J Mol Evol (1994) 38:282–294

Journal of Molecular Evolution

The Cytochrome *b* Region in the Mitochondrial DNA of the Ant *Tetraponera rufoniger:* Sequence Divergence in Hymenoptera May Be Associated with Nucleotide Content

Lars S. Jermiin, Ross H. Crozier

Department of Genetics and Human Variation, La Trobe University, Bundoora, Victoria 3083, Australia

Received: 19 October 1992 / Revised: 21 July 1993

Abstract. Polymerase chain reaction (PCR) followed by sequencing of single-stranded DNA yielded sequence information from the cytochrome *b* (cyt *b*) region in mitochondrial DNA from the ant *Tetraponera rufoniger.* Compared with the cyt *b* genes from *Apis mellifera, Drosophila melanogaster,* and *D. yakuba,* the overall A + T content (A + T%) of that of *T. rufoniger* is lower (69.9% vs 80.7%, 74.2%, and 73.9%, respectively) than those of the other three. The codon usage in the cyt *b* gene of *T. rufoniger* is biased although not as much as in *A. mellifera, D. melanogaster,* and *D. yakuba; T. rufoniger* has eight unused codons whereas *D. melanogaster, D. yakuba,* and *A. mellifera* have 21, 20, and 23, respectively. The inferred cyt *b* polypeptide chain (PPC) of *T. rufoniger* has diverged at least as much from a common ancestor with *D. yakuba* as has that of *A. mellifera* (~3.5 vs ~2.9). Despite the lower A + T%, the relative frequencies of amino acids in the cyt *b* PPC of *T. rufoniger* are significantly ($P < 0.05$) associated with the content of adenine and thymine (A + T%) and size of codon families. The mitochondrially located cytochrome oxidase subunit II genes (CO-II) of endopterygote insects have significantly higher average A + T% (~75%) than those of exopterygous (~69%) and paleopterous (~69%) insects. The increase in A + T% of endopterygote insects occurred in Upper Carboniferous and coincided with a significant acceleration of PPC divergence. However, acceleration of PPC divergence is not significantly correlated with the increase of the A + T% ($P > 0.1$). The high A + T%, the biased codon usage, and the increased PPC divergence of Hymenoptera can in that respect most easily be explained by directional mutation pressure which began in the Upper Carboniferous and still occurs in most members of the order. Given the roughly identical A + T% of the cyt *b* and CO-II genes from the other insects whose DNA sequences are known (*A. mellifera, D. melanogaster,* and *D. yakuba*), it seems most likely that the A + T% of *T. rufoniger* declined secondarily within the last 100 Myr as a result of a reduced directional mutation pressure.

Key words: Polymerase chain reaction — Mitochondrial DNA sequences —*Tetraponera rufoniger* — Extreme nucleotide content — Protein divergence — Molecular evolution — $tRNA^{Ser}_{UCN}$ gene — Cytochrome *b* gene — Cytochrome oxidase subunit II gene — Paleoptera — Exopterygota — Endopterygota

Correspondence to: L.S. Jermiin

Abbreviations: Myr = million years; mtDNA = mitochondrial DNA; scnDNA = single-copy nuclear DNA; A = adenine; C = cytosine; G = guanine; T = thymine; A + T% = content of A and T; PPC = polypeptide chain; cyt *b* = cytochrome *b;* CO-I = cytochrome oxidase sub-unit I; CO-II = cytochrome oxidase subunit II; ND1 = NADH dehydrogenase subunit 1; ND6 = NADH dehydrogenase subunit 6; $tRNA^{Ser}_{UCN}$ = transfer RNA for serine with a UCN anticodon

Introduction

Early studies of animal mtDNA (Brown et al. 1979, 1982; Miyata et al. 1982) suggested that the evolutionary rate of mtDNA was constant in time in all lineages

and that it evolved 5–10 times faster than scnDNA (reviewed by Moritz et al. 1987; Harrison 1989; Crozier 1990). Later, Powell et al. (1986) showed that mtDNA and scnDNA in *Drosophila* evolve at approximately the same rate whereas Vawter and Brown (1986) found extreme variation in evolutionary rates between mtDNA and scnDNA from primates and sea urchins and suggested that the rate variations were due to rate changes in scnDNA rather than in mtDNA. Variation in evolutionary rates of mammalian scnDNA has been attributed to DNA repair mechanisms (Britten 1986) or to differences in generation time (Wu and Li 1985; Li and Tanimura 1987).

Evolutionary rate variations of mtDNA genes received little attention before Gillespie (1986), who analyzed sequence data of mitochondrial and nuclear genes from several species and concluded that evolution at the DNA and protein levels is better explained by natural selection than neutrality. He further suggested that molecular evolution is episodic, i.e., with bursts of nucleotide substitutions interleaved with periods with little activity (Gillespie 1986). Nevertheless, the evolution of DNA is still believed to follow a clocklike behavior (Birley and Croft 1990; Preparata and Saccone 1991) and, thus, mtDNA might still provide data for estimation of species divergence time (Birley and Croft 1990).

Despite Gillespie's work, application of mtDNA for studies of phylogeny and evolution is common and has proven extremely useful (reviewed by Avise et al. 1987; Crozier 1992). Comparisons between protein-coding mtDNA genes from *Homo* (Anderson et al. 1981), *Mus* (Bibb et al. 1981), *Bos* (Anderson et al. 1982), *Xenopus* (Roe et al. 1985), *Drosophila* (Clary and Wolstenholme 1985), *Strongylocentrotus* (Jacobs et al. 1988), *Paracentrotus* (Cantatore et al. 1989), *Rattus* (Gadaleta et al. 1989), *Gallus* (Desjardins and Morais 1990), *Ascaris, Caenorhabditis* (Okimoto et al. 1992), and *Apis* (Crozier and Crozier 1993) have uncovered a pattern of differential evolutionary rates with homologous genes evolving at roughly the same rates and nonhomologous genes evolving at different characteristic rates. The evolutionary rates of homologous genes, however, are not constant, as Gillespie (1986) points out, but vary between 0.125 and 3.5% per million year (Honeycutt and Wheeler 1990; Birley and Croft 1990).

The interregional variation in the evolutionary rate of mtDNA is particularly useful because it facilitates the choice of appropriate genes for biogeographical (Avise et al. 1987) or phylogenetic studies (Crozier 1990, 1992). Sequence data from mitochondrial protein-coding (DeSalle et al. 1987; Crozier et al. 1989; Crozier 1990; Meyer and Wilson 1990; Irwin et al. 1991; Liu and Beckenbach 1992; Willis et al. 1992; Kusmierski et al. 1993) and ribosomal RNA genes (DeSalle et al. 1987; Miyamoto and Boyle 1989; Meyer and Wilson 1990; Cameron et al. 1992; Derr et al. 1992a, 1992b) have often indeed proven useful in clarifying phylogenetic relationships of both closely and distantly related species and have further shown that mtDNA genes can be informative about evolution spanning at least 600 Myr.

One striking observation from comparing mitochondrial and nuclear protein-coding genes is that the A + T% is higher in the former than in the latter, especially in insects (Crozier 1990). For example, in *Drosophila* nuclei the A + T% is 42% in the alcohol dehydrogenase gene (Bodmer and Ashburner 1984), 37% in the amylase gene (Boer and Hickey 1986), and 44% in the xanthine dehydrogenase gene (Keith et al. 1987). By contrast, the average A + T% in the mitochondrial protein-coding genes of *Drosophila* is markedly higher—77.4% in *D. yakuba* (Clary and Wolstenholme 1985). The excess rate of transitions (A ↔ G and T ↔ C), which has been observed within and between closely related *Drosophila* species (DeSalle et al. 1987) would be expected to force the high A + T% toward lower values. If, however, a high A + T% is maintained for a substantial length of time—e.g., by directional mutation pressure (Sueoka 1962, 1988, 1992)—proportionally more transversions (A ↔ T, A ↔ C, G ↔ T, and G ↔ C) must occur (Holmquist 1983) as observed in a comparison of mitochondrial genes from *D. yakuba* with those from *D. melanogaster* (Wolstenholme and Clary 1985). Because the structure of the mitochondrial genetic code for *Drosophila* (Clary and Wolstenholme 1985) is biased toward twofold degenerated codon families (60% of all amino acids are coded for by twofold degenerated codon families) and because codons within these codon families differ only by transitions at the third codon position, it may be expected that PPCs coded for by DNA sequences with a high A + T% will diverge relatively more from a common ancestor than those which coded for by DNA sequences with a slightly lower A + T%.

This may indeed explain results from recent analyses of the evolution of mitochondrial protein-coding genes in *A. mellifera.* Crozier et al. (1989) and Crozier and Crozier (1992) compared the CO-I, CO-II, and cyt *b* genes from *A. mellifera* with those from *D. yakuba* and showed that the corresponding PPCs of *A. mellifera* had diverged 1.9–3.5 times more from a common ancestor than had those of *D. yakuba.* The A + T% in the mitochondrial protein-coding genes of *A. mellifera* is 83.3% (Crozier and Crozier 1993) and, thus, higher than the 77.4% which has been reported from those of *D. yakuba* (Clary and Wolstenholme 1985).

This paper addresses two problems related to the above-mentioned results on evolution of protein-coding mitochondrial genes in *A. mellifera,* namely: (1) Is *A. mellifera* unique among Hymenoptera with respect to the high degree of PPC divergence and the high A + T%, and (2) is the increased PPC divergence of *A. mellifera* due to high A + T%?

For these purposes we sequenced the cyt *b* gene of an ant. Ants (Formicidae) and the honeybees (Apidae)

are both aculeate hymenopterans. Hymenopteran and dipteran fossils first appear in the Triassic (Carpenter and Burnham 1985) and probably diverged from one another ~280 Myr ago. The earliest-known fossil ant is 107–114 Myr old (Brandão et al. 1989; Brandão 1990) whereas the earliest-known fossil honeybee is 74–96 Myr old (Michener and Grimaldi 1988). Ants and bees are thus sufficiently closely related to have had a substantial period of shared evolution and yet distant enough to have had a long period of independence. The reasons for choosing the cyt *b* gene are that (1) cyt *b* is the best-known mitochondrially encoded PPC with respect to its structure and function (Hatefi 1985; Howell and Gilbert 1988; Howell 1989; di Rago et al. 1990; Tron et al. 1991; Crozier and Crozier 1992), (2) it is a relatively large protein-coding gene, (3) insect-specific oligonucleotide primers encompassing the gene are already known, and (4) it is the most frequently used mtDNA gene in phylogenetic and evolutionary studies (e.g., Irwin et al. 1991).

Materials and Methods

DNA Extraction. *Tetraponera rufoniger* workers were collected in Kuala Lumpur, Malaysia, and preserved in 95% ethanol. To reduce the likelihood of contamination with symbiont DNA, total DNA was extracted from only the head and the alitrunk of single individuals. Following a modification of the DNA extraction used by Crozier et al. (1991), dried specimens were homogenized with a pestle in a 1.5-ml Eppendorf tube and incubated at 65°C in 500 µl of 2× CTAB buffer [0.1 M Tris-HCl pH 8.3, 0.02 M EDTA pH 8, 1.4 M NaCl, and 0.2% β-mercaptoethanol (v:v)] for 2 h. Debris was removed after a 10-min centrifugation at 17,310*g* and total DNA was extracted from the supernatant using phenol-chloroform-isoamyl alcohol extraction (Sambrook et al. 1989). The extracted DNA was precipitated overnight at −20°C in 1.5 vol isopropanol, pelleted, washed twice with 100 µl 70% ethanol, dried, and finally resuspended in 30 µl TE.

Gene Amplification. DNA amplification was performed using the polymerase chain reaction (PCR) (Saiki et al. 1988) in a Perkin Elmer Cetus thermal cycler. Samples of 1× *Taq* polymerase buffer (Promega), 2 units *Taq* polymerase (Promega), two (0.5 µM) oligonucleotide primers (Table 1), dNTP (0.2 mM), and 2 µl template DNA were mixed to a final volume of 50 µl. Sealed by 70 µl mineral oil, the samples were then brought through 35 cycles of dissociation (92°C, 1 min), annealing (40–52°C, 1 min), and extension (70°C, 1–3 min). Excess oligonucleotide primer was removed from the double-stranded PCR product using 30,000 NMWL filter units (Millipore, Ultrafree-MC) and the PCR product subsequently was used as template in single-stranded PCR. The single-stranded PCR procedure was as above except that only one oligonucleotide primer was used and the total volume was doubled. The single-stranded PCR product was precipitated in 30 µl 7 M NH_4Ac and 120 µl 95% ethanol for at least 1 h at −20°C. The single-stranded DNA was then pelleted, washed twice in 100 µl 70% ethanol, dried, and resuspended in 8–16 µl TE.

Gene Sequencing. Single-stranded PCR products were sequenced directly using the Sequenase sequencing kit 2.0 (United States Biochemical Corp.) following a modification of the protocol used by Koulianos and Crozier (1991). The single-stranded DNA template was first annealed to an appropriate oligonucleotide primer (Table 1), then labeled with α-^{35}S-dATP or α-^{33}P-dATP while being extended, and finally subjected to dideoxy chain termination (Sanger et al. 1977). The DNA subfragments were then separated in 4% or 6% polyacrylamide gels at 55°–60°C. After electrophoresis, the polyacrylamide gels were fixed in 10% glacial acetic acid/10% ethanol, rinsed in H_2O, and dried overnight. The DNA sequences were finally visualized on single-sided x-ray-sensitive film (Fuji) after 1–2 weeks of exposure.

Data Analysis. Sequence data were aligned and analyzed using the program MacVector (IBI) on an Apple Macintosh IIci computer. We used the *Drosophila* mtDNA genetic code (Clary and Wolstenholme 1985) to infer PPCs from protein-coding genes in *T. rufoniger* because so far no inconsistencies have been found when it is applied to other insect mitochondrial genes (Crozier et al. 1989; Liu and Beckenbach 1992; Willis et al. 1992; Crozier and Crozier 1992, 1993). Nucleotide content and codon usage were determined using programs developed by R.H.C. Evolutionary divergences and phylogenetic relationships of inferred PPCs were determined using four-species Wagner network analysis as described by Crozier and Crozier (1992). Bootstrapped (Efron and Gong 1983) estimates of the mean and of its 95% confidence limit of the A + T% in genes and the relative divergence of their corresponding PPCs for major insect groups were calculated using a program developed by L.S.J.

Results and Discussion

The Cytochrome b *Region*

Exclusion of the Possibility of Duplicated Sequences

Double-stranded PCR products for both specimens were made with the PCR primer combinations CB1-CB2, CB1-ND1, CB1-tR^S, ND4-CB2, and ND4-CB5 (Table 1). DNA sequences from these fragments showed no inconsistency in overlapping regions and homologous sequences from the two specimens were identical. Finally, because the lengths of the open reading frames are almost equal to those of *A. mellifera* (Crozier and Crozier 1993), *D. melanogaster* (Garesse 1988), and *D. yakuba* (Clary and Wolstenholme 1985) and because the structure and function of the cyt *b* PPC and the $tRNA^{Ser}_{UCN}$ appear to be preserved (below), we conclude that the DNA sequence presented in Fig. 1 originates from the *T. rufoniger* mitochondrial genome.

Genetic Code and Gene Order

The 1,450 nucleotides presented in Fig. 1 determine the structures of the terminal part of the ND1 PPC, cyt *b* PPC, $tRNA^{Ser}_{UCN}$ and the terminal part of the ND6 PPC. Application of the mtDNA genetic code for *D. yakuba* (Clary and Wolstenholme 1985) showed no inconsistencies when applied to the mtDNA sequence of *T. rufoniger*. As this genetic code appears to apply to 10 orders of insects (Liu and Beckenbach 1992; Willis et al. 1992), we suggest that it arose well before the origin of the class Insecta.

The gene order in the DNA sequence of *T. rufoniger* is identical to those of *A. mellifera* (Crozier and Crozier 1992, 1993), *D. melanogaster* (Garesse 1988), and *D. yakuba* (Clary and Wolstenholme 1985). Compared

Table 1. Oligonucleotide primers for amplification and sequencing of the cytochrome *b* region in *Tetraponera rufoniger*

Primer	Sequences	Purpose of primer	Positions[a]
CB1[b]	5′–TATGTACTACCATGAGGACAAATATC-3′	PCR/Sequencing	11400–11425→
CB2[b]	5′–ATTACACCTCCTAATTTATTAGGAAT-3′	PCR/Sequencing	←11859–11884
CB3[c]	5′–CCTATTCATATTCAACC-3′	Sequencing	11802–11818→
CB5[d]	5′–GAAAATCCCCCCCA(A/C)A(T/C)TCA-3′	PCR/Sequencing	←11496–11515
CB6[e]	5′–AGTA(G/C)ATTCCTCGA(C/G)C-3′	Sequencing	←11304–11319
ND1[b]	5′–GTAGCATTTTTAACTTTATTAGAACG-3′	PCR	←13143–13168
ND4[b]	5′–GGAGCTTCAACATGAGCTTT-3′	PCR	9335–9354→
ND6/1[f]	5′–CTATTTTC(C/A)ATCATTAAAATCTC-3′	PCR/Sequencing	10885–10907→
tR[s][g]	5′–TATTTCTTTATTATGTTTTCAAAAC-3′	PCR/Sequencing	←12227–12251

[a] Position of primers are given according to the *Apis mellifera* mitochondrial genome (Crozier and Crozier 1993). The 3′ ends of primers are indicated by arrows

[b] Conserved sequence between *A. mellifera* (Crozier and Crozier 1993) and *Drosophila yakuba* (Clary and Wolstenholme 1985)

[c] Conserved sequence between the ants *Camponotus* sp. and *Myrmecia pilosula* (Jermiin and Crozier, unpublished)

[d] Conserved sequence between *Camponotus* sp., *Tetraponera rufoniger,* and the wasp *Ropalidia* sp. (Jermiin and Crozier, this paper and unpublished)

[e] Conserved sequence between the ants *T. rufoniger* and *Dorylus helvolus* (Jermiin and Crozier, this paper and unpublished)

[f] Conserved sequence between *D. helvolus* and *T. rufoniger* (Jermiin and Crozier, this paper and unpublished)

[g] Conserved sequence between *D. yakuba* (Clary and Wolstenholme 1985) and *T. rufoniger* (Jermiin and Crozier, this paper)

with the nematode gene order (Okimoto et al. 1992), the insects appear to be very different. Conservation of the gene order therefore cannot be expected between major invertebrate groups. The stop codon TAG, which terminates the *T. rufoniger* ND1 open-reading frame, is not used at all in the *A. mellifera* (Crozier and Crozier 1993) or *D. yakuba* (Clary and Wolstenholme 1985) mtDNA.

Transfer RNA Gene

The structure of the tRNA $^{Ser}_{UCN}$ gene (Fig. 2) appears similar to those of *A. mellifera* (Crozier and Crozier 1993) and *D. yakuba* (Clary and Wolstenholme 1985). Several mutations in stem, interstem, and loop regions have occurred. A comparison of the tRNA $^{Ser}_{UCN}$ of *T. rufoniger* with that of *A. mellifera* (Crozier and Crozier 1993) shows that the mutations include substitutions in the anticodon stem region (one pair), in the variable arm region (one), between the anticodon and dihydrouridine stem regions (one), in the dihydrouridine stem region (one), in the dihydrouridine loop (one), as well as insertions in the TψC loop (one) and the dihydrouridine loop (three). One mismatch between nucleotides (A vs C) was observed in the dihydrouridine stem region toward the interstem region.

Nucleotide Content and Codon Usage

Nucleotide Content

Given the higher similarity of the cyt *b* DNA and amino acid sequences between *T. rufoniger* and *A. mellifera* than between *T. rufoniger* and two *Drosophila* species (Table 2), it might be expected that the A + T% in the cyt *b* gene of *T. rufoniger* is similar to that of *A. mellifera.*

Unexpectedly, this is not so. Although the A + T% in the cyt *b* gene of *T. rufoniger* is generally as high (Table 3) as those reported for the CO-II genes of other insects (Table 4), and it has the generally observed bias against G in the third codon position (Brown 1985; Crozier 1990; Irwin et al. 1991; Liu and Beckenbach 1992; Crozier and Crozier 1992, 1993), the overall A + T% of the cyt *b* gene in *T. rufoniger* is markedly lower (69.9%) than those reported from *D. melanogaster* (74.2%) (Garesse 1988), *D. yakuba* (73.9%) (Clary and Wolstenholme 1985) and *A. mellifera* (80.7%) (Crozier and Crozier 1992). The A + T% in the cyt *b* genes from these four species can be placed in perspective by considering the distribution of nucleotide compositions in other insects (Table 4), derived mostly from Liu and Beckenbach's (1992) study of insect CO-II genes.

The mean A + T% of the CO-II gene of the paleopterous Pterygota and the neopterous Exopterygota are significantly lower than that in the neopterous Endopterygota (Fig. 3).

With the exception of *Sitophilus granarius,* all Endopterygota fall above the lower 95% confidence limit for the mean (Table 4, Fig. 3); Siphonaptera, Lepidoptera, and Hymenoptera, in particular, show A + T% that are well above average for Endopterygota. If the insect phylogeny by Carpenter and Burnham (1985) is accepted, the data (Table 4) suggest that the A + T% increased within the lineage which gave rise to Endopterygota. This increase may have continued after the origin of lineages giving rise to Siphonaptera, Lepidoptera, and Hymenoptera. As the A + T% of the known insect cyt *b* and CO-II genes (*D. yakuba, D.*

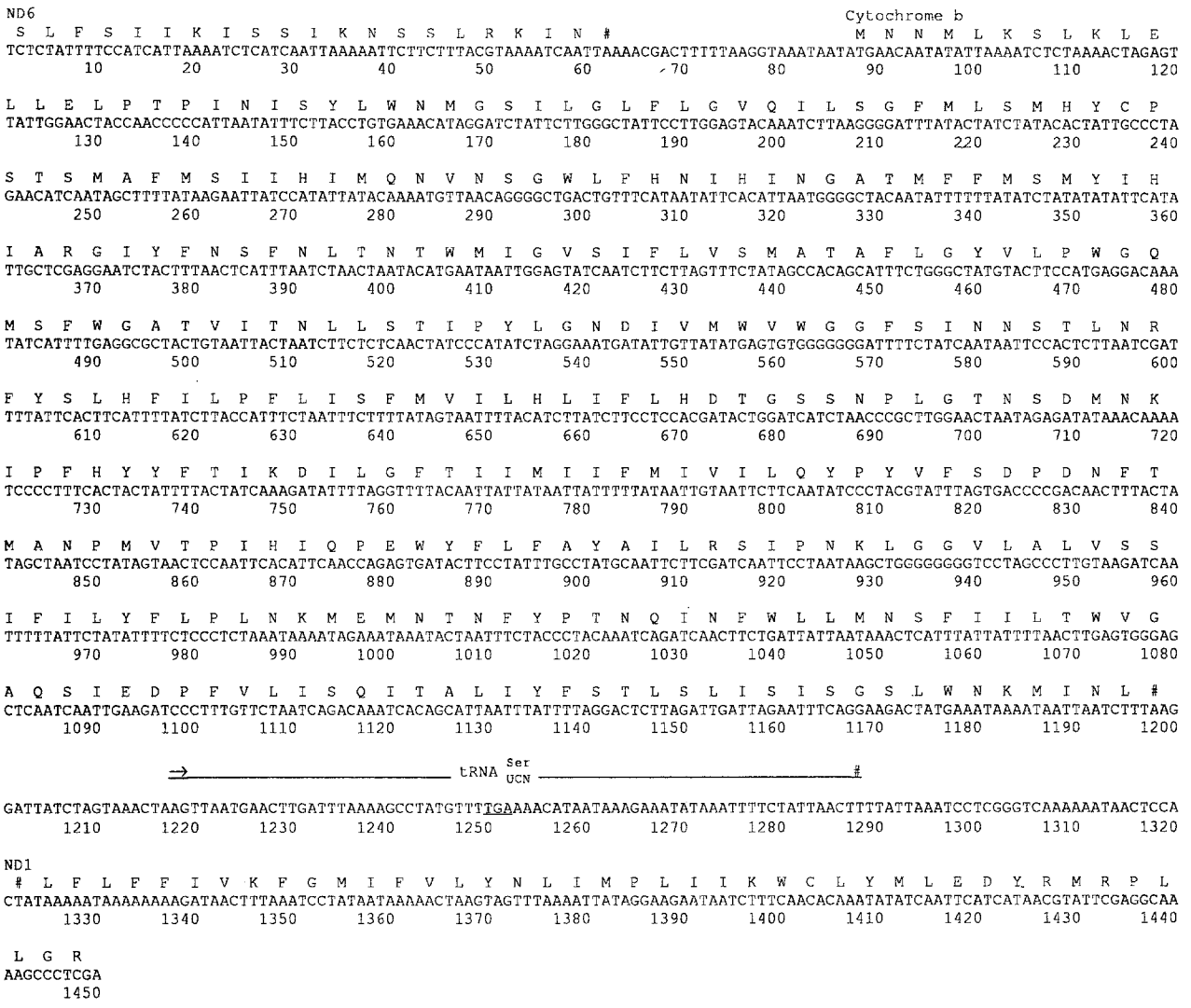

Fig. 1. Sequences of genes within the cytochrome *b* region from *Tetraponera rufoniger,* together with the inferred amino acid sequences for cytochrome *b* and subunits 1 and 6 from the NADH dehydrogenase system (ND1 and ND6). The tRNA for serine with the anticodon UCN is denoted tRNA $^{Ser}_{UCN}$. "#" = end of gene, "→" = start and direction of tRNA $^{Ser}_{UCN}$ gene.

melanogaster, and *A. mellifera*) is roughly identical for each species, we may extend our interpretation of the evolutionary events by adding that the high A + T% must have been lost secondarily and independently in lineages giving rise to *T. rufoniger* (as evaluated by the cyt *b* gene) and *S. granarius* (as shown by the CO-II gene). While our interpretation, to our knowledge, has no recorded equivalent in Metazoa, a similar clustering of related organisms according to the A + T% of rRNA genes occurs in Eubacteria (Hori and Osawa 1987). Such an association indicates that directional mutation pressure determines the A + T% within phylogenetic lineages (Muto and Osawa 1987; Osawa et al. 1992) and, hence, may indicate a relatively recent change in such pressures on the CO-II gene of *S. granarius* and the cyt *b* gene of *T. rufoniger.* As the A + T% in the CO-II gene of *Lasius* sp. is as high as that in other endopterygote insects, the cessation of mutation pressure on the cyt *b* gene of *T. rufoniger* may have occurred after a separation of the formicid subfamilies Formicinae and Pseudomyrmecinae.

Codon Usage

Given the relatively low A + T% of the cyt *b* gene of *T. rufoniger* compared to those of *A. mellifera, D. melanogaster,* and *D. yakuba,* a difference in codon usage between the four species is expected. This indeed is the case. Several codons with G or C in third position (CTC, CTG, ATC, CCC, CCG, AGG, GGC, and GGG) occur more often in *T. rufoniger* than in *A. mellifera, D. melanogaster,* and *D. yakuba* (Table 5).

Similarly, some codons with A or T in the third position (TTA, ATT, CCA, GGT, and GGA) occur less often in *T. rufoniger* than in *A. mellifera, D. melanogaster,* and *D. yakuba* (Table 5). The differences in codon usage in the cyt *b* gene of *T. rufoniger,* when compared to the other three species, however, do not follow simple changes in their nucleotide contents. Some codons

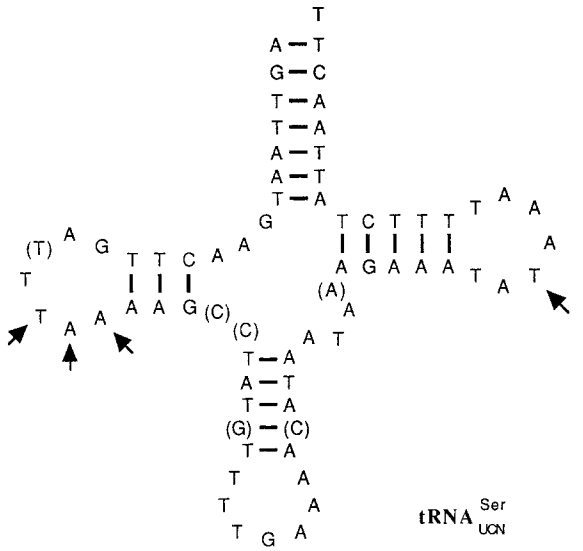

Fig. 2. The two-dimensional structure of the tRNA $^{Ser}_{UCN}$ gene in *Tetraponera rufoniger*. With reference to the corresponding *Apis mellifera* gene (Crozier and Crozier 1993), nucleotides in *parentheses* denote substitutions; insertions are marked with *arrows*.

which, on the basis of A + T%, were expected to occur less frequently in the cyt *b* gene of *T. rufoniger* than in those of *A. mellifera, D. melanogaster,* and *D. yakuba* (CTT, CTA, TCT, AGA and ACT) actually occurred more frequently in the first species than in the other three (Table 5), whereas the opposite also was true for at least one codon (TTC). In spite of these differences, the skewed codon usage reported from the cyt *b* gene in *A. mellifera* (Crozier and Crozier 1992) occurs also in *T. rufoniger,* but to a lesser extent. For example, the number of unused codons (except the stop codon) in the cyt *b* gene of *T. rufoniger* is eight as compared to 21 in *D. melanogaster,* 20 in *D. yakuba,* and 23 in *A. mellifera* (Table 5).

Stepwise multiple regression analysis (Sokal and Rohlf 1981, pp. 661–671) allows examination of the association between the relative frequency of amino acids in a PPC and various independent variables (e.g., codon family size, A + T% in codon families and proportion of A, T, C, and G within codon families). The relative frequency of amino acids in the cyt *b* PPC of *T. rufoniger* is significantly ($P < 0.05$) associated with the A + T% in codon families as well as with the codon family size, whereas those of *D. melanogaster* and *D. yakuba* are significantly ($P < 0.05$) associated with the T content in codon families and the codon family size. By contrast, the relative frequency of amino acids in the cyt *b* PPC of *A. mellifera* is significantly ($P < 0.05$) associated only with the T content in the codon families.

To assess whether a pattern of association between the relative frequency of amino acids in protein-coding genes and the six variables mentioned in the paragraph above is general among insects, we analyzed the CO-II genes presented by Liu and Beckenbach (1992). The relative frequency of amino acids in the CO-II PPCs from Paleoptera and the neopterous Exopterygota (except Orthoptera) is not significantly correlated with any of the six variables (Table 4). The relative frequency of amino acids in neopterous Endopterygota (except *S. granarius* and Lepidoptera), on the other hand, is significantly ($P < 0.05$) correlated with both the size and the A + T% of codon families. In Coleoptera, Siphonaptera, and Diptera, the codon family size is relatively more important than the codon family A + T%, whereas the opposite holds true for Hymenoptera (Table 4). Following the discovery that the A + T% in the CO-II genes of Endopterygota is significantly higher than in those of Paleoptera and Exopterygota, it is not surprising to find a significant correlation between the relative frequency of amino acids in the CO-II PPCs and the A + T% in codon families. If, as suggested above, directional mutation pressure constantly enforces an increase of the overall A + T% in the CO-II genes of endopterygote insects, it is likely that the A + T% will approach a limiting value beyond which the function of the PPCs is impaired. This may explain the increasing importance of the A + T% in codon families as a predictor of the relative frequency of amino acids in the CO-II genes of Paleoptera—over Exopterygota (except Orthoptera) and Endopterygota (except Lepidoptera and Hymenoptera)—to Hymenoptera (Table 4).

There are some deviations from this picture. In spite of a relatively low overall A + T%, the relative frequency of amino acids in the CO-II PPCs from Orthoptera (Table 4) appears to be significantly correlated with the size and the A + T% in codon families. This correlation may indicate that stationarity changes (Preparata and Saccone 1991) of the gene have occurred within the lineage leading to Orthoptera. DNA repair mechanisms (Britten 1986), suggested to exist in *Drosophila* (DeSalle et al. 1987; Sander et al. 1991) and *Apis* (Crozier 1990), may also explain this observation. Absence of significant association between the relative frequency of amino acids and any of the six independent variables mentioned in Table 4 for the CO-II PPC from *S. granarius* (Table 4) may be a derived result caused by the secondary loss of overall A + T%. The concurrence of the secondary loss of overall A + T% and the absence of a significant association between the relative frequency of amino acids in the CO-II PPC and any of the six independent variables in Table 4 strengthen our suggestion that the evolution of the CO-II gene in *S. granarius* has become less constrained in more recent times.

Evolutionary Divergence and Phylogenetic Relationship

Evolutionary Divergences

Because the DNA sequences coding for cyt *b* and CO-II PPCs are the only completely sequenced protein-coding genes from a large number of insects, the analyses of PPC divergence and phylogeny (below) will

Table 2. Overall similarity between the cytochrome *b* DNA and amino acid sequences of *Tetraponera rufoniger* and those of *Apis mellifera, Drosophila melanogaster,* and *Drosophila yakuba*[a]

	Apis melllifera		*Drosophila yakuba*		*Drosophila melanogaster*	
Level	Observed	Expected	Observed	Expected	Observed	Expected
DNA	0.6773	0.3136	0.6259	0.3005	0.6168	0.3012
Amino acids[b]	0.6125	0.0758	0.5559	0.0723	0.5368	0.0715
Amino acids[c]	0.7615	0.2164	0.7411	0.2186	0.7248	0.2175

[a] Data from Clary and Wolstenholme (1985) (*Drosophila yakuba*), Garesse (1988) (*Drosophila melanogaster*), and Crozier and Crozier (1992) (*Apis mellifera*). Expected similarities were determined as the products of the base or amino acid proportions in the sequences compared
[b] Structural similarity
[c] Functional similarity. Functional groups of amino acids are adopted from French and Robson (1983), as per Crozier and Crozier (1992, 1993)

Table 3. Nucleotide content in mtDNA from *Tetraponera rufoniger.* The segments are: the cytochrome *b* gene (cyt *b*), a transfer RNA with a UCN anticodon for serine (tRNA $^{Ser}_{UCN}$), parts of the subunits 1 and 6 in the NADH dehydrogenase system (ND1, ND6), the intergenic regions, and the total segment presented in Fig. 1

Segment	Size (bp)	Position	A	T	G	C
cyt *b*	1113	1	0.3827	0.2588	0.1644	0.1941
		2	0.2264	0.4501	0.1294	0.1941
		3	0.3666	0.4124	0.0620	0.1590
		Overall	0.3252	0.3738	0.1186	0.1824
$tRNA^{Ser}_{UCN}$	71		0.4366	0.3803	0.0986	0.0845
ND1	130	1	0.4773	0.3182	0.0682	0.1364
		2	0.4186	0.2791	0.1628	0.1395
		3	0.5814	0.2093	0.0465	0.1628
		Overall	0.4923	0.2692	0.0923	0.1462
ND6	63	1	0.4762	0.4286	0.0000	0.0952
		2	0.2857	0.3810	0.0476	0.2857
		3	0.3810	0.4286	0.0000	0.1905
		Overall	0.3810	0.4127	0.0159	0.1905
Intergenic	73		0.4110	0.3150	0.1233	0.1507
Total	1450		0.3524	0.3635	0.1110	0.1731

be based on these genes. Figure 4 presents alignment of the inferred amino acid sequences of cyt *b* for *T. rufoniger, A. mellifera, D. melanogaster, D. yakuba, X. laevis,* and *M. musculus.*

The conserved regions observed in cyt *b* PPC from *T. rufoniger* (Fig. 4) are roughly identical to those reported from vertebrates (Irwin et al. 1991) and insects (Crozier and Crozier 1992). Areas with substantial amounts of changes (deletions or substitutions) in the DNA sequence from *T. rufoniger* are not within regions thought to be important in maintaining cyt *b*'s structure and function (Hatefi 1985; Howell and Gilbert 1988; Howell 1989; di Rago et al. 1990; Irwin et al. 1991; Tron et al. 1991).

Crozier and Crozier (1992) reported that the cyt *b* PPC of *A. mellifera* had diverged ~2.9 times more from a common ancestor than that of *D. yakuba.* A similar degree of divergence (~3.5 times) has occurred for the cyt *b* PPC of *T. rufoniger* (Fig. 5).

Comparison of branch lengths from 14 relevant four-species Wagner network analyses involving the data from Fig. 4 shows that a large part of the increased cyt *b* PPC divergence is shared between aculeate hymenopterans (Table 6), regardless of out-group composition.

If the evolutionary rate has been constant over the last ~320 Myr, the result above suggests that the PPC divergence of Formicidae and Apidae took place quite recently. Alternatively, the evolutionary rate may have fluctuated in the past and hence the lineage giving rise to aculeate hymenopterans may have diverged relatively more from a common ancestor at an early stage of their evolution than that of the dipterans. In order to determine which of these two explanations is the most likely, attention must be given to studies of PPC divergence in other insects.

The DNA sequences presented by Liu and Beckenbach (1992), of the relatively short CO-II gene, provide the only suitable data bearing on this problem. The PPC divergences of 15 different insects in relation to the paleopterous pterygote insect *Sympetrum striolatum* are presented in Table 4. The mean PPC divergence of neopterous Exopterygota is not significantly different from paleopterous Pterygota (Fig. 6).

The mean PPC divergence of neopterous Endopterygota, on the other hand, is significantly higher than those of the paleopterous Pterygota and the neopterous Exopterygota. With the exception of Diptera, all Endopterygota are above the lower 95% confidence limit for the mean value (Table 4, Fig. 6), and the Coleoptera and Hymenoptera, in particular, show PPC divergences that are above average for Endopterygota. This result, similar to that found from the study of nucleotide content, indicates that increased PPC divergences are not confined to Hymenoptera but also apply to Endopterygota (except Diptera) as a whole. The data (Figs. 3, 6) also indicate that changes in overall A + T% and PPC divergence are positively associated. Given the relatively high A + T% and the possibility of directional mutation pressure on the CO-II genes of Endopterygota, proportionally more replacements may be expected to have occurred in the PPCs of this group of insects, and this may ultimately have caused the observed increase in the PPC divergences. The data

Table 4. Overall A + T content, amino acid sequence divergence, and amino acid predictors of the CO-II gene in 10 orders of insects[a]

Division	Order	Species	Overall A + T (%)	Amino acid sequence divergence[b]	Amino acid predictors[c]
Paleoptera	Odonata	*Sympetrum striolatum*	68.6	1.0000	—
Exopterygota	Orthoptera	*Acheta domesticus*	69.9	1.0000	Size, A + T
		Schistocerca gregaria	67.7	1.0400	Size, A + T
	Dictyoptera	*Periplaneta americana*	70.6	1.2143	—
	Isoptera	*Zootermopsis angusticollis*	67.0	1.0000	—
	Hemiptera	*Oncopeltus faciatus*	72.3	0.9524	—
Endopterygota	Coleoptera	*Adalia bipunctata*	73.8	1.8333	Size, A + T
		Sitophilus granarius	69.8	1.1739	—
	Siphonaptera	*Ctenocephalides felis*	78.0	1.1905	Size, A + T
	Diptera	*Drosophila melanogaster*	73.8	0.8333	Size, A + T
		Drosophila pseudoobscura	73.4	0.9524	Size, A + T
		Drosophila yakuba	74.0	0.8636	Size, A + T
	Lepidoptera	*Galleria mellonella*	76.6	1.3125	G
	Hymenoptera	*Apis mellifera*	80.0	1.8636	A + T, size
		Exeristes roborator	77.2	1.3478	A + T, size
		Lasius sp.	75.8	1.7619	A + T, size

[a] Data from de Bruijn (1983) (*Drosophila melanogaster*), Clary and Wolstenholme (1985) (*Drosophila yakuba*), Crozier et al. (1989) (*Apis mellifera*), and Liu and Beckenbach (1992)

[b] Amino acid sequence divergence was analyzed using a Wagner network. Divergence is given as the fraction of the branch length to *Sympetrum striolatum* in an analysis with *Mus musculus* (Bibb et al. 1981) and *Xenopus laevis* (Roe et al. 1985) as out-groups. All best trees were significantly better ($P < 0.01$) than other trees

[c] Correlation of the relative frequency of amino acids in a polypeptide chain with six independent variables (size of codon family [Size], A + T content within codon families [A + T], and proportion of A, T, G, and C within codon families) were analyzed using stepwise multiple regression analysis (Sokal and Rohlf 1981). Variables that significantly ($P < 0.05$) predict the relative frequency of amino acids in the inferred polypeptide chain are listed according to their relative importance.

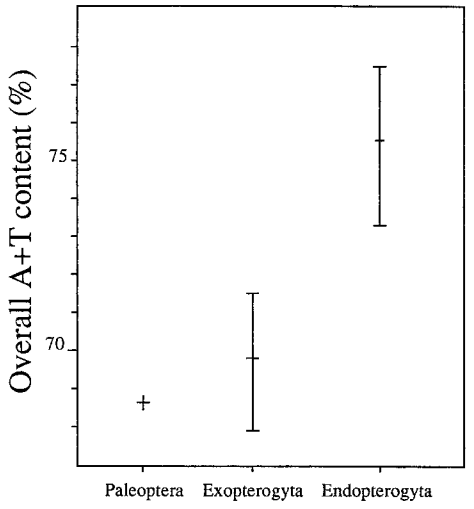

Fig. 3. The bootstrapped (10,000 replicates) mean and 95% confidence limits of the overall A + T content (%) in cytochrome oxidase subunit II of Paleoptera, Exopterygota, and Endopterygota. Orders were treated as units and we therefore averaged the A + T% of orders with more than one taxon. Data are those from Table 4.

in Table 4 do not support this association ($P > 0.1$; Fig. 7).

The nonsignificant association between the A + T% of the CO-II gene and the relative PPC divergence of the corresponding gene product among the 10 orders of insects may indicate that there exists no functional relationship between A + T% and relative PPC divergence but may also suggest that such a functional relationship may have been disguised by differential effects from the directional mutation pressure and/or selective constraints. The effect of directional mutation pressure and selective constraints on genomic A + T% was proposed by Sueoka (1962) and Freese (1962) and subsequently shown by several authors (e.g., Muto and Osawa 1987; Sueoka 1988, 1992). However, since the methods used by these authors are not directly applicable to the data analyzed in this paper, we cannot pursue this question further. Ideally, quantification of the directional mutation pressure, response to directional mutation pressure, and selective constraints from single DNA sequences would facilitate the study of association between directional mutation pressure and relative PPC divergence.

Phylogenetic Relationship

The evolution of A + T% and PPC divergence in the CO-II and cyt *b* genes of insects may have taken place according to the following scenario (Fig. 8). The two genes of the early insects (Paleoptera) probably had an A + T% of ~69% and their corresponding PPCs diverged from Apterygota at an unknown rate. Derived insects (Neoptera), except for the neopterous Endopterygota, maintained the A + T% at ~69% and their corresponding PPCs diverged no more from a common ancestor than did the paleopterous insects. In contrast to the situation for Paleoptera and Exopterygota, the lineage giving rise to the evolutionarily successful and speciose Endopterygota experienced an increase in A +

Table 5. Codon usage in the cytochrome *b* gene for *Tetraponera rufoniger* in comparison with those from *Apis mellifera, Drosophila melanogaster,* and *Drosophila yakuba*[a]

Codon	AA	*Te*	*Ap*	*Dm*	*Dy*	Codon	AA	*Te*	*Ap*	*Dm*	*Dy*
TTT	F	26	34	26	28	TCT	S	9	4	4	3
TTC		6	7	6	3	TCC		1	1	2	1
TTA	L	11	38	48	49	TCA		12	17	8	9
TTG		2	0	1	2	TCG		0	0	0	0
CTT	L	13	8	4	3	CCT	P	5	4	12	12
CTC		3	0	0	0	CCC		5	0	0	0
CTA		14	3	2	4	CCA		6	12	8	11
CTG		4	0	0	0	CCG		1	0	0	0
ATT	I	34	44	39	43	ACT	T	13	7	11	9
ATC		13	5	3	2	ACC		1	0	1	0
ATA	M	24	26	11	10	ACA		7	8	7	10
ATG		1	1	1	1	ACG		0	0	0	0
GTT	V	4	2	10	11	GCT	A	6	3	14	15
GTC		1	0	0	0	GCC		3	1	3	0
GTA		9	6	15	12	GCA		3	8	4	6
GTG		2	0	0	0	GCG		0	0	0	0
Codon	AA	*Te*	*Ap*	*Dm*	*Dy*	Codon	AA	*Te*	*Ap*	*Dm*	*Dy*
TAT	Y	10	24	14	13	TGT	C	0	2	3	2
TAC		6	0	6	5	TGC		1	0	0	1
TAA		1	1	1	1	TGA	W	10	13	13	13
TAG		0	0	0	0	TGG		1	0	0	0
CAT	H	5	8	10	6	CGT	R	0	1	0	1
CAC		5	1	0	5	CGC		0	0	0	0
CAA	Q	7	4	8	8	CGA		3	3	8	7
CAG		1	0	0	0	CGG		0	0	0	0
AAT	N	21	36	21	18	AGT	S	1	2	2	1
AAC		9	3	5	6	AGC		0	0	0	0
AAA	K	6	15	8	6	AGA		8	1	3	3
AAG		1	0	0	1	AGG		3	0	0	0
GAT	D	5	6	8	8	GGT	G	1	5	5	6
GAC		2	1	2	2	GGC		3	0	0	0
GAA	E	3	3	3	3	GGA		12	16	18	19
GAG		2	0	0	0	GGG		5	0	1	0

[a] Data from Crozier and Crozier (1992) (*Apis mellifera*), Garesse (1988) (*Drosophila melanogaster*), and Clary and Wolstenholme (1985) (*Drosophila yakuba*). Abbrevieations: AA = aino acid, *Te = Tetraponera rufoniger, Ap = Apis mellifera, Dm = Drosophila melanogaster, Dy = Drosophila yakuba*

T% from ~69% to ~75% and an acceleration of the divergences of their corresponding PPCs.

The fact that the concurrent elevation of the A + T% and PPC divergence is confined to Endopterygota only and that the endopterygous orders of insects probably already had separated in the Permian (245–285 Myr ago) (Carpenter and Burnham 1985) suggests that the transition to a higher A + T% may have occurred in the Upper Carboniferous (285–320 Myr ago). The scenario seems plausible, but must be regarded as tentative. Further surveys of slow-evolving mitochondrial or nuclear protein-coding genes are needed to test this hypothesis.

The evolution of the A + T% in the CO-II and cyt *b* genes of four insect species as well as the corresponding PPC divergence is not covered by the scenario presented above and therefore deserves a few comments. *Sitophilus granarius* (as shown by the CO-II gene) and *T. rufoniger* (as shown by the cyt *b* gene) both have A + T% significantly lower (Tables 3, 4) than average (Fig. 6) for neopterous Endopterygota and PPCs of the corresponding genes which have diverged relatively more from a common ancestor than those of paleopterous insects. They probably lost the high A + T% secondarily and independently after having had the high A + T% of other endopterous insects, perhaps due to a cessation in directional mutation pressure. For *T. rufoniger* (as shown by the cyt *b* gene) this change may have occurred within the last 100 Myr because the other formicid species, *Lasius* sp., resembles other endopterygous insects with respect to the A + T% of the CO-II gene and the corresponding PPC divergence. The three dipterans, on the other hand, have A + T% values which resemble those of other neopterous Endopterygota and yet their CO-II PPCs do not appear to have diverged markedly more from a common ancestor than have the paleopterous insects. Britten (1986) suggested that increased efficiency in DNA repair might explain the apparent slowdown in primate molecular evolutionary rates; similarly, we suggest that an improvement in mtDNA transversion repair mechanisms in Diptera

```
Tet   MNN-M--LKS ---LKLEL-- LELPTPINIS YLWNMGSILG LFLGVQILSG FMLSMHYCPS TSMAFMSIIH IMQNVNSGWL FHNIHINGAT    82
Api   .KKF.NFFS. NEF..mImST iY....v..N .m..F..... i..Mi..i.. .i.......N IDi..W..TN ..Kdm..... .rL..m...s    90
Dme   ..KPlRN-SH PLFKIAnNAL vd..a..... SW..F..l.. .C.Ii...t. LF.a...TAD INl..Y.vN. .CRd..Y... LrTl.A...s    89
Dya   .HKPlRN-SH PLFKIAnNAL vd..a..... SW..F..l.. .C.Ii...t. LF.a...TAD VNl..Y.vN. .CRd..Y... LrTl.A...s    89
Mus   .T-N.RK-TH PLFKIinHSF id..a.S... SW..F..l.. vC.M...it. LF.a...TSD .MT..S.vT. .CRd..Y... IrYm.A...s    88
Xen   .APNiRK-SH PLIKIinNSF id....S... S...F..l.. vC.IA..it. LF.a...TAD .....S.vA. .CFd..Y.L. Ir.l.A..Ls    89

Tet   MFFMSMYIHI ARGIYFNSFN LTNTWMIGVS IFLVSMATAF LGYVLPWGQM SFWGATVITN LLSTIPYLGN DIVMWVWGGF SINNSTLNRF   172
Api   Fy.lM..... s.NlfyC.yK .N.V.G..iM .L.m..:.a.. m......... .y........ ...a...i.d T..l.i.... ....a.....   180
Dme   F..iCi.l.v G....yG.yK F.P..l...I .LFlV.G... m......... ...V...... ..Ya.....M .l.Q.l....' avd.a..T..   179
Dya   F..iCi.l.. G....yG.yL F.P..lv..I .LFlV.G... m......... .......... ...a.....M .l.Q.l.... avd.a..T..   179
Mus   ...iClfl.v G..l.yG.yT FMe..N...L lLFAV..... m......... .......... ...a...i.T Tl.E.i.... .vdKa..T..   178
Xen   F..iCi.l.. G..l.yG..L YKe..N...I lLFlV..... v......... .......... ...aK..i.. Vl.Q.SL... .vd.a..T..   179

Tet   YSLHFILPFL ISFMVILHLI FLHDTGSSNP LGTNSDMNKI PFHYYFTIKD ILGFTIIMII FMIVILQYPY VFSDPDNFTM ANPMVTPIHI   262
Api   f.......L. .L.......F A..L..... ..s.FnNY.. S..P..s... l...Y..lF. ..FiNF.f.. HLG.....Ki ....N..T..   270
Dme   ftF......i vLA.Tmi..l ...q...N.. i.L..nid.. ...P...F.. .v..IvmiF. LiSlv.IS.N lLG.....YS ..SFSNTA..   269
Dya   ftF......i vLA.Tmi..l ...q...N.. i.L..nid.. ...P...F.. .v..IvmiF. LiSlv.IS.N lLG.....IP ...l...A..   269
Mus   faF......i .aAlA.v..l ...e...N.. T.L...Ad.. ...P.y.... ...IL.mFl. L.Tlv.Ff.D mLG....yMP ...lN..P..   268
Xen   faF..l...i .aGAS....l ...e...t.. T.L...Pd.v ...P..sY.. l...L.mlTA LTllAmFS.N lLG......P ...li..P..   269

Tet   QPEWYFLFAY AILRSIPNKL GGVLALVSSI FILYFLPL-- NKMEMNTNFY PTNQINFWLL MNSFIILTWV GAQSIEDPFV LISQITALIY   350
Api   K......... s...a..... ...iG..M.. L...ImIFYN .-KM..NK.N ML.K.Yy.mF i.N..l...l .K.L..Y..T N.NMlFtTT.   359
Dme   .......... .......... ...i...L.. A..MI..FYN LSKFRGIq.. .I..vM..Sm lVtV.l...i ..RPv.e.y. ..G..Ltvv.   359
Dya   .......... .......... ...i...L.. A..MI..FYN LSKFRGIq.. .I...L..Sm lVtV.l...i ..RPv.e.y. ..G..Lti..   359
Mus   K......... .......... ......iL.. L..ALm.FLH TSKqRSLM.R .IT..Ly.i. vANLl....i .G.Pv.H..i i.G.lasiS.   358
Xen   K......... .....m-... .......L.. L..ALm..LH TSKqRSLM.R .FT..M..A. vADTl....i .G.Pv...yT m.G.lasv..   358

Tet   FSTLSLISIS GSLWNKMI-N --L-                                                                          370
Api   .LYFF.NFYL SK..dNl.W. SP.N                                                                          383
Dme   .LYYLvNPlI TKW.dNl--- --.N                                                                          378
Dya   .LYYLiNPlV TKW.dNl--- --.N                                                                          378
Mus   ..IiLilMPI SGiIeDKmLK LYP-                                                                          381
Xen   ..IFIimFPL MGWVeNKlL. W---                                                                          379
```

Fig. 4. Inferred and aligned amino acid sequences of cytochrome *b* from *Tetraponera rufoniger, Apis mellifera* (Crozier and Crozier 1992), *Drosophila melanogaster* (Garesse 1988), *D. yakuba* (Clary and Wolstenholme 1985), *Xenopus laevis* (Roe et al. 1985), and *Mus musculus* (Bibb et al. 1981). Sequence differences are given in relation to *Tetraponera rufoniger.* Identical amino acids are indicated by *dots;* conservative replacements are given by *lowercase letters;* and nonconservative replacements are represented by *capital letters.* Conservative and nonconservative replacements follow the protocol of French and Robson (1983). *Hyphens* denote gaps inserted to improve alignment.

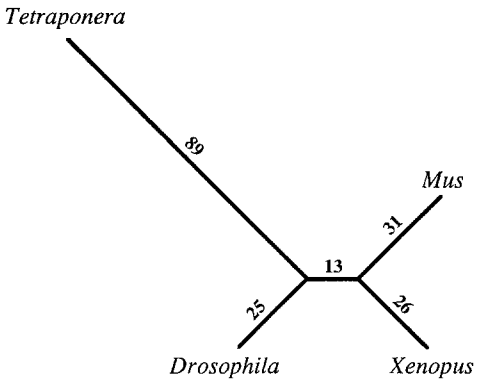

Fig. 5. Wagner network analysis of assignable amino acid replacements shows that the cytochrome *b* polypeptide chain of *Tetraponera rufoniger* has diverged ~3.5 times more from a common ancestor than has that of *Drosophila yakuba.* This topology is significantly better ($P < 0.05$) than that of other trees. Data from Fig. 4.

Table 6. Number of assignable amino acid replacements in cytochrome *b* from the best-supported four-species topology[a]

Terminal species							
Tetr.	*Apis*	*D.m.*	*D.y.*	*Mus*	*Xeno.*	Inter.[b]	Other
40[c]	42[c]	9	4	—	—	58	151
41[c]	33[c]	28	—	34	—	31	228
39[c]	34[c]	—	23	34	—	32	225
79[c]	—	13	3	30[c]	—	18	150
—	86[c]	11	4	33[c]	—	28	125
39[c]	40[c]	32	—	—	38	32	212
38[c]	41[c]	—	26	—	37	34	209
82[c]	—	13	4	—	32[c]	22	135
—	86[c]	11	4	—	35[c]	24	129
39[c]	37[c]	—	—	25	24	40	201
86[c]	—	31[c]	—	31	25	13	154¶
—	89[c]	36[c]	—	26	30	17	139
89[c]	—	—	25[c]	31	26	13	146
—	94[c]	—	32[c]	28	32	16	126

[a] Data from Fig. 4 were evaluated using a Wagner network analysis. Proportion of unvaried positions ≈0.4. Abbreviations: *Tetr.* = *Tetraponera rufoniger; Apis* = *Apis mellifera; D.m.* = *Drosophila melanogaster, D.y.* = *Drosophila yakuba; Mus* = *Mus musculus; Xeno.* = *Xenopus laevis;* Inter. = internode in a Wagner network; Other = unassignable positions

[b] All best trees are significantly better ($P < 0.05$) than other trees except one which is marked with ¶. The best tree from this particular four-species Wagner network analysis is significantly better than other trees at a 10% level

[c] Indicates which species pair grouped together in the best Wagner network tree

relative to other insects could explain the low rate of PPC divergence coupled with high A + T%. Such a mechanism would, with reference to our model, reduce the PPC divergence notably, even if all such mutations occurred at the third codon position only.

At this point we may return to the aim of the paper and answer the two questions which were raised in our introduction. While the A + T% in the CO-II gene of Endopterygota in general is significantly higher (~75%) than that observed in older and more primitive orders of

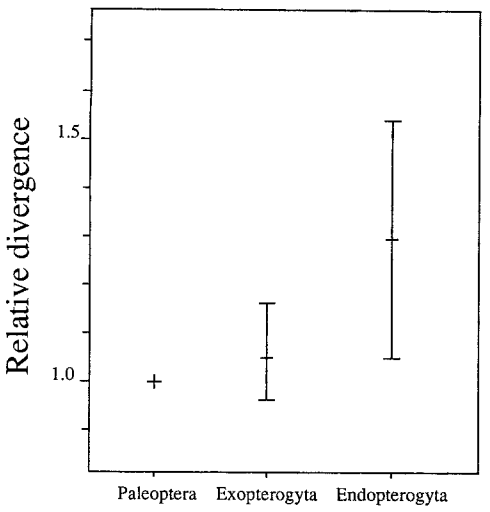

Fig. 6. The bootstrapped mean (10,000 replicates) and 95% confidence limits of the polypeptide chain divergences for cytochrome oxidase subunit II of Paleoptera, Exopterygota, and Endopterygota. Orders were treated as units and we therefore averaged the polypeptide chain divergences of orders with more than one taxon. Data are from Table 4.

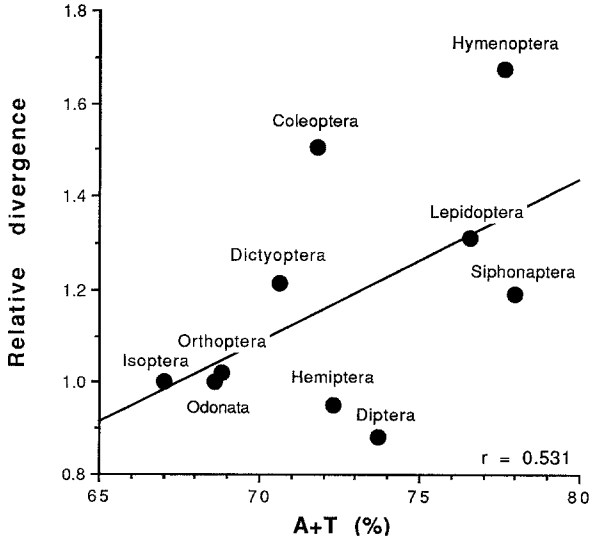

Fig. 7. Relative divergences of the cytochrome oxidase subunit II polypeptide chains for 15 species of insects are plotted against the A + T content (AT%) of their respective DNA sequences. The paleopterous insect *Sympetrum striolatum* was used as a reference for measures of divergence. Orders were treated as units. The simple regression line and the correlation coefficient (r) ($P > 0.1$) are superimposed. Data are from Table 4.

insects (~69%), the A + T% in Hymenoptera and particularly in *A. mellifera* is still unique in being the highest recorded mitochondrial A + T% among metazoans. The A + T% of the cyt *b* gene from *A. mellifera, D. melanogaster,* and *D. yakuba* closely follows the trend observed for the CO-II gene whereas that of *T. rufoniger* appears to be markedly lower.

Likewise, CO-II PPCs of Endopterygota are the only ones among insects that have diverged significantly more from a common ancestor than have those of Paleoptera. The CO-II PPCs of Hymenoptera have diverged relatively more than most other orders of insects and are equalled in this respect only by the beetle *Adalia bipunctata.* The CO-II PPC of *Apis mellifera* has indeed diverged further from a common ancestor than have those of other insects but is generally similar to the *Lasius* sp. and *Adalia bipunctata* sequences in this respect. The cyt *b* PPCs of *Apis mellifera, D. melanogaster* and *D. yakuba* closely follows the trend observed for the CO-II PPC and so does that of *T. rufoniger.* Comparisons between the cyt *b* PPCs of *A. mellifera* and *T. rufoniger* and between the CO-II PPCs of *A. mellifera, Lasius* sp., and *Exeristes roborator* finally suggest that the PPC divergence observed in previous analyses of *A. mellifera* (Crozier et al. 1989; Crozier and Crozier 1992) does not qualify the species to be called unique among hymenopterans with respect to PPC divergence.

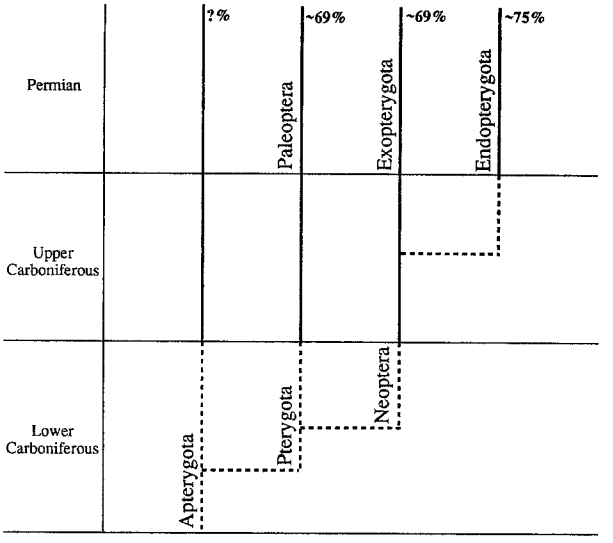

Fig. 8. Evolution of A + T content in the cytochrome oxidase subunit II gene of major insect groups according to temporal and phylogenetic affiliation.

Increased divergence of inferred PPCs is not significantly associated with an increase of the overall A + T% in the CO-II genes of insects. Marked deviations from such an association are noticeable; some species have a low A + T% but a high PPC divergence (*S. granarius* for the CO-II gene and *T. rufoniger* for the cyt *b* gene) whereas other species have a high A + T% and a relatively low PPC divergence (the three *Drosophila* CO-II genes). Directional mutation pressure, response to directional mutation pressure, and selective constraints can presently not be quantified for single DNA sequences and reasons for the nonsignificant association as well as the above-mentioned deviations therefore cannot be given. Furthermore, it is not known whether (1) a secondary reversal of mutation pressure resulting in lower A + T% could further increase PPC divergence following the initial attainment of high A + T%

and whether (2) a combination of strong mutation pressure toward a high A + T% coupled with a reduced substitution rate, possibly mediated by unusually effective mutation repair, can affect relative divergence.

Generation time has been suggested as a major determinant of molecular evolutionary rates (Wu and Li 1985; Li and Tanimura 1987). It is hard to evaluate generation time as a factor in influencing evolutionary rate differences between insect mtDNAs because of the great lengths of time involved during which life histories will have changed substantially (Crozier et al. 1989), but it is clear that changes in A + T% can have a major impact on the degree of divergence. The implication of nucleotide content in the determination of degree of PPC divergence may perhaps also explain some of the rate variation previously reported to occur at the protein level (Gillespie 1986) and hence indicate that such variation is, after all, consistent with the neutral theory of molecular evolution (Kimura 1983).

Acknowledgments. This paper constitutes a part of the research done by L.S.J. in partial fulfillment of the Ph.D. degree of La Trobe University. We thank A.T. Beckenbach and H. Liu for giving us their manuscript before publication and M. Villet for specimens of *Dorylus*. This research would not have been possible without support from the Danish Research Academy, the Australian Government's Overseas Postgraduate Research Scholarship, and La Trobe University's Postgraduate Scholarship to L.S.J. and support from the Australian Research Council, La Trobe University, and the Ian Potter Foundation to R.H.C. We thank Y.C. Crozier, N. Dobric, S. Koulianos, R. Kusmierski, S.H. Lawler, A. Poiani, and R.M. Lowe for constructive comments on the manuscript. The nucleotide sequence reported in this paper has been submitted to the GenBank with the accession number U 02458.

References

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulsen AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981) Sequence and organisation of the human mitochondrial genome. Nature 290:457–465

Anderson S, de Bruijn MHL, Coulsen AR, Eperon IC, Sanger F, Young IG (1982) Complete sequence of bovine mitochondrial DNA: conserved features of the mammalian mitochondrial genome. J Mol Biol 156:683–717

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

Bibb MJ, Van Etten RA, Wright CT, Walberg MW, Clayton DA (1981) Sequence and gene organisation of mouse mitochondrial DNA. Cell 26:167–180

Birley AJ, Croft JH (1990) Mitochondrial DNAs and phylogenetic relationships. In: Dutta SK (ed) DNA systematics. CRC Press, Boca Raton, pp 107–137

Bodmer M, Ashburner M (1984) Conservation and change in the DNA sequences coding for alcohol dehydrogenase in sibling species of *Drosophila*. Nature 300:425–427

Boer PH, Hickey DA (1986) The alpha-amylase gene in *Drosophila melanogaster:* nucleotide sequence, gene structure and expression motifs. Nucleic Acids Res 14:46–55

Brandão CRF (1990) Phylogenetic, biogeographic, and evolutionary inferences from the description of an early cretaceous South American Myrmeciinae. In: Veeresh GK, Mallik B, Viraktamath CA (eds) Social insects and the environment. Oxford and Ibh Publishing, New Delhi, pp 313–314

Brandão CRF, Martins-Neto RG, Vulcano MA (1989) The earliest known fossil ant (first southern hemisphere mesozoic record) (Hymenoptera: Formicidae: Myrmeciinae). Psyche 96:195–208

Britten RJ (1986) Rates of DNA sequence evolution differ between taxonomic groups. Science 231:1393–1398

Brown WM (1985) The mitochondrial genome of animals. In: MacIntyre RJ (ed) Molecular evolutionary genetics. Plenum, New York, pp 95–130

Brown WM, George M, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. Proc Natl Acad Sci USA 76:1967–1971

Brown WM, Prager EM, Wang A, Wilson AC (1982) Mitochondrial DNA sequences of primates: tempo and mode of evolution. J Mol Evol 18:225–239

Cameron SA, Derr JN, Austin AD, Woolley JB, Wharton RA (1992) The application of nucleotide sequence data to phylogeny of the Hymenoptera: a review. J Hym Res 1:63–79

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

Carpenter FM, Burnham L (1985) The geological record of insects. Annu Rev Earth Planet Sci 13:297–314

Clary DO, Wolstenholme DR (1985) The mitochondrial DNA molecule of *Drosophila yakuba:* nucleotide sequence, gene organization, and genetic code. J Mol Evol 22:252–271

Crozier RH (1990) From population genetics to phylogeny: uses and limits of mitochondrial DNA. Aust Syst Bot 3:111–124

Crozier RH (1992) Molecular methods for insect phylogenetics. In: Whitten MJ, Oakeshott JG (eds) Molecular approaches to pure and applied entomology. Springer-Verlag, Berlin, pp 164–221

Crozier RH, Crozier YC, Mackinley AG (1989) The CO-I and CO-II region of the honeybee mitochondrial DNA: evidence of variation in insect mitochondrial evolutionary rates. Mol Biol Evol 6: 399–411

Crozier RH, Crozier YC (1992) The cytochrome b and ATPase genes of honeybee mitochondrial DNA. Mol Biol Evol 9:474–482

Crozier RH, Crozier YC (1993) The mitochondrial genome of the honeybee *Apis mellifera:* complete sequence and genome organization. Genetics 113:97–117

Crozier YC, Koulianos S, Crozier RH (1991) An improved test for africanized honeybee mitochondrial DNA. Experientia 47:968–969

de Bruijn MHL (1983) *Drosophila melanogaster* mitochondrial DNA, a novel organization and genetic code. Nature 304:234–241

Derr JN, Davis SK, Woolley JB, Wharton RA (1992a) Variation and the phylogenetic utility of the large ribosomal subunit of mitochondrial DNA from the insect order Hymenoptera. Mol Phyl Evol 1:136–147

Derr JN, Davis SK, Woolley JB, Wharton RA (1992b) Reassessment of the 16S rRNA nucleotide sequence from members of the parasitic Hymenoptera. Mol Phyl Evol 1:338–341

DeSalle R, Freedman T, Prager EM, Wilson AC (1987) Tempo and mode of sequence evolution in mitochondrial DNA of Hawaiian *Drosophila*. J Mol Evol 26:157–164

Desjardin P, Morais R (1990) Sequence and gene organization of the chicken mitochondrial genome. A novel gene order in higher vertebrates. J Mol Biol 212:599–634

di Rago J-P, Netter P, Slonimski PP (1990) Pseudo-wild type revertants from inactive apocytochrome b mutants as a tool for the analysis of the structure/function relationships of the mitochondrial ubiquinol-cytochrome *c* reductase of *Saccharomyces cerevisiae*. J Biol Chem 265:3332–3339

Efron B, Gong G (1983) A leisurely look at the bootstrap, the jackknife and cross-validation. Am Stat 37:37–48

Freese E (1962) On the evolution of base composition of DNA. J Theor Biol 3:82–101

French S, Robson B (1983) What is a conservative substitution? J Mol Evol 19:171–175

Gadaleta G, Pepe G, De Candia G, Quagliariello C, Sbisà E, Saccone C (1989) The complete nucleotide sequence of the *Rattus norvegicus* mitochondrial genome: cryptic signals revealed by comparative analysis between vertebrates. J Mol Evol 28:497–516

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

Gillespie JH (1986) Variability of evolutionary rates of DNA. Genetics 113:1077–1091

Harrison RG (1989) Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. Trends Ecol Evol 4:6–11

Hatefi Y (1985) The mitochondrial electron transport and oxidative phosphorylation system. Annu Rev Biochem 54:1015–1069

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

Honeycutt RL, Wheeler WC (1990) Mitochondrial DNA: variation in humans and higher primates. In: Dutta SK, Winter WP (eds) DNA systematics. CRC Press, Boca Raton, pp 91–129

Hori H, Osawa S (1987) Origin and evolution of organisms as deduced from 5S ribosomal RNA sequences. Mol Biol Evol 4:445–472

Howell N (1989) Evolutionary conservation of protein regions in the protonmotive cytochrome *b* and their possible roles in redox catalysis. J Mol Evol 29:157–169

Howell N, Gilbert K (1988) Mutational analysis of the mouse mitochondrial cytochrome *b* gene. J Mol Biol 203:607–618

Irwin DA, Kocher TD, Wilson AC (1991) Evolution of cytochrome *b* gene of mammals. J Mol Evol 32:128–144

Jacobs HT, Elliott DJ, Math VB, Farquharson A (1988) Nucleotide sequence and gene organization of sea urchin mitochondrial DNA. J Mol Biol 202:185–217

Keith TP, Riley MA, Kreitman M, Lewontin RC, Curtis D, Chambers G (1987) Sequence of the structural gene for xanthine dehydrogenase (*rosy* locus) in *Drosophila melanogaster*. Genetics 116:67–73

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge, pp 1–367

Koulianos S, Crozier RH (1991) Two ancient mitochondrial alleles in Australian honeybees. Apidologie 22:621–626

Kusmierski R, Borgia G, Crozier RH, Chan BHY (1993) Molecular information on bower-bird phylogeny and the evolution of exaggerated male characteristics. J Evol Biol 6:737–752

Li W-H, Tanimura M (1987) The molecular clock runs more slowly in man than in apes and monkeys. Nature 326:93–96

Liu H, Beckenbach AT (1992) Evolution of the mitochondrial cytochrome oxidase II gene among 10 orders of insects. Mol Phyl Evol 1:41–52

Meyer A, Wilson AC (1990) Origin of tetrapodes inferred from their mitochondrial DNA affiliation to lungfish. J Mol Evol 31:359–364

Michener CD, Grimaldi DA (1988) The oldest fossil bee: apoid history, evolutionary stasis, and antiquity of social behaviour. Proc Natl Acad Sci USA 85:6424–6426

Miyamoto MM, Boyle SM (1989) The potential importance of mitochondrial DNA sequence data to eutherian mammal phylogeny. In: Fernholm B, Bremer K, Jörnvall H (eds) The hierarchy of life. Elsevier, Amsterdam, pp 437–450

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

Moritz C, Dowling TE, Brown WM (1987) Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Annu Rev Ecol Syst 18:269–292

Muto A, Osawa S (1987) The guanine and cytosine content of genomic DNA and bacterial evolution. Proc Natl Acad Sci USA 84:166–169

Okimoto R, Macfarlane JL, Clary DO, Wolstenholme DR (1992) The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. Genetics 130:471–498

Osawa S, Jukes TH, Watanabe K, Muto A (1992) Recent evidence for evolution of the genetic code. Microbiol Rev 56:229–264

Powell JR, Caccone A, Amato GD, Yoon C (1986) Rates of nucleotide substitution in *Drosophila* mitochondrial DNA and nuclear DNA are similar. Proc Natl Acad Sci USA 83:9090–9093

Preparata G, Saccone C (1991) DNA markov clocks in gene evolution. Evol Biol 5:45–58

Roe BA, Ma D-P, Wilson RK, Wong JF-H (1985) The complete nucleotide sequence of the *Xenopus laevis* mitochondrial DNA. J Biol Chem 260:9759–9774

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Millis KB, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487–491

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning. 2nd ed. Cold Spring Harbor Lab Press, Cold Spring Harbor

Sander M, Lowenhaupt K, Rich A (1991) *Drosophila* Rrp1 protein: an apurinic endonuclease with homologous recombination activities. Proc Natl Acad Sci USA 88:6780–6784

Sanger F, Nicklen S, Coulsen AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Sokal RR, Rohlf FJ (1981) Biometry, 2nd ed. Freeman, New York, pp 1–859

Sueoka N (1962) On the genetic basis of variation and heterogeneity of DNA base composition. Proc Natl Acad Sci USA 48:582–592

Sueoka N (1988) Directional mutation pressure and neutral molecular evolution. Proc Natl Acad Sci USA 85:2653–2657

Sueoka N (1992) Directional mutation pressure, selective constraints, and genetic equilibria. J Mol Evol 34:95–114

Tron T, Crimi M, Colson A-M, Degli Esposti M (1991) Structure/function relationships in mitochondrial cytochrome *b* revealed by the kinteic and circular dichroic properties of two yeast inhibitor-resistant mutants. Eur J Biochem 199:753–760

Vawter L, Brown WM (1986) Nuclear and mitochondrial DNA comparisons reveal extreme rate variation in the molecular clock. Science 234:194–196

Willis LG, Winston ML, Honda BM (1992) Phylogenetic relationships in the honeybee (genus *Apis*) as determined by the sequence of the cytochrome oxidase II region of mitochondrial DNA. Mol Phyl Evol 1:169–178

Wolstenholme DR, Clary DO (1985) Sequence evolution of *Drosophila* mitochondrial DNA. Genetics 109:725–744

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