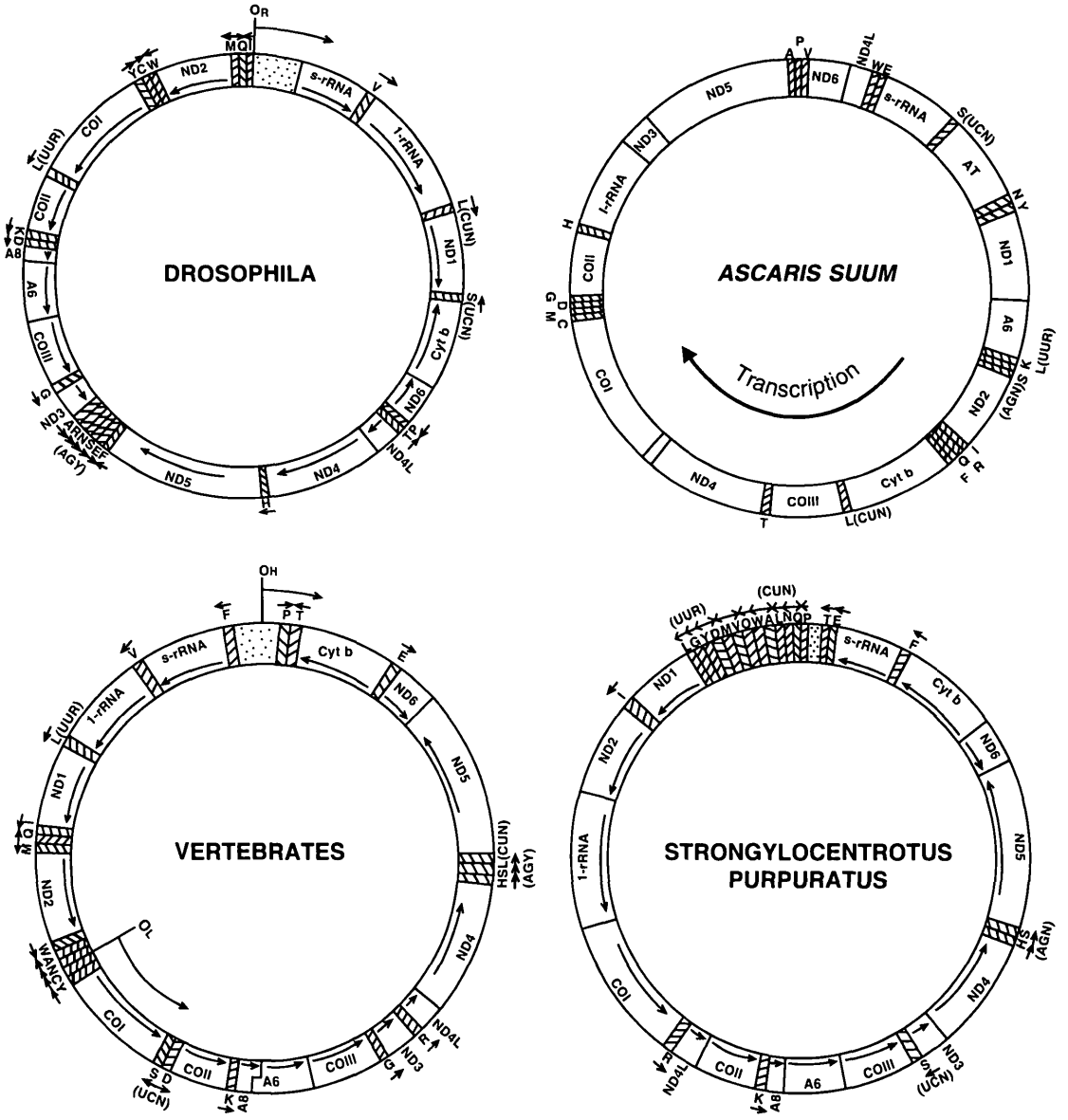


Fig. 1. Relative locations and direction of transcription of mitochondrial genes for *D. yakuba* (Clary & Wolstenholme 1985), *A. suum* (Okimoto et al. 1992), *H. sapiens* (Anderson et al. 1981), and *S. purpuratus* (Jacobs et al. 1988). Abbreviations of names of genes are listed in the footnotes of Table 1. Arrows indicate direction of transcription. Redrawn from the references cited.

somal RNA (rRNA) genes, 22 transfer RNA (tRNA) genes, and 13 protein coding genes (Fig. 1). With the exception of tRNAs, gene order appears to be highly conserved within vertebrates and within insects but not among animal phyla (Wolstenholme 1992).

Ribosomal, protein, and tRNA genes evolve in unique ways dictated by their structure and function. Below, we discuss the patterns of evolution of each class of gene. This information is important for the development of schemes for nucleotide genetic distance correction and weighting.

Ribosomal RNA Genes. Ribosomal genes have been studied intensively because of their critical role in protein assembly (Brimacombe et al. 1990, Noller et al. 1990). As a result of their universal occurrence, sequence and structural conservation, and abundance (caused by repetition), they have been used for phylogenetic analyses of a wide range of species and divergence levels including the deepest branches of *the tree of life* (reviewed in Mindell & Honeycutt 1990, Hillis & Dixon 1991, Hamby & Zimmer 1992).

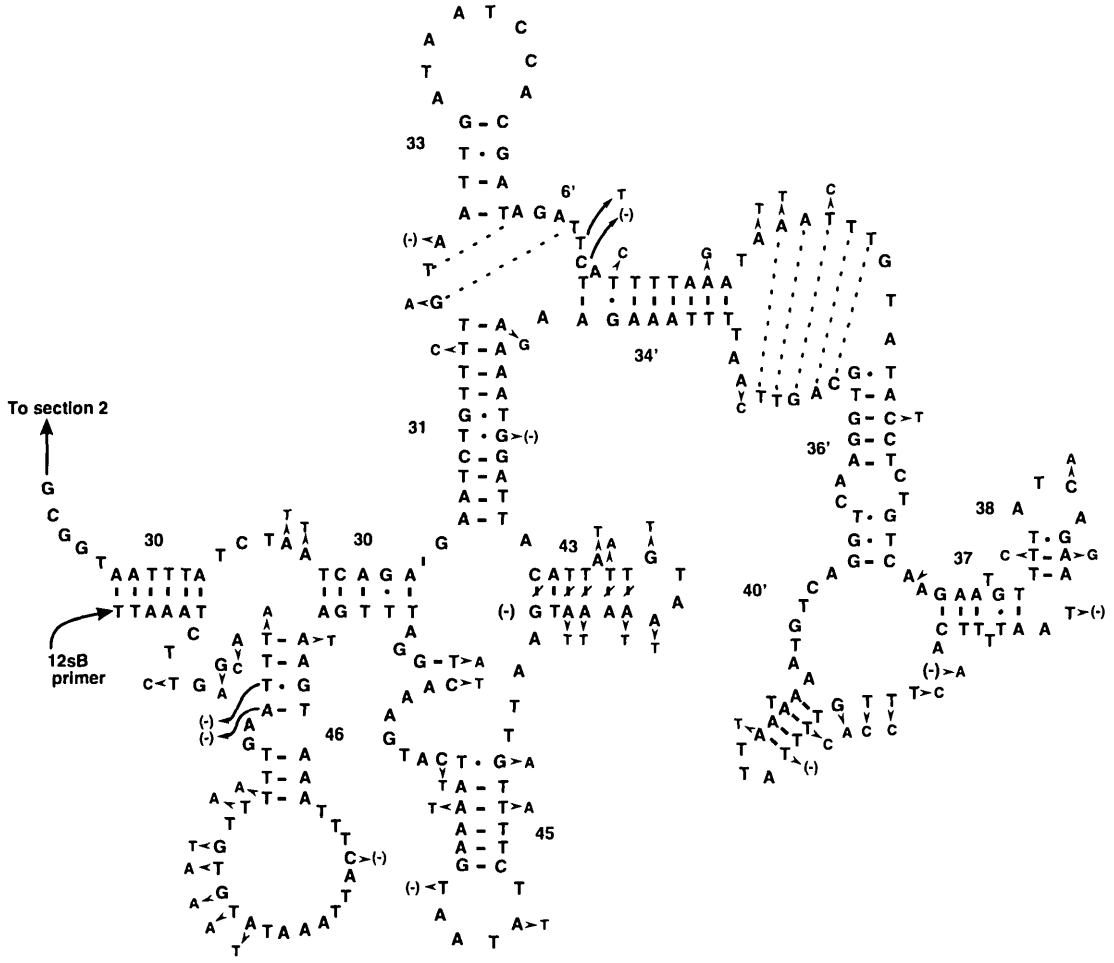


Fig. 2. The secondary structure of the third domain of the mitochondrial small ribosomal subunit (12S) gene of *M. tredecim*. Arrows point to substitutions (smaller letters) seen in the sequence of *O. vanduzeei* an out-group species in the same tribe (Tibicinini). Note that some stems are highly conserved (e.g., 30, 31, 33, 36), whereas others are highly variable (e.g., 43, 45, 46).

The rRNA genes produce a single strand of RNA that is later folded to pair with itself to form a secondary structure composed of helical *stems* connected by unpaired *loops* (Fig. 2). As one might expect from its functional importance, the secondary structural core of the ribosome is highly conserved across all genomes (De Rijk et al. 1993, Van de Peer et al. 1993). Peripheral secondary structural elements are less conserved, and it is these regions that are deleted or modified in the smaller, animal mitochondrial genomes. The rRNA secondary structure is in turn folded into a three-dimensional tertiary structure held together by a number of ribosomal proteins. Ribosomal protein attachment sites on RNA strands are also highly conserved and have been well characterized in *Escherichia coli* (Brimacombe et al. 1990, Noller et al. 1990). Together the ribosomal core and its accessory proteins form ribosomal subunits named according to their molecular weight.

Mitochondrial rRNA genes are much simpler than nuclear rRNA genes. Nuclear rRNA genes can have as many as five different subunits and may be duplicated as tandem arrays separated by transcribed and nontranscribed spacer regions. The many rRNA gene copies in the genome appear to evolve in concert, i.e., a substitution in one quickly spreads to the other copies (Hancock et al. 1988, Hillis & Dixson 1991 [review]). In animals, there are only two mitochondrial rRNA subunit genes and they are smaller, exist in only one copy per genome, and contain no spacer regions.

The rate of evolution of rRNA genes varies considerably along the length of the molecules (Hillis & Dixon 1991, Simon 1991). This variation is illustrated in Fig. 2 for a subsection of the

small rRNA subunit (12S) gene where nucleotide substitutions of the cicada *Okanagana vanduzeii* are plotted relative to the sequence and secondary structure of the cicada *Magicicada tredecim* (C.S. & S. Pääbo, unpublished data). Highly conserved nucleotide sites are associated with sites of ribosomal protein attachment, messenger RNA (mRNA) processing, tRNA attachment, and core helices (Noller et al. 1990). Conserved core helices (e.g., 30, 31, and 32 in Fig. 2) bind large subsections (domains) of the molecule and are called long-range stems because the bases on one side are separated by many bases in the primary sequence from the bases they pair with on the other side of the stem (Hixson & Brown 1986, Simon et al. 1990). In peripheral or short-range stems (e.g., 45 and 46), the bases on one side of the helix are separated from the bases with which they pair by only a few bases in the primary sequence. These stems and the loops that are associated with them tend to be less conserved, although there are notable exceptions (e.g., 33 in Fig. 2 as noted by Simon et al. 1990). Unpaired regions joining domains tend to be highly conserved. The pattern of substitution shown in Fig. 2 is typical of animal mitochondrial genomes and is most apparent at deep levels of divergence (Hixson & Brown 1986, Simon et al. 1990, R. Hickson, C.S., A. Cooper, J. Sullivan, G. Spicer & D. Penny, unpublished data).

The domains of rRNA evolve at different average rates dictated by their functional constraints. For example, in the 12S, the 5′ half (domains 1 and 2) has many fewer conserved nucleotide strings than the 3′ half (domains 3 and 4) (Clary & Wolstenholme 1985, De Rijk et al. 1993, Van de Peer et al. 1993). Similarly, in the 16S, domains 1 and 2, on average, are less conserved than domains 3 and 4 (Uhlenbush et al. 1987, Gutell et al. 1992).

Evolutionary rate differences in rRNA have been studied quantitatively by C.S., L. Nigro, J. Sullivan, A. Martin, A. Grapputo, A. Franke and C. McIntosh (unpublished data), who examined the small ribosomal subunit (12S) gene in six species of *Drosophila,* six species of periodical cicadas (genus *Magicicada*), and five species of hominoid primates. They found cicadas were evolving in a manner more like that of primates than that of *Drosophila*. In both cicadas and primates, the relative rate of evolution of the 5′ and 3′ halves of the 12S were similar. In *Drosophila*, the 3′ half was evolving much slower than the 5′ half. Even after 30–60 million yr of evolution, the average genetic distance among taxa was extremely low (<4%). They demonstrated that the low divergence was caused by severe structural or functional constraint (few sites free to vary) rather than a depressed mutation rate.

Transfer RNA Genes. Mitochondrial tRNA genes are less constrained by structure and function than their highly conserved nuclear counterparts and, therefore, evolve at a higher rate (Wilson et al. 1985). Nevertheless, mitochondrial tRNAs evolve more slowly than mitochondrial protein coding genes (Bibb et al. 1981, Wolstenholme & Clary 1985, Kumazawa & Nishida 1993), an indication that they are structurally or functionally more constrained.

In comparisons of mitochondrial tRNAs from five vertebrate species (human, rat, mouse, cow, and *Xenopus*), Gadaleta et al. (1989) showed that the degree of conservation of different functional regions within tRNAs varied, with the anticodon (AC) loop being the most conserved region, followed by the dihydrouridine (D) stem. The degree of conservation of the 5′ half was higher than that of the 3′ half. The D loop, the ribothymidine pseudouridine cytosine (T) loop, and the T stem were the most variable regions. A similar result was found in studies of closely related primate (Kondo et al. 1993) and salmon species (Thomas & Beckenbach 1989) and in a more distantly related drosophilid versus a lepidopteran species (Pashley & Ke 1992). The number of bases composing each mitochondrial tRNA stem is also highly conserved (Kumazawa & Nishida 1993).

Variations in size as well as sequence have been found in the D and T loops among animal mitochondrial tRNAs (Wolstenholme et al. 1987, Okimoto et al. 1992). In all animals, one of the two serine tRNAs lacks the D arm, whereas in nematodes, both serine tRNAs lack this arm (Okimoto et al. 1992). In the nematodes, *Ascaris suum* and *Caenorhabditis elegans*, the T loop is absent from all tRNAs (Wolstenholme et al. 1987, Okimoto et al. 1992).

Rate of evolution varies among as well as within mitochondrial tRNAs. Comparing a lepidopteran with a fly, Pashley & Ke (1992) found that the tRNA valine differed at only 20% of its sites, whereas tRNA leucine (CUN) differed at 35% of its sites. At a much closer level of divergence, Nigro et al. (1991) examined three tRNAs in seven species of *Drosophila* from the *melanogaster* group and found considerable differences in the rate of evolution among them. Rate differences could result from the fact that tRNAs adjacent to protein coding genes may play a punctuational role in protein processing (Ojala et al. 1981).

Like invertebrates, vertebrates show variability in substitution rate among mitochondrial tRNAs. A study of 11 tRNA genes in six hominoid species showed tRNA methionine and tRNA alanine to be substantially slower than the other 8 genes examined (Kondo et al. 1993). In comparisons of human, rat, mouse, cow, and frog tRNAs, those that coded for the most commonly used amino acids were the most conserved (Gadaleta et al. 1989). Comparisons among humans showed that tRNA genes that recognize 4-fold degenerate codons are generally more variable than

tRNA genes that recognize 2-fold degenerate codons (Cann et al. 1984). This observation does not hold for *Drosophila yakuba* compared with *Drosophila melanogaster* (Wolstenholme & Clary 1985); but the effect may be hidden by substitutional saturation and more recently diverged species should be analyzed.

The tRNA genes appear to be more mobile than protein coding genes (Wolstenholme 1992, Beard et al. 1993). Within insects, the relative positions of protein and rRNA genes in the mtDNA molecules sequenced so far do not vary, whereas a few tRNAs have been found to switch positions in locust (Haucke & Gellissen 1988), bee (Crozier et al. 1989), mosquito (HsuChen et al. 1984, Beard et al. 1993, Mitchel et al. 1993), and lepidopterans (Taylor et al. 1993) relative to *Drosophila* (Clary & Wolstenholme 1985). Some tRNA genes appear to be more stable in insects than others. H.L. & A.B. (unpublished data) sequenced the three tRNAs located between the ND2 and COI genes in cockroach, moth, dragonfly, cricket, locust, and flea: tRNAtrp, tRNAcys, and tRNAtyr. They found that the gene order and identity of the transcriptional strand were identical to that of *D. yakuba* but different from honey bee, which was ordered tRNAcys, tRNAtyr, tRNAtrp. In sea urchins, most tRNAs cluster in one area of the genome (Jacobs et al. 1988, Cantatore et al. 1989). Vertebrate and invertebrate tRNAs are less clumped than sea urchins and their relative positions are quite different (Clary & Wolstenholme 1985) (Fig. 1). In the sea anemone, *Metridium senile* there are only two tRNA genes present (Wolstenholme 1992). The situation in other cnidarians appears to be similar; tRNAs are thought to be transcribed in the nucleus and imported into the mitochondrion for protein processing (Wolstenholme 1992).

Mitochondrial tRNA genes have been known to change their function. In echinoderm evolution, a case is known where one of the two tRNA leucine genes was incorporated into a protein coding gene (ND5). The consequence of this shift was a change in the pattern of substitution to match that of the protein gene (Thomas et al. 1989).

Protein Coding Genes. The major difference between protein and RNA genes is that protein genes possess a triplet code for the assembly of proteins. This coding function places strong constraints on nucleotide changes at many first and all second codon positions. Because of the degenerate nature of the amino acid code, many third codon positions and some first positions are less constrained. Substitutions in these positions are termed *silent* (or *synonymous*) in that they do not cause amino acid replacement. These less constrained codon positions have been observed to evolve at a higher rate (Nei 1987, Irwin et al. 1991).

Silent Sites. In nuclear genes, although substitutions in many third positions of codons may have no effect on the identity of the amino acid, some bases are preferred. The strength of this codon preference or bias varies among genes. In general, codon bias is related directly to the rate of expression and inversely related to the rate of evolution of nuclear genes (Sharp & Li 1989); however, this is not the case for *Drosophila* histone genes, which are highly expressed and show low codon bias but are characterized by high rates of silent substitution (Fitch & Strausbaugh 1993). In fact, silent substitution rates in *Drosophila* histone genes are greater than in the Adh gene, which has a higher rate of amino acid substitution (Schaeffer & Aquadro 1987, Fitch & Strausbaugh 1993). However, among five histone genes, rates of silent substitution are lower in the most conserved gene, H4, as shown by comparisons of *D. melanogaster* to *Drosophila hydei* (Fitch & Strausbaugh 1993).

In mitochondrial genes, silent codon positions are more free to vary because an expanded codon recognition pattern allows a single tRNA species to decode as many as four codons (Gray 1989), although the efficiency of the process many not be equal for all four and the codon that matches the tRNAs with Watson-Crick pairings is not always preferred (Horai et al. 1992). Nevertheless, silent sites have been found to evolve at similar rates in different mitochondrial protein coding genes. Clary & Wolstenholme (1985) compared six protein coding genes from *D. yakuba* and *D. melanogaster* and found that the silent substitution rate was similar among genes even though the amino acid substitution rate varied substantially. Correspondingly, Nigro et al. (1991) found that silent sites evolved at a similar rate in portions of two mitochondrial protein coding genes (ND2 and COI) in seven species in the *melanogaster* and *oriental* species subgroups of *Drosophila;* these rates were similar to those calculated for four other mitochondrial protein coding genes in *Drosophila* (Sharp & Li 1989). Kessing (1991) found similar evolutionary rates for silent positions of four mitochondrial proteins in seven species of *Strongylocentrotus* sea urchins. Kondo et al. (1993), studying six primate species, found that, although the observed silent substitution rate appeared to differ substantially among genes, correction for base compositional biases among genes revealed that the actual silent substitution rate was not significantly different among genes. Furthermore, the silent substitution rate was not different from the rate of evolution of other unconstrained sites such as those in the control regions.

Amino Acid Replacement Rates. Rates of amino acid replacement substitutions vary considerably among proteins. In the evolution of nuclear as well as mitochondrial protein coding genes, the stringency of structural and functional

constraints varies tremendously. Nei (1987) compared rates of amino acid substitution in a wide variety of nuclear genes. Like nucleotide substitutions, amino acid substitutions are subject to multiple replacements so that the percentage of similarity may be overestimated in comparisons of distantly related species.

Amino acid substitution rate variation among mitochondrial protein genes is substantial and well documented. Nigro et al. (1991), studying seven species of *Drosophila* from two closely related subgroups, observed only one amino acid replacement out of 101 possible in the 5′ half of the COI gene but 30 out of 153 in the 3′ half of the ND2 gene. Although part of the difference could be caused by rate differences between the 3′ and 5′ halves of each molecule, in general, at the amino acid level, NADH dehydrogenase subunit genes are rapidly evolving and the cytochrome oxidase subunits are slowly evolving (Table 1). These relative levels of conservation indicate that the structural and functional requirements of the cytochrome oxidase proteins are far more rigid than those of the other genes. Shifts in relative degree of amino acid variability at progressively deeper levels of comparisons could indicate major adaptive shifts in protein structure and function. As noted by Crozier & Crozier (1993), there is a comparative acceleration in evolutionary rate in the *Apis* lineage with the result that *Drosophila* proteins are more similar in amino acid composition to those of *Locusta* than they are to those of *Apis* (Table 1).

An important conclusion from this discussion is that, for studies of closely related species that do not differ in amino acid sequence, there is theoretically no such thing as a rapidly evolving gene. One mitochondrial protein gene should be just as phylogenetically useful as another per length of DNA sequenced, with the qualifier that genes with more 4-fold than 2-fold degenerate sites should be more variable. At the level of amino acids, however, there is considerable variation among genes in substitution rates and, consequently, individual genes would have different phylogenetic utility. Some may even exhibit phylogenetically useful amino acid substitutions among closely related species (discussed later).

Structural and Functional Constraints. Evolutionary rate varies within protein coding genes as a result of structural and functional constraints (Fitch & Marcowitz 1970, Nei 1987, Shoemaker & Fitch 1989, Gillespie 1991, Li & Graur 1991, Britten 1993). As discussed above, at the DNA level, functional constraints related to codon position strongly affect the pattern of substitution. Also at the DNA level, mRNA secondary structure may play a role in patterns of substitution (Britten 1993). At the amino acid level, evolutionary rate is influenced by secondary, tertiary, and quaternary structural considerations. Some amino acid changes may be essentially neutral (French & Robson 1983). All factors at the DNA and amino acid level will interact in a complex fashion to determine the rate of nucleotide substitution.

At the amino acid level, structural and functional similarities in any given protein compared among diverse organisms are reflected in the proportion of hydrophobic versus hydrophilic amino acids in different regions of the gene. Crozier et al. (1989) demonstrated that COII gene hydrophobicity profiles in *Apis melifera* were similar to those in *Drosophila* and *Xenopus*. Such similarities can be used to identify protein coding genes even when amino acid similarities are very low, as in the A6, ND6, and ND4L genes of nematodes compared with *Drosophila* and mouse (Okimoto et al. 1992).

In certain proteins, the secondary, tertiary, and quaternary structure is well understood. In an elegant study that made use of the known secondary structure for cytochrome b, Irwin et al. (1991) plotted amino acid substitutions among 17 mammalian taxa onto the secondary structural model for the cytochrome b protein gene (Cytb). They found that the outer surface, which is important in redox function, is more constrained than the transmembrane (hydrophobic) segments and the surface that protrudes into the mitochondrial matrix. Martin & Palumbi (1993) found the same pattern and relative rates of substitution in the cytochrome b gene of 13 species of sharks even though the absolute rate of nucleotide substitution was six times slower and the shark cellular environment is highly modified by urea. Disotell et al. (1992) examined substitutions in relation to structure in the COII gene of primates and found similar higher levels of substitution in the transmembrane portion. Such an analysis has yet to be applied to insect protein coding genes, but observed patterns of variability suggest similar results are likely. Nigro et al. (1991) found that certain regions of the COI and ND2 genes of *Drosophila* are more variable than others and postulated that regions rich in A+T nucleotides were indicators of reduced selective constraints. Sperling & Hickey (1994) found similar heterogeneity within the COI and COII genes of spruce budworm.

Among Taxon Variation in Evolutionary Rate of Specific Protein Coding Genes. Although generalizations can be made about the comparative evolutionary rate of specific protein genes, a given protein gene may vary in rate from one taxon to the next (Wu & Li 1985, Nei 1987, Li & Graur 1991). Crozier et al. (1989) showed that the COI and COII genes exhibited significantly more amino acid substitutions in *A. melifera* than in *D. melanogaster* relative to mouse and *Xenopus*. Ruvolo and coworkers (Ruvolo et al. 1991, Disotell et al. 1992) demonstrated that the COII gene, which has a relatively constant rate of amino acid replacement among higher pri-

Table 1. Percentage of amino acid similarity at various divergence levels

Gene	*D. melanogaster*–*D. yakuba*[a] 3–19[b, c]	Human–cow[d] >90[b, e]	Apis–*Drosophilia*[f] >280[b, g]	*Drosophila*–Locust[h] >300[b, i]	*Xenopus*–Human[j] >350[b, k]	Sea star–sea urchin[l] >450[b, k]	Human–sea urchin[j] >700[b, k]	Mouse–*Drosophila*[m] >700[b, k]	*Drosophila*–Nematode[m] >700[b, k]	Mouse–Nematode[m] >700[b, k]
COI	99	91	70	82	86	89	75	75	62	59
COII	98	73	55	66	68	81	61	57	39	39
COIII	98	87	53	73	80	76	62	64	47	43
Cytb	96	79	53	76	71	—	62	68	43	42
ND1	96	78	47	68	63	67	55	45	36	35
ND4	95	74	45	58	59	—	43	42	30	28
ND3	95	74	49	62	58	67	48	43	30	26
ND5	89	70	42	64	55	—	44	40	31	25
ND4L	99	74	37	51	40	67	34	36	27	21
A6	96	78	47	73	53	69	37	34	22	22
ND2	94	63	27	47	50	57	39	34	17	21
A8	96	52	46	41	30	44	21	26	NF	NF
ND6	91	63	31	45	34	—	30	17	19	13

COI, cytochrome oxidase subunit I; COII, cytochrome oxidase subunit II; COIII, cytochrome oxidase subunit III; Cytb, cytochrome b; ND1, NADH dehydrogenase subunit 1; ND2, NADH dehydrogenase subunit 2; ND3, NADH dehydrogenase subunit 3; ND4, NADH dehydrogenase subunit 4; ND4L, NADH dehydrogenase subunit 4L; ND5, NADH dehydrogenase subunit 5; ND6, NADH dehydrogenase subunit 6; A6, ATPase subunit 6; A8, ATPase subunit 8; NF, A8 not found in nematode.

[a] deBruijn 1983, Wolstenholme & Clary 1985, Garesse 1988.
[b] Years in millions.
[c] Caccone et al. 1988, Beckenbach et al. 1993.
[d] Anderson et al. 1982.
[e] Novacek 1992.
[f] Crozier & Crozier 1993.
[g] Carpenter & Burnham 1985.
[h] McCracken et al. 1987, Ulhenbusch et al. 1987, Flook et al. 1995.
[i] Wootton 1981.
[j] Jacobs et al. 1988.
[k] Brusca & Brusca 1990.
[l] Smith et al. 1990.
[m] Okimoto et al. 1992.

mates, has a much higher replacement rate in higher primates than in rodents and other mammalian out-groups. Martin et al. (1992) found that shark cytochrome b genes evolved six times slower than those of mammals. Gu & Li (1992) in an extensive analysis of 54 nuclear encoded protein sequences for human, mice, rats, and chicken (out-group), demonstrated that 35 proteins evolved significantly faster in the rodent lineage and 12 proteins evolved faster in the human lineage. These differences are probably caused by differences in mutation rate rather than differences in structural and functional constraints (proportion of sites free to vary) because the structural and functional properties of proteins in general appear to remain relatively constant among taxonomic groups (Fitch 1971). Differences in mutation rate may be related to intrinsic properties of genes or chromosomes (e.g., Coulondre et al. 1978) or to metabolic-rate-related oxidative damage, polymerase errors, and repair efficiency (Martin et al. 1992, Martin & Palumbi 1993). As in rRNA genes, differences in rates of evolution among genes or gene segments are most clear at the deeper levels of divergence.

Natural Selection for Amino Acid Replacements. The fact that there are areas of molecules that accept amino acid substitutions more readily than others suggests that some amino acid substitutions may be selectively neutral. Thus, amino acid replacements at that position would be expected even in recently diverged species. Other amino acid substitutions may have a variety of effects on protein structure and function, ranging from subtle to drastic, depending on the chemical properties of the amino acids involved and the location of the substitution in relation to important structural and functional domains. Amino acid changes that have a structurally disadvantageous effect could be compensated for by another amino acid substitution which rescues the structure (major adaptive shift) or they could be transient in nature and selected from the population.

Amino acid substitutions with subtle effects on protein structure and function may provide a selective advantage to the organisms that possesses them. For example, substitution of an amino acid with different chemical characteristics could allow an enzyme to function more efficiently in a new temperature regime. Natural selection for altered temperature tolerance could favor such changes.

Sperling & Hickey (1994) found five amino acid substitutions (two in COI and three in COII) between two species of spruce budworm that differed in nucleotide sequence by only 1% at silent sites. Although two of these changes involved nonpolar residues unlikely to have major structural consequences, three involved changes to amino acids of different classes and may reflect naturally selected differences. Similarly, Willis et al. (1992) found three amino acid replacements between two honey bee species that differed by only three nucleotide substitutions. One of the changes involved chemically similar amino acids, the other two may have some selective value. A similar excess of replacement versus silent substitutions suggests that selection is operating at the molecular level in nuclear protein genes such as abalone sperm binding proteins (Lee & Vacquier 1992), major histocompatibility complex proteins of humans and mice (Hughes et al. 1990), and the Adh locus of *Drosophila* (McDonald & Kreitman 1991). Selection at the molecular level could lead to transient accelerations in the rate of amino acid substitution in some lineages but not in others, resulting in an episodic molecular clock (Gillespie 1991).

General Properties of Mitochondrial Noncoding Regions

Animal mitochondrial DNA has very few spacer or noncoding regions. The most consistently present spacer region is the *control region* that surrounds the origin of replication of the molecule. A few other small spacer regions occur sporadically in various taxa (Cann & Wilson 1983, Jacobs et al. 1988, Cantatore et al. 1989, Crozier et al. 1989, Cornuet et al. 1991, Pashley & Ke 1992, Kondo et al. 1993).

The control region is called the *A+T-rich* region in insects (Fauron & Wolstenholme 1980, Clary & Wolstenholme 1987) and the *D-loop* region in vertebrates (Upholt & David 1977, Crews et al. 1978, Aquadro & Greenberg 1983). As in nuclear noncoding regions (introns, spacer regions, pseudogenes) (Li et al. 1984, Hillis & Davis 1986), this lack of coding is accompanied by a hypervariability. This level of variability has made the control region useful for studies of relationships below the species level in vertebrates (Cann et al. 1987, Thomas et al. 1990). The only insect taxon for which control region sequences have been compared for a group of closely related species is the lepidopteran genus, *Jalmenus* (Taylor et al. 1993). The six species studied are believed to have recently undergone rapid speciation. Although more rapidly evolving than other regions, the control region exhibited too little sequence variation to provide useful phylogenetic information in these butterflies.

In invertebrates, the control region may be difficult to use for phylogenetic analysis. One problem is that it is even more A+T rich than the genome as a whole (e.g., 96 versus 85% in honey bees [Crozier & Crozier 1993]; 93.1 versus 75.9% in *C. elegans* [Okimoto et al. 1992; R. Crozier, personal communication]). With only two nucleotides available for substitutions, convergence will be high and difficult to detect. For this reason, caution should be taken in alignment, and

significance tests are needed to distinguish random matches from homology (Taylor et al. 1993). Second, this region is sometimes characterized by extreme length variability (Hale & Singh 1986, Solignac et al. 1986, Moritz et al. 1987, Boyce et al. 1989, Rand & Harrison 1989, Martin & Simon 1990). The largest and most variable control region so far recorded in insects is that of bark weevils, which varies between 9 and 13 kb (Boyce et al. 1989). Length variability is probably caused by the presence or absence of repeated sequence blocks (Rand & Harrison 1989). Number of repeats could perhaps provide useful phylogenetic information. Longer control regions are more difficult to amplify by PCR (see *Appendix*).

Length variation appears to be absent within and among species in the small control regions of the few lepidopterans that have been sequenced (*Jalmenus evaforas*, *Helicoverpa punctigera*, and *Strymon melinus*) (Taylor et al. 1993). Although equal in size (350+ bp), they showed little sequence similarity to each other or to *Drosophila*.

Average Rates of Evolution of Nuclear Versus Mitochondrial Genes

Early studies of mammals demonstrated that the average rate of evolution of the mitochondrial genome was 5–10 times higher than that of the nuclear genome (Brown et al. 1979, 1982; Miyata et al. 1982). A similar conclusion was reached in a study comparing 11 strains of the nematode *C. elegans* for two genes, nuclear calmodulin, and mitochondrial COII (Thomas & Wilson 1991). *Drosophila* (Powell et al. 1986, Solignac et al. 1986) and sea urchins (Vawter & Brown 1986), on the other hand, were found to exhibit average rates of mitochondrial gene evolution similar to those of nuclear genes. Sharp & Li (1989), using published sequence data, showed that *Drosophila* nuclear genes evolve faster than mammalian nuclear genes and that *Drosophila* mitochondrial genes evolve faster than nuclear genes with high codon bias and at approximately the same rate as nuclear genes with low codon bias. They found the rate of substitution at silent sites in *Drosophila* nuclear genes to be at least three times higher than the rate of substitution at silent sites in mammalian nuclear genes.

The average rate of evolution of the mitochondrial genome, although interesting from a theoretical standpoint, is less relevant for systematic studies than the rate of evolution of individual genes or nucleotide positions. When selecting characters for phylogenetic analysis, it is necessary to calculate or estimate site-specific or region-specific substitution rates so that appropriate characters can be excluded or downweighted according to their observed level of variability in the taxa under investigation. If genetic distances are to be calculated, level of variability can be used to correct for multiple substitutions. It is important to realize that weighting and multiple-substitution corrections are analogous in concept. These two procedures are discussed in detail in the next section.

Using Molecular Data for Phylogenetic Analysis

Alignment

Before nucleotide homologies can be assessed and genetic distances among taxa calculated, DNA sequences must be aligned. For distantly related species and rapidly evolving genes, this is not an easy task. Research into efficient and accurate alignment strategies is progressing rapidly (Feng & Doolittle 1987, 1990; Hein 1989a, b; Waterman et al. 1991, 1992; Chan et al. 1992; Higgins et al. 1992; Wheeler & Gladstein 1992). For protein coding genes, the presence of conserved amino acid sequences, strong selection against frame-shift mutations, and knowledge of biochemically likely amino acid substitutions can greatly aid, but does not guarantee, correct alignment (McClure et al. 1994). tRNA genes are small, conserved, and relatively easy to align. Alignment of rRNA is more difficult. Structurally important and highly conserved sequence blocks can be used as sporadic anchors for alignment; however, in large and small subunit gene sequences, confident alignment of variable bases located between these conserved anchors is time consuming and requires a comparative approach because of insertion and deletion events (Kjer et al. 1994, R. Hickson, C.S., A. Cooper, J. Sullivan, G. Spicer & D. Penny, unpublished data). The effect of alignment ambiguous data on phylogenetic analysis has been discussed recently by Waterman et al. (1992), Gatesy et al. (1993), and Collins et al. (1994b). The discussion below assumes that correct alignments have been obtained.

Phylogenetic Techniques

Phylogenetic algorithms can be divided into two basic categories, those that operate on pairwise distances among taxa and those that operate on individual character data. An excellent discussion of the assumptions of both kinds of analysis can be found in Felsenstein (1988), Swofford & Olsen (1990, 1995), and Penny et al. (1992). We will not discuss the assumptions of each class of algorithm here, but, for both, multiple substitutions at nucleotide positions pose serious problems.

When reading the following sections on distance correction and weighting, it must be kept in mind that properties of the data themselves can have far more influence on the success of

phylogenetic analysis than the algorithm used to construct the tree or the correction or weighting scheme applied. Successful results (well-supported relationships) obtained for one data set using a particular method do not guarantee that that method will be best for all data sets. Nonhomoplasious data will produce well-supported trees and extremely homoplasious data will produce poorly supported trees regardless of the methods used. Trees or subtrees with intermediate levels of homoplasy may be improved by data correction or weighting (or both).

Correcting Genetic Distances for Effects of Multiple Substitutions

If one could know true genetic distances and if there were no shared ancestral polymorphisms, phylogenetic trees could be constructed with certainty. As discussed above, multiple substitutions obscure the true genetic distances among taxa, and the rate at which they obscure these distances varies in a complex manner governed by specific molecular processes. True genetic distances are difficult to recover. A variety of algorithms has been devised to correct for multiple substitutions and those algorithms have become progressively more realistic. In this section we discuss them in rough chronological order.

Jukes-Cantor Equation and Its Assumptions. A discussion of the various methods suggested for correcting genetic distances for multiple substitutions will highlight the molecular processes involved in nucleotide substitution. The first and simplest method to be proposed, the Jukes-Cantor method (Jukes & Cantor 1969), uses a Poisson model to calculate the probabilities of multiple substitutions and makes the following three assumptions: (1) any base may change to any other base with equal probability; (2) the numbers of G, A, T, and C nucleotides are equal (no nucleotide bias); and (3) all sites along a DNA molecule change with equal probability. Comparative studies of nucleotide differences among taxa demonstrate that all these assumptions are seriously violated.

Assumption of Equally Probable Base Substitutions. The first assumption of the Jukes-Cantor equation is that any base changes to any other base with equal probability. For several reasons this assumption is never true. In biased genomes, such as the insect mitochondrial genomes discussed below, substitutions favor the common nucleotides (e.g., DeSalle et al. 1987). In addition, transitions—base changes from purines to purines or pyrimidines to pyrimidines—are favored over transversions—base changes involving a purine and a pyrimidine (Derancourt et al. 1967, Brown et al. 1979); this bias is caused by poor or deficient mtDNA repair mechanisms and a tautomeric base pairing chemistry that allows non-Watson-Crick G–T and A–C bonds (Topal & Fresco 1976, Wilson et al. 1985). For example, Kondo et al. (1993) in their study of six primate species found that the transition rate was 17 times faster than the transversion rate. Furthermore, the number of transitions involving purines may not equal the number of transitions involving pyrimidines (Tamura & Nei 1993).

The observed ratio of transition to transversion substitutions may not be constant from one taxon to the next, but within any given lineage, recently diverged species pairs always have a higher proportion of transitional differences than more distantly related pairs. This is because, with multiple substitutions, transversions accumulate and erase the record of transitions. Thus, even though transitions always occur more frequently in evolution, observed transitions exceed transversions only when recently diverged species or slowly evolving gene regions are compared; graphs of percentage of transitions versus time show an exponential decrease to a constant low value between 35 and 45% (Brown et al. 1979, 1982; Holmquist 1983; Aquadro et al. 1984; DeSalle et al. 1987; Miyamoto et al. 1990; Simon et al. 1990; Irwin et al. 1991; Simon 1991; Beckenbach et al. 1993). In some cases, the percentage of transitions observed in comparisons of distantly related arthropods can be even lower than 35% because of heavily A+T-biased genomes (W. Black, personal communication). A few studies have revealed that transitions may be obscured by transversions even in comparisons of supposed closely related species, but comparisons of populations or strains within these species will reveal a preponderance of transitions (Clary & Wolstenholme 1985, Xiong & Kocher 1993a; F.F. & C.S., unpublished data). In such cases, species may be older than previously supposed or rates of evolution may be accelerated.

In highly conserved protein genes, the percentage of transition substitutions can be high even between distantly related species such as bee and fly. This is because all two-codon families are partially constrained such that the alternative codons differ by a transition in the third position (Crozier & Crozier 1993).

In slowly evolving gene regions, transitions should predominate, indicating a lack of multiple substitutions. Simon et al. (1990) compared the third domain of the 12S rRNA gene in a cicada, a *Drosophila,* and a human, and found that in regions conserved among the three distantly related taxa, transitions were 9 times more common than transversions. Failure to find a transition bias in such a case would be an indicator of multiply substituted hot spots in otherwise conserved regions.

Because transversions (TV) obscure transitions (TS), the TS:TV ratio of a particular piece of DNA can be used as an indicator of the degree of multiple substitutions. However, caution should be taken in this approach because substitution

bias and rate variation among sites can be conflated in simple pairwise comparisons (Wakeley 1994). To get an accurate picture of the TS:TV ratio, base compositional biases (Kondo et al. 1993) plus differential probability of substitution among sites (Wakeley 1994) must be corrected. In addition, the statistical error associated with the calculation of TS:TV ratios is high, especially if the number of sites sampled is low.

Assumption of No Nucleotide Bias. Many genomes, including prokaryote (Hasegawa & Hashimoto 1993), vertebrate nuclear (Bernardi & Bernardi 1986), and insect mitochondrial (deBruijn 1983, Clary & Wolstenholme 1985, DeSalle et al. 1987, Simon 1991) have biased nucleotide compositions. That bias is often unevenly distributed among and within genes (Bernardi 1993) and tends to accumulate in the most variable positions (Hancock et al. 1988, Simon et al. 1990), for example, in the mitochondrial control region of insects (Taylor et al. 1993).

Insect mitochondrial genomes in general tend to be highly A+T biased. An examination of complete mitochondrial genome sequences reveals that *D. yakuba* has 78.6% A+T nucleotides (Clary & Wolstenholme 1985), whereas honey bee has 84.9% A+T (Crozier & Crozier 1993). For other insects, individual mitochondrial genes have been sequenced (reviewed below) and all reveal high A+T content. High A+T bias may be characteristic of many but not all protostome invertebrates if the COII gene and the third domain of the 12S rRNA gene are good indicators (Tables 2 and 3). These preliminary data suggest that there may be a trend within insects of increasing A+T bias in the more recently derived orders. Because G–C bonds are stronger than A–T bonds (Zucker 1989, Zucker et al. 1991), we suspect that structural and functional constraints put an upper limit on A+T richness, with spiders and honey bees at the pinnacle. Within A+T-rich groups such as *Drosophila,* slight reversals of this trend toward A+T richness can be found (Nigro et al. 1991). DeSalle et al. (1987) suggested that regardless of whether A+T richness arose in insect mtDNA as a result of chance or selection, these insect genomes are now adapted to high A+T conditions and selection acts to maintain this. Jermiin & Crozier (1994) have found systematic differences in A+T content in insect mitochondrial genes suggesting changes in cell metabolism during insect evolution.

Nucleotide bias can have drastic effects on phylogenetic reconstructions, especially if the bias is different for individual taxa. Such bias in phylogenetic analysis can result in the grouping of taxa based on nucleotide composition rather than on shared history (Hasegawa & Hashimoto 1993, Steel et al. 1993). In addition, empirical and simulation studies (Collins et al. 1994a) suggest that nucleotide compositional bias can result in systematic errors in character state reconstruction in parsimony analyses. This effect increases in severity with increasing rate of nucleotide evolution and results in an underestimation of the number of rare to common changes.

Table 2. A+T content of insects compared with selected invertebrate and vertebrate taxa for the third domain of the mitochondrial 12S gene

Common name	Genus/species	% A+T
Nematodes	*Ascaris, Caenorabditis*[a]	72 –73.5[b]
Prosobranch gastropod	*Cellana tramoserica*[c]	70
Polyplacophoran mollusk	*Ischonochiton australis*[c]	71
Earthworm	*Aporrectodea rosea*[c]	59
Onycophoran	*Atherton, Euperipatoides, Plicatoperipatus*[c]	79.5–83
Spider	*Tetragnatha* spp.[c]	83
Scorpion	*Liocheles waigiensis*[c]	77
Decapod crustacean	*Cherax quadricarinatus*[c]	70
Centipede	*Allothereura, Cormocephasus*[c]	70
Insects		
Silverfish	*Ctenolepisma longicaudata*[c]	67
Locust	*L. migratoria*[d]	73
Cicada	*M. tredecim*[e]	78
Honey bee	*A. mellifera*[c]	81.5
Drosophilid fly	*Drosophila* spp.[c, g]	77 –79
Other flies	*Australofannia, Austrosimulium, Ceromya, Lucilia, Musca, Macropodexia, Scaptia*[c]	73 –80
Mosquito	*Anopheles spp.*[c, h]	76 –77
Sea urchin	*Paracentrotus, Stronglyocentrotus*[i]	54 –58
Fish	*Cyprinus carpio*[j]	52
Frog	*X. laevis*[k]	54
Mammals	*Balaneoptera, Mus, Phoca*[l]	57.5–64
Human	*H. sapiens*[m]	53

[a] Okimoto et al. 1992.
[b] These values refer to the complete 12S sequence because in nematodes it is impossible to determine unequivocally the position of the third domain.
[c] Ballard et al. 1992.
[d] Flook et al. 1995.
[e] Simon et al. 1990.
[f] Crozier & Crozier 1993.
[g] Clary & Wolstenholme 1985.
[h] Beard et al. 1993, Mitchell et al. 1993.
[i] Thomas et al. 1989, Cantore et al. 1989.
[j] Chang et al. 1994.
[k] Roe et al. 1985.
[l] Bibb et al. 1981, Arnason et al. 1991, Arnason & Johnsson 1992.
[m] Anderson et al. 1981.

Assumption that All Sites Change with Equal Probability. Golding (1983) demonstrated that large genetic distances will be substantially underestimated given the assumption that all sequence positions evolve at equal rates. The probability of substitution of an individual nucleotide is strongly affected by its structural and functional importance in the gene product. This phenomenon was discussed in detail above, in the section on general properties of the three types of genes. Thus, patterns of substitution in ribo-

Table 3. A+T content in the COII gene of insects compared with selected invertebrate and vertebrate taxa

Common name	Genus/species	% A+T
Nematodes	*Ascaris, Caenorabditis*[a]	69 –74
Insects		
Collembolan	*Isotomurus, Orchesella, Tetrodontophora, Thaumanura*[b]	60 –71
Dragonfly	*Sympetrum striolatum*[c]	68.5
Cockroach	*P. americana*[c]	70.5
Termite	*Zootermopsis angusticollis*[c]	67
Orthopteran	*Abracris, Acheta, Aeropus, Caconemobius, Locusta, Schistocerca, Zoniopoda*[c,d,e]	68 –74
Hemipteran	*O. fasciatus*[c]	72.5
Coleopteran	*Adalia, Sitophilus*[c]	70 –74
Green lacewing	*Chrysoperla plorabunda*[f]	72
Hymenoptera	*Apis, Exeristes, Lasium*[c,g,h]	76 –80
Flea	*C. felis*[c]	78
Fly	*Drosophila* spp.[i,j]	73.5–74
Mosquito	*Anopheles* spp.[k]	72.5–73
Lepidopteran	*Choristoneura, Galleria*[c,l]	76 –79
Sea urchin	*Paracentrotus, Strongylocentrotus*[m]	57 –59
Fish	*C. carpio*[n]	57.5
Frog	*X. laevis*[o]	61
Mammals	*Balaneoptera, Mus, Phoca*[p]	58 –61.5
Human	*H. sapiens*[q]	54

[a] Okimoto et al. 1992.
[b] F.F., unpublished data.
[c] Liu & Beckenbach 1992.
[d] A. Phillips, personal communication.
[e] Flook et al. 1995. P.F., unpublished data.
[f] M. Martinez-Wells, personal communication.
[g] Crozier et al. 1989.
[h] Willis et al. 1992.
[i] Clary & Wolstenholme 1985.
[j] Beckenbach et al. 1993.
[k] Beard et al. 1993, Mitchell et al. 1993.
[l] Sperling & Hickey 1994.
[m] Jacobs et al. 1988, Cantatore et al. 1989.
[n] Chang et al. 1994.
[o] Roe et al. 1985.
[p] Bibb et al. 1981, Arnason et al. 1991, Arnason & Johnsson 1992.
[q] Anderson et al. 1981.

somal, transfer RNA, and protein coding genes are highly nonrandom and rates of evolution vary substantially within and among genes at the nucleotide and protein level in a continuous fashion. In the vertebrate D loop, substitutions accumulate according to a negative binomial (gamma distributed rates of evolution) rather than the Poisson model (Kocher & Wilson 1991; Tamura & Nei 1993; Wakeley 1993, 1994). Yang, Goldman, and Friday (Yang 1993, Yang et al. 1994) applied a similar negative binomial model to a maximum likelihood phylogenetic analysis of rRNA and protein coding gene sequences. More than 20 yr ago, Uzzell & Corbin (1971) demonstrated that substitutions at the amino acid level could best be described by a negative binomial distribution.

Distance Corrections Attempting To Improve on Jukes-Cantor. For closely related taxa, all distance correction methods will give approximately the same estimate because multiple substitutions will be few. As the genetic distance among taxa increases, it becomes increasingly important to incorporate realistic assumptions into genetic distance correction methods. Unfortunately even the most realistic correction methods do not appear to work well enough to recover true genetic distances among distantly related taxa, as discussed below.

Distance Corrections Incorporating Biased Base Substitutions. The distance correction method known as the Kimura two-parameter method (Kimura 1980) extended the method of Jukes and Cantor to incorporate information on transition and transversion probabilities. More elaborate correction methods have been described using a substitution matrix with up to 12 independent parameters (Nei 1987). In practice, however, these methods are seldom employed because of the difficulty of obtaining reliable estimates for these parameters. Even for the case of Kimura's two-parameter method, where only transitions and transversions are distinguished, there is disagreement concerning the best method for estimating the two parameters. PHYLIP (Felsenstein 1993) defaults to a constant TS:TV ratio of 2:1 (although other values may be used), whereas others use an empirical estimate based on the sequences themselves. Neither method is entirely satisfactory; a constant 2:1 ratio is inaccurate, whereas estimates based on the data are biased with time as multiple hits involving transversions erase previous transitions (Brown et al. 1982, Beckenbach et al. 1993). To get an accurate estimate of the TS:TV ratio and the true genetic distance among taxa, long stretches of DNA must be sequenced (Martin et al. 1990), and nucleotide bias and differential probabilities of substitution must be taken into account (Yang et al. 1993, Wakeley 1994).

Distance Corrections Incorporating Nucleotide Bias. Tajima & Nei (1984), developed a method of correcting genetic distances that allowed nucleotide frequencies to deviate from 0.25. Tamura (1992) presented a general method that corrected for nucleotide and TS:TV bias at the same time. This method works well if the nucleotide bias is the same for all taxa being compared and if the taxa are not too distantly related. For insect mitochondrial genes, the equal bias assumption may be met (but see apteryogotes and odonates in Tables 2 and 3). For other genomes or for more diverse taxonomic assemblages such as prokaryotes, nucleotide bias may vary from one taxon to the next and, as mentioned earlier, cause severe problems in estimating relationships (Hasegawa & Hashimoto 1993, Steel et al. 1993, Collins et al. 1994a). Even if existing nucleotide bias corrections had realistic assumptions, this type of correction alone would not be sufficient.

Distance Corrections Incorporating Differential Probabilities of Substitution. This form of distance correction is intuitively appealing because the most variable characters are those most likely to have experienced multiple hits. Specific corrections have been devised for both protein and rRNA genes. Other corrections are applicable to any region of DNA sequence.

a) Correcting Distances Calculated from Protein Coding Genes. Although any of the above methods may be used to correct distances in protein coding sequences, the triplet code superimposes additional structure onto sequence variation. This structure can be exploited in three different ways: by classifying the sites according to degeneracy levels (4-fold, 2-fold, or nondegenerate), by counting and perhaps weighting the possible nucleotide substitution pathways that could give rise to observed or inferred amino acid replacements, or by weighting replacement substitutions by the chemical differences in the amino acids.

Li et al. (1985, Li 1993) devised a method to classify sites according to degeneracy level and to apply Kimura (1980)-type corrections for each of the categories. This method divides sites into discrete probability of change classes corresponding to 2-fold degenerate, 4-fold degenerate, and replacement sites. It is a major improvement over methods that assume that the probability of substitution is the same at all sites. It is limited, however, in more distant comparisons, in that it assumes that a given site will remain in the same probability of substitution class over time (Nei 1987). In fact, we know that in protein genes 4-fold degenerate sites can, for example, become 2-fold degenerate because of an amino acid replacement (Lewontin 1989). In addition, this measure does not take into account the influence of protein secondary structural and functional considerations on nucleotide variability.

Lewontin (1989) identifies problems associated with determining sequence divergence for protein coding sequences, which are significantly but not highly constrained. Where selection is involved, an amino acid replacement substitution may alter the probabilities of substitutions at other sites within the codon. For example, a second codon transversion, CAA → CTA (glutamine → leucine), changes the degeneracy level at both the first and third codon positions. For highly constrained sequences, he discourages the correction of nucleotide and amino acid sequences, and recommends instead that the codon is the appropriate unit of analysis. He provides a method for estimating divergence.

Nei & Gojobori (1986) focused on the codon level, classifying nucleotide substitutions as synonymous or nonsynonymous. Where two or more substitutions are observed in the homologous codon of two sequences, this classification may depend on the order in which the substitutions took place. The actual classification is then accomplished by adding up all the pathways. The pathways themselves can be weighted according to the chemical effect of the amino acid replacements required for each (Miyata & Yasunaga 1980). Jukes-Cantor or Kimura two-parameter corrections are then applied separately to the synonymous and nonsynonymous substitutions.

Finally, distance measures can be applied to the translated sequences. Dayhoff et al. (1978) devised an amino acid replacement matrix, based on comparisons of large numbers of nuclear encoded amino acid sequences, that reflects the chemical effect of each replacement. A distance measure can be achieved by summing up the probabilities for all observed differences.

b) Correcting Distances Calculated from rRNA and tRNA Genes. In genes or DNA regions which lack a triplet code, structural and functional constraints that influence probability of substitution are more obvious. As mentioned earlier, there is considerable variation in probability of substitution among sites in rRNA, tRNA, and even noncoding control regions.

Structural, functional, and comparative sequence information could be used to create a simple a priori weighting scheme for ribosomal genes. In such a scheme the DNA sequence could be divided into a few discrete classes—quickly, moderately, and slowly evolving—corresponding to specific structural and functional parts of the molecule as determined by comparative studies of taxa whose relationships are known (Simon 1991). Like the discrete approach for protein coding genes (Li et al. 1985), this method would be limited in that a given site will not remain in the same probability of substitution class over time. For example, an rRNA nucleotide, which is a member of a highly conserved helix, could slip out into a bulge or loop and subsequently accept substitutions freely. Even more of a problem in rRNA genes is the observation that structural and functional constraints on a particular nucleotide site can differ drastically among taxa (C.S., L. Nigro, J. Sullivan, A. Martin, A. Grapputo, A. Franke & C. McIntosh, unpublished data). In addition, the patterns of variation clearly observable in distant comparisons (De Rijk et al. 1993, Van de Peer et al. 1993) may be totally irrelevant to variation among species with close or intermediate levels of divergence.

c) Correction Methods Applicable to Both rRNA and Protein Sequence Data. Nucleotide variability on a finer scale can be estimated from the data themselves and used to weight genetic distances. Gutell et al. (1985) and Manske & Chapman (1987) used Shannon's (1948) information theory formula to measure relative nucleotide variability at each site in a set of aligned sequences. Manske & Chapman (1987) went on

to use these Shannon weights to compute Jukes-Cantor (1969) genetic distances between all pairs of taxa. They then used these corrected distances as input to tree-building algorithms best described as cladistic modifications of simple unweighted paired-group arithmetic average (UPGMA) clustering (similar to Farris' [1977] special similarity method). They concluded that variability corrections were useful in revealing major patterns in the early evolution of life.

Another data-based distance weighting method was recently proposed by Van de Peer et al. (1993). They weighted discrete sets of rRNA positions according to their level of variability. To estimate variability, sequences were aligned, a pairwise distance matrix calculated, and pairs of taxa were grouped into five distance intervals. For each of the distance intervals, variability was calculated on a site-by-site basis by counting the percentage of pairwise comparisons differing at that position. For each site, this percentage was then plotted against distance interval and a regression used to obtain its average variability across distance intervals. This average variability was used to weight each nucleotide position and calculate a second distance matrix, which was then analyzed by the neighbor-joining (NJ) tree-building algorithm (Saitou & Nei 1987). Using this method, Van de Peer et al. (1993) analyzed a large number of sequences from their small subunit rRNA database (De Rijk et al. 1983) and were able to resolve traditional relationships among taxa, which unweighted NJ analysis failed to recognize. All methods in which the initial data are used to estimate nucleotide site-specific variability are dependent on a large sample of taxa. The larger and more representative the sample, the better the variability estimate.

As opposed to the methods discussed above that divide nucleotide sites into discrete classes according to their level of variability or estimate site-specific weights based on the data, the remaining correction methods assume that probability of substitution is a continuous function that varies along the molecule according to a Poisson, log normal, or gamma distribution. Olsen (1987) recognized the problem of unequal probability of substitution among sites and developed a distance correction based on a log normal distribution. He then used the least squares optimality criterion to find the best distance-based tree. Later studies suggested that the gamma distribution is an appropriate model for nucleotide sequence evolution (Kocher & Wilson 1991; Tamura & Nei 1993; Wakeley 1993, 1994; Yang 1993, Yang et al. 1994). A gamma distribution may fit nucleotide data better because it can accommodate a large class of nonvariable sites, whereas the log normal can not.

Tamura & Nei (1993) proposed a distance correction based on the gamma distribution in which excess transitions, unequal nucleotide frequencies, and variation of substitution rate among different sites are all taken into account (Tamura & Nei 1993). This substitution matrix model is similar to the maximum likelihood model of Hasegawa et al. (1985) but with the more realistic assumption that the rate of purine transitions is allowed to differ from the rate of pyrimidine transitions. In addition, Tamura & Nei (1993) derived an analytical formula for estimating the number of nucleotide substitutions. The restrictive assumptions of their method is that purine and pyrimidine transition biases and general nucleotide bias, although allowed, are relatively similar across lineages and that the nucleotide composition of the two DNA strands is equal. This assumption may often be violated and invalidate the correction (Collins et al. 1994a). In addition, Tamura & Nei (1993) cautioned that their simulation tests suggest that the application of this equation should be confined to cases involving relatively closely related sequences and relatively large numbers of nucleotides. For closely related species, however, genetic distance corrections have little effect. This point is illustrated in Table 4 using the data of Beckenbach et al. (1993). Despite an average G–C content of about 26%, the Tamura (1992) and Tamura & Nei (1993) estimates give very little improvement over the Kimura two-parameter model, and none give much improvement in distance estimates over the Jukes-Cantor model. Part of the reason can be seen in the right hand columns of Table 4. Both Kimura's and Tamura & Nei's methods provide separate estimates of transition and transversion rates. The ratio of these estimates drops off quickly with divergence time and stabilizes between 0.35 and 0.45 for distant species. This change in TS:TV ratio is clearly artifactual and is primarily caused by the fact that transversions mask both prior and subsequent transitions (Brown et al. 1982, DeSalle et al. 1987). Pairwise genetic distances corrected using these methods would still be underestimates such that if plotted against time, they would not increase linearly.

Nei (in Kumar et al. 1993) suggests, for distance-based phylogenetic analysis, that if two distance correction measures give similar results, the simpler method should be used because it has a smaller variance. He suggests that for low genetic distance values the use of simple methods such as Jukes-Cantor or the uncorrected distance will increase the likelihood of a correct tree as long as substitution rates do not vary among lineages.

Using Molecular Character Data for Phylogenetic Analysis

The most difficult task in systematics is to construct a single tree that depicts relationships among both closely and distantly related species.

Table 4. Comparison of genetic distance correction methods using data from Beckenbach et al. (1993) for the *D. obscura* group.

Taxa	Distance from *D. affinis* (×100)				TI:TV		
	JC	K2P	Tamura	Tam & Nei	K2P	Tamura	Tam & Nei
D. algonquin	0.88	0.88	0.88	0.88	5.04	5.05	5.05
D. athabasca	1.18	1.18	1.18	1.18	7.07	7.10	7.10
D. azteca	2.37	2.38	2.38	2.39	1.69	1.70	1.70
D. narragansett	2.67	2.69	2.71	2.72	8.19	8.25	8.29
D. tolteca	4.05	4.08	4.12	4.12	5.95	6.02	6.02
D. miranda	6.08	6.08	6.11	6.13	0.83	0.84	0.85
D. pseudoobscura	6.24	6.25	6.29	6.33	1.08	1.09	1.10
D. lowei	6.40	6.42	6.46	6.51	1.38	1.39	1.41
D. ambigua	8.66	8.70	8.78	8.91	1.50	1.53	1.56
D. subobscura	8.98	9.03	9.13	9.14	1.49	1.51	1.52
D. bifasciata	9.98	10.03	10.15	10.17	1.44	1.47	1.48
D. yakuba	10.53	10.53	10.60	10.61	0.56	0.57	0.57
D. melanogaster	11.38	11.39	11.50	11.52	0.83	0.84	0.85

JC, Jukes-Cantor correction; K2P, Kimura two-parameter correction; Tamura, Tamura correction (Tamura 1992); Tam-Nei, Tamura-Nei correction (Tamura & Nei 1993); TI, transitions; TV, transversions. These values were calculated using the MEGA program (Kumar et al. 1993).

Sites that are useful for relating recently diverged species would almost certainly have experienced multiple substitutions in comparisons of distantly related species. Weighting against highly variable sites would ensure that they would not be used in phylogeny construction at deep levels and would only be used if no other informative sites were available (i.e., when recently diverged species are compared). As mentioned previously, weighting in character-based analyses is analogous to correcting genetic distances for multiple hits in distance-based analyses.

Phylogenetic algorithms that operate on character data, in this case individual nucleotide sites, include parsimony methods, maximum-likelihood methods, and spectral-analysis methods. Maximum-likelihood and spectral-analysis methods show great promise for phylogenetic analysis but are computationally intensive and currently work well only for a limited number of taxa (Penny et al. 1992). Because this article focuses on properties of the data rather than on properties of the phylogenetic tree construction algorithms, a discussion of weighting using character-based parsimony as an example will serve to illustrate a few important points about the effects of multiple substitutions on character-based analyses. Similar considerations can be incorporated into maximum-likelihood and spectral-analyses methods.

Shared derived character states are the source of phylogenetic information in parsimony analysis. Multiple substitutions at single nucleotide sites can obscure shared derived character states and reduce the reliability of a site for establishing relationships among taxa. Multiple substitutions can also create false signals because of chance convergence. Small amounts of convergence will not obscure a strong phylogenetic signal because all convergent characters are unlikely to suggest the same incorrect phylogeny (Farris et al. 1970). Problems arise, however, when the phylogenetic signal is not strong enough to overcome the noise. This problem will be severe when the rate of evolution varies substantially among lineages (Felsenstein 1988) or when some older lineages are represented in the analysis by one or a few taxa (one or a few long branches), while other older lineages are represented by many taxa (many shorter branches) (Maddison et al. 1984, Hendy & Penny 1989). When the phylogenetic signal is weak, there will be a tendency for long branches to join each other regardless of the true tree (Felsenstein 1988, Hendy & Penny 1989). Thus, nucleotide sites that have experienced multiple substitutions should not be used to establish phylogenetic branching order and long branches, which undoubtably harbor many multiple substitutions, should be avoided. These two pitfalls can often be circumvented by proper weighting and by prudent choice of taxa.

Weighting schemes for character-based parsimony analysis can be divided into two groups: methods that assume a particular set of discrete costs based on extrinsic criteria or models of evolution and methods that estimate weights or probabilities of substitution from the data itself; these can be iterative or noniterative (Swofford & Olsen 1990, 1995; Goloboff 1993). Both approaches attempt to identify the most conservative characters (least likely to have experienced homoplasy) and give them greater weight. The various approaches are discussed below.

Discrete Cost Weighting Approaches Based on Extrinsic Criteria. Knowledge of molecular processes can be used to predict which sites will evolve most rapidly and, therefore, have the highest probability of experiencing multiple hits. As discussed previously, we could expect 4-fold degenerate sites, unimportant sites in

rRNA, and the majority of sites in noncoding regions to evolve fastest. The same cautions would apply as in distance analysis corrections; a particular nucleotide site will not necessarily remain in the same probability-of-change class over time and rRNA domains can vary in degree of constraint among taxa of similar ages.

a) Extrinsic Criteria for Discrete Weighting of Protein Gene Sequences. Down-weighting of rapidly evolving third-codon position sites could improve phylogenetic analysis because the less frequent, second- and first-position changes will be used preferentially to establish relationships among distantly related taxa. Because closely related taxa will not usually differ by second- and first-position substitutions, the lesser weighted third positions will by default be used to establish relationships among them. Weighting according to first-, second-, and third-codon positions is, of course, an oversimplification. Substitutions at many third codon positions (those decoded by less than four tRNAs) can result in an amino acid replacement, as can substitutions at most, but not all, first-codon positions (e.g., leucine is an exception). Realistic weighting according to degeneracy of positions would be too complex. Simplified approximations can be made, for example, by using only 4-fold degenerate third positions or omitting first positions of leucine codons (Irwin et al. 1991).

If sufficient differentiation exists, translation into amino acids could provide a more realistic picture of the results of nucleotide substitution, but after millions of years of divergence, amino acid replacement sites also experience many multiple hits. To compensate for this, less conserved amino acids, unimportant in secondary and higher structures, could be weighted less heavily. In addition, substitutions involving chemically similar amino acids could be given less weight. Although we know many of the chemical rules governing amino acid substitutions, creating a realistic weighting scheme based on these rules would be a difficult task.

b) Extrinsic Criteria for Discrete Weighting of Ribosomal Gene Sequences. For phylogenetic studies of distantly related species using 5.8 S ribosomal genes, Wheeler & Honeycutt (1988) suggested differential weighting of stems versus loops. Simon (1991) pointed out that this weighting scheme is not directly applicable to larger, more complex rRNA molecules where some stems evolve at the same rate as some loops. Instead, if discrete weighting were desirable, it would have to take the form of down-weighting, rapidly evolving, short-range stems in relation to more conserved long-range stems. Similarly, rapidly evolving, unpaired regions could be down-weighted relative to more slowly evolving, functionally important, unpaired regions (Simon 1991). Such patterns of variation in rRNA genes are clearly observable in comparisons of distantly related species (e.g., Fig. 2). In fact, studies of very distantly related organisms, such as the tree-of-life studies reviewed in Hamby & Zimmer (1992) make such corrections partially by default, because rapidly evolving regions cannot be aligned and are, therefore, omitted from phylogenetic analyses although the decision as to where to draw the line in discarding tenuous alignments is not easy (Gatesy et al. 1993). One problem, however, in using the discrete method for weighting stems and loops in phylogenetic studies of distantly related taxa is that rapidly evolving sites could exist in otherwise conserved regions (C.S., L. Nigro, J. Sullivan, A. Martin, A. Grapputo, A. Franke & C. McIntosh, unpublished data).

For species with recent and intermediate divergence times, it is not clear how to apply discrete weighting methods because exact patterns of rate variation are less apparent. Discrete weighting of ribosomal regions was employed in a recent study of New Zealand skinks (Hickson 1993), but results were equivocal because this group had apparently experienced a burst of speciation, which resulted in many short internal branches on the phylogenetic tree and, therefore, low resolution no matter which weighting or phylogenetic method was employed.

Wheeler & Honeycutt (1988) also proposed that bases in paired rRNA stem regions should be down-weighted by 50% compared with unpaired loop regions, because mutations on one side of a stem would require a compensatory mutation on the other in order to maintain pairing. Simon (1991) pointed out that this view was an oversimplification because G–T bonds were biochemically acceptable and common in rRNA and because mutations on one side of a stem could also be compensated for by stem slippage (with and without bulge formation). This theoretical argument was supported empirically by Dixon & Hillis (1993), who demonstrated that stem bases experienced a greater number of compensatory substitutions than would have been expected at random, but that the number was <40% of that expected by Wheeler & Honeycutt (1988) with perfect one-to-one compensation.

c) Transversion Weighting. Because transitions are more common than transversions, the suggestion has been made that they be given less weight in phylogenetic analysis (Kocher et al. 1989). In the mammalian study of Irwin et al. (1991) mentioned above, the investigators found that for divergences of >5 million yr, silent transversions appeared to be as useful as amino acid replacements for phylogenetic inference. In addition they found that transversion substitutions were approximately linear with time for divergences ≤80 million yr. DeSalle et al. (1987) used linear accumulation of transversions in sequences of parts of the ND1, ND2, ND5, and 16S

genes to order deep branch points in their Hawaiian *Drosophila* phylogeny. For close relatives they used all changes (largely transitions), and for more distantly related comparisons, only transversions. The resulting tree agreed with relationships predicted on the basis of larval hemolymph proteins (Beverly & Wilson 1985). Transversion weighting was also found to improve the support for phylogenetic relationships constructed using COII gene sequences for subgroups of the *Drosophila obscura* species group (Beckenbach et al. 1993).

Transversion weighting may not work as well for ribosomal genes because, as predicted by Simon (1991) and demonstrated empirically by Allard & Miyamoto (1992), weighting transversions in rRNA favors the use of sites that have experienced multiple substitutions (i.e., transversions accumulate in the most variable positions). In addition, in highly biased genomes such as insect mitochondrial DNA, A–T transversions predominate over others. DeSalle (1992a) sequenced an rRNA gene segment comprising 665 bp on the 3′ end of the 16S gene to examine relationships within the genus *Drosophila*. Most of the substitutions observed were transversions and so transversion weighting gave results identical to the unweighted data set, but these results conflicted with traditional expectations in many cases.

Methods that Estimate Weights or Probabilities of Substitution from the Data. Character state variability can be assessed from aligned sequence data before tree building (Gutell et al. 1985; Penny & Hendy 1985, 1986; Manske & Chapman 1987; Knight & Mindell 1993) or after tree building (Farris 1969, Sankoff & Cedergren 1983, Olsen 1987, Williams & Fitch 1989, Fitch & Ye 1991, Van de Peer et al. 1993). In the latter case, a complete or partial initial tree is used to identify highly variable sites, which are downweighted in the next round of tree building.

a) Data-Based Weighting Schemes that Operate Before Tree Building. Weighting based on character compatibility was first used for parsimony methods by Penny & Hendy (1985, 1986). It involves counting the number of observed incompatibilities between each character and comparing this with the number of incompatibilities expected by chance. The greater the compatibility of a character, the greater the weight assigned to that character (Swofford & Olsen 1990, 1995). The justification for this method has been that maximizing compatibility minimizes homoplasy, although this assumption has been disputed (Goloboff 1993). This method has not been explored in detail.

The simplest data-based weighting scheme is the method of Knight & Mindell (1993) in which variable positions are identified by comparisons of aligned DNA sequences. Their method seeks to assess the degree to which each of six substitutional classes occurs more or less frequently than expected. Drawbacks of their approach include a lack of a specified model to determine expected substitutions, failure to adjust for compositional bias, and an underestimation of homoplasy from the use of pairwise comparisons to estimate character state changes (Collins et al. 1994b). Recognizing that pairwise comparisons will seriously underestimate homoplasy because the maximal number of changes counted at any position is constrained by the presence of only four nucleotide character states, other authors have evaluated variability in subsets of related taxa as determined by partially (Olsen 1987, Van de Peer et al. 1993) or completely resolved phylogenetic trees (Farris 1969, Sankoff & Cedergren 1983, Williams & Fitch 1989, Fitch & Ye 1991). These are discussed below.

b) Noniterative Data-Based Methods that Rely on an Initial Unweighted Tree. Olsen (1987) devised a weighting scheme for character-based parsimony in which variability at individual nucleotide sites was estimated with respect to a partial phylogenetic tree, that is, within previously defined taxonomic groups of bacteria. Because he was interested in deep-level relationships, he excluded the sites that were least conserved within groups. The method of Van de Peer et al. (1933), developed for distance comparisons, is conceptually similar except that it uses information present in the data itself to establish five hierarchical groups (distance classes). These groups are in essence a partially resolved phylogeny that is used to estimate site specific variability. This weighting scheme could be used for character-based parsimony analysis. The logical extension of these partial-tree variability analyses is successive weighting.

c) Iterative Data-Based Methods that Rely on an Initial Unweighted Tree. Farris (1969) first proposed the method of successive weighting. In his method, homoplasy is estimated on a fully resolved initial phylogenetic tree. Characters exhibiting low homoplasy are given more weight and a new tree is built. This procedure continues iteratively until the observed homoplasy stabilizes and the tree topology does not change with successive iterations. Philosophically similar successive weighting methods were developed specifically for nucleotide sequences (Sankoff & Cedergren 1983, Williams & Fitch 1989, Fitch & Ye 1991). This approach is not completely circular because the final tree may differ from the starting tree (Fitch & Ye 1991); however, it does suffer from the problem that the final result is highly dependent on the initial tree. This aspect of the successive weighting method could be improved by using other systematic data (e.g., morphological, chromosomal, and allozyme) to assess the reliability of the initial tree before proceeding. A second problem with successive weighting is that weights for a particular tree are

not derived in a way that allows valid comparison of alternative trees. Thus, there is no standalone criterion for choosing an optimal tree (D. Swofford, personal communication).

Goloboff (1993) recently designed a noniterative method that overcomes the problem of starting-point dependency. The reliability of characters is estimated during the analysis. Preferred trees have fewer steps for the characters that are most consistent in the sense of Kluge & Farris (1969). Exploration of this method will be useful.

How Should DNA Data Be Corrected or Weighted?

Philosophically, the best solution to the problem of multiple hits is to choose the most realistic correction or weighting method available. Lack of a sufficient number of exploratory studies and difficulties involved in creating methods without unrealistic simplifying assumptions, however, means there is no good answer currently. The most realistic genetic distance corrections take into account nucleotide bias, transition-transversion bias, and differential probabilities of substitution among sites, yet these corrections do not completely remove the effects of multiple substitutions, as discussed above. The most realistic character-based weighting schemes also attempt to account for all three of these within and among site substitution biases without complete success. Methods that rely on the data themselves, among other problems, require a large sample of nucleotides and taxa to obtain realistic assessments of variability. This is not always possible. Methods that rely on extrinsic criteria are based on an incomplete knowledge of molecular structure and function and can only partially solve the problem as illustrated by examples in this article. Although most researchers agree that conservative positions should be favored, others have abandoned every method except uniform weighting because of the sampling-associated problems involved in assessing within- and among-site variability. Uniform weights, however, can be more arbitrary than weights based on theoretical considerations (Swofford & Olsen 1990, 1995).

We recommend experimentation with a variety of weighting schemes. Information from other character sets, such as morphological, chromosomal, or allozyme data, can be extremely useful in assessing the strengths of individual nodes supported by molecular characters. In addition, partial phylogenetic information provided by nonmolecular data can be invaluable in assessing variability among nucleotide sites.

Information from diverse character sets can be combined in phylogenetic analysis but extreme caution should be taken in this exercise because of concerns with data heterogeneity as discussed in detail by Bull et al. (1993). Enthusiastic supporters of the combined data approach, still analyze morphological and molecular data sets independently as well as in combination (e.g., Wheeler et al. 1993).

Because properties of the data can have far more effect on the results of phylogenetic analysis than the method used to analyze the data, empirical studies involving a variety of data, taxa, and methods are needed. Some examples are presented in the next section. Many studies are currently in progress.

In addition to studies on taxa whose relationships are yet to be established; simulation, experimental, and model–taxon studies will be invaluable in assessing data correction or weighting methods. Simulation studies can be useful in identifying certain properties of phylogenetic methods and weighting schemes but have the problem that a model of evolution must be used to generate the data. Therefore, the best method to correct the data will be a method whose assumptions match the generating model (Huelsenbeck & Hillis 1993). Experimental studies with laboratory organisms have also provided valuable insights (Hillis et al. 1992, 1994) but are limited in that they do not exactly mirror the evolution of genes in natural populations of animals.

Studies are needed in which sequence data are gathered for a variety of genes (nuclear as well as mitochondrial) for model taxa whose phylogenetic relationships are well understood based on a variety of other types of data. These sequence data could then be weighted using various methods and subjected to phylogenetic analysis. The resulting trees could be evaluated by comparison to the expected tree. Empirical studies that follow this plan could eventually provide another way to evaluate phylogenetic tree-building methods. It is hoped that methods preferred on philosophical grounds will not be rejected because of difficulties in adapting a realistic correction or weighting scheme for the data. Studies of the kind described here are in progress in our laboratories.

Assessing the Phylogenetic Usefulness of Individual Mitochondrial Genes for Insect Taxa

Choosing the Right Gene or Genome

Choosing the correct gene for a given level of divergence is a form of weighting. Genes are favored whose level of variability will minimize the problem of multiple hits while maximizing the number of nonhomoplasious shared character states. For closely-related species, few nucleotide positions are likely to vary; therefore, genes or regions that contain the highest proportion of unconstrained sites should be chosen

(Kocher et al. 1989). For distantly-related species, the opposite reasoning would apply. Gene choice and weighting of sites within genes is philosophically similar to the process of character choice undertaken by morphological systematists as discussed by Novacek & Wheeler (1992).

The final section of this article assesses the phylogenetic usefulness of individual genes—those for which we have the most information—and discusses the level(s) at which each is most useful. This information is relevant to distance-based as well as character-based analyses. Primers useful for sequencing these genes and others are described in the *Appendix*. Because the properties of a data set can have far more influence on the outcome of phylogenetic analysis than the weighting scheme or tree-building algorithm used to manipulate the data, gene choice is of prime importance.

Ribosomal Genes. The 3′ regions of both the small (12S) and large (16S) rRNA genes have been sequenced frequently because of the convenient location of conserved primers (SR-N-14588 + SR-J-14233 or LR-N-13398 + LR-J-12887, respectively) ≈400–500 bp apart (Table 5; see *Appendix*). The 5′ halves of these molecules are less well studied but some data are available.

12S Gene. The 12S rRNA gene, especially the third domain (amplified by the universal primers mentioned above) appears to be phylogenetically useful for distant taxa but can in some instances be problematic for assessing relationships among recently diverged species. C.S., L. Nigro, J. Sullivan, A. Martin, A. Grapputo, A. Franke & C. McIntosh (unpublished data) studied the 12S gene from six species of *Drosophila:* two in the *melanogaster* group, two in the *obscura* group, one in the *willistoni* group, and one in the *virilis* group. They found that variation was extremely low in the 3′ half of the molecule. Because the phylogenetic relationships among the *Drosophila* species they studied are generally well-supported by a variety of other types of data, they were able to demonstrate that, of the five sites that varied, four were homoplasious. They also examined homoplasy in the 12S for six periodical cicadas species and five primate species. Homoplasy was low enough in the latter two comparisons that phylogenetic analysis of the data produced the well-supported relationships. In general, the 12S molecule contained few informative substitutions. The 5′ half of the molecule appears to have more useful information and less homoplasy than the 3′ half. Of special concern, is the fact that low percentage of sequence divergence (as in the *Drosophila* 3′ half) does not guarantee freedom from multiple hits.

Other studies of closely related species using the frequently sequenced third domain of the 12S have demonstrated that this segment has limited phylogenetic usefulness. Ballard (1994) sequenced 301 bp of the 3′ half of the 12S in two species of Australian blackflies (5 and 17 individuals, respectively). The among-species differentiation was low but useful for species identification. The low number of variable sites argues against using this gene for establishing relationships among blackfly species.

Similar low levels of phylogenetic information can be seen in the 12S gene of other invertebrates. Thomas et al. (1989) studied the 3′ half of five sea urchin taxa in the genus *Stronglyocentrotus.* Although they found 48 variable positions (out of 500), only 6 of these were informative (shared with at least one other taxon). If we plot these six onto a well supported tree based on five mitochondrial protein genes (Thomas et al. 1989, Kessing 1991), we find that three are homoplasious (Fig. 3).

Ironically, homoplasy in the 12S caused by multiple substitutions may be less of a problem for distantly related species comparisons than for closely related species. A study by Simon et al. (1990) suggests that highly conserved 12S sequence blocks seem to be evolving at a slower rate and that the few sites in these conserved blocks that do vary are not plagued by multiple hits. Simon et al. (1990) compared the 12S third domain of primates, *Drosophila,* and cicadas and found that the most conserved areas had 90% transition substitutions and low A+T bias, indicating low levels of multiple substitutions compared with the most variable regions. Outside of these conserved sequence blocks, few sites could be aligned with confidence. Based on these results, the general agreement between classical and rRNA-based trees for animal phyla (Field et al. 1988) could be explained by the fact that rRNA data came from the more conserved regions. The more variable regions—those that have experienced many multiple substitutions—were discarded because they could not be aligned.

Given that information useful for assessing deep relationships appears to be present in the more conserved domains of the mitochondrial ribosomal subunits, the question then becomes, is there enough information present to make a reliable tree, especially when only parts of the molecule are sequenced. The answer depends on the degree of divergence and the number of taxa under investigation. Using only sequence produced by amplification of the 3′ half of the molecule, Hickson (1993) was able to construct an unrooted network for 10 vertebrate taxa, which agreed with traditional expectations. Ballard et al. (1992), using this same 3′ segment, constructed a tree for 40 metazoan taxa, most of which were arthropods. Character-based parsimony analysis (Swofford 1993) of these 40 taxa produced 144 equally parsimonious trees indicating considerable ambiguity in the data.

Table 5. Published sequences of insect mitochondrial genes

Gene taxa sequenced	Reference
Complete mtDNA sequences	
L. migratoria	P.F. et al., 1995
A. gambiae	Beard et al. 1993
A. quadrimaculatus	Mitchell et al. 1993
A. mellifera	Crozier & Crozier 1993
D. yakuba	Clary & Wolstenholme 1985
All mitochondrial protein coding genes	
D. melanogaster	Wolstenholme & Clary 1985, Garesse 1988, deBrujin 1993
Single genes or parts of genes	
COI	
Complete[c]	
Calliphoridae (3 genera)	Sperling et al. 1994[a]
COI	
Partial	
A. mellifera (10 subspp.)	Garnery et al. 1992[a]
A. cerana	Garnery et al. 1992[a]
D. melanogaster group (3 spp.)	Nigro et al. 1991[a]
D. melanogaster subgroup (4 spp.)	Satta et al. 1987[a]
D. melanogaster subgroup (4 spp.)	Satta & Takahata 1990[a, b]
Choristoneura (6 spp.)	Sperling & Hickey 1994[a, b]
Greya (15 spp.)	Brown et al. 1994[a]
Prodoxus quinquepunctellus	Brown et al. 1994[a]
Tetragma gei	Brown et al. 1994[a]
Heliconius (sensu lato, 37 spp.)	Brower 1994[a]
Dione juno	Brower 1994[a]
Dryadula phaetusa	Brower 1994[a]
Dryas iulia	Brower 1994[a]
Philaethria dido	Brower 1994[a]
Podotricha telesiphe	Brower 1994[a]
COII	
Complete[c]	
S. striolatum	Liu & Beckenbach 1992
A. domesticus	Liu & Beckenbach 1992
S. gregaria	Liu & Beckenbach 1992
P. americana	Liu & Beckenbach 1992
Z. angusticollis	Liu & Beckenbach 1992
O. fasciatus	Liu & Beckenbach 1992
Adalia bipunctata	Liu & Beckenbach 1992
Sitophilus granarius	Liu & Beckenbach 1992
Lasius sp.	Liu & Beckenbach 1992
Apis (5 spp.)	Willis et al. 1992
Excristes roboratus	Liu & Beckenbach 1992
C. felis	Liu & Beckenbach 1992
Drosophila pseudoobscura	Liu & Beckenbach 1992
Drosophila (12 spp.)	Beckenbach et al. 1993
Calliphoridae (3 genera)	Sperling et al. 1994[a]
Choristoneura (6 spp.)	Sperling & Hickey 1994[a, b]
G. mellonella	Liu & Beckenbach 1992
COII	
Partial	
A. mellifera (10 subspp.)	Garnery et al. 1992[a]
A. cerana	Garnery et al. 1992[a]
Greya (15 spp.)	Brown et al. 1994[a]
P. quinquepunctellus	Brown et al. 1994[a]
T. gei	Brown et al. 1994[a]
Heliconius (sensulato, 37 spp.)	Brower 1994[a]
D. juno	Brower 1994[a]
D. phaetusa	Brower 1994[a]
D. iulia	Brower 1994[a]
P. dido	Brower 1994[a]
P. telesiphe	Brower 1994[a]

Table 5. Continued

Gene taxa sequenced	Reference
COIII	
Partial	
Cicindela dorsalis (4 subspp.)	Vogler & DeSalle 1993[a, b]
Cicindela puritana	Vogler & DeSalle 1993[a, b]
C. dorsalis (2 subspp.)	Vogler et al. 1993a[a, b]
Cicindela (6 spp.)	Vogler et al. 1993b[a, b]
Cytb	
Complete[c]	
Tetraponera rufoniger	Jermiin & Crozier 1994[a]
ND1	
Partial	
C. dorsalis (4 subspp.)	Vogler & DeSaslle 1993[a, b]
C. puritana	Vogler & DeSalle 1993[a, b]
C. dorsalis (2 subspp.)	Vogler et al. 1993[a, b]
Cicindela (6 spp.)	Vogler et al. 1993[a, b]
T. rufoniger	Jermiin & Crozer 1994[a]
A. albopictus	HsuChen et al. 1984[a]
Hawaiian *Drosophila* (9 spp.)	DeSalle et al. 1987[b]
Hawaiian *Drosophila* (7 spp.)	DeSalle 1992a[a]
Drosophilidae (17 spp.)	DeSalle 1992b[a]
Cosmosoma myodora	Pashley & Ke 1992[a]
Phoebis sennae	Pashley & Ke 1992[a]
Pseudoplusia includens	Pashley & Ke 1992[a]
S. frugiperda	Pashley & Ke 1992[a]
Symmerista albifrons	Pashley & Ke 1992[a]
ND2	
Partial	
D. melanogaster group (3 spp.)	Nigro et al. 1991[a]
D. melanogaster subgroup (4 spp.)	Satta et al. 1987[a]
D. melanogaster subgroup (4 spp.)	Satta & Takahata 1990[a, b]
Hawaiian *Drosophila* (9 spp.)	DeSalle et al. 1987[b]
ND3	
Partial	
A. albopictus	HsuChen & Dubin 1984[a]
ND5	
Partial	
C. capitata	Frohlich et al. 1993[a]
Hawaiian *Drosophila* (9 spp.)	DeSalle et al. 1987[b]
ND6	
Partial	
T. rufoniger	Jermiin & Crozier 1994[a]
12S rDNA	
Partial	
C. longicaudata	Ballard et al. 1992
M. tredecim	Simon et al. 1990, Simon et al. 1991
O. vanduzeii	C.S. (Fig. 1)
Anopheles hilli	Ballard et al. 1992
Australofannia sp.	Ballard et al. 1992
Austrosimulium (2 spp.)	Ballard et al. 1992
Ceromya sp.	Ballard et al. 1992
D. virilis	Clary & Wolstenholme 1987
D. melanogaster group (10 spp.)	Nigro & Grapputo 1993
L. cuprina	Ballard et al. 1992
Macropodexia sp.	Ballard et al. 1992
Musca (2 spp.)	Ballard et al. 1992
Scaptia sp.	Ballard et al. 1992
J. evagoras	Taylor et al. 1993
16S rDNA	
Complete[c]	
A. albopictus	HsuChen et al. 1984[a]
S. frugiperda	Pashley & Ke 1992[a]

Table 5. Continued

Gene Taxa sequenced	Reference
16S rDNA	
Partial	
Blaberus sp.	DeSalle et al. 1992
Mastotermes (2 spp.: one fossil)	DeSalle et al. 1992
Cicadellidae (29 spp.)	Fang et al. 1993
C. dorslis (4 subspp.)	Vogler & DeSalle 1993[a, b]
C. puritana	Vogler & DeSalle 1993[b]
C. dorsalis (2 subspp.)	Vogler et al. 1993a[a, b]
Cicindela (6 spp.)	Vogler et al. 1993b[a, b]
Cnephia dacotensis	Xiong & Kocher 1991
Tribe Apini (5 spp.)	Cameron, 1993
Tribe Bombini (3 spp.)	Cameron, 1993
Tribe Meliponini (4 spp.)	Cameron, 1993
Tribe Euglossini (2 spp.)	Cameron, 1993
Tribe Xylocopini (1 sp.)	Cameron, 1993
Tribe Allodapini (1 sp.)	Cameron, 1993
C. capitata	McPheron et al. 1994
Hawaiian *Drosophila* (9 spp.)	DeSalle et al. 1987
Hawaiian *Drosophila* (7 spp.)	DeSalle 1992b[b]
D. melanogaster group (10 spp.)	Nigro & Grapputo 1993
Drosophilidae (17 spp.)	DeSalle 1992a[a]
Prosimulium (2 spp.)	Xiong & Kocher 1991
Simulium (3 spp.)	Xiong & Kocher 1991
Simulium (5 spp.)	Xiong & Kocher 1993a[a]
Simulium (6 spp.)	Xiong & Kocher 1993b
Stegopterna mutata	Xiong & Kocher 1991
Alabagrus stigma	Derr et al. 1992
Allorhogas pyralophagus	Derr et al. 1992
Aphytis (2 spp.)	Derr et al. 1992
Cotesia flavipes	Derr et al. 1992
Digonogastra kimballi	Derr et al. 1992
Polistes versicolor	Derr et al. 1992
Tremex columba	Derr et al. 1992
Xanthopimpla stemmator	Derr et al. 1992

Additional information concerning the identify of many species that have been sequenced but not yet published is provided in the primer compilation *Appendix*.

[a] Studies that include tRNA sequences. Published studies of insect tRNAs not mentioned above include mosquitoes (Dubin et al. 1986) and black flies (Pruess et al. 1992).

[b] These works include intrapopulation and intraspecific variability analysis.

[c] See also complete mtDNA genome sequences at beginning of this table.

16S Gene. Patterns of variation in 16S appear to be similar to those in the 12S. The 3′ half of the 16S (domains 3 and 4) is highly conserved; many sections are identical between *Drosophila* and locust, and some smaller sections are identical with *E. coli* (Uhlenbusch et al. 1987, Gutell et al. 1992 [compilation]). Pashley & Ke (1992) sequenced the entire 16S gene of *Spodoptera frugiperda* and compared it with published sequences for *Drosophila, Aedes, Apis,* and *Locusta* The conserved 3′ half was easy to align for all species; the 5′ half of the more divergent *Locusta* could not be aligned to the other three taxa.

In addition to mammalian 16S studies (Miyamoto et al. 1990, Kraus & Miyamoto 1991, Allard et al. 1992, Milinkovitch et al. 1993), nucleotide data for the 16S gene have been used to determine phylogenetic relationships in several invertebrate taxa. These studies have either used the 3′ (domain 3) section amplified by the conserved primers LR-J-12887 + LR-N-13398, or both 16S domains 3 and 4 (Table 5).

DeSalle (1992a) used a 905-bp piece comprising 665 bp on the conserved 3′ end of the 16S gene, as well as the tRNAleu gene, and 141 bp on the 5′ end of the ND1 gene, to examine relationships at various levels within the genus *Drosophila.* He generally found that the 665 bp of 16S gene produced results similar to the 905-bp data set as a whole and that the data appeared problematic for determining relationships among species within the subgenera *Sophophora* and *Drosophila.* Most nodes were poorly supported by bootstrap analysis. Relationships at the deeper nodes within *Drosophila* were supported more strongly but did not agree with studies based on other data. The phylogenetic results were even more problematic using only transversions in that the bootstrap value increased for groupings that were in conflict with relationships based on morphology, larval hemolymph, or 2-D electrophoresis.

DeSalle (1992b) used the same 905-bp piece to study the placement of the Hawaiian *Drosophila* with respect to the rest of the family Drosophilidae. Although it was clear that the Hawaiian *Drosophila* were part of the subgenus *Drosophila* radiation, relationships within this radiation were resolved poorly as indicated by low bootstrap percentages. Again, the deeper nodes of the tree were supported more strongly than the majority of the shallower nodes.

Xiong & Kocher (1991) sequenced the 16S from four genera of blackflies containing seven species. Sequence divergence (uncorrected) ranged between ≈5–8% among genera and 4–6% among species within genera. Transversions predominated, with 65% transversional changes between genera and 56% between species within genera. The A+T bias of this region—77% compared with 82% in the same region of *Drosophila*—exaggerates the tendency toward transversions. The high transversional bias suggests that multiple hits within this region will reduce its phylogenetic usefulness at and above the species level for blackflies. This conclusion has been borne out by later studies.

Xiong & Kocher (1993b) used sequence data from the 3′ half of the 16S to investigate relationships in a complex of five morphologically indistinguishable species of black flies recognizable only by chromosomal analysis. The five cryptic species differed at only 24/460 bp (5.2%) of the sites sequenced. Analysis of intraspecific variation (10–30 individuals per species) showed even lower variation—most species containing only two common haplotypes. High levels of transversions observed in the comparisons among the five cryptic species contrasted with

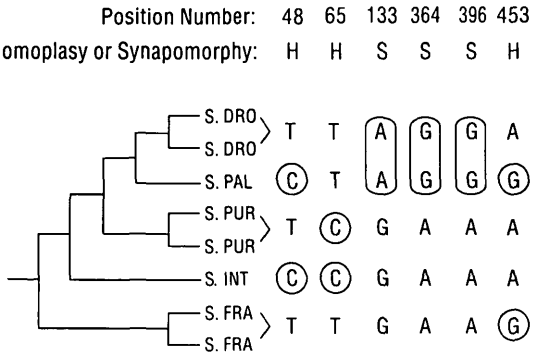


Fig. 3. The most parsimonious tree for selected sea urchin species in the genus *Strongylocentrotus* estimated based on five mitochondrial protein genes (Thomas et al. 1989, Kessing 1991) showing the character state distribution and presence or absence of homoplasy for all informative substitutions in a 500-bp segment from the 3′ half of the 12S rRNA gene (data from Thomas et al. 1989). Ovals identify synapomorphic character states.

low levels of transversions observed among individuals within species suggest that multiple substitutions have occurred. This in turn suggests either that these species may be more distantly related than was previously suspected or that the rate of evolution of the sites that are free to change is higher than in other species.

One other insect study has examined within-species nucleotide sequence variation in detail. Vogler et al. (1993a) and Vogler & DeSalle (1993) sequenced the same 3′ 16S segment as well as parts of the COIII and ND1 gene in four morphologically distinct subspecies of North American *Cincindela dorsalis* (tiger beetle) and an out-group species *Cincindela puritana* in another subgenus. They examined 78 individuals from 28 widespread populations. Comparisons among populations within species showed that transitions predominated and phylogenetic analysis of these populations was characterized by low homoplasy. On the other hand, among-species comparisons of the in-group to the out-group taxon showed that only 10 sites varied in the 16S gene, yet these sites appeared to contain high numbers of multiple hits as revealed by high levels of transversions among species. Thus, although these data appeared useful at the level of populations, they suggest that the 16S gene will not be useful for establishing relationships among species of *C. dorsalis*.

The above studies suggest that, like the 3′ half of the 12S, the 3′ half of the 16S is not very useful for phylogenetic studies of recently diverged populations within species because few sites vary. In many taxa, these same 3′ regions are also useless for studies of relationships among species, because of the few sites that do vary, many are plagued by multiple substitutions.

On the other hand, the conserved 3′ region of the 16S appears to be useful for studies of more distantly related taxa. DeSalle (1992a) found that among the Drosophilid genera *Chymomyza, Scaptodrosophila,* and *Hirtodrosophila,* there was general agreement between molecular (16S) and morphological data sets in the placement of these most basal taxa but not in the placement of the more derived taxa (as discussed above). Similarly, the same 3′ 16S region appears to be phylogenetically useful for tribes and families of Hymenoptera but may be less so for species in the genus *Apis* (see below) (Cameron 1991, 1993; Cameron et al. 1992; Derr et al. 1992). The same 3′ 16S segment was found to be useful in distinguishing tribes of *Deltocephalus*-like leafhoppers; however, other than six closely related couplets, relationships within tribes were not well resolved (Fang et al. 1993). In leafhoppers, as in blackflies, transversion rates were high, indicating that multiple substitutions may have been a problem at many sites.

Combining 12S and 16S Data. The strengths of the ribosomal genes appear to be at deep levels of divergence. Several studies have combined data from the 12S and 16S genes. In phylogenetic trees created for bovid tribes using entire 12S and 16S rRNA genes, many of the deeper nodes were strongly supported (Miyamoto et al. 1990, Allard et al. 1992). A study of 16 whale taxa using only the most conserved domains of these two ribosomal genes (Milinkovitch et al. 1993) showed deep branches of the tree to be resolved with high bootstrap values, using a variety of phylogenetic algorithms. As we would predict, many of the branches at the tips of the tree were collapsed into an unresolved multifurcation.

Not all deep-level comparisons will be resolvable even if conserved sequences are available. For example, phylogenetic relationships can be obscured if a group experiences a burst of speciation. This is caused by the shared inheritance of ancestral polymorphisms and the random extinction of these polymorphisms across species. This explanation was invoked by Hickson (1993) where sequence data from the third domain of the 12S gene failed to completely resolve phylogenetic relationships in New Zealand *Leiolopisma* skinks. A similar conclusion was reached by Kraus & Miyamoto (1991) for Pecoran ruminants using the entire 12S and 16S and adjacent transfer RNAs.

In summary, ribosomal genes are most likely to be useful at the population level where highly variable sites have not yet experienced multiple substitutions and at deep levels of divergence where the more conserved sites, which are likewise free of homoplasy, supply useful phylogenetic information. At intermediate levels of divergence, the sites that do vary are likely to have experienced multiple substitutions and phyloge-

netic signal may be obscured. These generalizations may not hold for all insects because rates and patterns of evolution of mitochondrial ribosomal genes can vary greatly among taxa (C. S., L. Nigro, J. Sullivan, A. Martin, A. Grapputo, A. Franke & C. McIntosh, unpublished data; W. Black, personal communication); however, if a researcher is choosing a gene de novo, a protein coding gene might be a better first choice for studying relationships among closely related species.

Transfer RNA Genes. Transfer RNA genes have been used less frequently for phylogenetic analysis because they are small and highly conserved. Because of this extreme conservation, they could prove useful in studies of distantly related species if data from many tRNA genes are combined. Kumazawa & Nishida (1993) demonstrated that phylogenetic analysis of transversions from highly conserved tRNA stem regions produced well-supported trees of vertebrate classes, which agreed with the accepted phylogeny based on anatomical, biochemical, and fossil evidence. Sea urchin was used as an out-group. In contrast, phylogenies constructed from Cytb and COI sequences did not agree or agreed only weakly with accepted relationships. They found that in these conserved tRNA genes transitions evolved linearly with time for nearly 100 million yr, while transversions evolved linearly with time for more than 350 million yr.

Protein Coding Genes. *COII Gene.* This gene is the most widely used mitochondrial protein coding gene in insects (Table 5). Beckenbach et al. (1993) found that nucleotide sequence data from the COII gene was useful for studying relationships within the *D. obscura* group when transitions were excluded from the nucleotide sequence analysis. In their analysis, two of the three subgroups were well-supported although relationships within subgroups in general were supported poorly. In contrast to morphological analyses but in agreement with mtDNA restriction site (Latorre et al. 1988) and DNA–DNA hybridization analyses (Goddard et al. 1990), the two North American subgroups were more closely related to each other than either was to the European.

In the *D. obscura* group, nucleotide sequence divergence (Jukes-Cantor corrected) ranged from 1 to 11% and amino acid divergence ranged from 0 to 4%. A comparison of these data with the COII gene of six species of the genus *Apis* shows similar levels of nucleotide sequence divergence (0.4–10%, uncorrected) but very different levels of amino acid sequence divergence (1–10%) (Willis et al. 1992), supporting earlier conclusions that the rate of evolution of *A. melifera* is rapid compared with *Drosophila* (Crozier et al. 1989).

F.F. and C.S. (unpublished data) sequenced the complete COII gene for seven species representing four collembolan genera; each genus belongs to a different family but all are in the same suborder. All comparisons showed Kimura (1980) corrected nucleotide sequence divergence >20% among species within genera and >30% among species belonging to different families. Approximately 56% of the variable sites were third-codon positions. Amino acid similarity among species within a genus ranged from 98% (genus *Isotomurus*, two species) to 88–96% (genus *Orchesella*, five species). Amino acid similarity among the four families ranged from 62 to 82%. These COII amino acid similarity values are as low as the COII values for orders of insects (Liu & Beckenbach 1992; see below). All comparisons indicate that collembolan genera, families, and some species are either ancient or evolve very rapidly. Comparing these results with those obtained from comparisons among other insect species belonging to a single genus (Willis et al. 1992, Beckenbach et al. 1993), we see an excellent illustration of the well-known fact that genera (or any taxon above species) are not genetically equivalent and, thus, no blanket generalizations can be made about the taxonomic level at which specific genes are useful.

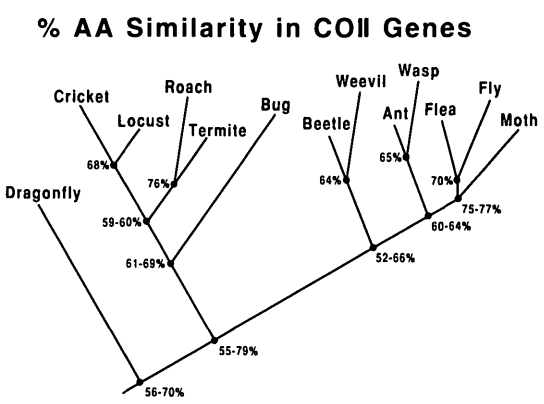


Fig. 4. Percentage of amino acid similarity of COII genes from pairwise comparisons of species from different orders of insects (Liu & Beckenbach 1992) superimposed on a phylogenetic tree hypothesizing insect ordinal relationships as suggested by Hennig 1981.

At deeper levels in insects, Liu & Beckenbach (1992) demonstrated that amino acid substitutions in the COII gene were useless for constructing relationships among 10 insect orders. Amino acid similarity values ranged from 56 to 77% and were uncorrelated with degree of taxonomic divergence (Fig. 4). For example, a dragonfly and a fly shared 70% of COII gene amino acids, whereas, an ant and a wasp were only 65% similar. A cricket and a locust shared 68% of their amino acids, whereas, the same cricket and a moth shared 79%. These results suggest that convergence has taken place in proteins that have diverged by only 30%. An examination of the

amino acid variability represented in Table 1 suggests that all protein coding genes other than COI evolve faster than COII at the amino acid level. Thus, even slowly evolving proteins, those in which the majority of amino acid positions are highly constrained, are not useful for discovering deep-level relationships. The few amino acid positions that have changed appear to have changed rapidly (see also Lewontin 1989).

Phylogenetic analysis of insect orders could perhaps be improved by sampling more species per order. This would lessen the problem of long, unbranched lineages grouping together based on convergent character states (Hendy & Penny 1989) and perhaps identify the precise point between genera and orders where phylogenetic information becomes obscured. The depth at which the information disappears will vary for nucleotides versus amino acids and we suspect will vary among amino acid positions and among taxa.

COI Gene. The COI gene has been sequenced less frequently than the COII gene (Table 5). It is the most conserved gene in terms of amino acid evolution (Table 1). Nigro et al. (1991) sequenced 503 bp at the 5′ end of the COI gene for two species of *Drosophila* and 263 bp of a third species for this same gene. These data were combined with 1,000 bp from the ND2 gene to create a phylogenetic tree (discussed below in the ND2 section). The percentage of polymorphic sites was 14.4% for the COI gene and 18.8% for the ND2 gene. As discussed earlier, variation compared between the two genes was similar at silent sites but very different at replacement sites with only 1% of the amino acids replaced in COI versus 19.6% replaced in ND2.

F.F. (unpublished data) sequenced the 3′ half (522 bp) of the COI gene from six species (one–six populations each) in two genera of Collembola. The two genera belonged to different families. Kimura-corrected nucleotide sequence divergence values ranged from 1 to 6% among populations within species and from 13 to 25% among species within a genus. The percentage of divergence between genera was considerably lower in the COI gene (23–28%) than it was in the COII gene (40–43%). The fact that the transition bias, which is apparent when populations are compared, is significantly reduced in among-species comparisons suggests that species are old or that rapid evolution has taken place. Phylogenetic analysis is in preparation.

B.C. (unpublished data) sequenced a 412-bp piece of COI for seven genera of Thysanoptera in four families, plus a cicada, two aphids, and the hemipteran *Oncopeltus fasciatus.* Among three genera within the family Thripidae, nucleotide divergences ranged from 17 to 21%. In contrast, divergences ranged from 23 to 34% both between the four families within Thysanoptera (average 30%) and between the Thysanoptera and the Hemiptera–Homoptera (average 28%) (B.C. & L. Vawter, unpublished data). Similarly, amino acid divergences were 10% within the Thripidae, 16–37% among families of Thysanoptera (average 29%), and 20–36% between the two orders (average 27%). Phylogenetic analysis demonstrated good support for the monophyly of the family Thripidae, but relationships among the families remained largely obscure. These results from COI are consistent with those of F.F. (unpublished data) and Liu & Beckenbach (1992) for COII, in supporting a general lack of usefulness of mitochondrial protein-coding genes for phylogenetics at the among-family and among-order levels.

Combined COI and COII Studies. Sequence data from the 3′ half of the COI gene and the entire COII gene collected by Sperling & Hickey (1994) proved very useful in studying the evolution of six spruce budworm species. The two genes showed similar levels of silent and replacement variation among species. A parsimony analysis (PAUP, Swofford 1993) of 10 specimens from six named taxa produced a well-supported tree (CI = 0.814; bootstraps all≥85%), which revealed two unexpected well-differentiated haplotypes possibly representing unidentified cryptic species. Evidence was also provided for very recent, incomplete differentiation in three of the taxa formerly considered to be good species. Forty-seven specimens representing these same taxa were sequenced over a 470-bp region of the COI gene to examine within-species variation. Variation was found in all species and was characterized by individuals with one or two substitutional differences from each other. In the two species with sample sizes of 9 or more individuals, the pattern of variation was very different: one species was represented by seven genotypes (two of which had 2 individuals each) and the other by four genotypes with 15 individuals of a common haplotype plus 3 individuals representing three rare haploptypes each differing by one substitution from the common type. Thus, like vertebrates (Avise et al. 1987), insects show a variety of species specific patterns of geographic variability.

Brower (1994) sequenced ≈800 bases from the 3′ end of the COI gene, the tRNA leucine gene and most of the COII gene for 37 species in the genus *Heliconius* (sensu lato) and five Heliconiine out-group taxa. Pairwise sequence divergence ranged from 0.5% between conspecific races to 13% between in-group and out-group taxa. Although uniformly weighted character-based parsimony analysis produced 70 equally short trees, the strict consensus of these trees still showed considerable structure. His analyses demonstrated that third positions contained many homoplasious substitutions at deeper levels of the tree, nevertheless, weighting in favor of second and first positions changed the topol-

ogy only slightly. This result indicated that the data contained a strong phylogenetic signal that was not overcome by the homoplasy present. A comparison with morphologically-based phylogenetic analysis found that, although the traditional view of relationships in *Heliconius* was largely supported, only 13 of the 33 comparable nodes were concordant. Several new hypotheses of relationships were suggested than can be tested with further data.

Brown et al. (1994) sequenced a 765-bp segment spanning ≈159 bp on the 3′ end of the COI gene, the tRNA leucine gene, and 545 bp on the 5′ end of the COII gene for 16 species in the prodoxid moth genus *Greya* and two out-group genera. Corrected percentage of sequence divergence between pairs of the in-group ranged up to 16%. Again, parsimony analysis of these mitochondrial data resulted in substantial congruence with previously proposed morphologically based relationships. The majority of nodes were strongly supported by bootstrap analysis. The exceptions were largely at deep levels. Five of the six most basal nodes showed bootstraps <50% suggesting homoplasy at the amino acid level. As in the Heliconiine analysis, trees produced by successive weighting, transversion weighting (2:1), or weighting second positions more heavily were not different from uniformly weighted trees. However, in this study and that of Brower (1994), complete elimination of third positions destroyed much of the resolution.

ND1 Gene. Pashley & Ke (1992) sequenced a 314-bp fragment at the 5′ end of the ND1 gene in five species (four families) of Lepidoptera. Parsimony analysis conducted with several progressively closer out-groups and with different subsets of characters, although maintaining the correct relationships among the orders (Orthoptera, Diptera, and Lepidoptera) was never able to recover the accepted phylogeny below the ordinal level. Data analysis showed a strong transversional bias (70–76% of total substitutions) and a high level of homoplasy.

DeSalle et al. (1987) compared partial ND1, ND2, and ND5 sequences from several species of Hawaiian and continental *Drosophila* and a mosquito (*Aedes albopictus*). They estimated a 16:1 TS:TV bias between closely related species and demonstrated saturation of the sequences over time (a bias of 1:1 in more distant taxa). They concluded that in *Drosophila* silent sites are saturated at 8% sequence divergence. Their parsimony-based phylogenetic analysis of Hawaiian *Drosophila* using the continental *Drosophila* species as an out-group was consistent with phylogenies based on morphological characters and antigenic comparisons.

DeSalle (1992a, b) also sequenced the ND1 gene (in conjunction with the 16S rDNA) in other drosophilid species, but the small size of the fragment sequenced did not provide any phylogenetic resolution because of the lack of informative sites.

ND2 Gene. Nigro et al. (1991) sequenced the 3′ end of the ND2, three tRNAs, and the 5′ end of the COI gene from *Drosophila erecta* and *Drosophila eugracilis* (>1,300 bp) and gathered data from the literature for this same region in four other species: *D. melanogaster* (DeBruijn 1983), *D. yakuba* (Clary & Wolstenholme 1985), and *Drosophila simulans* and *Drosophila mauritiana* (Satta et al. 1987). Character-based parsimony (PAUP) and distanced based analysis (NJ) produced a tree consistent with data from morphological characters and DNA–DNA hybridization distances (Lemeunier et al. 1986, Caccone et al. 1988). Analysis of first and second positions alone (all substitutions but one in the ND2 gene) produced the same tree. No information was given on the level of support for each node nor for the entire tree.

Other Protein Coding Genes. The cytochrome b gene has been used extensively in vertebrates because of the availability of highly conserved primers that work well to amplify a wide variety of species (Kocher et al. 1989, Irwin et al. 1991). To date, this gene has not been used for phylogenetic studies of insects, although some are in progress (D. R. Frohlich, personal communication). The COIII gene and the ATPase genes have been sequenced in only a few taxa (Table 5) and have not been used for phylogenetic studies published to date. It will be interesting to study the phylogenetic usefulness of amino acid substitutions in the ATPase genes for studies of taxa exhibiting close and intermediate levels of divergence, because in these genes more amino acid sites are free to vary (Table 1).

Comparisons of rRNA and Protein Coding Genes. Several factors argue that protein coding genes may be more appropriate than ribosomal genes for phylogenetic analysis for close and intermediate levels of divergence. These factors are related to the fact that we have a better understanding of how the nucleotide positions of codons evolve than we do of how individual ribosomal stems and loops evolve. Furthermore, at intermediate levels of divergence, amino acid sequence data can be useful even when silent sites of codons and variable rRNA positions are saturated. As discussed above, patterns of conservation of rRNA stems and loops are difficult to define at intermediate and close levels of divergence. In addition, protein coding genes are easier to align than ribosomal genes because they have amino acids that aid the process and because they have fewer large deletions.

For analysis of relationships at deep levels of divergence, protein coding genes may be less useful than rRNA genes because highly conserved proteins are likely to have only two classes of sites: the less constrained amino acid sites that have many multiple substitutions and

the highly constrained amino acid sites that have no substitutions. Empirical studies are needed for more genes and taxa to establish which amino acid positions are likely to be slowly evolving.

Comparisons of ribosomal and protein coding genes for the same taxa can allow assessments of relative phylogenetic usefulness, as in the urchin comparison discussed above (Fig. 3); however, it is not always clear which data set provides the most reliable tree. For example, phylogenetic relationships of honey bee species have been studied using both ribosomal (16S) and protein coding (COII) genes. Willis et al. (1992) demonstrated that in the COII gene, nucleotide sequence data contained considerable homoplasy when used in a phylogenetic analysis of six species in the genus *Apis*. This result was reflected in low bootstrap values at all nodes and is probably caused by the accelerated rate of evolution in *Apis* (Crozier et al. 1989). Amino acid substitution data, on the other hand, produced a well-supported tree with little homoplasy and high bootstrap values at most nodes. Cameron et al. (1992) examined the commonly sequenced 3′ segment of the 16S gene for five of these same *Apis* species. The tree produced from these data was not congruent with the COII amino-acid-based phylogeny. The 16S tree indicated strong bootstrap support for two nodes not found on the amino acid tree and weak support for a third node. Although the use of a very distant outgroup (wasp) in the COII analysis could partially explain the disagreement, other factors must be operating because even unrooted trees are not congruent for well-supported nodes. The high level of nucleotide sequence variability found for *Apis*, the high A+T bias (>80% on average in the 3′ 16S segment), the low bootstrap support for one of the nodes, and the general problematic nature of the 3′ 16S region (discussed above), suggest that the 16S tree may be incorrect. However, the 16S tree agrees completely with that obtained from analyses of morphological characters (Alexander 1991) and, therefore, cannot be dismissed so easily. In fact, a reanalysis of the Willis et al. (1992) COII data using the spectral analysis method (Penny et al. 1992, Steel et al. 1993) with a new *log determinant* transformation designed to correct for unequal nucleotide composition, produced a tree congruent with 16S and morphological data (Lockhart et al. 1994).

Phylogenetic Usefulness of Mitochondrial Genes

The mtDNA-sequence data have proven to be valuable tools in phylogenetic analysis. In general, silent sites of mitochondrial protein coding genes appear to be useful for studying relationships of recently diverged taxa, and these sites appear to evolve at the same rate in all mitochondrial genes when artifactual biases are removed. In addition, unconstrained amino acid positions may provide information that is useful for studying relationships among recently separated species or populations. Mitochondrial rRNA genes, especially the conserved 3′ halves, have fewer sites that vary among closely related species and, therefore, are less useful. In addition, strong constraints that inhibit substitutions in rRNA can give a false impression of the degree of divergence among species.

For intermediate levels of divergence where silent sites contain many multiple hits, amino acid changes or nonsynonymous nucleotide sites can be useful for constructing phylogenetic relationships. Some mitochondrial protein coding genes are more highly constrained than others, limiting the number of amino acid substitutions. Strong amino acid conservation does not guarantee that the few sites that change are free of multiple amino acid substitutions. The same multiple hits problems that face nucleotides can affect amino acids and must be corrected or avoided.

For deep levels of divergence, mitochondrial protein coding genes may be saturated at the amino acid level. Highly conserved regions of mitochondrial rRNA and tRNA genes may be useful because they contain few multiple substitutions. The problem is to identify with certainty the conservative changes. Preliminary studies will then be needed to determine whether there are sufficient numbers of conservative changes to produce a well-supported tree.

Because of the arbitrariness of taxonomic categories, no sweeping generalizations can be made about the taxonomic rank at which particular mitochondrial genes are useful. Divergence time, if it were known, would be a better predictor of degree of genetic divergence. In addition, some insect species may exhibit high levels of nucleotide sequence divergence among populations, whereas others may show little variation. In combination with the age of the species, population genetic factors such as population size can influence variation.

As more and more DNA-sequence data accumulate we will be able to gain an even better understanding of the way in which genes and species evolve. Rapidly advancing sequencing technology promises abundant data in the near future. The attached *Appendix* of conserved primers will allow sequencing of most mitochondrial genes in most insect taxa.

Acknowledgments

We thank A. Brower, W. Black, S. Cameron, R. Crozier, R. DeSalle, C. Dietrich, J. Felsenstein, N. Goldman, D. Hillis, D. Higgins, K. Holsinger, D. Norris, G. Olsen, D. Pashley, J. Sullivan, D. Swofford, and J. Wakeley for helpful comments on parts of the text. S. Bogdanowicz, W. Black, A. Brower, J. M. Brown, S.

Cameron, Y. C. Crozier, R. Crozier, B. Farrell, R. Harrison, L. Jermiin, T. King, T. Kocher, B. Normark, D. Pashley, A. Phillips, T. Powers, R. Roehrdanz, F. Sperling, G. Spicer, G. Wallis, S. Weller, and M. M. Wells generously contributed primers, data, notes or comments for the *Appendix*. J. Brown, D. Frohlich, D. Lunt, B. McPheron, J. Smith, J. Strassman, and D. Wesson kindly responded to queries concerning tables. M. J. Spring, J. Dooley, and C. Champ assisted with illustrations. The late A. C. Wilson provided C.S. with support and the opportunity to apply PCR to insect species. This research was supported in part by grants from the National Science Foundation (to C.S. and to B.C.), Ministero dell'Universita e Della Ricerca Scientifica e Tecnologica and Consiglio Nazionale delle Ricerche (to F.F.), and Natural Sciences and Engineering Research Council of Canada (to A.B. and to B.C.).

References Cited

Alexander, B. A. 1991. Phylogenetic analysis of the genus *Apis* (Hymenoptera: Apidae). Ann. Entomol. Soc. Am. 84: 137–149.

Allard, M. W. & M. M. Miyamoto. 1992. Testing phylogenetic approaches with empirical data as illustrated with the parsimony method. Mol. Biol. Evol. 9: 778–786.

Allard, M. W., M. M. Miyamoto, L. Jarecki, F. Kraus & M. R. Tennant. 1992. DNA systematics and evolution of the artiodactyl family Bovidae. Proc. Natl. Acad. Sci. U.S.A. 89: 3972–3976.

Anderson, S., A. Bankier, B. Barrell, M. DeBruijn, A. Coulson, J. Drouin, I. Speron, D. Nierlich, B. Roe, F. Sanger, P. Schreier, A. Smith, R. Staden & I. Young. 1981. Sequence and organization of the human mitochondrial genome. Nature (Lond.) 290: 457–465.

Anderson, S., M. DeBruijn, A. Coulson, I. Eperon, F. Sanger & I. Young. 1982. Complete sequence of bovine mitochondrial DNA. Conserved features of the mammalian mitochondrial genome. J. Mol. Biol. 156: 683–717.

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

Aquadro, C. F., N. Kaplan & K. J. Risko. 1984. An analysis of the dynamics of mammalian mitochondrial DNA sequence evolution. Mol. Biol. Evol. 1: 423–434.

Archie, J. 1989. randomization test for phylogenetic information in systematic data. Syst. Zool. 38: 239–252.

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

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

Attardi, G. & G. Schatz. 1988. Biogenesis of mitochondria. Annu. Rev. Cell Biol. 4: 289–333.

Avise, J. C. 1989. Gene trees and organismal histories: a phylogenetic approach to population biology. Evolution 43: 1192–1208.

Avise J. C., J. Arnold, R. M. Ball, E. Bermingham, T. Lamb, J. E. Neigel, C. A. Reeb & N. C. Saunders. 1987. Intraspecific phylogeography: the mitochondrial bridge between population genetics and systematics. Annu. Rev. Ecol. Syst. 18: 489–522.

Ballard, J.W.O. 1994. Sequence data resolves a morphological conundrum in *Austrosimulium* (Diptera: Simuliidae). J. Aust. Entomol. Soc. (in press).

Ballard, J.W.O., G. J. Olsen, D. P. Faith, W. A. Odgers, D. M. Rowell & P. W. Atkinson. 1992. Evidence from 12S ribosomal RNA sequences that onychophorans are modified arthropods. Science (Washington, DC) 258: 1345–1348.

Beard, C. B., D. Mills Hamm & F. H. Collins. 1993. The mitochondrial genome of the mosquito *Anopheles gambiae:* DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects. Insect Mol. Biol. 2: 103–124.

Beckenbach, A. T., Y. W. Wei & H. Liu. 1993. Relationships in the *Drosophila obscura* species group, inferred from mitochondrial cytochrome oxidase II sequences. Mol. Biol. Evol. 10: 619–634.

Bernardi, G. 1993. The vertebrate genome: isochores and evolution. Mol. Biol. Evol. 10: 186–204.

Bernardi, G. & G. Bernardi. 1986. Compositional constraints and genome evolution. J. Mol. Evol. 24: 1–11.

Beverly, D. M. & A. C. Wilson. 1985. Ancient origins for Hawaiian Drosophilinae inferred from protein comparisons. Proc. Natl. Acad. Sci. U.S.A. 82: 4753–4757.

Bibb, M. J., R. A. Van Etten, C. T. Wright, M. W. Walberg & D. A. Clayton. 1981. Sequence and gene organization of mouse mitochondrial DNA. Cell 26: 167–180.

Boyce, T. M., M. E. Zwick & C. F. Aquadro. 1989. Mitochondrial DNA in the bark weevils: size, structure and heteroplasmy. Genetics 123: 825–836.

Brimacombe, R., B. Greuer, P. Mitchell, M. Osswald, J. Rinke-Appel, D. Schuler & K. Stade. 1990. Three-dimensional structure and function of *Escherichia coli* 16S and 23S rRNA as studied by cross-linking techniques, pp. 73–92. *In* W. E. Hill, A. Dahlberg, R. A. Garrett, P. B. Moore, D. Schlesinger & J. R. Warner [eds.], The ribosome: structure, function and evolution. American Society of Microbiology, Washington, DC.

Britten, R. J. 1993. Forbidden synonymous substitutions in coding regions. Mol. Biol. Evol. 10: 205–220.

Brower, A.V.Z. 1994. Phylogeny of *Heliconius* butterflies inferred from mitochondrial DNA sequences (Lepidoptera: Nymphalidae). Mol. Phylogenet. Evol. 3: 159–174.

Brower, A.V.Z. & R. DeSalle. 1994. Practical and theoretical considerations for choice of a DNA sequence region in insect molecular systematics, with a short review of published studies using nuclear gene regions. Ann. Entomol. Soc. Am. 87: 702–716.

Brown, J. M., O. Pellmyr, J. N. Thompson & R. G. Harrison. 1994. Phylogeny of *Greya* (Lepidoptera: Prodoxidae) based on nucleotide sequence variation in mitochondrial cytochrome oxidase I and II: congruence with morphological data. Mol. Biol. Evol. 11: 128–141.

Brown, W. M. 1985. The mitochondrial genome of animals, pp. 95–130. *In* R. J. MacIntyre [ed.], Molecular evolutionary genetics. Plenum, New York.

Brown, W. M., M. George, Jr., & A. C. Wilson. 1979. Rapid evolution of animal mitochondrial DNA. Proc. Natl. Acad. Sci. U.S.A. 76: 1967–1971.

Brown, W. M., E. M. Prager, A. Wang & A. C. Wilson. 1982. Mitochondrial DNA sequences of primates: tempo and mode of evolution. J. Mol. Evol. 18: 225–239.

Brusca, R. C. & G. J. Brusca. 1990. Invertebrates. Sinauer, Sunderland, MA.

Bull, J. J., J. P. Huelsenbeck, C. W. Cunningham, D. L. Swofford & P. J. Waddell. 1993. Partitioning and combining data in phylogenetic analysis. Syst. Biol. 42: 384–397.

Caccone, A., G. D. Amato & J. R. Powell, Jr. 1988. Rates and patterns of scnDNA and mtDNA divergence within the *Drosophila melanogaster* subgroup. Genetics 118: 671–683.

Cameron, S. A. 1991. A new tribal phylogeny of the apidae inferred from mitochondrial DNA sequences, pp. 71–87. *In* D. R. Smith [ed.], Diversity in the genus *Apis*. Westview, Boulder, CO.

1993. Multiple origins of advanced eusociality in bees inferred from mitochondrial DNA sequences. Proc. Natl. Acad. Sci. U.S.A. 90: 8687–8691.

Cameron, S. A., J. N. Derr, A. D. Austin, J. B. Woolley & R. A. Wharton. 1992. The application of nucleotide sequence data to phylogeny of the Hymenoptera: a review. J. Hymenopteran Res. 1: 63–79.

Cann, R. & A. C. Wilson. 1983. Length mutations in human mitochondrial DNA. Genetics 104: 699–711.

Cann, R., W. M. Brown & A. C. Wilson. 1984. Polymorphic sites and the mechanism of evolution in human mitochondrial DNA. Genetics 106: 479–499.

Cann, R. L., M. Stoneking & A. C. Wilson. 1987. Mitochondrial DNA and human evolution. Nature (Lond.) 325: 31–36.

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

Cantatore, P., M. Roberti, G. Rainaldi, M. N. Gadaleta & C. Saccone. 1989. The complete nucleotide sequence, gene organization, and genetic code of the mitochondrial genome of *Paracentrotus lividus*. J. Biol. Chem. 264: 10965–10975.

Carpenter, F. M. & L. Burnham. 1985. The geological record of insects. Annu. Rev. Planetary Sci. 13: 297–314.

Chan, S. C., A.K.C. Wong & D.K.Y. Chiu. 1992. A survey of multiple sequence comparison methods. Bull. Math. Biol. 54: 563–598.

Chang Y.-S., F.-L. Huang & T.-B. Lo. 1994. The complete nucleotide sequence and gene organization of carp (*Cyprinus carpio*) mitochondrial genome. J. Mol. Evol. 38: 138–155.

Clary, D. O. & D. R. Wolstenholme. 1985. The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. J. Mol. Evol. 22: 252–271.

1987. *Drosophila* mitochondrial DNA: conserved sequence in the A+T rich region and supporting evidence for a secondary structure model of the small ribosomal RNA. J. Mol. Evol. 25: 116–125.

Collins, T. M., F. Kraus & G. Estabrook. 1994a. Compositional effects and weighting of nucleotide sequences for phylogenetic analysis. Syst. Biol. 43: 449–459.

Collins, T. M., P. H. Wimberger & G.J.P. Naylor. 1994b. Compositional bias, character-state bias, and character-state reconstruction using parsimony. Syst. Biol. (in press).

Cornuet, J.-M., L. Garnery & M. Solignac. 1991. Putative origin and function of the intergenic region between COI and COII of *Apis mellifera* L. mitochondrial DNA. Genetics 128: 393–403.

Coulondre, C., J. Miller, P. Farabaugh & W. Gilbert. 1978. Molecular basis of base substitution hotspots in *Escherichia coli*. Nature (Lond.) 274: 775–780.

Crews S., D. Ojala, J. Posakony, J. Nishiguchi & G. Attardi. 1978. Nucleotide sequence of a region of human mitochondrial DNA containing the precisely identified origin of replication. Nature (Lond.) 277: 192–198.

Croom, H. B., R. G. Gillespie & S. R. Palumbi. 1991. Mitochondrial DNA sequences coding for a portion of the RNA of the small ribosomal subunits of *Tetragnatha hawaiensis* (Araneae, Tetragnathidae). J. Arachnol. 19: 210–214.

Crozier, R. H. 1990. From population genetics to phylogeny: Uses and limits of mitochondrial DNA. Aust. Syst. Bot. 3: 111–124.

1993. Molecular methods for insect phylogenetics, pp. 164–221. *In* J. Oakeshott & M. J. Whitten [eds.], Molecular approaches to fundamental and applied entomology. Springer, Berlin.

Crozier, R. H. & Y. C. Crozier. 1992. The cytochrome b and ATPase genes of honeybee mitochondrial DNA. Mol. Biol. Evol. 9: 474–482.

1993. The mitochondrial genome of the honeybee *Apis mellifera:* complete sequence and genome organization. Genetics 133: 97–117.

Crozier, R. H., Y. C. Crozier & A. G. Mackinlay. 1989. The CO-I and CO-II region of honeybee mitochondrial DNA: evidence for variation in insect mitochondrial evolutionary rates. Mol. Biol. Evol. 6: 399–411.

Dayhoff, M. O., R. M. Schwartz & B. C. Orcutt. 1978. A model of evolutionary change in proteins, pp. 345–352. *In* M. O. Dayhoff [ed.], Atlas of protein sequence and structure, vol. 5, suppl. National Biomedical Research Foundation, Silver Spring, MD.

DeBruijn, M.H.L. 1983. *Drosophila melanogaster* mitochondrial DNA, a novel organization and genetic code. Nature (Lond.) 304: 234–241.

De Rijk, P, J.-M. Neefs, Y. Van de Peer & R. De Wachter. 1993. Compilation of small ribosomal subunit RNA sequences. Nucleic Acids Res. 20: 2075–2089 (suppl.).

DeSalle, R. 1992a. The phylogenetic relationships of flies in the family Drosophilidae deduced from mtDNA sequences. Mol. Phylogenet. Evol. 1: 31–40.

1992b. The origin and possible time of divergence of the Hawaiian Drosophilidae: evidence from DNA sequences. Mol. Biol. Evol. 9: 905–916.

DeSalle, R., T. Freedman, E. M. Prager & A. C. Wilson. 1987. Tempo and mode of sequence evolution in mitochondrial DNA of Hawaiian *Drosophila*. J. Mol. Evol. 26: 157–164.

DeSalle, R., J. Gatesy, W. Wheeler & D. Grimaldi. 1992a. DNA sequences from a fossil termite in

Oligo-Miocene amber and their phylogenetic implications. Science (Washington, DC) 257: 1933–1936.

DeSalle, R., A. Williams & M. George. 1992b. Isolation and manipulation of animal mitochondrial DNA. Methods Enzymol. 224: 176–203.

Derancourt, J., A. S. Lebor & E. Zukerkandl. 1967. Sequence nucleotides et evolution. Bull. Soc. Chim. Biol. 49: 577–603.

Derr, J. N., S. K. Davis, J. B. Woolley & R. A. Wharton. 1992. Reassessment of the 16S rRNA nucleotide sequence from members of the parasitic Hymenoptera. Mol. Phylogenet. Evol. 1: 338–341.

Disotell, T. R., R. L. Honeycutt & M. Ruvolo. 1992. Mitochondrial DNA phylogeny of the old-world monkey tribe Papionini. Mol. Biol. Evol. 9: 1–13.

Dixon, M. T. & D. M. Hillis. 1993. Ribosomal RNA secondary structure: compensatory mutations and implications for phylogenetic analysis. Mol. Biol. Evol. 10: 256–267.

Donoghue, M. J. & M. J. Sanderson. 1992. The suitability of molecular and morphological evidence in reconstructing plant phylogeny, pp. 14–35. *In* P. S. Soltis, D. E. Soltis & J. J. Doyle [eds.], Molecular systematics of plants. Chapman & Hall, New York.

Downie, S. R. & J. D. Palmer. 1992. Use of chloroplast DNA rearrangements in reconstructing plant phylogeny, pp. 14–35. *In* P. S. Soltis, D. E. Soltis & J. J. Doyle [eds.], Molecular systematics of plants. Chapman & Hall, New York.

Doyle, J. J. 1992. Gene trees and species trees: molecular systematics as one-character taxonomy. Syst. Bot. 17: 144–163.

Dubin, D. T., C.-C. HsuChen & L. E. Tillotson. 1986. Mosquito mitochondrial transfer RNAs for valine, glycine, and glutamate: RNA and gene sequences and vicinal genome organization. Curr. Genet. 10: 701–707.

Faith, D. P. 1991. Cladistic permutation tests for monophyly and nonmonophyly. Syst. Zool. 40: 366.

Fang, Q., W. C. Black, Jr., H. D. Blocker & R. F. Whitcomb. 1993. A phylogeny of New World Deltocephalus-like leafhopper (Homoptera: Cicadellidae) Genera based on mitochondrial 16S ribosomal DNA sequences. Mol. Phylogenet. Evol. 2: 119–131.

Farris, J. S. 1969. A successive approximations approach to character weighting. Syst. Zool. 18: 374–385.

1977. On the phenetic approach to vertebrate classification, pp. 823–850. *In* M. K. Hecht, P. C. Goody & B. M. Hecht [eds.], Major patterns in vertebrate evolution. Plenum, New York.

Farris, J. S., A. G. Kluge & M. J. Eckardt. 1970. A numerical approach to phylogenetic systematics. Syst. Zool. 19: 172–189.

Fauron, C.M.-R. & D. R. Wolstenholme. 1980. Intraspecific diversity of nucleotide sequences within the adenine + thymine-rich region of mitochondrial DNA molecules of *Drosophila mauritiana, Drosophila melanogaster* and *Drosophila simulans.* Nucleic Acids Res. 8: 5391–5410.

Felsenstein, J. 1988. Phylogenies from molecular sequences: inference and reliability. Annu. Rev. Genet. 22: 521–565.

1993. Phylogeny inference package (PHYLIP), version 3.5c. Department of Genetics, University of Washington, Seattle.

Felsenstein, J. & H. Kishino. 1993. Is there something wrong with the bootstrap on phylogenies? A reply to Hillis and Bull. Syst. Biol. 42: 193–200.

Feng D.-F. & R. F. Doolittle. 1987. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J. Mol. Evol. 25: 351–360.

1990. Progressive alignment and phylogenetic tree construction of protein sequences, pp. 375–387. *In* R. F. Doolittle [ed.], Molecular evolution: computer analysis of protein and nucleic acid sequences. Methods in enzymology 183. Academic, New York.

Field, K. G., G. J. Olsen, D. J. Lane, S. J. Giovannoni & M. T. Ghiselin. 1988. Molecular phylogeny of the animal kingdom. Science (Washington, DC) 239: 748–752.

Fitch, D.H.A. & L. D. Strausbaugh. 1993. Low codon bias and high rates of synonymous substitution in *Drosophila hydei* and *D. melanogaster* histone genes. Mol. Biol. Evol. 10: 397–413.

Fitch, W. M. 1971. Rate of change of concomitantly variable codons. J. Mol. Evol. 1: 84–96.

1986. The estimate of total nucleotide substitutions from pairwise differences is biased. Philos. Trans. R. Soc. (Lond.) B Biol. Sci. 312: 317–324.

Fitch, W. M. & E. Markowitz. 1970. An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution. Biochem. Genet. 4: 579–593.

Fitch, W. M. & J. Ye. 1991. Weighted parsimony: does it work? pp. 147–154. *In* M. M. Miyamoto & J. Cracraft [eds.], Phylogenetic analysis of DNA sequences. Oxford University Press, Oxford.

Flook, P. K., C.H.F. Rowell & G. Gellissen. 1995. The sequence, organization, and evolution of the *Locusta migratoria* mitochondrial genome. J. Mol. Evol. (in press).

French, S. & B. Robson. 1983. What is a conservative substitution. J. Mol. Evol. 19: 171–175.

Frohlich, D. R., A. S. Robinson & M. A. Wells. 1993. Mediterranean fruit fly, *Ceratitis capitata* (Wiedemann), mitochondrial DNA: genes and secondary structures for six tRNAs. Insect Mol. Biol. 1: 165–169.

Gadaleta, G., G. Pepe, G. DeCandia, C. Quagliariello, E. Sbisa & C. Saccone. 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.

Garesse, R. 1988. *Drosophila melanogaster* mitochondrial DNA: gene organization and evolutionary considerations. Genetics 118: 649–663.

Garnery, L., J.-M. Cornuet & M. Solignac. 1992. Evolutionary history of the honey bee *Apis mellifera* inferred from mitochondrial DNA analysis. Mol. Ecol. 1: 145–154.

Gatesy, J., R. DeSalle & W. Wheeler. 1993. Alignment-ambiguous nucleotide sites and the exclusion of systematic data. Mol. Phylogenet. Evol. 2: 152–157.

Gillespie, J. H. 1991. The causes of molecular evolution. Oxford University Press, New York.

Goddard, K., A. Caccone & J. R. Powell. 1990. Evolutionary implications of DNA divergence in the *Drosophila obscura* group. Evolution 44: 1656–1670.

Gojobori, T., W.-H. Li & D. Graur. 1982. Patterns of nucleotide substitution in pseudogenes and functional genes. J. Mol. Evol. 18: 360–369.

Golding, G. B. 1983. Estimates of DNA and protein sequence divergence: an examination of some assumptions. Mol. Biol. Evol. 1: 125–142.

Goloboff, P. A. 1993. Estimating character weights during tree-search. Cladistics 9: 83–91.

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

Gu, W. & W.-H. Li. 1992. Higher rates of amino acid substitution in rodents than in humans. Mol. Phylogenet. Evol. 1: 211–214.

Gutell, R. R., B. Weiser, C. R. Woese & H. F. Noller. 1985. Comparative anatomy of 16-S-like ribosomal RNA. Prog. Nucleic Acid Res. Mol. Biol. 32: 155–216.

Gutell, R. R., M. N. Schnare & M. W. Gray. 1992. A compilation of large subunit (23S- and 23S-like) ribosomal RNA structures. Nucleic Acids Res. 20: 2095–2109 (suppl.).

Guyer, R. L. & D. E. Koshland, Jr. 1990. The molecule of the year. Science (Washington, DC) 246: 1543–1544.

Hale, L. R. & R. S. Singh. 1986. Extensive variation and heteroplasmy in size of mitochondrial DNA among geographic populations of *D. melanogaster*. Proc. Natl. Acad. Sci. U.S.A. 83: 8813–8817.

Hamby, R. K. & E. A. Zimmer. 1992. Ribosomal RNA as a phylogenetic tool in plant systematics, pp. 50–91. *In* P. S. Soltis, D. E. Soltis & J. J. Doyle [eds.], Molecular systematics of plants. Chapman & Hall, New York.

Hancock, J. M., D. Tautz & G. A. Dover. 1988. Evolution of the secondary structures and compensatory mutatuins of the ribosomal RNAs of *Drosophila*. Mol. Biol. Evol. 5: 393–414.

Hasegawa, M. & T. Hashimoto. 1993. Ribosomal RNA trees misleading? Nature (Lond.) 361: 23.

Hasegawa, M., H. Kishino & T. Yano. 1985. Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22: 160–174.

Haucke, H.-R. & G. Gellissen. 1988. Different mitochondrial gene orders among insects: exchanged tRNA gene positions in the COII/COIII region between an orthopteran and a dipteran species. Curr. Genet. 14: 471–476.

Hein, J. 1989a. A new method that simultaneously aligns and reconstructs ancestral sequences for any number of homologous sequences, when phylogeny is given. Mol. Biol. Evol. 6: 649–668.

1989b. A tree reconstruction method that is economical in the number of pairwise comparisons used. Mol. Biol. Evol. 6: 669–684.

Hendy, M. D. & D. Penny. 1989. A framework for the quantitative study of evolutionary trees. Syst. Zool. 38: 297–309.

Hennig, W. 1981. Insect systematics. Wiley, New York.

Hickson, R. 1993. Sequences and signals: evolutionary histories of New Zealand skinks. Ph.D. dissertation, Massey University, Palmerston North, New Zealand.

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

Hillis, D. M. & J. J. Bull. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42: 182–192.

Hillis, D. M. & S. K. Davis. 1986. Evolution of ribosomal DNA: fifty million years of recorded history in the frog genus Rana. Evolution 40: 1275–1288.

Hillis, D. M. & M. T. Dixon. 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. Q. Rev. Biol. 66: 411–453.

Hillis, D. M. & J. P. Huelsenbeck. 1992. Signal, noise, and reliability in molecular phylogenetic analyses. J. Hered. 83: 189–195.

Hillis, D. M. & C. Moritz. 1990. Molecular systematics. Sinauer, Sunderland, MA.

1995. Molecular systematics II. Sinauer, Sunderland, MA.

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

Hillis, D. M., J. P. Huelsenbeck & C. W. Cunningham. 1994. Application and accuracy of molecular phylogenies. Science (Washington, DC) 264: 671–677.

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

Hoffman, R. J., J. L. Boore & W. M. Brown. 1992. A novel mitochondrial genome organization for the blue mussel *Mytilus edulis*. Genetics 131: 397–412.

Holmquist, R. 1983. Transitions and transversions in evolutionary descent: an approach to understanding. J. Mol. Evol. 19: 134–144.

Horai, S., Y. Satta, K. Hayasaka, R. Kondo, T. Inoue, T. Ishida, S. Hayashi & N. Takahata. 1992. Man's place in the hominoidea revealed by mitochondrial DNA genealogy. J. Mol. Evol. 35: 32–43.

HsuChen, C.-C. & D. T. Dubin. 1984. A cluster of four transfer RNA genes in mosquito mitochondrial DNA. Biochem. Int. 8: 385–391.

HsuChen, C.-C., R. M. Kotin & D. T. Dubin. 1984. Sequences of the coding and flanking regions of the large ribosomal subunit RNA gene of mosquito mitochondria. Nucleic Acids Res. 12: 7771–7785.

Huelsenbeck, J. P. & D. M. Hillis. 1993. Success of phylogenetic methods in the four-taxon case. Syst. Biol. 42: 247–264.

Hughes, A. L., T. Ota & M. Nei. 1990. Positive Darwinian selection promotes charge profile diversity in the antigenbinding cleft of class I major-histocompatibility-complex molecules. Mol. Biol. Evol. 7: 515–524.

Innes, M. A., D. H. Gelfand, J. J. Sninsky & T. J. White. 1990. PCR protocols. Academic, San Diego.

Irwin, D. M., T. D. Kocher & A. C. Wilson. 1991. Evolution of the cytochrome b gene of mammals. J. Mol. Evol. 32: 128–144.

Jacobs, H. T., D. J. Elliott, V. B. Math & A. Farquharson. 1988. Nucleotide sequence and gene organization of sea urchin mitochondrial DNA. J. Mol. Biol. 202: 185–217.

Jermiin L. & R. Crozier. 1994. The cytochrome b region in the mitochondrial DNA of the ant *Tetraponera rufoniger*: sequence divergence in Hymenoptera may be associated with nucleotide content. J. Mol. Evol. 38: 282–294.

Jukes, T. H. & C. R. Cantor. 1969. Evolution of protein molecules, pp. 21–120. *In* H. W. Munro [ed.], Mammalian protein metabolism. Academic, New York.

Kessing, B. H. 1991. Stronglyocentrotid sea urchin mtDNA: phylogenetic relationships and patterns of molecular evolution. M.S. thesis, University of Hawaii, Honolulu.

Kessing, B., H. Croom, A. Martin, C. McIntosh, W. O. McMillan & S. Palumbi. 1989. The simple fool's guide to PCR, version 1.0. Department of Zoology, University of Hawaii, Honolulu.

Kimura, M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16: 111–120.

Kishino, H. & M. Hasegawa. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J. Mol. Evol. 29: 170–179.

Kjer, K. M., G. D. Baldridge & A. M. Fallon. 1994. Mosquito large subunit ribosomal RNA: simultaneous alignment of primary and secondary structure. Biochim. Biophys. Acta 1217: 147–155.

Knight, A. & D. P. Mindell. 1993. Substitution bias, weighting of DNA sequence evolution, and the phylogenetic postition of Fea's viper. Syst. Biol. 42: 18–31.

Kocher, T. D. & A. C. Wilson. 1991. Sequence evolution of mitochondrial DNA in human and chimpanzees: control region and a protein-coding region, pp. 391–413. *In* S. Ozawa & T. Honjo [eds.], Evolution of life: fossils, molecules and culture. Springer, Tokyo.

Kocher, T. D., W. K. Thomas, A. Meyer, S. V. Edwards, S. Pääbo, F. X. Villablanca & A. C. Wilson. 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc. Natl. Acad. Sci. U.S.A. 86: 6196–6200.

Kondo, R., S. Horai, Y. Satta & N. Takahata. 1993. Evolution of hominoid mitochondrial DNA with special reference to the silent substitution rate over the genome. J. Mol. Evol. 36: 517–531.

Kraus, F. & M. M. Miyamoto. 1991. Rapid cladogenesis among the pecoran ruminants: evidence from mitochondrial DNA sequences. Syst. Zool. 40: 117–130.

Kumar, S., K. Tamura & M. Nei. 1993. MEGA: molecular evolutionary genetic analysis, version 1.0. Pennsylvania State University, University Park.

Kumazawa, Y. & M. Nishida. 1993. Sequence evolution of mitochondrial tRNA genes and deep-branch animal phylogenetics. J. Mol. Evol. 37: 380–398.

Latorre, A., E. Barrio, A. Moya & F. J. Ayala. 1988. Mitochondrial DNA evolution in the *Drosophila obscura* group. Mol. Biol. Evol. 56: 717–728.

Lee, Y.-H. & V. D. Vacquier. 1992. The divergence of species-specific abalone sperm lysins is promoted by positive Darwinian selection. Biol. Bull. 182: 97–104.

Lemeunier, F., J. R. David, L. Tsacas & M. Ashburner. 1986. The *melanogaster* species group, pp. 147–256. *In* M. Ashburner, H. L. Carson & J. N. Thompson [eds.], The genetics and biology of *Drosophila*, vol. 3e. Academic, London.

Lewontin, R. C. 1989. Inferring the number of evolutionary events from DNA coding sequence differences. Mol. Biol. Evol. 6: 15–32.

Li, W.-H. 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J. Mol. Evol. 36: 96–99.

Li, W.-H. & D. Graur. 1991. Fundamentals of molecular evolution. Sinauer, Sunderland, MA.

Li, W.-H., C.-I. Wu & C.-C. Luo. 1984. Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudo genes and its evolutionary implications. J. Mol. Evol. 21: 58–71.

1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitutions considering the relative likelihood of nucleotide and codon changes. Mol. Biol. Evol. 2: 150–174.

Liu, H. & A. T. Beckenbach. 1992. Evolution of the mitochondrial cytochrome oxidase II gene among ten orders of insects. Mol. Phylogenet. Evol. 1: 41–52.

Lockhart, P. J., M. A. Steel, M. D. Hendy & D. Penny. 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. Mol. Biol. Evol. 11: 605–612.

Maddison, W. P. & D. R. Maddison. 1992. MacClade, version 3.0 manual. Sinauer, Sunderland, MA.

Maddison, W. P., M. J. Donoghue & D. R. Maddison. 1984. Outgroup analysis and parsimony. Syst. Zool. 33: 83–103.

Manske, C. L. & D. J. Chapman. 1987. Nonuniformity of nucleotide substitution rates in molecular evolution: computer simulation and analysis of 5S ribosomal RNA sequences. J. Mol. Evol. 26: 226–251.

Martin, A. P. & S. R. Palumbi. 1993. Protein evolution in different cellular environments: cytochrome b in sharks and mammals. Mol. Biol. Evol. 10: 873–891.

Martin, A. P. & C. M. Simon. 1990. Differing levels of among-population divergence in the mitochondrial DNA of 13-versus 17-year periodical cicadas related to historical biogeography. Evolution 44: 1066–1088.

Martin, A. P., B. D. Kessing & S. R. Palumbi. 1990. Accuracy of estimating genetic distance between species from short sequences of mitochondrial DNA. Mol. Bio. Evol. 7: 485–488.

Martin, A. P., G.J.P. Naylor & S. R. Palumbi. 1992. Rates of mitochondrial DNA evolution in sharks are slow compared with mammals. Nature (Lond.) 357: 153–155.

McClure, M. A., T. K. Vasi & W. M. Fitch. 1994. Comparative analysis of multiple protein-sequence alignment methods. Mol. Biol. Evol. 11: 571–592.

McCracken, A., I. Uhlenbusch & G. Gellissen. 1987. Structure of the cloned *Locusta migratoria* mitochondrial genome: restriction mapping and sequence of its ND-1 gene. Curr. Genet. 11: 631–638.

McDonald, J. H. & M. Kreitman. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. Nature (Lond.) 351: 652–654.

McPheron, B. A., G. E. Gasparich, H-Y. Han, G. J. Steck & W. S. Sheppard. 1994. Mitochondrial DNA restriction map for the Mediterranean fruit fly, *Ceratitis capitata*. Biochemical Genetics 32: 25–33.

Milinkovitch, M. C., G. Orti & A. Meyer. 1993. Revised phylogeny of whales suggested by mitochondrial ribosomal DNA sequences. Nature (Lond.) 361: 346–348.

Mindell, D. P. & R. L. Honeycutt. 1990. Ribosomal RNA in vertebrates: evolution and phylogenetic applications. Annu. Rev. Ecol. Syst. 21: 541–566.

Mitchell, S. E., A. F. Cockburn & J. A. Seawright. 1993. The mitochondrial genome of *Anopheles quadrimaculatus* species A: complete nucleotide sequence and gene organization. Genome 36: 1058–1073.

Miyamoto, M. M. & J. Cracraft. 1992. Phylogenetic analysis of DNA sequences. Oxford University Press, Oxford.

Miyamoto, M. M., F. Kraus & O. A. Ryder. 1990. Phylogeny and evolution of antlered deer determined from mitochondrial DNA sequences. Proc. Natl. Acad. Sci. U.S.A. 87: 6127–6131.

Miyata, T. & T. Yasunaga. 1980. Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. J. Mol. Evol. 16: 23–36.

Miyata, T., H. Hayashida, R. Kikuno, M. Hagesagawa, M. Kobayashi & K. Koike. 1982. Molecular clock of silent substitution: at least sixfold preponderance of silent changes in mitochondrial genes over those in nuclear genes. J. Mol. Evol. 19: 28–35.

Moritz, C., T. E. Dowling & W. M. Brown. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Annu. Rev. Ecol. Syst. 18: 269–292.

Mullis, K., F. Faloona, S. Scharf, R. Saiki, G. Horn & H. Erlich. 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harbor Symp. Quant. Biol. 51: 263–273.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

Nei, M. & T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3: 418–426.

Neigel, J. E. & J. C. Avise. 1986. Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation, pp. 515. *In* S. Karlin & E. Nevo [eds.], Evolutionary processes and theory. Academic, New York.

Nigro, L. & A. Grapputo. 1993. Evolution of the mitochondrial rRNA in the *oriental* species subgroup of *Drosophila*. Biochem. Syst. Ecol. 21: 79–83.

Nigro, L., M. Solignac & P. Sharp. 1991. Mitochondrial DNA sequence divergence in the *melanogaster* and *oriental* species subgroups of *Drosophila*. J. Mol. Evol. 33: 156–162.

Noller, H. F., D. Moazed, S. Stern, T. Powers, P. N. Allen, J. M. Robertson, B. Weiser & T. Triman. 1990. Structure of rRNA and its functional interactions in translation, pp. 73–92. *In* W. E. Hill, A. Dahlberg, R. A. Garrett, P. B. Moore, D. Schlesinger & J. R. Warner [eds.], The ribosome: structure, function and evolution. American Society of Microbiology, Washington, DC.

Novacek, M. J. 1992. Mammalian phylogeny: shaking the tree. Nature (Lond.) 356: 121–125.

Novacek, M. J. & Q. D. Wheeler. 1992. Extinction and phylogeny. Columbia University Press, New York.

Ojala, D., J. Montoya & G. Attardi. 1981. tRNA punctuation model of RNA processing in human mitochondria. Nature (Lond.) 290: 470–474.

Okimoto, R., J. L. Macfarlane, D. O. Clary & D. R. Wolstenholme. 1992. The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. Genetics 130: 471–498.

Olsen, G. J. 1987. Earliest phylogenetic branchings: comparing rRNA-based evolutionary trees inferred with various techniques. Cold Spring Harbor Symp. Quant. Biol. 52: 825–837.

Pääbo, S. 1990. Amplifying ancient DNA, pp. 159–166. *In* M. A. Innes, D. H. Gelfand, J. J. Sninsky & T. J. White [eds.], PCR protocols. Academic, San Diego.

Palumbi, S. 1995. Nucleic acids II: the polymerase chain reaction. *In* D. M. Hillis & C. Moritz [eds.], Molecular systematics II. Sinauer, Sunderland, MA (in press).

Palumbi, S., A. Martin, S. Romano, W. O. McMillan, L. Stice & G. Grabowski. 1991. The simple fool's guide to PCR, version 2.0. Department of Zoology, University of Hawaii, Honolulu.

Pamilo, P. & M. Nei. 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5: 568–583.

Pashley, D. P. & L. D. Ke. 1992. Sequence evolution in the mitochondrial ribosomal and ND-1 genes in Lepidoptera: implications for phylogenetic analyses. Mol. Biol. Evol. 9: 1061–1075.

Penny, D. & M. D. Hendy. 1985. Testing methods of evolutionary tree construction. Cladistics 1: 266–272.

1986. Estimating the reliability of evolutionary trees. Mol. Biol. Evol. 3: 403–417.

Penny, D., M. D. Hendy & M. A. Steele. 1992. Progress with methods for constructing evolutionary trees. TREE 7: 73–79.

Powell, J. R., A. Caccone, G. D. Amato & C. Yoon. 1986. Rates of nucleotide substitution in *Drosophila* mitochondrial DNA and nuclear DNA are similar. Proc. Natl. Acad. Sci. U.S.A. 83: 9090–9093.

Pruess, K. P., X. Zhu & T. O. Powers. 1992. Mitochondrial transfer RNA genes in a black fly, *Simulium vittatum* (Diptera: Simuliidae), indicate long divergence from mosquito (Diptera: Culicidae) and fruit fly (Diptera: Drosophilidae). J. Med. Entomol. 29: 644–651.

Rand, D. M. & R. G. Harrison. 1989. Molecular population genetics of mitochondrial DNA size variation in crickets. Genetics 121: 551–569.

Roe, B., D. Ma, R. Wilson & J. Wong. 1985. The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome. J. Biol. Chem. 260: 9759–9774.

Roehrdanz, R. L. 1993. An improved primer for PCR amplification of mitochondrial DNA in a variety of insect species. Insect Mol. Biol. 2: 89–91.

Ruvolo, M., T. R. Disotell, M. W. Allard & W. M. Brown. 1991. Resolution of the African hominoid trichotomy by use of a mitochondrial gene sequence. Proc. Natl. Acad. Sci. U.S.A. 88: 1570–1574.

Saiki, R. K., S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Erlich & N. Arnheim. 1985. Enzymatic amplification of B-globin genomic sequences and restriction site analysis for diagnosis of sickle

cell anemia. Science (Washington, DC) 230: 1350–1354.

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

Sankoff, D. 1993. Analytical approaches to genomic evolution. Biochimie (Paris) 75: 409–413.

Sankoff, D. & R. J. Cedergren. 1983. Simultaneous comparison of three or more sequences related by a tree, pp. 253–263. *In* D. Sankoff & J. B. Kruskal [eds.], Time warps, string edits, and macromolecules: the theory and practice of sequence comparison. Addison-Wesley, Reading, PA.

Sankoff, D., G. Leduc, N. Antoine, B. Pacquin, B. F. Lang & R. Cedergren. 1992. Gene order comparisons for phylogenetic inference: evolution of the mitochondrial genome. Proc. Natl. Acad. Sci. U.S.A. 89: 6575–6579.

Satta, Y. & N. Takahata. 1990. Evolution of *Drosophila mitochondrial* DNA and the history of the *melanogaster* group. Proc. Natl. Acad. Sci. U.S.A. 87: 9558–9562.

Satta, Y., I. Ishiwi & S. I. Chigusa. 1987. Analysis of nucleotide substitutions of mitochondrial DNAs in *Drosophila melanogaster* and its sibling species. Mol. Biol. Evol. 4: 638–650.

Schaeffer, S. W. & C. F. Aquadro. 1987. Nucleotide sequence of the Adh gene region of *Drosophila pseudoobscura:* evolutionary change and evidence for an ancient gene duplication. Genetics 117: 61–73.

Shannon, C. E. 1948. The mathematical theory of communication. Bell System Technology. J. 27: 379–423, 623–656.

Sharp, P. & W.-H. Li. 1989. On the rate of DNA sequence evolution in *Drosophila*. J. Mol. Evol. 28: 398–402.

Shoemaker, J. S. & W. M. Fitch. 1989. Evidence from nuclear sequences that invariable sites should be considered when calculating sequence divergence. Mol. Biol. Evol. 6: 270–289.

Simon, C. 1988. Evolution of 13- and 17-year periodical cicadas (Homoptera: Cicadidae: Magicicada). Bull. Entomol. Soc. Am. 34: 163–176.

1991. Molecular systematics at the species boundary: exploiting conserved and variable regions of the mitochondrial genome of animals via direct sequencing from enzymatically amplified DNA, pp. 33–71. *In* G. M. Hewitt, A.W.B. Johnston & J.P.W. Young [eds.], Molecular techniques in taxonomy. NATO Advanced Studies Institute, H57. Springer, Berlin.

Simon, C., A. Franke & A. Martin. 1991. The polymerase chain reaction: DNA extraction and amplification, pp. 329–355. *In* G. M. Hewitt [ed.], Molecular techniques in taxonomy. NATO Advanced Studies Institute, H57. Springer, Berlin.

Simon, C., C. McIntosh & J. Deniega. 1993. Standard restriction fragment length analysis of the mitochondrial genome is not sensitive enough for phylogenetic analysis or identification of 17-year periodical cicada broods (Hemiptera: Cicadidae): the potential for a new technique. Ann. Entomol. Soc. Am. 86: 228–238.

Simon, C., S. Pääbo, T. Kocher & A. C. Wilson. 1990. Evolution of the mitochondrial ribosomal RNA in insects as shown by the polymerase chain reaction, pp. 235–244. *In* M. Clegg & S. O'Brien [eds.], Molecular evolution. UCLA Symposia on Molecular and Cellular Biology, New Series, vol. 122. Liss, New York.

Smith, M. J., D. K. Banfield, K. Dateval, S. Gorski & D. Kowbel. 1990. Nucleotide sequence of nine protein coding genes and 22 tRNAs in the mtDNA of the sea star *Pisaster ochraceus*. J. Mol. Evol. 31: 195–204.

Smith, M. J., A. Arndt, S. Gorski & E. Fajber. 1993. The phylogeny of Echinoderm classes based on mitochondrial gene arrangements. J. Mol. Evol. 36: 545–554.

Solignac, M., M. Monnerot & J.-C. Mounolou. 1986. Mitochondrial DNA evolution in the *melanogaster* species subgroup of *Drosophila*. J. Mol. Evol. 23: 31–40.

Sperling, F.A.H. & D. A. Hickey. 1994. Mitochondrial DNA sequence variation in the spruce budworm species complex (Choristoneura: Lepidoptera). Mol. Biol. Evol. 11: 656–665.

Sperling, F.A.H., G. S. Anderson & D. A. Hickey. 1994. A DNA-based approach to the identification of insect species used for postmortem interval estimation. J. Forensic Sci. 39: 418–427.

Sprinzl, M., J. Moll, F. Meissner & T. Hartman. 1987. Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Res. 15: r53-r188 (suppl.).

Steel, M. A., P. J. Lockhart & D. Penny. 1993. Confidence in evolutionary trees from biological sequence data. Nature (Lond.) 364: 440–442.

Swofford, D. L. 1993. Phylogenetic analysis using parsimony. Illinois Natural History Survey, Urbana, IL.

Swofford, D. L. & G. J. Olsen. 1990. Phylogeny reconstruction, pp. 411–501. *In* D. M. Hillis & C. Moritz [eds.], Molecular systematics. Sinauer, Sunderland, MA.

1995. Phylogeny reconstruction. *In* D. M. Hillis & C. Moritz [eds.], Molecular systematics II. Sinauer, Sunderland, MA (in press).

Tajima, F. & M. Nei. 1984. Estimation of evolutionary distance between nucleotide sequences. Mol. Biol. Evol. 1: 269–285.

Tamura, K. 1992. Estimation of the number of nucleotide substitutions when there are strong transition–transversion and G+C-content biases. Mol. Biol. Evol. 9: 678–687.

Tamura, K. & M. Nei. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10: 512–526.

Taylor, M.F.J., S. W. McKechnie, N. Pierce & M. Kreitman. 1993. The lepidopteran mitochondrial control region: structure and evolution. Mol. Biol. Evol. 10: 1259–1272.

Thomas, W. K. & A. T. Beckenbach. 1989. Variation in salmonid mitochondrial DNA: evolutionary constraints and mechanisms of substitution. J. Mol. Evol. 29: 233–245.

Thomas, W. K. & A. C. Wilson. 1991. Mode and tempo of molecular evolution in the nematode Caenorhabditis. Genetics 128: 269–279.

Thomas, W. K., J. Maa & A. C. Wilson. 1989. Shifting constraints on tRNA genes during mitochondrial DNA evolution in animals. New Biologist 1: 93–100.

Thomas, W. K., S. Pääbo, F. X. Villablanca & A. C. Wilson. 1990. Spatial and temporal continuity of kangaroo rat populations shown by sequencing mitochondrial DNA from museum specimens. J. Mol. Evol. 31: 101–112.

Topal, M. D. & J. R. Fresco. 1976. Complementary base pairing and the origin of substitution mutations. Nature (Lond.) 263: 285–289.

Uhlenbusch, I., A. McCracken & G. Gellissen. 1987. The gene for the large (16S) ribosomal RNA from the *Locusta migratoria* mitochondrial genome. Curr. Genet. 11: 631–638.

Upholt, W. B. & I. B. David. 1977. Mapping of mitochondrial DNA of individual sheep and goats: rapid evolution in the D loop regions of mitochondrial DNA. Cell 11: 571–583.

Uzzell, T. & K. W. Corbin. 1971. Fitting discrete probability distributions to evolutionary events. Science (Washington, DC) 172: 1089–1096.

Van de Peer, Y., J.-M. Neefs, P. De Rijk & R. De Wachter. 1993. Reconstructing evolution from eukaryotic small-ribosomal-subunit RNA sequences: calibration of the molecular clock. J. Mol. Evol. 37: 221–232.

Vawter, L. & W. M. Brown. 1986. Nuclear and miotochondrial DNA comparisons reveal extreme rate variation in the molecular clock. Science (Washington, DC) 234: 194–196.

Vogler, A. P. & R. DeSalle. 1993. Phylogeographic patterns in coastal North American tiger beetles, *Cincindela dorsalis,* inferred from mitochondrial DNA seqeunces. Evolution 47: 1192–1202.

Vogler, A. P., R. DeSalle, T. Assmann, C. B. Knisley & T. D. Schultz. 1993a. Molecular population genetics of the endangered tiger beetle *Cincindela dorsalis* (Coleoptera: Cicindelidae). Ann. Entomol. Soc. Am. 86: 142–152.

Vogler A. P., C. B. Knisley, S. B. Glueck, J. M. Hill & R. DeSalle. 1993b. Using molecular and ecological data to diagnose endangered populations of the puritan tiger beetle *Cicindela puritana.* Mol. Ecol. 2: 375–383.

Wakeley, J. 1993. Substitution rate variation among sites in hypervariable region of human mitochondrial DNA. J. Mol. Evol. 37: 613–623.

1994. Substitution rate variation among sites and the estimation of transition bias. Mol. Biol. Evol. (in press).

Waterman, M. S., M. Eggert & E. Lander. 1992. Parametric sequence comparisons. Proc. Natl. Acad. Sci. U.S.A. 89: 6090–6093.

Waterman, M. S., J. Joyce & M. Eggert. 1991. Computer alignment of sequences, pp. 59–72. *In* M. M. Miyamoto & J. Cracraft [eds.], Phylogenetic analysis of DNA sequences. Oxford University Press, Oxford.

Weller, S. J., D. P. Pashley, J. A. Martin & J. L. Constable. 1994. Phylogeny of noctuoid moths and the utility of combining independent nuclear and mitochondrial genes. Syst. Biol. 43: 194–211.

Wheeler, W. C. & D. L. Gladstein. 1992. A heuristic multiple sequence alignment procedure based on parsimony. Computer program. American Museum of Natural History, New York.

Wheeler W. & R. Honeycutt. 1988. Paired sequence difference in ribosomal RNAs: evolutionary and phylogenetic implications. Mol. Biol. Evol. 5: 90–96.

Wheeler, W. C., P. Cartwright & C. Y. Hayashi. 1993. Arthropod phylogeny: a combined approach. Cladistics 9: 1–39.

Williams, P. L. & W. M. Fitch. 1989. Finding the weighted minimal change in a given tree, pp. 453–470. *In* B. Fernholme, K. Bremer & H. Jornval [eds.], Nobel symposium on the hierarchy of life. Elsevier, Cambridge.

Willis, L. G., M. L. Winston & B. M. Honda. 1992. Phylogenetic relationships in the honey bee (Genus *Apis*) as determined by the sequence of the cytochrome oxidase II region of mitochondrial DNA. Mol. Phylogenet. Evol. 1: 169–178.

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 & M. Stoneking. 1985. Mitochondrial DNA and two perspectives on evolutionary genetics. Biol. J. Linn. Soc. 26: 375–400.

Wolstenholme, D. R. 1992. Animal mitochondrial DNA: structure and evolution, pp. 173–216. *In* K. W. Jeon & D. R. Wolstenholme [eds.], Mitochondrial genomes. Int. Rev. Cytol. 141: 173–216.

Wolstenholme, D. R. & D. O. Clary. 1985. Sequence evolution of *Drosophila* mitochondrial DNA. Genetics 109: 725–744.

Wolstenholme, D. R., J. L. MacFarlane, R. Okimoto, D. O. Clary & J. A. Wahleithner. 1987. Bizarre tRNAs inferred from DNA sequences of mitochondrial genomes of nematode worms. Proc. Natl. Acad. Sci. U.S.A. 84: 1324–1328.

Wooten R. J. 1981. Paleozoic insects. Annu. Rev. Entomol. 26: 319–344.

Wu, C.-I. & W.-H. Li. 1985. Evidence for higher rates of nucleotide substittution in rodents than in man. Proc. Natl. Acad. Sci. U.S.A. 82: 1741–1745.

Xiong, B. & T. D. Kocher. 1991. Comparison of mitochondrial DNA sequences of seven morphospecies of black flies (Diptera: Simuliidae). Genome 34: 306–311.

1993a. Intraspecific variation in sibling species of *Simulium venustum* and *Simulium verecundum* complexes (Diptera: Simuliidae) revealed by the sequence of the mitodhondrial 16S rRNA gene. Can. J. Zool. 71: 1202–1206.

1993b. Phylogeny of sibling species of *Simulium venustum* and *S. verecundum* (Diptera: Simuliidae) based on sequences of the mitochondrial large subunit rRNA gene. Mol. Phylogenet. Evol. 293–303.

Yang, Z. 1993. Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. Mol. Biol. Evol. 10: 1396–1401.

Yang, Z., N. Goldman & A. Friday. 1994. Comparison of models for DNA substitutions used in maximum likelihood phylogenetic inference. Mol. Biol. Evol. 11: 316–324.

Zucker, M. 1989. Computer prediction of RNA structure. Methods Enzymol. 180: 262–288.

Zucker, M, J. A. Jaeger & D. H. Turner. 1991. A comparison of optimal and suboptimal RNA secondary structures predicted by free energy minimization with structures dertermined by phylogenetic comparison. Nucleic Acid Res. 19: 2707–2714.

Received for publication 3 January 1994; accepted 3 June 1994.

Appendix

Compilation of Conserved PCR Primers Useful for Amplifying and Sequencing Mitochondrial DNA

This compilation of conserved mitochondrial DNA primers will facilitate the sequencing of all mitochondrial genes in the majority of animal taxa and owes much to the seminal work begun in the laboratory of A. C. Wilson especially that of T. D. Kocher in designing the first *universal* PCR primers (Kocher et al. 1989). Many investigators generously contributed to this *Appendix* and each is acknowledged below in the notes accompanying each primer. Many useful primers and notes were supplied by Richard Harrison and his students, Steve Bogdanowitz, Andrew Brower, Jackie Brown, and Ben Normark; they are acknowledged throughout as *the Harrison laboratory*. This *Appendix* is a substantially enlarged update of the published compilation by Simon et al. 1991, which was an enlarged update of primers originally contained in the *Simple Fool's Guide to PCR* (Kessing et al. 1989, revised by Palumbi et al. 1991 and Palumbi 1995). This compilation adds many new primers, eliminates those primers that were useful only in a limited number of species, and concentrates on highly conserved primers that are likely to be useful in a wide range of taxa. Emphasis is placed on the usefulness of these primers in insects, but many are applicable to a much broader range of taxa as illustrated by the comparative format in which they are presented. Primers are compared with *D. yakuba* (fly [Clary & Wolstenholme 1985]), *A. melifera* (bee [Crozier & Crozier 1993]), *Locusta migratoria* (locust [Flook et al. 1995]), *Homo sapiens* (human [Anderson et al. 1981]) or *Bos tarus* (cow [Anderson et al. 1982]), *Xenopus laevis* (toad [Roe et al. 1985]), *Strongylocentrotus purpuratus* (sea urchin [Jacobs et al. 1988]), and, where possible, *C. elegans* (roundworm [Okimoto et al. 1992]), and a sampling of available insects. This comparative format will facilitate evaluation or modification of primers to suit individual needs or taxa. Note that, as discussed in the text, the bee is extremely A+T rich and has undergone considerable evolution relative to *D. yakuba*. It often matches the primer (usually fly) sequence poorly compared with other taxa. A similar poor fit is often observed for nematode sequences.

Notes. In addition to the names of the primer designers, notes below each primer identify publications in which the primers are used, representative taxa in which they have been tried (noting success or failure), and tips for use.

How Primers are Named. Until now, many primer names were not standardized (and sometimes duplicated) among laboratories. We have standardized names but we have also included the *nickname* of each primer from the laboratory in which it originated. Nicknames are useful because they are shorter, easier to remember, and are already familiar to a large number of people.

Primer names are standardized in a format similar to that used by Kocher et al. (1989) for vertebrates where the strand and the location of the 3′ base in a reference sequence are identified. Because we focus on insects, we use *D. yakuba* (Clary & Wolstenholme 1985) as our reference sequence. A two-letter or digit code is used to identify each gene. The next letter in the primer name identifies the strand within which each primer can be found in *Drosophila*. Because insects, unlike vertebrates, have no distinct light (L) or heavy (H) strand (Brown 1985), we identify strands according to the number of genes transcribed from them. In the insects studied to date, most mitochondrial genes are transcribed from one strand (Fig. 1). We identify this strand by a 'J' to signify maJority. We call the other strand, the 'N' or miNority strand. Had we named primers as 'L' or 'H,' we would have had to arbitrarily designate the two strands in *Drosophila* as 'L' and 'H' or to have used the location in the human sequence to determine the primer name. These options would be unsatisfactory because they would sometimes result in complementary primer pairs having the same strand designation (because the strand from which genes are transcribed can vary in insects compared with vertebrates). We chose not to use 'S' (Sense) or 'A' (Antisense), because this nomenclature could also result in opposing primers with the same strand designation (because in insects approximately one-third of the genes are transcribed from the minority strand). Because rRNA and protein gene order and transcription direction appear to be largely conserved in insects, the N and J designation will restrict the problem of opposing primers with similar strand designations to primers located in occasional transposed tRNAs. When one primer lies in a tRNA or when noninsect taxa are to be amplified, gene locations and direction of transcription (Fig. 1) should be checked for the taxa in question to make sure the primers are facing each other. Of course, this is never a problem when opposing primers are located within the same gene.

The two-letter or digit code for each gene is listed below. Transfer RNAs are identified by a 'T' followed by their standard one-letter code. Exceptions are tRNA leucine and tRNA serine, which exist in two copies per mitochondrial genome. The two copies are numbered 1 and 2. Subunits of protein genes are numbered with the Arabic numerals corresponding to the standard Roman or Arabic numerals.

AT, A+T-rich region
SR, small-subunit rRNA (12S)
LR, large-subunit rRNA (16S)

N1, NADH dehydrogenase (ND) 1
CB, cytochrome b
N6, ND6
NL, ND4L
N4, ND4
N5, ND5
N3, ND3
C3, cytochrome c oxidase (CO) III
A6, ATPase (ATP) 6
A8, ATP8
C2, COII
C1, COI
N2, ND2
TA, tRNA alanine
TR, tRNA arginine
TN, tRNA asparagine
TD, tRNA aspartic acid
TC, tRNA cysteine
TQ, tRNA glutamine
TE, tRNA glutamic acid
TG, tRNA glycine
TH, tRNA histidine
TI, tRNA isoleucine
TL1, tRNA leucine (CUN)
TL2, tRNA leucine (UUR)
TK, tRNA lysine
TM, tRNA methionine
TF, tRNA phenylalanine
TP, tRNA proline
TS1, tRNA serine (UCN)
TS2, tRNA serine (AGY)
TT, tRNA threonine
TW, tRNA tryptophan
TY, tRNA tyrosine
TV, tRNA valine

Estimating the Size of the Segment To Be Amplified. If gene order is the same as in *D. yakuba,* approximate sizes of pieces to be amplified can be obtained by subtracting location numbers of opposing primers. For this reason, the position numbers given for each reference taxon mark the location of the 5′ base (even though the primer is named according to the location of its 3′ base). For unpublished sequences and published partial sequences, numerical locations are not given. A map of each gene (or group of genes) is provided immediately preceding the relevant primers. This will facilitate visualization of the size and location of the fragment to be amplified.

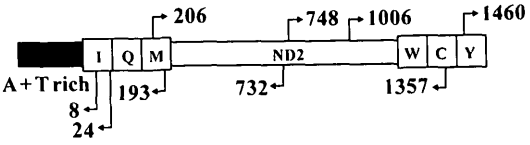
All primers are written in the 5′ to 3′ direction. Primers are listed in the order in which the genes in which they are located occur in the published *D. yakuba* sequence. The alias or nickname is given in parentheses. Degenerate primers—those with a mixture of two or more bases at one position—are indicated by one or more bases identified by 'Y' (pyrimidine), 'R' (purine), or 'N' (all four bases).

It should be noted that we have carefully checked each primer and comparative sequence, but in an article of this size errors may occur. Researchers wishing to modify primers to match other taxa are encouraged to check the reference sequences in the original publication as cited.

Additional information concerning primer design and performance is distributed by way of the insect molecular systematics computer network BUG-NET, which currently comprises more than 165 members. To join BUG-NET, send your e-mail address to CRESPI@SFU.CA. Comments and corrections related to the following compilation are welcome.

Compilation of Conserved PCR Primers

The first three primers can be used to amplify the A+T-rich region in insects. If the A+T-rich region is very long (>2.5 kb), or if tRNAs isoleucine or methionine are transposed, the amplification may fail (See text for cautions relevant to the use of this region for phylogenetic analysis.).



tRNA-Isoleucine

TI-N-8 (alias t-Ile)

(28mer)	5′ CTATCAAGGTAACCCTTTTTATCAGGCA 3′	
Fly	.....	35
Bee	T.....AA..TTT....–.....A..	466
Locust	.....A.....–A.....	34
Human	.....A.....T....–.....A..	4296
Toad	.....A..G.T..C.AC.T..G....	5805

Designed by J. M. Brown, it can be used in *Drosophila* to amplify the A+T region in conjunction with 12S primers. A partially overlapping primer (moved 16 bases downstream), which avoids the bee, locust, and human deletion, is given below. Urchin and roundworm are not shown because they match poorly.

TI-N-4 (alias t-Iso)

19 mer	5′ ATTTACCCTATCAAGGTAA 3′	
Fly	.....	42
Bee	TAA..TTT.....AA..T	473
Locust	.....T.....A...	41
Human	.....T.....A....	4303
Toad	T..C..T.....A..G.	5813

Designed by C.S. and C. Orego and appeared in Simon et al. (1991), it was used in conjunction with 12SB to amplify the A+T-rich region in *Drosophila virilis.* It failed to amplify genomic DNA from the *Magicicada* species probably because the A+T-rich region is too large. In *Apis,* the isoleucine tRNA and methionine tRNA exchange positions compared with *Drosophila,* but both primers should still be useful for the A+T-rich region. In vertebrates, the tRNA isoleucine

is not adjacent to the control region. The ATC in the middle of the primer is the isoleucine mRNA codon. Urchin and roundworm are not shown because they match poorly.

tRNA-Methionine

TM-J-206 (alias Frank)

(24mer)	5′ GCTAA—ATAAAGCTAACAGGTTCAT 3′	
Fly	.....—T...-....CTG.....	184
Bee	.....—.....	231
Locust	.....—TC..-.....TG.....	141
Human	.....—....-....T.G..CC...	4412
Toad	.....—.A..-...TTTG..CC...	5920
Sea urchin	.....GTGC...-...TTTG..C....	1805
Roundworm	.A...—T...GT..GTA.....	9488

Designed by the Harrison laboratory, it is partially complementary to TM-N-193. They report that it is useful for amplifying the whole ND2 gene and part of COI (in combination with C1-J-2191), e.g., in the milkweed bug, *Oncopeltus*. In honey bee, three tRNAs (glycine, alanine, and isoleucine) are inserted between the tRNA methionine and the beginning of ND2.

TM-N-193 (alias met-20)

(20mer)	5′ TGGGGTATGAACCCAGTAGC 3′	
Fly	.....	212
Bee	CA.....TGT....	260
Locust	.....T....	169
Human	C.....GG...GA....	4440
Toad	.....GG....AA...	5948
Sea urchin	.....G....AA...	1836
Roundworm	AA.....TTAC..A	9517

Designed by M. Taylor, it is partially complementary to TM-J-206. Taylor notes that it amplifies DNA from butterflies but not from aphids. He used this primer in conjunction with a slightly modified version of the primer SR-J-14612 to amplify and sequence the A+T-rich region of *H. punctigera* (Noctuidae), *S. melinus* (Lycaenidae), and six species in the genus *Jalmenus* (Lycaenidae) (Taylor et al. 1993). The tRNA methionine does not flank the control region in vertebrates, urchins, and blue mussels (*Mytilus*) (Hoffmann et al. 1992).

NADH Dehydrogenase Subunit 2

Because of the paucity of primers in this region, we include two conserved primers, which we have not yet tested. Anyone testing these primers is urged to report results via BUG-NET.

N2-N-732 (untested primer-1)

(20mer)	5′ GAAGTTTGGTTTAAACCTCC 3′	
Fly	.....	751
Mosquito	.....A..C..T.....	711
Mouse	TGT.....A.GT.....	4434
Human	TGG..C.....T..A..	4990

This untested primer was designed by H.L. for use with *Drosophila* species. The mosquito is *Anopheles gambiae* (Beard et al. 1993).

N2-J-748 (untested primer-2)

(20mer)	5′ ATTGGAGGTTTAAACCAAAC 3′	
Fly	.....	731
Mosquito	T.A.....A..G..T.....	689
Mouse	TGA.....AC.T.....	4411
Human	TGA..T..A.....G..	4968

This untested primer was designed by H.L. for use with *Drosophila* species. The mosquito is *A. gambiae* (Beard et al. 1993).

N2-J-1006 (alias Ind2)

(25mer)	5′ TAGGTGGACTACCTCCATTTTYAGG 3′	
Fly	.....T.....T...	982
Bee	ATTCAAT.TAT..AATT...CTTTC	1220
Locust	.....G..G..A...A.ACT...	939
Human	....A..C..G..C..GC.AACC..	5221
Toad	.....C..T..A..TC...C...	6730
Sea urchin	....A..C..T..C.....AACC..	3952
Roundworm	TTTTAAATA.T..ATTTAGAGTTTC	4034

Designed by H.L. for the flea (*Ctenocephalides felis*), it has been used to amplify the gene junction between COI and ND2 successfully for five orders of insects with the primer C1-N-1560. These insects include greater wax moth, *Galleria mellonella;* house cricket, *Acheta domesticus;* locust, *Schistocerca gregaria;* American cockroach, *Periplaneta americana;* dragonfly, *Sympetrum* striolatum; and flea, *C. felis*. It works well for sequencing double-stranded DNA. The poor match for bee and roundworm suggests that the alignments for bee and roundworm may be incorrect; primer nucleotides were aligned according to the published amino acid alignment.

tRNA-Tyrosine

TY-J-1460

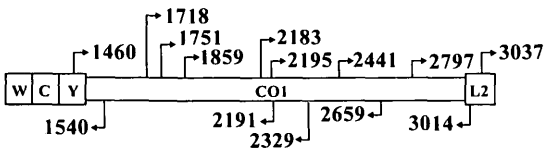
(26mer)	5′ TACAATTTATCGCCTAAA-CTTCAGCC 3′	
Fly	.....C.....-.....	1460
Locust	.....C..A.A.....TAA.....	1388
Cow	.TT.CAG.C.AATGCTTT-GC.....	5677
Toad	.GT.CACGGAGCTAC..T-.CG.C...	7374

This primer was designed by F. Sperling and has been shown to amplify species of lepidopterans, dipterans, coleopterans, and hemipterans (Sperling & Hickey 1994, Sperling et al. 1994). Urchin, bee, and roundworm are not given because the tRNA is transposed.

Cytochrome c Oxidase Subunit I

For all COI primers, *springtail* in the primer comparison refers to the collembolan *Orchesella villosa* (F.F., unpublished data). The thrips is a *Holothrips* and the milkweed bug is an *Oncopeltus* (B.C., unpublished data). In the primer notes,

collembolans refers to the genera, *Orchesella* spp., *Isotomurus* spp., *Tetrodontophora* spp., and *Thaumanura* spp. (F.F., unpublished data).



C1-N-1560

(21mer)	5′ TGTTCCTACTATTCCGGCTCA 3′	
Fly	.....	1540
Bee	..A.....G.....A.A...	1860
Locust	.....A.....	1519
Sea urchin	...G.....C..G..A.....	5877
Cow	.....A.....C..	5779
Toad	G.....G..G.GC..T.....	7489
Roundworm	A..A..A..C..A..A.A...	7961

Designed by H.L., it has been found to amplify the gene junctions between COI and ND2 successfully for five orders of insects (listed under N2-J-1006) with the primer N2-J-1006.

C1-J-1718

(26mer)	5′ GGAGGATTTGGAAATTGATTAGTTCC 3′	
Fly	..G..G.....G..	1693
Bee	.....GC.TA....	2013
Locust	.....C.....A..	1652
Cow	.....C..T..C...C.T.....	5912
Toad	..T.....G..C.....	7622
Sea urchin	..T.....G.....C.CA....	6010
Roundworm	..T..T.....T..C..A...T.A..	8093

Designed by B. Farrell for phytophagous beetles, it also works well for other coleopterans, thysanopterans, homopterans, and hemipterans. It works well in combination with C1-N-2191.

C1-J-1751 (alias Ron)

(23mer)	5′ GGATCACCTGATATAGCATTCCC 3′	
Fly	...G.T.....C.....	1729
Bee	.....	2049
Locust	...G.T..A.....T..T..	1688
Cow	..TG.T..C.....T..	5948
Toad	...G.C..A.....T..	7658
Sea urchin	..TG.G..A.....G..C.....	6046
Roundworm	...G.....AG...T..	8130

Designed by the Harrison laboratory where they report that it works well with virtually all insects tested, including lepidopterans, dipterans, coleopterans, thysanopterans, hemipterans, and homopterans. It also works well with collembolans (F.F., unpublished data). F. Sperling (personal communication) notes that it fails in some cases because the third C from the 3′ end frequently mutates to a T. He suggests constructing degenerate primer with a mixture of C and T at this position. This primer works well with C1-N-2191 (Nancy) and L2-N-3014 (Pat). The single-stranded product of this primer usually does not sequence well.

C1-J-1859 (alias RonII)

(26mer)	5′ GGAACIGGATGAACAGTTTACCCICC 3′	
Fly	..T..A..T.....T.....T..	1834
Bee	.....T.....A..A..T..A..	2154
Locust	..T..A.....A..T..T..	1793
Springtail	.....C.....T..C.....C..	
Cicada	.....C..T.....T..T..	
Cow	.....A..C.....C..G.....T..	6053
Toad	..C..A..T.....T..G.....G..	7763
Sea urchin	.....T..C.....TA.C.....T..	6151
Roundworm	..G..TAGG.....C.....A..	8235

Designed by B.C. for insects in general, it works well in thysanopterans, coleopterans, homopterans, hemipterans, lepidopterans, and psocopterans. Single-stranded amplifications using this primer usually sequence well. I = inosine. This degenerate primer has a mixture of A and T at position 15 and a mixture of C and T at position 21. It works well in combination with C1-N-2191 (Nancy) and L2-N-3014 (Pat).

C1-J-2183 (alias Jerry)

(23mer)	5′ CAACATTTATTTTGATTTTTTGG 3′	
Fly	.....	2161
Bee	.....	2481
Locust	.....C.....C.....	2120
Thrips	.....	
Milkweed bug	.....C.....	
Springtail	.....C.....C..	
Brine shrimp	.....C.....G..C.....	253
Cow	.....C.....C.....C.....	6380
Toad	.....CC.G..C.....C.....	8090
Sea urchin	.....CC....C..GC.....	6478
Roundworm	.....G.....	8559

Designed by B.C. for insects in general, it can be used to amplify coleopterans, thysanopterans, homopterans, psocopterans, lepidopterans, and hemipterans. It works well in combination with L2-N-3014 (Pat). The brine shrimp, *Artemia*, sequence is from GeneBank accession number X67958.

C1-N-2191 (alias Nancy)

(26mer)	5′ CCCGGTAAAATTAAAATATAAACTTC 3′	
Fly	.....	2216
Bee	..A.....	2536
Locust	..T.....	2175
Springtail	.....C.....T.....	
Brine shrimp	..G.....G....G.....C.....	308
Cow	..A.....G.....	6435
Toad	..T.....G..A.G...G..C.....	8145
Sea urchin	.....G..A.G.....C..C..	6533
Roundworm	G.T.....C.....T.....	8614

Designed by the Harrison laboratory where they report that it works very well with virtually all insects tested, including lepidopterans, dipterans, coleopterans, thysanopterans, hemipterans, homopterans. It also works well with collembolans (F.F., unpublished data). It usually makes good single-stranded product. It works well in combination with C1-J-1718, C1-J-1751 (Ron), and C1-J-1859 (Ron II). The brine shrimp, *Artemia*, sequence is from GeneBank accession number X67254.

C1-J-2195 (alias CO1-RLR)

(24mer)	5' TTGATTTTTTGGTCATCCAGAAGT 3'	
Fly	.....C..T.....	2172
Bee	.....	2492
Locust	.....C.....C.....	2131
Springtail	.....C..C..C..T.....	
Brine shrimp	...G..C.....C..T.....	264
Cow	C.....C.....A..C..C.....	6391
Toad	C.....C.....G..C.....	8101
Sea urchin	C..GC.....A..C..C..G..	6489
Roundworm	.....T.....	8570

Designed by Roehrdanz (1993) for insects in general, it works well on some coleopterans, hymenopterans, dipterans, and lepidopterans. The brine shrimp, *Artemia*, sequence is from GeneBank accession number X67958.

C1-N-2329 (alias K525)

(23mer)	5' ACTGTAAATATATGATGAGCTCA 3'	
Fly	.....	2351
Bee	.....G.....T.....	2671
Locust	.....G.....	2310
Springtail	.....G.....G..G.....	
Cow	..A..G.....G..G.....	6570
Toad	..C.....G.....G.....	8280
Sea urchin	.....C.....G..G..C..	6668
Roundworm	.....T.C.....G.....	8749

Designed by F. Sperling, it works very well with many insects, including some lepidopterans, dipterans, and coleopterans.

C1-J-2441 (alias Dick)

(32mer)	5' CCAACAGGAATTAAAATTTTTAGATGATTAGC 3'	
Fly	..T.....	2410
Bee	.....G.....	2730
Locust	.....GG.A..C.....A....	2369
Springtail	.....T..GG.....T...A.T..	
Cow	.....C..GG.A...G.C..C..C.....G..	6629
Toad	..T.....TG.....G.A.....C.....	8339
Sea urchin	.....T.A..GG....C.....A.G..	6727
Roundworm	.....TG.....G.G.....G..	8808

This primer, designed for insects in general by B. C., works well in some hemipterans and some thysanopterans in combination with L2-N-3014 (Pat).

C1-N-2659 (alias Mila1)

(20mer)	5' GCTAATCCAGTGAATAATGG 3'	
Fly	.TC.....A.....	2678
Bee	AA.....A.T.....	2998
Locust	.T.....T..A.....A..	2637
Springtail	CTAC....T.....G..	
Cow	.TAT.A..T.A.....G...	6897
Toad	.TAT.A.....A.....C..	8607
Sea urchin	CTAT.A....A...G.GG..	6995
Roundworm	A.AT.C..T..A.TA...CT	9076

Designed by F. Sperling (Sperling & Hickey 1994), it works with some lepidopterans, dipterans, and coleopterans. This primer, or more taxon-specific modifications of it, is useful as an internal sequencing primer for the segment amplified by the combination C1-J-2183 (Jerry) and L2-N-3014 (Pat). Sperling also suggests an alternative primer to be used in place of this one: C1-J-2578 (k741), a 23mer. It has been found to work well on a wide range of lepidopterans.

C1-J-2797

(22mer)	5' CCTCGACGTTATTCAGATTACC 3'	
Fly	.....	2776
Bee	..A.....C..T.	3096
Locust	..A.....A.....T.....T.	2735
Cow	.....A..C..C..C....	6995
Toad	.....A..C..T..C....	8705
Sea urchin	..A.....G..C..A..C..T.	7093
Roundworm	..A..TAAA....T.....	9174

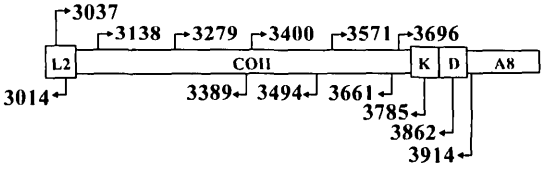
Designed by Tom Powers, it works with dipterans and coleopterans. Base position 2798 might be more effective if substituted with a 'C' or made degenerate.

tRNA-Leucine (UUR)

TL2-N-3014 (alias Pat)

(25mer)	5' TCCAATGCACT———AATCTGCCATATTA 3'	
Fly	....T.....———.....	3038
Bee	.T.....T———.T.....	3382
Locust	..T.....T.———.....	2997
Human	.TTT....GA.TACCGGGC.....C...	3262
Toad	..TTTC..AT.AGCCAGGC.....CGC..	4756
Sea urchin	C.TTC....T.———..CCACTTTGC.AC	2117
Roundworm	A.A.....T———CT.A...T....C.	3334

Designed by the Harrison laboratory, where they report that it works for gerrids, weevils, mosquitoes, flies, and lepidopterans. It is complementary to TL2-J-3037. It works well in combination with C1-J-1718, C1-J-1751 (Ron), C1-J-1859 (Ron II), C1-J-2183 (Jerry), and C1-J-2441 (Dick).



TL2-J-3037 (alias A-tLEU)

(20mer)	5' ATGGCAGATT———AGTGCAATGG 3'	
Fly	.....———.....	3018
Bee	.....A.———A.....T..A	3381
D. melanogaster	.....———.....A.	
Locust	T.....———.A....T.A.	2977
Human	.....GCCCGGTA.TC...TAAA	3236
Toad	G.....GCCTGGCT.A...G.AA.	4730
Sea urchin	CAAAGT.G..———.A....GAA.	2097
Roundworm	.TA...T.AG———A.....T.T.	3333

Designed by Liu & Beckenbach (1992) and used to amplify and sequence individuals from five different orders of insects (listed under N2-J-1006). It has worked for several species of collembolans (F.F., unpublished data), cicadas (*Magicicada;* C.S., unpublished data), green lacewings (*Chrysoperla;* M. Martinez-Wells, personal communication) and many orthopterans, including representatives of the nemobiine

genus *Caconemobius* (A. Phillips, personal communication) and species in the following families: Acrididae, Eumastacidae, Lentulidae, Mimnermidae, Ommexechidae, Pauliniidae, Proscopiidae, Rhipidistiidae, and Romaleidae (P.F., unpublished data). Beckenbach et al. (1993) used it to amplify 12 species of the *D. obscura* group.

Simon et al. (1993) modified this primer to make it more universal for insects by removing the four 3′ bases and adding four bases at the 5′ end. This primer (A-tLEU mod) was then used in combination with C3-N-5460 (COIIIb) to amplify a 2,463-bp piece of DNA for high resolution restriction fragment length polymorphism (RFLP) analysis of periodical cicadas. The same strategy works for New Zealand wetas and chafer beetles (T. King & G. Wallis, personal communication). The modified version has been found to work well in combination with the modified version of TK-N-3785 (B-tLYS mod) to amplify and sequence the COII gene in collembolas, cicadas, and crickets. In vertebrates, the tRNA leucine (UUR) is transposed compared with insects. This primer is complementary to a part of TL2-N-3014 (Pat). The *D. melanogaster* sequence is from Sprinzl et al. 1987.

Cytochrome Oxidase II

Eight conserved primers located within the COII gene are provided below. These have been found to work well in insects. If additional primers are needed, the published COII gene alignment for 10 orders of insects provided by Liu & Beckenbach (1992) is a convenient aid to primer design. For all COII primer comparisons: cricket refers to *Caconemobius* spp. (A. Phillips, personal communication); cicada, *Magicicada* spp. (C.S., unpublished data); springtail (collembolan), *O. villosa* (F.F., unpublished data). In the primer notes, *collembolans* refers to the genera, *Orchesella* spp., *Isotomurus* spp., *Tetrodontophora* spp., and *Thaumanura* spp. (F.F., unpublished data).

C2-J-3138 (alias Pierre)

(23mer)	5′ AGCGCCTCTCCTTTAATAGAACA 3′	
Fly	..A..T.....G.....	3116
Bee	TCAAAT..ATA..ATGCT..TA.	3651
Locust	G.A..T..A..A.....	3074
Cicada	GCTA.A.....A.....	
Cricket	..TT.....AC.T.....	
Springtail	G.A..A..A.....	
Human	GCTA.T.....A.C.....G.	7619
Toad	GCA.....AA.T.....G.	9142
Sea urchin	GCAT....C...C.T..G..GG.	7744
Roundworm	CAGCATAG.TTA..TGCTAGTT.	9685
Roundworm (mod.)	..TTTAG..AG..AT.....TTG	9691

Designed by F. Sperling for budworms, it works for various flies, weevils, and lepidopterans. The sequence seems conserved at the 3′ end in all insects examined (Liu & Beckenbach 1992), except ants and bees (but it is well conserved in wasps). The second roundworm sequence is our modified amino acid alignment, which seems to match a little better than that of Okimoto et al. (1992).

C2-J-3279 (alias A-171)

(26mer)	5′ GGTCAAACAATTGAGTCTATTTGAAC 3′	
Fly	..A...CTT.....AATA.....	3254
Bee	AA...T.AT.....AAT.....	3789
Locust	.....TTT.....AA....C.....	3212
Cricket	..A.....TA...G....	
Springtail	TCC...C.TT.A..A.TAT.....	
Human	.C...GGA...A..AA.CG.C.....	7757
Toad	.CA...GAG..C..AATAG.G.....	9280
Sea urchin	..A...GAGT.A..AA.A.....	7882
Roundworm	TA...GTTTGG...A.TAT.G..T.G	9826

Designed by Liu & Beckenbach (1992), it is well conserved at the 3′ end in collembolans and in the 10 orders of insects they studied. The 3′ end is conserved in vertebrates as well but not in roundworms. P.F. has found that it works successfully in representatives of nine different orthopteran families (listed in TL2-J-3037).

C2-N-3389 (alias Marilyn)

(20mer)	5′ TCATAAGTTCARTATCATTG 3′	
Fly	.....C....G.....	3408
Bee	.....T.A...A.....	3943
Locust	.....C....G.....	3366
Cricket	.....TT....A.....	
Springtail	.....A...G..C.....	
Human	..G..G.....G..C.....	7911
Toad	..G..GC....G.....	9434
Sea urchin	..G..TC....G.....C..	8036
Roundworm	.....TC... A.....C..	9983

Designed by the Harrison laboratory, where they report that it works for weevils, lepidopterans, flies, and mosquitoes. The sequence is partially complementary to C2-J-3400. Note that it is a degenerate primer (base 9 mixed A plus G). It works well for internal priming. A slightly modified version (six bases longer at the 3′ end, eight bases shorter at the 5′ end) has been successfully used as internal sequencing primer in collembolans (F.F., unpublished), cicada (*Magicicada;* C.S., unpublished data), and lacewings (*Chrysoperla;* M. Martinez-Wells, personal communication).

C2-J-3400 (alias A-298)

(21mer)	5′ ATTGGACATCAATGATATTGA 3′	
Fly	.....T.....C...	3380
Bee	.....T.....	3915
Locust	.....GA.....C...	3338
Cricket	.....	
Springtail	..A.....G..C...	
Human	.....C..C.....G..C...	7883
Toad	..C..C..C.....C...	9406
Sea urchin	T.C..T.....G.....C...	8008
Roundworm	.CA.....G.....	9955

Designed by Liu & Beckenbach (1992), it is very well conserved in all insects studied. This primer works for the cricket Caconemobius (A. Phillips, personal communication) and in representatives of nine different orthopteran families

(listed in TL2-J-3037). It is partially complementary to C2-N-3389. A slightly modified version (one base shorter with the 5′ end shifted three bases downstream) was used by Beckenbach et al. (1993) for 12 species of the *D. obscura* group.

C2-N-3494 (alias B-434)

(23mer)	5′ GGTAAAACTACTCGATTATCAAC 3′	
Fly	.....T.....	3516
Bee	..A.TT....T.....GT	4045
Locust	.....TG..GT.....	3474
Cicada	.....GC..GT.....C..	
Cricket	.....T.GT.....T..	
Springtail	..A.TT..TGTA..G.....T..	
Human	..G.GT.....G.....	8019
Toad	..G.CT....T.A.....	9542
Sea urchin	..A.GG..C.A...G..G..C..	8144
Roundworm	..A.T...ACAA.....	10091

Designed by Liu & Beckenbach (1992) and used to amplify five orders of insects (listed under N2-J-1006). Removal of the three 3′ bases might increase its versatility.

C2-J-3571 (alias A-470)

(19mer)	5′ AGATGTTCTTCACTCATGA 3′	
Fly	.....AA....T..T...	3553
Bee	.....AA....T.....	4082
Locust	T.....A..C.....	3511
Cicada	T.....CA....T..T...	
Cricket	T.....AA....T..T...	
Springtail	...C..AT.A..T..T...	
Human	...C..CT.G.....	8057
Toad	...C..C..C.....G...	9579
Sea urchin	.....A..A.....C...	8181
Roundworm	T.....A....TG.T..G	10128

Designed by Liu & Beckenbach (1992) and used to amplify and sequence insects from five orders (listed under N2-J-1006). It is reported to work successfully in representatives of nine different orthopteran families (listed in TL2-J-3037). Three different versions (five more bases at the 5′ end and one less at the 3′ end) have been designed by F. Sperling (personal communication) and reported to work in all lepidopterans tested (including microlepidopterans), and the weevil *Pissodes*.

C2-N-3661 (alias Barbara)

(24mer)	5′ CCACAAATTTCTGAACATTGACCA 3′	
Fly	..G.....	3684
Bee	.....	4213
Locust	.....G.....G.....G	3642
Cicada	.....A.....	
Cricket	.....A.....	
Springtail	..G.....A..A.....	
Human	.....G.....A..G.....	8187
Toad	..G.....T..G	9710
Sea urchin	..G.....G..G..C..G...	8312
Roundworm	.....C.....	10259

Designed by the Harrison laboratory where they report that it works for weevils, looper, *Ostrinia*, and papilionid butterflies. A modified version (with the 3′ base deleted, and 'R' instead of 'A' at the new third base position) has been successfully used in honey bees and *Drosophila* (R. H. Crozier, personal communication). Another modified version (three bases longer at the 5′ end and seven bases shorter at the 3′ end; alias C2B-605) works in several families of collembolans (F.F., unpublished data), periodical cicadas (C.S., unpublished data), *Caconemobius* crickets (A. Phillips, personal communication), and lacewings (*Chrysoperla;* M. Martinez-Wells, personal communication) for amplifications with TL2-J-3037 and sequencing. C2-N-3661 is in a very conserved part of the sequence, but the 3′-end base is substituted in some species of the *D. obscura* group (Beckenbach et al. 1993) and might be better omitted. The primer is partially complementary to C2-J-3696 (A-611).

C2-J-3696 (alias A-611)

(23mer)	5′ GAAATTTGTGGAGCAAATCATAG 3′	
Fly	.....C..G..T.....	3674
Bee	.....TAT.....	4203
Locust	.....C.....T.....	3632
Cicada	.....	
Cricket	.....T.....	
Springtail	.....C.....	
Human	.....C.....C..C..	8177
Toad	.....C.....C..C..	9700
Sea urchin	.....C..G..T..C.....	8302
Roundworm	..G.....	10249

Designed by Liu & Beckenbach (1992) (mistakenly labeled A-592) and used in insects from five different orders (listed under N2-J-1006). P.F. has found that it works successfully in representatives of nine different orthopteran families (listed in TL2-J-3037). It is partially complementary to C2-N-3661 (Barbara). This primer might be useful for amplifying the COII-ATPase8 junction (containing the tRNA lysine and the tRNA aspartic acid). A modified version (shifted nine bases upstream) is published in the *Simple Fool's Guide to PCR* (Palumbi et al. 1991), but it is specific for sea urchin. C2-J-3696 has been used in combination with C3-N-5460 (CO3b) to produce reliably a 1700+ bp segment for PCR/RFLP of periodical cicadas, which excludes most of the COII gene. This fragment, because it is smaller, is amplified more reliably than the PCR/RFLP fragment described in Simon et al. (1993) amplified by TL2-J-3037 and C3-N-5460.

tRNA-Lys

TK-N-3785 (alias B-tLYS)

(20mer)	5′ GTTTAAGAGACCAGTACTTG 3′	
Fly	.....	3804
Bee	A.....A...T..ACT...A	4407
Locust	.....T.....A	3826
Human	AC....A..GTT.A.G..AA	8310
Toad	.C....A..G.TGTCG..AA	9819
Sea urchin	AA.....T.T.AAG...T	8442
Roundworm	AA....A..TTTG.C....T	3283

Designed by Liu & Beckenbach (1992) and used to amplify five orders of insects (listed under N2-J-1006) and 12 species of the *D. obscura* group (Beckenbach et al. 1993). P.F. has found

that it works successfully in representatives of nine different orthopteran families (listed in TL2-J-3037). A slightly modified, more universal version (one base shorter at the 3′ end) designed by C.S. (B-tLYS mod) works successfully in collembolans (F.F., unpublished data), periodical cicadas (C.S., unpublished data), *Caconemobius* crickets (A. Phillips, personal communication), and green lacewings (M. Martinez-Wells, personal communication) for amplification and sequencing of the entire gene in combination with TL2-J-3037 (A-tLEU mod). The overlapping Harrison laboratory primer 'Eva' has it's 3′ end at 3782 and is reported to work very well for some weevils, carabids, blowflies, mosquitoes, and a wide range of lepidopterans but not for gerrids (F. Sperling, personal communication). It should be noted that tRNA lysine is transposed in honey bees (Crozier et al. 1989) and inverted in locust (Haucke & Gellisen 1988) and the tRNA aspartic acid lies between the end of the COII gene and the tRNA lysine. Willis et al. (1992) use a primer in the tRNA aspartic acid to amplify the honey bee COII gene (TD-N-3862; see below).

tRNA-Asp

TD-N-3862 (alias tRNA-Asp)

(24mer)	5′ GGCCGTCTGACAAACTAATGTTAT 3′	
Fly	TTTA..T.....T.....	3885
Bee	TTTA..T.....	4335
Locust	TTTA.AT.....TT.....	3763
Human	TTAAC.T.....G.T...AA..	7558
Toad	TT.GCCT.....GGCGTGC.A..	9078
Sea urchin	TTTG.C.....C.C..A.....	1896
Roundworm	TTTT.AT.A...GT.A.CAA..C.	1689

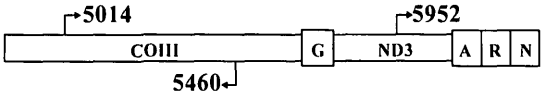
Designed by Willis et al. (1992) for honey bees and used to amplify the entire COII gene. In vertebrates, the tRNA aspartic acid is on the 5′ side of the COII gene.

ATPase 8

A8-N-3914

(24mer)	5′ TCATCTTATAGGTACTATTTGAGG 3′	
Fly	.....A..T...G.....T..	3937
Bee	....T.....T.....	4470
Locust	....A....T...GA.....	3904
Sea urchin	.....AAGC.AA.T...G...T..	8502
Cow	....G..GAC.TGT...G...C..	8155
Toad	....GGGCCT..GTT..AC..T..	9899

Designed by H.L. for the ants in the genus *Lasius*, it has been found to amplify the COII gene successfully for the ant with the primer TL2-J-3037. Both primers also work for the flea (*C. felis*) (H.L., unpublished data).



Cytochrome Oxidase Subunit III

Two conserved primers that have been found to work well in insects are presented below. To design additional primers for the COIII and for A6, A8, and COII genes of insects we recommend using the locust and *Drosophila* sequence alignment provided in Haucke & Gellissen (1988).

C3-J-5014 (alias CO3a)

(20mer)	5′ TTATTTATTGCATCAGAAGT 3′	
Fly	.....TT.....	4995
Bee	.....TT.....T.	5534
Locust	.....	4970
Human	C.....A.C.....	9459
Toad	C.....A.....	10963
Sea urchin	C.....AA.C.....G..	9582
Roundworm	.....G.GTTTAGG..GT.	5918

Designed by C.S. (Simon et al. 1991) and used in conjunction with C3-N-5460 (CO3b) to amplify and sequence periodical cicadas and collembolans (C.S., unpublished data; F.F., unpublished data). A slightly modified version (shifted four bases upstream) was used by Vogler & DeSalle (1993) to amplify and sequence a 220-bp segment of the COIII gene of tiger beetles in conjunction with the primer 5′-AGGCTCGAATC-CACACTCAAAAGGTGA-3′, located at position 5215 of *D. yakuba*.

C3-N-5460 (alias CO3b)

(20mer)	5′ TCAACAAAGTGTCAGTATCA 3′	
Fly	.....A.....	5479
Bee	..T.....A.....A.....	6018
Locust	.....	5454
Human	..T.....A..C.....	9943
Toad	.....G.....	11444
Sea urchin	..T.....C.....	10066
Roundworm	.....A.....A...A.	6396

Designed by C.S. (Simon et al. 1991) and used in conjunction with C3-J-5014 to amplify and sequence periodical cicadas and collembolans (C.S., unpublished data; F.F., unpublished data). It can be used in combination with TL2-J-3037 (A-tLEU mod) or C2-J-3696 for PCR/RFLP analysis (Simon et. al. 1993). This same PCR/RFLP strategy works well for New Zealand wetas (T. King & G. Wallis, personal communication).

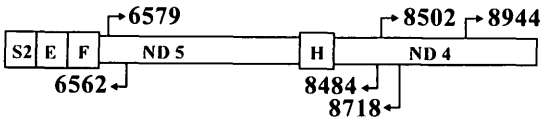
NADH Dehydrogenase Subunit 3

N3-J-5952

(21mer)	5′ GAATGAAATCAAGGTATATTA 3′	
Fly	.....	5932
Bee	.....TAA.TAA.TAT...	6503
Locust	.....AGC.C.T	5882
Sea urchin	..G....TAA....GGGCC..	10516
Cow	..G....C....AAAGG.C..	10139
Toad	.....CT.....AGGCC..	11871
Roundworm	..G...TGAT.T....A....	11656

Designed by H.L. for the flea (*C. felis*), it has been found to amplify the large cluster of tRNAs in the gene junction between ND3 and ND5 in

the flea when used in combination with N5-N-6562 (H.L., unpublished data).



NADH Dehydrogenase Subunit 5

N5-N-6562

(18mer)	5′ GATCAAGGTTGGTCAGAA 3′	
Fly	.....	6579
Bee	T.....AAAAT.T.T.T	7019
Locust	...TCT.....AGGT..G	6471
Sea urchin	..C.G...A...CA....	13828
Cow	..C.T.ATC..ACT....	13768
Toad	....T.TCC..A.AT...	15517
Roundworm	T..TT.TAAAATAA....	13104

Designed by H.L. for the flea (*C. felis*), it has been found to amplify the gene junction between ND5 and ND3 successfully for the flea in combination with the primer N3-J-5952 (H.L., unpublished data).

N5-J-6579

(18mer)	5′ TTCTGACCAACCTTGATC 3′	
Fly	.....	6562
Bee	A.A.A.ATTTT.....A	7002
Locust	C..ACCT.....AGA...	6454
Sea urchin	....TG...T...C.G..	13845
Cow	....AGT..GAT.A.G..	13785
Toad	...AT.T..GGA.A....	15534
Roundworm	....TTATTTTA.AA..A	13121

Designed by H.L. for the flea (*C. felis*), it has been found to amplify the gene junction between ND5 and ND4 for the flea in combination with the primer N4-N-8484 (H.L., unpublished data).

NADH Dehydrogenase Subunit 4

N4-N-8484

(20mer)	5′ GCTAATATAGCAGCTCCTCC 3′	
Fly	.....	8503
Bee	T.A.....G.T....AGT	8913
Locust	T.....G..TT.....	8391
Sea urchin	..C..AT.G.G.TTA.....	11664
Cow	A.C..CT....TCTA..C..	11621
Toad	.....CCTA..A..	13362
Roundworm	T.A....GG.GT.TA..A..	7481

Designed by H.L. for the flea (*C. felis*), it has been found to amplify the gene junction between ND4 and ND5 successfully for the flea in combination with the primer N5-J-6579. It works well for double-stranded sequencing. This primer could be made more versatile for insects by removing the three 3′ nucleotides.

N4-J-8502 (alias ND4R2)

(22mer)	5′ GTAGGAGGAGCTGCTATATTAG 3′	
Fly	..T.....	8481
Bee	.A.ACT....A.C.....T.	8891
Locust	.AC.....AA..C.....	8369
Human	..G..G..TAAG..G.GG....	11874
Toad	.AC..T..TAGG.....	13384
Sea urchin	.AG.....TAA.C.C.AT..G.	11687
Roundworm	.A...T..TA.AC.CC....T.	7503

Designed by T. Powers and used in conjunction with N4-N-8718 to amplify *Diuraphis* spp. aphids.

N4-N-8718 (alias ND4F3)

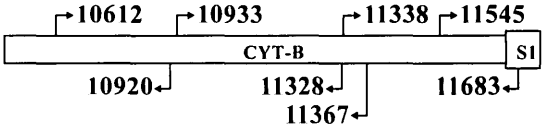
(20mer)	5′ GCTTATTCATCGGTTGCTCA 3′	
Fly	..A.....A.....	8737
Bee	..AAT.....TA...T...	9147
Locust	.....T..T.....	8625
Human	..A..C..T..AA.CAGC..	11618
Toad	..C..C.....T..AAGC..	13122
Sea urchin	..C..C..T..A....GG..	11431
Roundworm	.....T..A..CA....	7247

Designed by T. Powers and used in conjunction with N4-J-8502 to amplify *Diuraphis* spp. aphids.

N4-J-8944 (alias ND4)

(20mer)	5′ GGAGCTTCAACATGAGCTTT 3′	
Fly	.....	8925
Bee	.....	9335
Locust	.....C..	8813
Human	..G.....G.....G.....	11430
Toad	..G.....T.....	12934
Sea urchin	.....T..C..	11240
Roundworm	.....C..T.....	7068

Designed by Y. C. Crozier (used in Jermiin & Crozier 1994), it has been used successfully with bees, some ants, and *Drosophila* in conjunction with N1-N-12595 and CB-N-11367 to amplify a region including the 5′ half of ND4, ND4L, tRNA threonine, tRNA proline, ND6, Cytb, tRNA serine, and most of ND1 (Jermiin & Crozier 1994).



Cytochrome b

CB-J-10612 (alias CB1L)

(26mer)	5′ CCATCCAACATCTCAGCATGATGAAA 3′	
Fly	...ATT..T..T...AG.....	10587
Bee	...GTA..T..TAATTATAT.....	11079
Locust	...A.A.....T...AT.....	10468
Human	.....C.....	14816
Toad	.....A.....T...T...T.....	16321
Sea urchin	..C.....C.T..CATT..G.....	14581
Roundworm	AG.AAA.CTT.AA..TT.A.....	4564

This shorter version of the universal cytochrome B primer (Kocher et al. 1989) was included in Palumbi et al. (1991). It works in most verte-

brates in conjunction with CB-N-10920. It should work well in insects.

CB-J-10933 (alias CB1)

(26mer)	5′ TATGTACTACCATGAGGACAAATATC 3′	
Fly	.....TT....T.....	10908
Bee	.....	11400
Locust	..C..TT.....T.....	10789
Human	.....C..C..G.....C.....	15137
Toad	.....T.....	16642
Sea urchin	.....CT..GTC..G..G.....G..	14902
Roundworm	.....TT..GTT....CT.....AG	4885

Designed by Y. C. Crozier (Crozier et al. 1991), it works with bees, ants, and *Drosophila* in conjunction with N1-N-12595 and CB-N-11367 to amplify a region including the 3′ half of Cytb, tRNA serine and most of ND1 (Jermiin & Crozier 1994). It is partially complementary to CB-N-10920.

CB-N-10920 (alias CB2-H)

(24mer)	5′ CCCTCAGAATGATATTTGTCCTCA 3′	
Fly	T.....A.....	10943
Bee	A.....AT.....	11435
Locust	T.....A.....A.....	10824
Human	.....G.....	15172
Toad	.....A..A.....	16677
Sea urchin	AG....A..G..C.....C..C..	14937
Roundworm	TG....A...CT.....AG....	4920

This shorter version of the universal cytochrome b primer (Kocher et al. 1989) was included in Palumbi et al. (1991). It works in most vertebrates in conjunction with CB-J-10612. It is well conserved in *Drosophila* and *Apis* and partially complementary to CB-A-10933.

CB-J-11338 (alias CB3R-L)

(23mer)	5′ CATATTAAACCCGAATGATAYTT 3′	
Fly	..C...C....A.....T..	11316
Bee	.....T.....T..	11808
Locust	..C...C....A.....T..	11197
Human	..C..C..G.....T..	15545
Toad	.....A.....C..	17050
Sea urchin	..C...C.G..A.....G..C..	15310
Roundworm	.....GTT..A..G.....T..	5290

Designed by Palumbi et al. (1991), it is reported to work for many vertebrates groups. Note that the third position includes both T and C bases. It is very well conserved in *Drosophila, Locusta,* and *Apis*. It is partially complementary to CB-N-11328.

CB-N-11328 (alias CB3-H)

(21mer)	5′ GGCAAATAGGAARTATCATTC 3′	
Fly	A.....AA..A.....	11348
Bee	T.....A.....	11840
Locust	T.....G..A.....	11229
Human	...G.....A.....	15577
Toad	...G.....G.....	17082
Sea urchin	...G....A...G..C.....	15342
Roundworm	A.....AA..A.....C..	5322

Designed by Palumbi et al. (1991) and reported to work in most vertebrates. It is very well conserved in *Drosophila, Locusta,* and *Apis*. It is partially complementary to CB-J-11338.

CB-N-11367 (alias CB2)

(26mer)	5′ ATTACACCTCCTAATTTATTAGGAAT 3′	
Fly	..A..T.....	11392
Bee	.....	11884
Locust	..A..T.....	11273
Human	.GG..G.....G...G.....G.C	15621
Toad	.AC..T..G....G...G..——T..	17123
Sea urchin	.....C.....C..G..G..G..	15386
Roundworm	.....C...AAG.C.....T.....	5366

Designed by Y. C. Crozier (Crozier et al. 1991). R. H. Crozier (personal communication) reports that it works in bees, ants, and *Drosophila* in conjunction with CB-J-10933 and N4-J-8944 to amplify a region including the 5′ half of ND4, ND4L, tRNA threonine, tRNA proline, ND6, and the 3′ half of Cytb.

CB-J-11545 (alias CTBA)

(23mer)	5′ ACATGAATTGGAGCTCGACCAGT 3′	
Fly	.....	11523
Bee	.....T.A..TAAA.A.TT.AT	12012
Locust	.....AAA.....	11404
Human	..C.....C....GA.A.....	15752
Toad	..C.....G..A.....	17254
Sea urchin	.....G..A..CAGA.A...C..	15517
Roundworm	.G....C.A..TCAAT.CA.T..	5488

Designed by Pruess et al. (1992), it could be useful in conjunction with N1-N-11841 to sequence the ND1-Cytb junction (including the tRNA serine) and with a modified version of LR-N-12866 (LR-N-12854) to amplify and sequence the entire ND1 gene.

tRNA-Ser

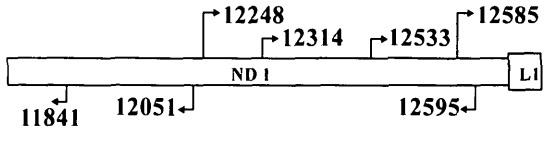
TS1-N-11683 (alias TRs)

(25mer)	5′ TATTTCTTTATTATGTTTTCAAAAC 3′	
Fly	A.A....A.C.....	11707
Bee	.....A.....A.....T	12250
Locust	AC..CT...C.....	11585
Human	A.CCC.CCA.AGC..G.....GC.	7464
Toad	A.CCC.CC..C.C..A.....GT.	8970
Sea urchin	A.CC...A.CAA.C.G.....GC.	10132
Roundworm	..CC..AAAA–C.....	1635

Designed by L. Jermiin for the ant *Tetraponera rufoniger,* it has been used successfully with CB-J-10933 and N4-J-8944 (Jermiin & Crozier 1994).

NADH Dehydrogenase Subunit 1

For all ND1 primer comparisons, moth refers to fall armyworm (*S. frugiperda*) Pashley & Ke 1992). Mosquito refers to *A. albopictus* (Hsu Chen et al. 1984). Tiger beetle refers to *C. dorsalis* (Vogler et al. 1993a, Vogler & DeSalle 1993).



N1-N-11841 (alias CTBB)

(27mer)	5′ GGTACATTACCTCGGTTTCGTTATGAT 3′	
Fly	.....	11867
Bee	....TT.....AA....A.....	12421
Locust	.....T.....T.....	11749
Human	ACAG...AC..C..A..C..C..C..C	4129
Toad	.CAT...AC..A..A.....A.....	5636
Sea urchin	.C.G.C.AC..A..A..C..G.....	2996
Roundworm	A..T...AC.....T.A.....	2528

Designed by Pruess et al. (1992), it is used in conjunction with CB-J-11545 to sequence the ND1-Cytb junction (including the tRNA serine) in the black fly *Simulium vittatum*.

N1-N-12051

(23mer)	5′ GATTTTGCTGAAGGTGAATCAGA 3′	
Fly	.....	12074
Bee	.....AAT.....A.....T..	12628
Locust	.....A..G..C..	11956
Sea urchin	..CC.AA.A.....A.....T..	2767
Cow	.....AA.....A.....	3695
Toad	..CC..A.A.....A.....T..	5408
Roundworm	.....T.....AGG..	2339

Designed by H.L. for the flea (*C. felis*), it has been found to amplify the gene junction between ND1 and Cytb for the flea in conjunction with the primer CB-J-11545. It works well for double-stranded sequencing.

N1-J-12248 (alias FawND1B)

(21mer)	5′ AAGCTAATCTAACTTCATAAG 3′	
Fly	.....	12227
Bee	..AA...AT...T....A.T.	12781
Locust	.....T.....	12109
Moth	.....	
Human	TG....GGG.G.....T.	3748
Toad	.TC.A.GAG.T.....G..T.	5255
Sea urchin	GT..C.GG....TC.....T.	2615
Roundworm	...AA..AGC..T.....	2210

Designed by Pashley & Ke (1992) for the fall armyworm moth, to be used with LR-N-12866. It works with Papilionoidea and Noctuoidea-(*Spodoptera, Pseudoplusia, Cosmosoma, Symmerista,* and *Phoebis,* [Pashley & Ke 1992, Weller et al. 1994]). In all primer comparisons moth refers to fall armyworm.

N1-J-12314 (alias FawND1A)

(22mer)	5′ TAGAATTAGAAGATCAACCAGC 3′	
Fly	.....G.....	11294
Bee	..C.....AT...T...A.	12847
Locust	.T.....	12175
Moth	.....	
Human	.T..G..T..T.C...C..T.A	3682
Toad	.T.....G..T.....T..G.A	5189
Sea urchin	....T..T..G.C.....A	2549
Roundworm	.T..T..TCTTACGAT...GCT	2144

Designed by Pashley & Ke (1992) to amplify the fall armyworm moth in conjunction with a modified version of LR-N-12866. It works with Papilionoidea and Noctuoidea (*Spodoptera, Pseudoplusia, Cosmosoma, Symmerista,* and *Phoebis,* [Pashley & Ke 1992, Weller et al. 1994]).

N1-J-12533

(20mer)	5′ GCATCACAAAAAGGCTGAGG 3′	
Fly	.....T.....	12514
Bee	.....T.....T...AA	13062
Locust	.....T.....T.....	12393
Mosquito	.T.....T.....T..CAA	
Moth	.....TG.....T..TAA	
Human	..G...GCG..G..T..TA.	3461
Toad	C....TGC..TT..T...AT	4968
Sea urchin	C....TGC...G..T..TAA	2328
Roundworm	C....CA.T....CT..T.C	1923

Designed by Vogler et al. (1993a) for the tiger beetle (*Cicindela*) and used in conjunction with LR-N-12866 to amplify several populations and subspecies of *C. dorsalis* (Vogler et al. 1993a, Vogler & DeSalle 1993). The 3′ end is not well conserved in the available insect sequences, but could be easily modified to make it more generally applicable.

N1-J-12585 (alias ND1A)

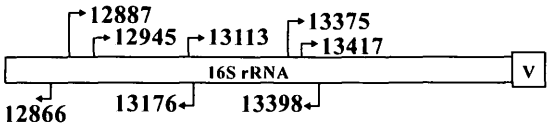
(27mer)	5′ GGTCCCTTACGAATTTGAATATATCCT 3′	
Fly	.....T.....C...	12559
Bee	..A..T.....TCC.....A...	13107
Mosquito	..C.....	
Tiger beetle	..G.....T.....C.....	
Locust	..A.....T.....G.....	12438
Moth	..A..T.....A...	
Human	..G..T..G..T.G...T.....G...	3416
Toad	..G..T.....GTG...T.....A	4923
Sea urchin	..A..T..T....A...C.....G..C	2283
Roundworm	..C..TAG....T....TC.TCT...C	1878

Designed by Pruess et al. (1992), it was used in conjunction with a modified version of LR-N-12866 (LR-N-12854) to amplify the 5′ end of the ND1 gene and the 3′ end of the 16S gene (including the tRNA leucine [CUN]) in nine species of blackflies. It is well conserved in insects.

N1-N-12595 (alias ND1)

(26mer)	5′ GTAGCATTTTTAACTTTATTAGAACG 3′	
Fly	.....T.....	12620
Bee	.....	13168
Mosquito	.....T.....G.....A....	
Tiger beetle	.....T.....A.....	
Locust	..T..T.....A.G..G..	12499
Moth	..T..T.A.....G..G..	
Human	A.G.....CC...TGC.TACC.....	3355
Toad	.....CC.C...C.TA.C.....	4861
Sea urchin	.....T...C.C..AC..G....G..	2222
Roundworm	ATC..T...A.T...C...AT..G..	1817

Designed by Y. C. Crozier and used by Jermiin & Crozier (1994) to amplify hymenopterans. It works in conjunction with N4-J-8944 and CB-J-10933 in some ants and bees to amplify a region including the 5′ half of ND4, ND4L, tRNA threonine, tRNA proline, ND6, Cytb, tRNA serine, and most of ND1 (Jermiin & Crozier 1994).



Large Ribosomal Subunit (16S)

For all 16S primer comparisons, moth refers to fall armyworm (*S. frugiperda*) Pashley & Ke 1992). Mosquito refers to *A. albopictus* (Hsu Chen et al. 1984). Tiger beetle refers to *C. dorsalis* (Vogler et al. 1993a, Vogler & DeSalle 1993). Cricket refers to species in the genus *Gryllus* (D. M. Rand & R. G. Harrison, personal communication).

LR-N-12866 (16Sbi)

(22mer)	5′ ACATGATCTGAGTTCAAACCGG 3′	
Fly	.....	12875
Bee	.....T.....T..A	13415
Moth	.....	
Mosquito	.....	
Locust	.....G.....	12761
Human	..G.....G.....	3059
Toad	..G.....G.....	4551
Sea urchin	..G.....G.....	5661
Roundworm	..G....A.T....T..TT.AT	11232

C.S. designed this reversed version of universal primer LR-J-12887 that is modified to match *Drosophila*. It is very well conserved in vertebrates and insects. Note that the 3′ base is different in honey bee. This primer has been found by H.L. to be useful for amplifying the gene junction between lrRNA and Cytb successfully for the flea with the primer CB-J-11545 and to work well for double-stranded sequencing. This primer was used by Vogler et al. (1993a) to amplify and sequence tiger beetles in conjunction with N1-J-12533. A partially overlapping primer (shifted 11 bases downstream) has been reported to work in *Spodoptera* (Pashley & Ke 1992) in conjunction with N1-J-12314. A similar but longer version of the Pashely & Ke primer is reported by W. Black to be useful for amplification and sequencing of the entire ND1 gene and flanking region in plant hoppers in conjunction with CB-J-11545. Pruess et al. (1992) used an overlapping primer (LR-N-12854) in combination with N1-J-12585 to amplify nine species of blackflies.

LR-J-12887 (alias 16Sbr, 16Sb)

(22mer)	5′ CCGGTCTGAACTCAGATCACGT 3′	
Fly	.....T.....T..	12866
Bee	T..A.T.....A....T..	13394
Moth	.....T.....T..	
Mosquito	.....T.....T..	
Locust	.....	12740
Human	.....	3080
Toad	.....	4572
Sea urchin	.....	5682
Roundworm	AT.AAT.A....A.T.....	11253

Designed by T. Kocher (Wilson laboratory primer 16S1), it can be used in conjunction with LR-N-13398 to amplify a 500–650-bp 16S fragment. It was used by Vogler et al. 1993 to amplify and sequence tiger beetles. A slightly modified version (moved 4 bases upstream) was used by Xiong & Kocher (1991) to amplify black flies in conjunction with LR-N-13398.

LR-N-12945 (alias Faw16S2)

(21mer)	5′ GCGACCTCGATGTTGGATTAA 3′	
Fly	.....	12965
Bee	.....A.....	13473
Moth	.....	
Mosquito	.....	
Locust	.....	12838
Human	A.....C.G	2981
Toad	A.....A.C.	4472
Sea urchin	.....CGG	5583
Roundworm	TTT..A.....T...C.	11152

Designed by Pashley & Ke (1992) to amplify the fall armyworm moth in conjunction with N1-J-12314 and N1-J-12248, it works with Papilionoidea and Noctuoidea (*Spodoptera, Pseudoplusia, Cosmosoma, Symmerista,* and *Phoebis,* [Pashley & Ke 1992, Weller et al. 1994]).

LR-N-13176 (alias 16S2)

(20mer)	5′ ACGAGAAGACCCTATGGAGC 3′	
Fly	.....AA.T.	13195
Bee	....T.....A..AT	13724
Moth	.....A...T	
Mosquito	.....A..T.	
Locust	.....A....	13067
Tiger beetle	.....A..T.	
Human	.....	2717
Toad	.....-.....	4181
Sea urchin	.....G.....	5298
Roundworm	.A..T...T.T.C—...AA	10988

Designed in the Wilson laboratory, it could be easily modified to match most insects. The round worm is very difficult to align for this region.

LR-J-13331 (alias 16Smid2 reverse)

(23mer)	5′ TGATTATGCTACC–TTTGCACAGT 3′	
Fly	.....-.....	13309
Bee	.T.....-....T.....	13835
Moth	.....-....T.....	
Mosquito	.....-..C.....	
Locust	.....G....A.....	13179
Human	.....-.....G..	2604
Toad	.....C.....-.....G..	4076
Sea urchin	.....-.CC.-..G..	5187
Roundworm	.T...T.....-..AATGTCC.	10879

Designed in the Templeton laboratory by J. Patton, it was used in conjunction with SR-N-14588

(12Sai) by K. Shaw to amplify and sequence 17 species in the Hawaiian cricket genus *Laupala*.

LR-J-13375 (alias 16Sc)

(18mer)	5′ TCAGT-GAGCAGGTTAGAC 3′	
Fly	.....-.G.....	13358
Bee	...T.-.....A.C-...	13885
Moth	.....-.G....A.....	
Mosquito	C....-.G.....	
Locust	.....-.....	13229
Human	...C.-.G.....CGGTG.	2554
Toad	...C.T.G.....C.G...	4026
Sea urchin	...C.-.G.....CAG...	5137
Roundworm	...-.A.G....AAGCC.G	10832

Designed by C.S, it can be used to amplify the 5′ half of the 16S gene in combination with SR-N-14588 (12Sai). The roundworm is difficult to align in this region.

LR-J-13417 (alias 16Sa)

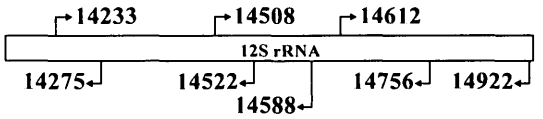
(20mer)	5′ ATGTTTTTGTTAAACAGGCG 3′	
Fly	.....	13398
Cricket	.....	
Bee	.....A.....T.	13924
Moth	.....A.....A.T.	
Mosquito	.....A.....T.	
Locust	.....A.....	13269
Human	.....G.....	2510
Toad	.....G.....	3997
Sea urchin	.....G.....	5092
Roundworm	.A..C...AAA.....TTT.	10798

Designed by C.S. to match *Drosophila* and *Gryllus* (D. M. Rand & R. G. Harrison, personal communication), it is a reversed and slightly modified version of a universal Wilson laboratory primer designed by T. Kocher. It works well for cicadas in combination with SR-N-14588 (12Sai) to amplify the 5′ half of the 16S gene and tRNA valine. See note about hairpin loop below.

LR-N-13398 (alias 16Sar)

(20mer)	5′ CGCCTGTTTAACAAAAACAT 3′	
Fly	.....	13417
Bee	.A.....T.....	13943
Moth	.A.T.....T.....	
Mosquito	.A.....T.....	
Locust	.....T.....	13288
Human	.....C.....	2491
Toad	.....C.....	3977
Sea urchin	.....C.....	5073
Roundworm	.AAA.....TTT...G..T.	10779

Designed by T. Kocher, it is complementary to LR-J-13417. It was used by Xiong & Kocher (1991) to amplify black flies in conjunction with a modified version of LR-J-12887 and by Vogler et al. (1993a) for the tiger beetles in conjunction with LR-J-12887 (16Sbr). It is very conserved over a broad range of organisms and has been used successfully to amplify DNA from crustaceans, sea urchins, vertebrates, insects, and corals. It is surprising that this primer works so well because it is capable of forming a hairpin loop with a 5-bp stem (Uhlenbush et al. 1987).



Small Ribosomal Subunit (12S)

For all 12S primer comparisons, cicada refers to *M. tredecim* (Simon et al. 1991) and cricket refers to species in the genus *Gryllus* (Rand & Harrison, unpublished data). Chimpanzee and gorilla sequences come from Hixson & Brown (1986). The silver fish, *Ctenolepisma longicauta*, and the gastropod, *Celana tramoserica*, are from Ballard et al. 1992.

SR-J-14233 (alias 12Sbi)

(20mer)	5′ AAGAGCGACGGGCGATGTGT 3′	
Fly	.....	14214
Cicada	.....	
Cricket	.....	
Bee	G.A.TT.....T...	14788
Locust	G....T.....	14109
Human	G..G.T.....G.....	1478
Toad	G..G.T.....G.....	2916
Sea urchin	G....T.....	874

Designed by T. Kocher (Kocher et al. 1989) and modified by C. Simon to match insects using cricket (D. M. Rand & R. G. Harrison, personal communication) and *Drosophila* sequences, it can be used to amplify a wide variety of taxa in conjunction with SR-N-14588. The insect primer will amplify humans and the human primer will amplify insects (Simon et al. 1990). It is very conserved and broadly used (Ballard et al. 1992). Because the 3′ end of this primer also matches a sequence in the nuclear rRNA gene of many animal species (De Rijk et al. 1993), another primer was designed (12Sfi; Simon et al. 1991) that is shifted eight bases downstream. This primer amplifies periodical cicadas, crickets, and planthoppers. The roundworm 12S gene is difficult to align with other phyla.

SR-N-14275 (alias 12Sc)

(20mer)	5′ AAGGTGGATTTGGTAGTAAA 3′	
Fly	.....	14294
Bee	..A..A..-..AA-.....T	14868
Locust	.....A.....T	14187
Cicada	..AC.....AA.....	
Silverfish	.....AA-.....T	
Gastropod	..ATA...C..AAA.....T	
Human	.....A.C.....	1435
Toad	....C.....A.C.....	2834
Sea urchin	..AT.....CA.C.....G	794

This primer was designed in the Wilson laboratory and modified by C.S. to match *Drosophila*. Sea urchin has been modified to match Thomas et al. 1989. Roundworm is difficult to align with other phyla. Deletion of the 3′ base would make it more versatile.

SR-J-14508 (alias 12Sj)

(19mer)	5′ TACAAAACAGATTCCTCTG 3′	
Fly	.....G.....	14490
Bee	..ATCG...A.CAT....A	15056
Locust	C.TG.G.T..G.....	14383
Cicada-*Okanagana*	..T.G.....	
-*Diceroprocta*	..T.G.....	
-*Magicicada*	.....	
Silverfish	C..GGGG...G.....A	
Human	....G.....GC.....A	1213
Toad	....G.....GC.....A	2624
Sea urchin	.CA.TGG..AGC.....CA	597
Gastropod	..A.GG....G....C..A	

Designed by C.S. for cicadas, it is a highly reliable primer and has been used successfully in combination with SR-N-14756 (12Sj) to amplify and sequence a 260+ bp segment in periodical cicadas. It could be made more versatile by omitting the 3′ base. Cicada sequences are from *M. tredecim, O. vanduzeii,* and *Diceroprocta apache* (C.S., S. Pääbo & C. McIntosh, unpublished data). Urchin has been modified to match Thomas et al. 1989.

SR-N-14522 (alias 12Se)

(19mer)	5′ ATTCAAAGAATTTGGCGGT 3′	
Fly	.C.T...A.....	14540
Bee	.A.T...CG..A.....C	15107
Locust	.CC.....	14433
Cicada	.....	
Human	.C.....G.CC.....	1162
Toad	.CC.....G.C.....	2573
Sea urchin	.....G.....	547

Designed by C.S., it is located at the beginning of the third domain of the 12S gene and works well for periodical cicadas. The GGCGG motif at the 3′ end is one of the most highly conserved motifs in the 12S gene. The T on the 3′ end could be dropped to make it slightly more universal. Seven bases on the 3′ end of this primer constitute the 5′ end of a primer designed by H. Croom to sequence spiders (Croom et al. 1991); it is not included here because of lack of conservation with other species.

SR-N-14588 (alias 12Sai)

(25mer)	5′ AAACTAGGATTAGATACCCTATTAT 3′	
Fly	.....	14612
Bee	.....C...	15179
Locust	.G.....	14507
Cicada	.....	
Cricket	.....	
Human	.....G.....C.C...	1067
Toad	.....G.....C.C...	2485
Sea urchin	....C.....G....	491

Designed by T. Kocher (Kocher et al. 1989) and modified by C.S. to match insects using cricket (D. M. Rand & R. G. Harrison, personal communication) and *Drosophila* sequences, it can be used to amplify a wide variety of taxa in conjunction with SR-J-14233. The insect primer will amplify humans and the human primer will amplify insects (Simon et al. 1990). It is very conserved and broadly used (Ballard et al. 1992). It is complementary to SR-J-14612.

SR-J-14612 (12Sair)

(20mer)	5′ AGGGTATCTAATCCTAGTTT 3′	
Fly	.....	14593
Bee	.....	15160
Locust	.....C.	14488
Cicada	.....	
Human	G.....C.....	1086
Toad	G.....C.....	2504
Sea urchin	.....G....	510

Designed by C.S. for cicadas, it is the complement to SR-N-14588 (above). It has been used in combination with SR-N-14922 (12SG) to amplify and sequence the 5′ half of the 12S gene in periodical cicadas and *Drosophila* (Simon et al. 1991). Taylor et al. (1993) used a nearly identical primer (shifted one base downstream) to amplify the A+T-rich region of the following lepidopterans: *H. punctigera* (Noctuidae), *S. melinus* (Lycaenidae), and six species in the genus *Jalmenus* (Lycaenidae).

SR-N-14756 (alias 12Sh)

(21mer)	5′ GACAAA-ATTCGT-GCCAGCAGT 3′	
Fly	...C..-...G..-.....	14776
D. virilis	...C..-...G..-.....A.	
Bee	..T...T..GA..C.....A.	15353
Locust	.G..T.-T.....-.....C	14662
Cicada	.....-.....-.....	
Human	.GTC..-T.....-.....CAC	877
Toad	.GTC..-TC....-.-....C.C	2294
Sea urchin	.TA...--A...T.....CAC	315
Chimpanzee	.GTC..-T.....-..T...CAC	
Gorilla	.GTC..-T.....-.....CAC	

Designed by C.S. for cicadas, it works extremely well in combination with SR-J-14508 (12Sj) to amplify and sequence a 260+ bp segment. The removal of three bases from the 3′ end would make it more versatile. Note that honey bee has two insertions compared with the other insects. *D. virilis* is from Clary & Wolstenholme (1987).

SR-N-14922 (alias 12Sgi)

(18mer)	5′ AAGTTTTATT-TTGGCTTA 3′	
Fly	.....-.....	14939
Bee	.TTA.....A.AAAT...	15524
Locust	.....-C.T.....	14776
Human	.G....GG.C-C.A..C.T	651
Cow	.G....GG.C-C.A..C.T	434
Toad	.G....GG.C-C.A..C.T	2208
Sea urchin	CGA...GG.C-C.A.TCCC	76

Designed by C.S. for cicadas based on an angiosperm rRNA sequencing primer designed by E. Zimmer, it is located only a few bases away from the beginning of domain I of the small ri-

bosomal subunit RNA gene. Although not highly conserved between vertebrates and insects it could be modified easily to match vertebrates. In honey bee, part of the primer is located outside the 5′ end of the 12S gene. This primer can be used to amplify and sequence the 5′ half of the 12S gene. It has been used successfully in combination with SR-J-14612 (12Sair) to amplify and sequence periodical cicadas. L. Nigro and A. Grapputo used a modified version of this primer (with the two 3′ bases removed) in combination with SR-J-14612 to amplify this same region in five species of *Drosophila* (L. Nigro, personal communication).