

The Mitochondrial Genome of the Honeybee *Apis mellifera*: Complete Sequence and Genome Organization

R. H. Crozier and Y. C. Crozier

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

Manuscript received June 12, 1992
Accepted for publication October 3, 1992

ABSTRACT

The complete sequence of honeybee (*Apis mellifera*) mitochondrial DNA is reported being 16,343 bp long in the strain sequenced. Relative to their positions in the *Drosophila* map, 11 of the tRNA genes are in altered positions, but the other genes and regions are in the same relative positions. Comparisons of the predicted protein sequences indicate that the honeybee mitochondrial genetic code is the same as that for *Drosophila*; but the anticodons of two tRNAs differ between these two insects. The base composition shows extreme bias, being 84.9% AT (*cf.* 78.6% in *Drosophila yakuba*). In protein-encoding genes, the AT bias is strongest at the third codon positions (which in some cases lack guanines altogether), and least in second codon positions. Multiple stepwise regression analysis of the predicted products of the protein-encoding genes shows a significant association between the numbers of occurrences of amino acids and %T in codon family, but not with the number of codons per codon family or other parameters associated with codon family base composition. Differences in amino acid abundances are apparent between the predicted *Apis* and *Drosophila* proteins, with a relative abundance in the *Apis* proteins of lysine and a relative deficiency of alanine. *Drosophila* alanine residues are as often replaced by serine as conserved in *Apis*. The differences in abundances between *Drosophila* and *Apis* are associated with %AT in the codon families, and the degree of divergence in amino acid composition between proteins correlates with the divergence in %AT at the second codon positions. Overall, transversions are about twice as abundant as transitions when comparing *Drosophila* and *Apis* protein-encoding genes, but this ratio varies between codon positions. Marked excesses of transitions over chance expectation are seen for the third positions of protein-encoding genes and for the gene for the small subunit of ribosomal RNA. For the third codon positions the excess of transitions is adequately explained as due to the restriction of observable substitutions to transitions for conserved amino acids with two-codon families; the excess of transitions over expectation for the small ribosomal subunit suggests that the conservation of nucleotide size is favored by selection.

ANIMAL mitochondrial DNA (mtDNA) occurs as a single circular molecule, generally about 16,000 bp long. This organelle genome contains 13 (or sometimes 12—WOLSTENHOLME *et al.* 1987; OKIMOTO *et al.* 1992) protein-encoding genes, the genes for 22 tRNAs and two ribosomal RNA subunits, and a noncoding region containing the origin of replication for the heavy strand in vertebrates (BROWN 1985) and both strands in *Drosophila* (CLARY and WOLSTENHOLME 1987). Other than this latter control region, noncoding nucleotides are rare in animal mtDNA (HARRISON 1989).

Gene order variation occurs both between and within phyla. Among vertebrates, birds differ from mammals and amphibians in gene order, suggesting that the move of a segment containing one tRNA and one protein-encoding gene occurred in the ancestor of class Aves (DESJARDINS and MORAIS 1990). Within

mammals, marsupials differ from eutherians in the positions of some tRNAs (PÄÄBO *et al.* 1991). Within insects, tRNA genes are known to vary in position both between the orders Diptera (CLARY and WOLSTENHOLME 1985) and Hymenoptera (CROZIER, CROZIER and MACKINLAY 1989) and within the order Diptera, with differences observed between *Aedes* (HSUCHEN, KOTIN and DUBIN 1984; HSUCHEN and DUBIN 1984) and *Drosophila yakuba* (CLARY and WOLSTENHOLME 1985). However, prior to the present study, a complete sequence was known from only one insect species, *D. yakuba* (CLARY and WOLSTENHOLME 1985), laying open the possibility of more extensive rearrangements.

Drosophila mtDNA is highly AT-rich (CLARY and WOLSTENHOLME 1985); sequences of several honeybee mtDNA genes show that *Apis* mtDNA is even more so (CROZIER, CROZIER and MACKINLAY 1989; CROZIER and CROZIER 1992). In very AT-rich genomes the general bias toward transitions in substitutions might be expected not to occur, because the

The sequence data presented in this article have been submitted to the EMBL/GenBank Data Libraries under the accession number L06178.

high AT content would then be eroded. In agreement with this expectation WOLSTENHOLME and CLARY (1985) reported an excess of transversions over transitions in comparing *D. yakuba* and *Drosophila melanogaster*, but SATTA, ISHIWA and CHIGUSA (1987) and DESALLE *et al.* (1987) found an excess of transitions in some comparisons between (and within) more closely related *Drosophila* species.

The five honeybee protein gene sequences reported previously indicate that a much higher number of substitutions occurred in the honeybee lineage than in that for *Drosophila* since their divergence from their common ancestor. Whether this degree of change resulted from a higher instantaneous rate of evolution, or systematic pressures resulting from an extreme AT bias in base composition, could not be resolved.

We here report the complete sequence of a honeybee mtDNA, and extend the evolutionary and molecular insights begun previously.

MATERIALS AND METHODS

mtDNA was prepared from *Apis mellifera ligustica* worker bees from a single hive as described previously (CROZIER, CROZIER and MACKINLAY 1989), and restriction mapped with single and double digests and by comparison with published data (SMITH and BROWN 1988). *Eco*RI, *Bcl*I and *Bgl*II restriction fragments of honeybee mtDNA were cloned into pUC8 or pUC18. In addition, mtDNA was cut with *Acc*I, end-filled using the Klenow fragment of DNA polymerase I and was then cut with either *Bcl*I or *Eco*RI for cloning into pUC18. Recombinant plasmids were identified by hybridization with labeled honeybee mtDNA and by sizing of insert DNA. Clones were obtained covering all of the honeybee mtDNA except for the smallest *Acc*I fragment (527 bp).

For sequencing by the dideoxy chain termination method of SANGER *et al.* (1980), the pUC clones were subcloned into M13mp8, M13mp18 and M13mp19. Where overlapping clones were not available, sequencing across restriction sites was achieved through direct sequencing of polymerase chain reaction (PCR) fragments spanning those sites [KOULIANOS and CROZIER (1991), modified]. Synthetic oligonucleotide primers based on sequence already obtained were used for sequence extension and PCR amplification. The sequencing strategy is shown in Figure 1.

The MacVector (International Biotechnologies) package was used for data entry and sequence analysis. Also used in analysis were "DottyPlot for Mac" (GILBERT 1989), "MultiDNA" (R. GONZALEZ), and programs available from the authors. Sequences were aligned using the programs cited and those alignments maximizing sequence similarity were used for further analysis.

Stepwise multiple regression analysis was carried out according to standard principles (SOKAL and ROHLF 1981) using Statview II (Abacus Concepts) on a Macintosh SE/30 computer. Except where stated otherwise, all statistical tests used the 5% significance level.

RESULTS AND DISCUSSION

Genome organization: Genes were identified from structural considerations and by similarity to the

mtDNA genes of other organisms. The honeybee mtDNA gene map is given in Figure 2 and the complete sequence in Figure 3, both drawn so as to facilitate comparison with that of *D. yakuba* (CLARY and WOLSTENHOLME 1985). The protein-encoding genes, the ribosomal RNA genes, and the presumptive control region are all at the same positions relative to each other as in the *Drosophila* map, but 11 of the 22 tRNA genes are in altered positions. All the tRNA genes with positions different to those they have in the *Drosophila* map are located in the upper left quadrant of the honeybee map, between the A + T-rich region and the ATPase8 gene.

The *A. mellifera* mitochondrial genome is slightly longer (16,343 bp) than that of *D. yakuba* (16,019 bp), has a shorter control region (826 cf. 1,077 bp), and tends to have longer intergenic, noncoding sequences than seen in *Drosophila* mtDNA. The longest such noncoding sequence, apart from the A + T-rich region, is between the tRNA^{Leu}_{UUR} and COII genes (CROZIER, CROZIER and MACKINLAY 1989) (Figures 2 and 3) and comprises 193 nucleotides. This region occurs as longer variants in some other honeybee strains, and in these cases sequence analysis supports an origin involving duplications of the tRNA^{Leu}_{UUR} gene and the 3' end of the COI gene (CORNUET, GARNERY and SOLIGNAC 1991). Apart from the A + T-rich region and the noncoding sequence between the tRNA^{Leu}_{UUR} and COII genes, there are 618 noncoding nucleotides in the sequence described, resulting from intergenic gaps ranging from one to 95 nucleotides long (Figure 2). This figure is much larger than that for *D. yakuba* of 183 (CLARY and WOLSTENHOLME 1985).

The ND2 and tRNA^{Cys} genes overlap, as do those for ATPase8 and ATPase6. Only one of the three cases of overlaps between mitochondrial genes reported for *D. yakuba* (CLARY and WOLSTENHOLME 1985) also occurs in *A. mellifera* (that between the ATPase subunit genes). The COI and COII genes would also be involved in overlaps (with the tRNA^{Leu}_{UUR} and tRNA^{Asp} genes, respectively) if terminated with TAA codons. However, OJALA, MONTOYA and ATTARDI (1981) sequenced a number of mRNAs produced from HeLa cell mtDNA and found that these were polyadenylated at the 3' ends, leading to completion of otherwise incomplete termination codons, and suggested that precise excision of mitochondrial tRNAs occurs from a polycistronic transcript. WOLSTENHOLME (1992) notes that incomplete termination codons, such as those proposed here for the COI and COII genes, are a general feature of those animal mtDNAs which have been completely sequenced, except that of the cnidarian *Metridium senile*, and points out that this indicates an early origin of the cleavage-polyadenylation mechanism.

Overall, the honeybee mtDNA is 43.2% A, 41.7%

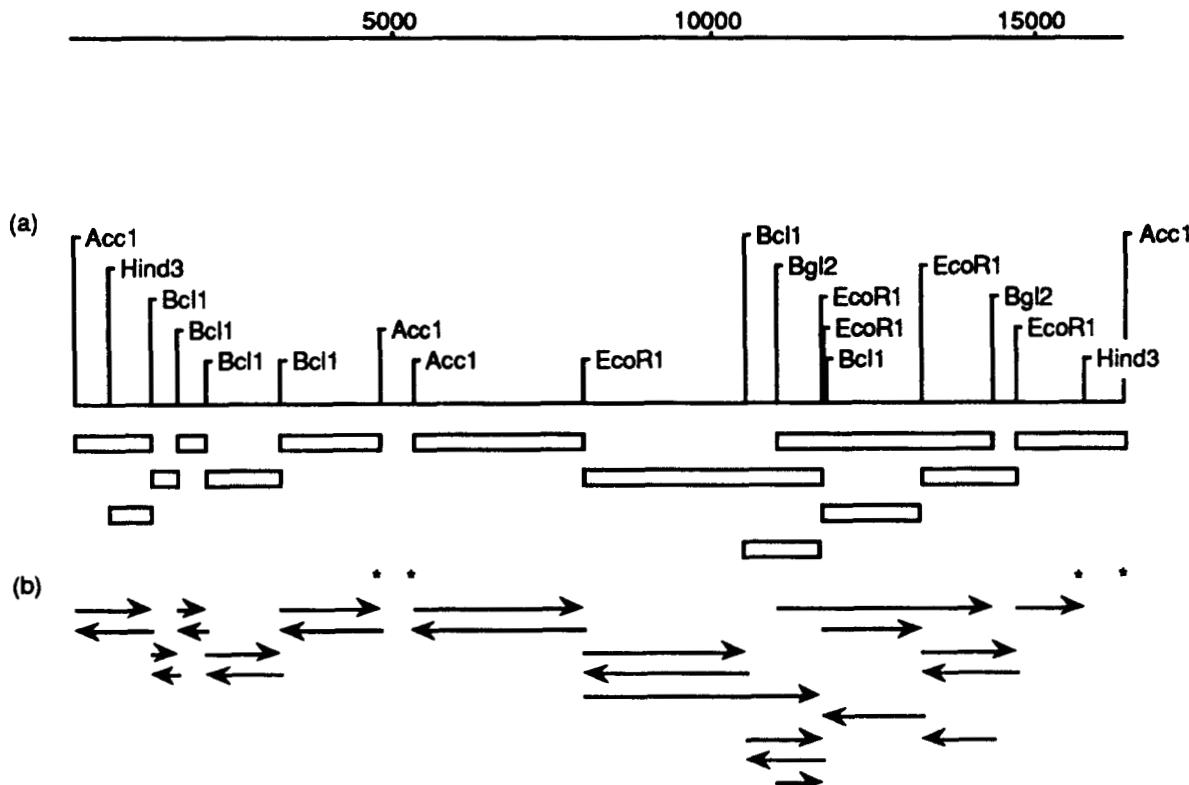


FIGURE 1.—Restriction map of the *A. mellifera ligustica* mitochondrial sequence (a) showing restriction enzymes used in cloning and sequencing strategy (b). Horizontal bars under (a) denote mtDNA fragments cloned into pUC8 or pUC18 plasmids. Arrows in (b) denote lengths and directions of M13 clones used for sequencing. Asterisks show adjacent sites between which sequence was obtained by direct sequencing of PCR fragments overlapping these sites. PCR sequencing was also used to confirm sequence across restriction sites for which there were abutting but not overlapping clones.

T, 5.5% G, and 9.6% C. The genome is thus particularly AT-biased, with 84.9% of the nucleotides being either A or T. In having guanine as the rarest nucleotide, honeybee mtDNA resembles that of the echinoderm *Paracentrotus lividus* and not that of the other nonvertebrate genomes which have been completely sequenced, namely the fly *D. yakuba*, the nematodes *Ascaris suum* and *Caenorhabditis elegans*, and the cnidarian *M. senile*, in all of which cytosine is the rarest nucleotide (WOLSTENHOLME 1992).

Genetic code: Use of the *Drosophila* genetic code to infer the translations of *A. mellifera* mtDNA genes (reported below) yields results consistent with this code also being applicable to the honeybee. However, as discussed below, two honeybee tRNAs appear to use different anticodons to their *Drosophila* counterparts.

Initiation and termination: Based on the translations using the *Drosophila* mitochondrial code, all initiation codons in honeybee mtDNA protein-encoding genes are either methionine (three ATG, three ATA) or isoleucine (one ATC, six ATT). No anomalous initiation codons, such as ATAA reported for *D. yakuba* (CLARY and WOLSTENHOLME 1985), are apparent. All stop codons shown in Figure 3 are TAA, save that the COI and COII genes may each terminate

in single T by analogy with findings from mammalian mitochondrial systems. Five of the stop codons of *D. yakuba* mtDNA are reported to be incomplete, either T or TA (CLARY and WOLSTENHOLME 1985).

The A + T-rich region, which is 96.0% AT, lacks any apparent signals for the initiation of replication such as those of vertebrates (DESJARDINS and MORAIS 1990; SACCOME, PESOLE and SBISÁ 1991). CORNUET, GARNERY and SOLIGNAC (1991) argue that the duplicate region between the tRNA^{Leu}_{UUR} and COII genes may contain an additional, or replacement, origin of replication, on the basis of the likely effects on DNA helix stability of variation in nucleotide content. This region is 92.2% AT which, while representing a high bias, is less than that of the A + T-rich region.

It is unclear that any replicative activities of the gap between the tRNA^{Leu}_{UUR} and COII genes represent essential functions of the region, because in other bees the region is often reduced and sometimes absent (CORNUET, GARNERY and SOLIGNAC 1991). In addition, gaps of the same size as in the sequence in Figure 2 are seen in some attine ants but in others the region is reduced or absent (J. WETTERER, personal communication), indicating that region may be prone to repeated duplicative events during hymenopteran evolution.

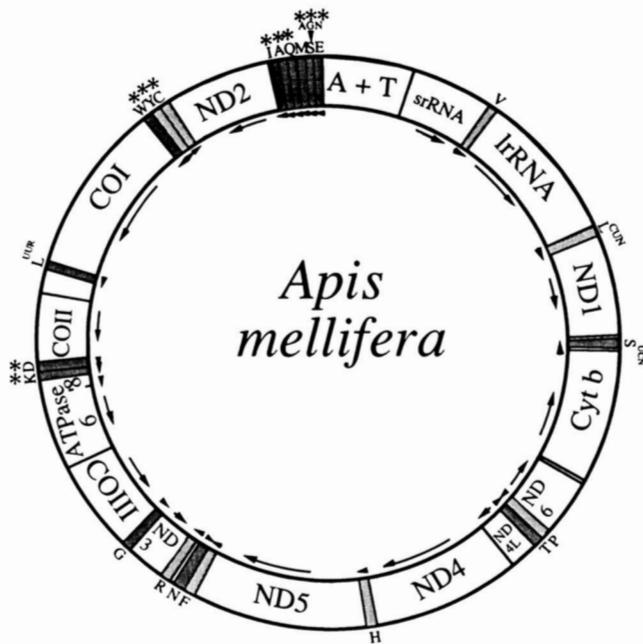


FIGURE 2.—Map of the circular genome of the honeybee *A. mellifera*. Genes for tRNAs are denoted by the one-letter code for their corresponding amino acids; tRNA genes bearing asterisks are in different relative positions to their counterparts in the mitochondrial genome of *D. yakuba*. Protein-encoding genes are denoted COI, COII, COIII for the genes encoding subunits one, two and three of cytochrome *c* oxidase, Cyt *b* for the cytochrome *b* gene, and ND1–6 and ND4L for the genes encoding subunits 1–6 and 4L of the NADH dehydrogenase system. The AT-rich region believed to contain the origin of replication on the basis of the organization of *Drosophila* mitochondrial DNA (GODDARD and WOLSTENHOLME 1978, 1980) is denoted “A + T.” The direction of transcription for each coding region is shown by arrows.

tRNA genes: The 22 tRNA genes found in honeybee mtDNA are shown in Figure 4, folded into the configurations inferred for the corresponding tRNAs. All except two appear to use the same sequence in the anticodon region as reported for *Drosophila* (CLARY and WOLSTENHOLME 1985). The exceptions are tRNA^{Lys}, with TTT as in *Xenopus* (ROE *et al.* 1985), *Gallus* (DESJARDINS and MORAIS 1990) and *Caenorhabditis* (OKIMOTO *et al.* 1992) as against CTT in *Drosophila*, and tRNA^{Ser}_{AGN}, with GCT in *Drosophila*, *Xenopus* and *Gallus* as against TCT in *Caenorhabditis* (OKIMOTO *et al.* 1992) and *Apis*.

The proportions of the four nucleotides in the honeybee tRNA genes (G, 7.2%; A, 45.0%; T, 42.1%; C, 5.7%) yield a higher AT-content (87.1%) than the genome as a whole (84.9%), and than the mean (83.3%) for the protein-encoding genes (see Table 2).

Among insects, mismatched base pairs have been reported in mitochondrial tRNAs inferred for *Drosophila* (CLARY *et al.* 1982; CLARY and WOLSTENHOLME 1983a,b, 1984; DE BRUIJN 1983; CLARY, WAHLTHNER and WOLSTENHOLME 1984), *Aedes* (HSUCHEN and DUBIN 1984), and *Locusta* (UHLENBUSCH, RIPPE and GELLISSSEN 1987). Examples of these are

seen in various of the *Apis* genes according to the configurations of Figure 4, including the inferred formation of a pyrimidine-pyrimidine pair in the tRNA^{Val} gene.

The tRNA genes for tRNA^{Asp}, tRNA^{Leu}, tRNA^{Thr} and tRNA^{Val} all present the unusual feature of having TΨC stems with 6 nucleotide pairs, compared with the 4–5 usual among animal mtDNA tRNA genes (with the exception of the tRNA^{Ser}_{AGN} gene). In addition, the TΨC loop of the inferred tRNA^{Thr} is exceptionally large.

Mitochondrial tRNAs present unusual features in various species, including the lack or abridgement of the TΨC loop in some nematode tRNAs and the likely universal lack of a dihydrouridine arm in tRNA^{Ser}_{AGN} (WOLSTENHOLME 1992). The *Apis* tRNA^{Ser}_{AGN} structure in Figure 4 conforms to this interpretation.

Ribosomal RNA genes: The region of the large ribosomal subunit gene (l-rRNA) is bounded by the tRNA genes tRNA^{Leu} and tRNA^{Val}, the latter also marks one end of the small ribosomal subunit gene (s-rRNA). The boundaries shown in Figure 2 are inferred from similarity comparisons with the *Drosophila* and *Aedes* genes, and from the identification of the 3' end of the honeybee l-rRNA gene by VLASAK, BURGSCHWAIGER and KREIL (1987). The other end of the s-rRNA gene abuts the A + T-rich region; the s-rRNA boundary indicated was selected on the basis of similarity to the *Drosophila* equivalent.

The ribosomal genes as indicated in Figure 2 have lengths of 786 and 1371 bases, compared to their *Drosophila* equivalents of 789 and 1326 (CLARY and WOLSTENHOLME, 1985). The base composition of the s-rRNA gene is 41.9% A, 39.5% T, 12.7% G and 5.8% C, which is slightly more AT-rich (81.45%) than that for *Drosophila* (79.3%), but less AT-rich than the honeybee mitochondrial genome as a whole. The base composition for the l-rRNA gene is 40.4% A, 44.9% T, 10.7% G, and 5.0% C, which parallels the *Drosophila* case in being more AT-rich (84.3%) than the s-rRNA gene. The honeybee l-rRNA gene is also more AT-rich than its *Drosophila* equivalent (83.3%).

Alignment of the two genes with their *Drosophila* equivalents (Figure 5) yields similarities of 68.3% (s-rRNA) and 71.5% (l-rRNA). This similarity between the l-rRNA genes is slightly lower than can be derived by comparing the honeybee sequence with that reported for *Aedes* by HSUCHEN, KOTIN and DUBIN (1984).

Protein-encoding genes: The numbers of codons used in the 13 protein-encoding genes of *A. mellifera* and *D. yakuba* mtDNAs are compared in Table 1. Codon bias in the honeybee is more extreme than is the case in *Drosophila*, which is in turn more extreme than in vertebrates. For example, TTT and TTC occur 125 and 103 times, respectively, among *Xeno-*

-----tRNA_{Glu}-----
 ATTTATAGTTAAAAACATTATTTCAATATAAAATAATTAAATTATAAATAATTAAAGTCAAATTAAATTAAATCTAAAATTATTTATAAATAAGAAA
 10 20 30 40 50 60 70 80 90 100 110 120
 HindIII XbaI Hinfl
 -----tRNA_{Ser}_{AGN}-----
 TATAAATAAAAGCTTCAACTTAACCTAGATTCAAATAATCTATTTCTATTATAAATTAAATTAGATAAAATTAAATTATAATAAGCTAAATAAA
 130 140 150 160 170 180 190 200 210 220 230 240
 -----tRNA_{Gln}-----
 GCTAACAGGTTCAACCCGTGCGATAAAATAATTAAATTATAATTAAATTAAATTAGTTAGCTTAAAGCACATAAAATTGAAATTTATAGTAACTAAATTAAATT
 250 260 270 280 290 300 310 320 330 340 350 360
 -----tRNA_{Ala}-----
 GGATATTAGTTAATAAAATAAACATTAATTCGATTTAAATTAAATTATTTATATCTAAAAACTAATATGCTGATAAAAGAAATTGGATAAAATTATAATGATA
 370 380 390 400 410 420 430 440 450 460 470 480
 ND2 =>
 -----* I F F M N F K Y H W F I Y F L I T I F V L M M N S N N I F I Q W M
 ATTTATATACTATTACTTCTTCATAAAATTAAATACCCTGATTATTATTAAATTACTATTGGTATTAAATAATAATTCCAATAATTTTATTCAATGAATA
 490 500 510 520 530 540 550 560 570 580 590 600
 L M E F G T I I S I S L I N I K S T N K T P S L I Y Y S V S V I S S I F L F F M
 TTAATAGAATTGGTACAATCATTAGAATTAGATAATTAAATACCCTGATTATTATTAAATTACTATTGGTATTAAATAATAATTCCAATAATTTTATTCAATGAATA
 610 620 630 640 650 660 670 680 690 700 710 720
 AccI
 I I V Y L S S I S F T K T D T F N F M V Q M M F F L K I G T F P F H F W M I Y S
 ATTATTGATACTTATCATCCATTAGATTTACTAAACAGATACTTTAATTAGTCAATAATATTTTTAAATTGGAACTTCCCCTTCATTGGATAATAATTCT
 730 740 750 760 770 780 790 800 810 820 830 840
 Y E M M N W K Q I F L M S T L I K F I P I Y M M V S M T K I N S W T L Y F L I T
 TATGAAATAAAATTGAAAGCAAATTTTAAATATCAACATTAAATTAAATTCTCAATTATAATAGTTCAATAACTAAATTACATGAACATTATTTTAAATTACA
 850 860 870 880 890 900 910 920 930 940 950 960
 -----Hinfl-----
 F I A M I I L Y S F N Y F L L I S F L N K F N I Q N F N F M F Y N K Y Q M Y T F
 TTTATTGCTATAATTATTTATTCATTAAATTATTTTATTAAATTAGTTCAAAATTAAATTCAAAATTAAATTCAAAATTAAATTCAAAATTAAATTCAAAATTACATT
 1090 1100 1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
 -----HindIII-----
 L T L M F N Y S M Y P I F L S F V I K W N L I F M M M V S V K A Y N W I L F L L M
 TTAACATTAATTAAATTCAATATCCAATTCTCTTCATTGTAATTAAATGAAATCTAATTAAATTAGTAAGGTTAAAGCTTATAATTGAATTATTCTTAAATA
 1210 1220 1230 1240 1250 1260 1270 1280 1290 1300 1310 1320
 XbaI
 I S S M L M I W N Y I I I L K R V F L K M N F Y K N N F I D D K D N K Y M Y H S
 ATTCTAGAAATTAAATTGAAATTATTTAAATTCTGATGATAAGATAATAATTATCATAGA
 1330 1340 1350 1360 1370 1380 1390 1400 1410 1420 1430 1440
 -----*-----
 Y F A L T L L S F N I S F F I T L N F L *
 TATTTGCACCTACCTCTTCATTAAATTTCATTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAA
 1450 1460 1470 1480 1490 1500 1510 1520 1530 1540 1550 1560
 tRNA_{Cys}-----*-----<-tRNA_{Tyr}-----*
 ACTATAATAATTAAACATGATAATAATAGTATTAGAAAAATTAAATTCAATTAAATTCAATTAAACTCTTATTAGATAACTAAATAAGATTAAAAACCT
 1570 1580 1590 1600 1610 1620 1630 1640 1650 1660 1670 1680
 COI =>
 -----tRNA_{Trp}-----* M M
 TTTTATTTTTATTTTTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAA
 1690 1700 1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
 -----Hinfl-----
 K W F M S T N H K N I G I L Y I I L A L W S G M L G S S M S L I I R M E L S S P
 AGTGATTCAATCAACCAATCATAAAAATTGGATCTGTATATTCTAGCTTATGATCTGAATAGGATCATCAATGAGACTTATTTCGAATAGAATTAGATCCCCAG
 1810 1820 1830 1840 1850 1860 1870 1880 1890 1900 1910 1920
 BcII
 G S W I S N D Q I Y N T I V T S H A F L M I F F M V M P F L I G G F G N W L I P
 GATCATGAATTAGCAATGATAAAATTATAACATTGTTACTAGTCATGCATTCTAAATTGGGATTTAGTTACATCATTGGGAGATTGGAAATTGGCTTATTCCCT
 1930 1940 1950 1960 1970 1980 1990 2000 2010 2020 2030 2040
 L M L G S P D M A F P R M N N I S F W L L P P S L F M L L L S N L F Y P S P G T
 TAATACTAGGATCACCTGATATAGCATTCCCCGAAATAATAATTAGATTGATTACTCTCCCTCATTTTATCTTAAAGAAATTATTTATCCAAAGACCGAGAACTG
 2050 2060 2070 2080 2090 2100 2110 2120 2130 2140 2150 2160
 G W T V Y P P L S A Y L H S S P S V D F A I F S L H M S G I S S I M G S L N L
 GATGAACAGTATCCACCAATTATCAGCATATTATCATTCTCACCTCAGTAGATTGCAATTCTCTCATATCAGGAATTCTCAATTAGGATCATTAAACTTAA
 2170 2180 2190 2200 2210 2220 2230 2240 2250 2260 2270 2280
 BcII
 M V T I M M M K N F S M N Y D Q I S L F P W S V F I T A I L L I M S L P V L A G
 TAGTTACAATTATAATAATAAAATTCTATAAAATTATGACCAATTTCATGATCAGTTTTATTACAGCAATTAAATTATCATACCTGTATTAGCTGGAG
 2290 2300 2310 2320 2330 2340 2350 2360 2370 2380 2390 2400
 A I T M L L F D R N F N T S F F D P M G G G D P I L Y Q H L F W F F G H P E V Y
 CAATTACTACTATTATGATCGAAATTAAATCATCATTCTGATAGGAGGGAGATCCAATTCTTATCAACATTATTTGATTTGGTCACTCCAGAAAGTTATA
 2410 2420 2430 2440 2450 2460 2470 2480 2490 2500 2510 2520

FIGURE 3.—The sequence of the *A. mellifera ligustica* mitochondrial genome. Genes are read from left to right except for the genes for NADH reductase subunits 1, 4, 4L, and 5, the two ribosomal RNA genes, and the tRNA genes for valine, proline, phenylalanine, arginine, tyrosine, and cysteine, as indicated.

Fig. 3 part 2.

I M S F Y I M K S F F F M E I I L D K S Y Y G V L F P F G C L S L I S F I L I M
ATTATTGAAAAATAATTATTTCTAAAAAAATTTCAATAATTAAATCTTGTAAATAACCAACTAAAAAGGAAACCATATACTTAAATTGAAAAATTAAATTAT
7450 7460 7470 7480 7490 7500 7510 7520 7530 7540 7550 7560

S K M P Y I Y Y M G Y Y M R I D Q N S Y M Y H M Y S G V C M F M L S K F M A H I
TCTTTTATTGGATAATAATATAACCATAATATTCGAATATCTGATTACTATATATAATGTATATACTACCAACACATATAACATTAATGATTAAATTGATGAA
7570 7580 7590 7600 7610 7620 7630 7640 7650 7660 7670 7680

F L H L F V L E T S G I S L M S M M F G L Q S L T S Y A V V K K L D L E F N A V
AAATAATGAAAAATACTAATTCACTGGAAACCAATTGATAATATTCTTATAATCCTAATTGCTTAAAGTAGAAATAGCAACACTTTTTAACTAATTCAAAATTGCAAC
7690 7700 7710 7720 7730 7740 7750 7760 7770 7780 7790 7800

L G A F L M T L S A I L M I Y N K Y N F D L L N V Y R I L L Y I G A T V L T S S
TAAACAGCAAATAATTGTTAACATCAATATAATTAAATTAAATTAAATTCTAATAATTACATCGAATTAAATAAAATTCCAGCAGTAACAAATTGATGA
7810 7820 7830 7840 7850 7860 7870 7880 7890 7900 7910 7920

H V L S S V P T P A M M A M P L W T S F P I Q A S K T F A M L L I Y I M M F E N
ATGAACTAAAGATGAAACAGGAGTGGAGCTATTGCTATTGCTAAGTGAAAGGAAATTGCTCTTTACTAAAGCTATCAATAAAATATAATCATTAAATTCAATT
7930 7940 7950 7960 7970 7980 7990 8000 8010 8020 8030 8040

ClaI

M K Y F S L N W S G Y Y T M L G M I L L L G I D G L R N L L I T V M G S T F S K
TATTTATAAAACTTAAATTCTCCATAATATGTTATTACCCATAATTAAATAACCAATATCCTAATGTTAAATAATGAACTATACCTGAAGTAATGATT
8050 8060 8070 8080 8090 8100 8110 8120 8130 8140 8150 8160

M K M Y Y I V L C Y S I L G L G D W G L I I S L M N P S L I L M Y M S I L F L I
TATTTATATAAAATTACAAGACAATAAGAAATTCTAACCTCTAACCTAAATTGATAATATTTGCTTAAATAACATATAATAGAAAATTAAATTAAAT
8170 8180 8190 8200 8210 8220 8230 8240 8250 8260 8270 8280

M L Y L F R D M K L E S L D M Y S I S Y I I I M S F I M S V L F I F M L S K Y D
TATTAATATAAAACGATCCATTTCATTCACTAAATCTATATCTAAATTCTATAAAATAATATAGAAAATTATTCTAACTAAAAATAATTAATGTTATAATC
8290 8300 8310 8320 8330 8340 8350 8360 8370 8380 8390 8400

I L L L F N F K M S N F T Y I N W E F F F E K N L Y L L Y L S M L M M L F S F E
AATTAATAACAAAAAAACTTATTGAAATTGATAATATTCTATTCAAAATTCCATTAAATAATCTTAAATTAAATAATATAATCTTAACTTAAATTAAACCTAAATC
8410 8420 8430 8440 8450 8460 8470 8480 8490 8500 8510 8520

← start ND5*-----<-tRNA^{His}-----*

F L L I G C V M M K I I
AAATAATAAAATTCCACAAACTATTATTTAATAATGATTTAAATTATTCATTGATCCCAGAAATCAATATTAACTTAACTTAAATCTAATTAAACTTT
8530 8540 8550 8560 8570 8580 8590 8600 8610 8620 8630 8640

end ND4

EcoRI

* I F Y L K L F M L N L P I W H L L L V F Y E V L I G N K I K F M I F I K G H
AATTTAAATAAAATTAAAGTTTAAACATTTAAAGGAAATTCAATGTAATAATAACAAAATTCAACTAAATTCCATTTTAAACATAATAAAATTCTCATG
8650 8660 8670 8680 8690 8700 8710 8720 8730 8740 8750 8760

N I F M F L Y I S Y I F S F L C Y M M L I L M M F K L W S I M G I L L M V E S I
ATTAATAATAATAATAAAATTGAAATAAAATCTAAATAACATATTAAATTAAATTAAATTAAATTAAACACAGAAATTATCCAAATAATAACCTCGCTAAT
8770 8780 8790 8800 8810 8820 8830 8840 8850 8860 8870 8880

L N L S V P S G M N S S C L M F W L L S M S P M F N I M G K N I F M L R S N T Q
TAAATTAAAGAAACTGGAGATCTATATTGATGAACATAACATAATGATATTGAAGGTATAAAATTAAATTACCTTATTAAATAATTAATCGTCTATTGTTG
8890 8900 8910 8920 8930 8940 8950 8960 8970 8980 8990 9000

S Y I V N V L F F L G S S S L G H S I M M L Y G G I L S I K L F T M M S M I M L
TCTATAATTACATTAACTAAAAACCTGAAGAATTAAACCATGAGAAATTCTAAATGATATTCTAAATCTAAATTCTTAAATTCTTAAATTCTTAAATTCTTAA
9010 9020 9030 9040 9050 9060 9070 9080 9090 9100 9110 9120

G M H V I S S I A I I S K M D F Q S L C M L S L I L V G F S N I M V L I K Q I L
TCCTATGAAACAATAGATGAAATTGCAATAATTGATTTATCTAAATTGTGATAACATATTCTAAATTAAACCCCAAAGAATTATTACCAAAATTGAAATTAA
9130 9140 9150 9160 9170 9180 9190 9200 9210 9220 9230 9240

I F E N K Y I I M L R L M G Y G G L K L M I S A L I M S G Y Y P A E V H A K L L
AATAAAATTCTTTATAATAATTATCAATGTAATTCCATCTCTAAATTAAATTGAAAGCTAAATTGATCCATAAGGAGCTCAACATGAGCTTTAA
9250 9260 9270 9280 9290 9300 9310 9320 9330 9340 9350 9360

W G H F L Y I P I K V L F S M L L Y I F L M M N L N L N L M E M L M F N L S Y D
CCAACCATGAAATAATAAAATTAGGAATTAAACCAAAATTGATATTAAATAATAATTATTTAAATTAAATTCTCTATTAAATAATAAAATTCTATAATC
9370 9380 9390 9400 9410 9420 9430 9440 9450 9460 9470 9480

ClaI

I L Y I Y Y I I Y L M P L S F I M T Y F M L Y F G S L W R N E S Y G W K V V L Y
AATTAATAATAATAATAATAATTGTTAAAGGAAATTCTGATAATGAAATTAAATAACATCGATTTCACTATAACCTCATTTACAACTAAATA
9490 9500 9510 9520 9530 9540 9550 9560 9570 9580 9590 9600

F I L L L G F E Y F L Y F L L L N M S L F V L L L S I M L L L N M F L C N L S N
AAAAATAATAACAAACCAAAATTCAATAATAATAATAATAATTATAGATAAAAAACTATAATAAAAGAATTATAATAATAACAAATTCTATT
9610 9620 9630 9640 9650 9660 9670 9680 9690 9700 9710 9720

NdeI

N N L S I F I L G F I W L T L M I L G Y S Y M N F S L N C F I Y I W D I W N L N
ATTATTAAAGAAATAAAATTAAATCATAAAATTCTAAATGAAATTCTAAATCTAAATTAAATAACAAATAATTGATATTAAACATATTCA
9730 9740 9750 9760 9770 9780 9790 9800 9810 9820 9830 9840

F L N L L L N I I I L N G I I L N L N N K M K N M F L M F L Y I M S M M L M N L
AAATAATTAAATAATAATAATTAAATTCTAAATGAAATTCTAAATCTAAATTAAATAACAAATAATTGATATTAAACATATTCA
9850 9860 9870 9880 9890 9900 9910 9920 9930 9940 9950 9960

Fig. 3 part 4.

EcoRI EcoRI BclI

L M E T W C M N M L K D Y R I R P L I G R I W I I L C I H F L Y I L I F K I S W											
TTAATATTCAGTTCAACATATATTATAATTCATATCGAATTGAGGTAAATACCTCGAATTCAAAATTAATTAACAAATATGAATAATAATTAAAATATTAAATTAATTCATGC											
12370 12380 12390 12400 12410 12420 12430 12440 12450 12460 12470 12480											
Y K F G Y F M L S L I V S M F M I N M Y E S L F I L V F M S S H Y E I N F G S V											
ATATTTAAATCCAAAAATATACTTAAATATACTCTTAAATATAATTTATATTCAGATAAAAATTAATACAAATATTCTCTATGATATTCAATATTAAACAGATA											
12490 12500 12510 12520 12530 12540 12550 12560 12570 12580 12590 12600											
HinfI											
L E S E G E I L D F P T R N L E I L M S T F M M M L Y L P Y L L I A F K I N N Q F											
CTAATTCAAGATTCTCCTCAATTAAATCAAATGGACTCGATTAAATCAATTAAATTCAGTAAACATCATCAAAATATAATGGATAACATAAAATTGCAAAATTAAATATTCTGAA											
12610 12620 12630 12640 12650 12660 12670 12680 12690 12700 12710 12720											
F F F E N F S F S E V M M M L S F V L F F L N I E F S I M T S V L R M S G L I A											
AAAAAAAATTCATTAAATGAAAAACTCTCAACTATTATACTAAATACTAAAAAAATRAATAATTCAAAATGAAATTATTGTTGATACTAACCGTATTGATCCTAAATAG											
12730 12740 12750 12760 12770 12780 12790 12800 12810 12820 12830 12840											
Y N C N S I W G V F L V P Y V S L G L V L L M F L I S F E I Y Y M F G F W P Y L											
CATAATTACAATTGAAATTCATCCAAACAAATAAACAGGATAAACTCTTAAACCTAAACATAACAAACAAATACTAAATCAATTAAATATAATATAATCCAAATCAAGGATAACA											
12850 12860 12870 12880 12890 12900 12910 12920 12930 12940 12950 12960											
I W M V L S L F F M L M P S Y I F L N S Y N F F F W E K S L L K L A D S F P Q F											
AAATTCATATAACTAAATGAAAAAAATATAATAGGTCTATAATAATAATTGAAATAATTAAAAAAATCTATTCTTAGATAATAACTTAAATGCATCACTAAAGGTTGAA											
12970 12980 12990 13000 13010 13020 13030 13040 13050 13060 13070 13080											
M G F L M I K N P G K R D Q I Y G L I K R E L L T L F A V S I L V M I M L I L L											
ATATACCAACAAATATAATTTTATTAGGACCTTACGATCTGAATAACCTAAATTCTGTTCTATAAAAGTTAAATGCTACTCTAAATAACTATGATTATCAAATCAATA											
13090 13100 13110 13120 13130 13140 13150 13160 13170 13180 13190 13200											
= start ND1											
N I L V W I *-----<-tRNA ^{Leu} _{CUN} -----*											
AATTAATTAATCACTAAATTATTATTTAAAGAAATTAAATAATTAAATCCTAAATTAAATGCACTATTATGCTAAATAACTATATAATATTTATTAAATTATAAA											
13210 13220 13230 13240 13250 13260 13270 13280 13290 13300 13310 13320											
<hr/>											
TAATTTAAAAATTAAAGTCCTTTCGTACAATTAAATTAAATTAGATAGAACCAATCTGACTTACGTCGATTGAACTCAAATCATGTAAGATTAAAGTCGAACAGACT											
13330 13340 13350 13360 13370 13380 13390 13400 13410 13420 13430 13440											
<hr/>											
AAAAACTTAAACTACTGCGCTTAATTCTCATTAATTCAACATCGAGGTCGCAAACATCTTATCTATGATCTATCAAAGATATTACGCTGTTACCTAAAGTAATTCTTT											
13450 13460 13470 13480 13490 13500 13510 13520 13530 13540 13550 13560											
<hr/>											
TAATTACAATTATAATTCAAAAAATTATCTTATATCAAAATTAAATCTAAATAAGTTATTAAATTACCAATCCCAATCAAATTAAATCTAAATTATATAATTAAATTAT											
13570 13580 13590 13600 13610 13620 13630 13640 13650 13660 13670 13680											
<hr/>											
AAATAAATTAAAAATTAAATTAAATTCTATAGGGCTTACGTCCTCATTAATTAAATTAGAATTAAACTAAATTAACTTAAATTAATTAGACAGTTATTCTTCAATTAA											
13690 13700 13710 13720 13730 13740 13750 13760 13770 13780 13790 13800											
PstI											
<hr/>											
ATTCTTCATACAATTCTCAATTAAAGACAATTATTGCTACCTTGACAGTCACATCTGAGCTATTAAATTCAATTGAGCAGATCGACCTAAATTATAATCAACAG											
13810 13820 13830 13840 13850 13860 13870 13880 13890 13900 13910 13920											
EcoRI											
<hr/>											
GACATGTTTGTAAACAGGTGAATAATATTTGCGAATTCTTAAATTATAACAAATATCATTATAAAACTTTACTAATCTAATCACTATTCTATATTAA											
13930 13940 13950 13960 13970 13980 13990 14000 14010 14020 14030 14040											
<hr/>											
TTAATTAAATATAATTATAGAAAAATAAAATAAATTAAATCATTATAATTAAATTAAATTAAATTAAAGAAAAATTATCTATAACATTAA											
14050 14060 14070 14080 14090 14100 14110 14120 14130 14140 14150 14160											
<hr/>											
AAATTAAATTAATTAAATTCTTAAATTATAGATTATCCCATAATTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAA											
14170 14180 14190 14200 14210 14220 14230 14240 14250 14260 14270 14280											
<hr/>											
TTCTAAAAAAACTAGATATAAAAGTCGTTAACATTAAATTCTAAATTCTATATTATAATTGCTACAAAAAAATAATATAATTAGCTCCCTTATTCGAGATATT											
14290 14300 14310 14320 14330 14340 14350 14360 14370 14380 14390 14400											
<hr/>											
AAATCATTAAATAATTAAATCAACCTGATACAAAGGTACAAAAATTATTCTACTTTTCACATTAAATTTCATTACATTCTAAACTATAAAATTAAATTCTAA											
14410 14420 14430 14440 14450 14460 14470 14480 14490 14500 14510 14520											
<hr/>											
ATAATTAACTTAAACCTAAATAAAATCTTATAAAACTTTAAAAAAATAATTATTAAATTATATAATTATATAATTATATAATTATAATTAAACCAAATA											
14530 14540 14550 14560 14570 14580 14590 14600 14610 14620 14630 14640											
<hr/>											
<-1srRNA-----*-----<-tRNA ^{Val} -----*											
TTTTATAATAATTAAATTATAAAATTAAATTCTTAAACTTCTTCACTGTAAGAAATTTACTTAAACTAAATTCTAAATCTAAACACCCCGGTACTCTTAC											
14650 14660 14670 14680 14690 14700 14710 14720 14730 14740 14750 14760											
<hr/>											
TATGTTACGACTTATTCAATTAAAGGGCGATTGTACACTTTAACTATCTCAATTAAAGAAATTACTTTAAATTCTCTTCAATTAA											
14770 14780 14790 14800 14810 14820 14830 14840 14850 14860 14870 14880											

Fig. 3 part 6.

Fig. 3 part 7.

pus codons (ROE *et al.* 1985), but in *Apis* these figures are 355 and 26 and in *Drosophila* 313 and 17. Excluding termination codons, the percentage of codons ending in A or T is 93.8% in *Drosophila* and 95.2% in *Apis*.

Base composition differs markedly between codon positions (Table 2), evidence of the operation of a strong mutational bias in the molding of the current bee mtDNA. At 83.3% AT, the overall composition of the protein encoding genes is slightly less AT-biased than the mitochondrial genome as a whole, but the AT content at the third codon position, at which most (but not all) substitutions will not lead to replacements is 92.1%. The second position is least AT-biased (76.4%), but this varies between genes (see below).

The honeybee protein-encoding genes are significantly more AT-biased than those of *D. yakuba* for each codon position ($\chi^2_1 = 86.9, P < 0.001; 70.5, P < 0.001; 6.6, P < 0.02$; for the first, second and third codon positions, respectively).

The inferred amino acid sequences of the proteins encoded by honeybee mtDNA are given in Figure 6. The protein genes of *Apis* and *Drosophila* are compared in Table 3. The order of similarities between bee and fly genes and proteins (Table 3) is similar but

not identical to that between fly and mouse (CLARY and WOLSTENHOLME 1985). However, because many of the divergences are not significant from each other for each insect (χ^2 tests, not shown), in fact the only significant difference is the relative placement of the ND2 divergence. In several instances, Drosophila proteins (but only marginally that for COIII among the genes) are more similar to those of Mus than they are to those of *Apis* [e.g., the similarity between the Drosophila and Mus cytochrome *b* amino acid sequences is 67% (CLARY and WOLSTENHOLME 1985), whereas between the Drosophila and *Apis* sequences the similarity is 53%].

The extreme base composition and codon biases of honeybee mtDNA genes leads naturally to the question of whether or not these are associated with the amino acid composition of the respective proteins. Multiple stepwise regression was performed using the number of occurrences of amino acids as the Y variable and the following potential X variables pertaining to each codon family: %A, %T, %G, %C, %AT, and size of family (number of codons). The same analyses were carried out for the *Drosophila*, *Mus*, and *Xenopus* mitochondrial genomes. Following previous results with the cytochrome *b* gene (CROZIER and CROZIER

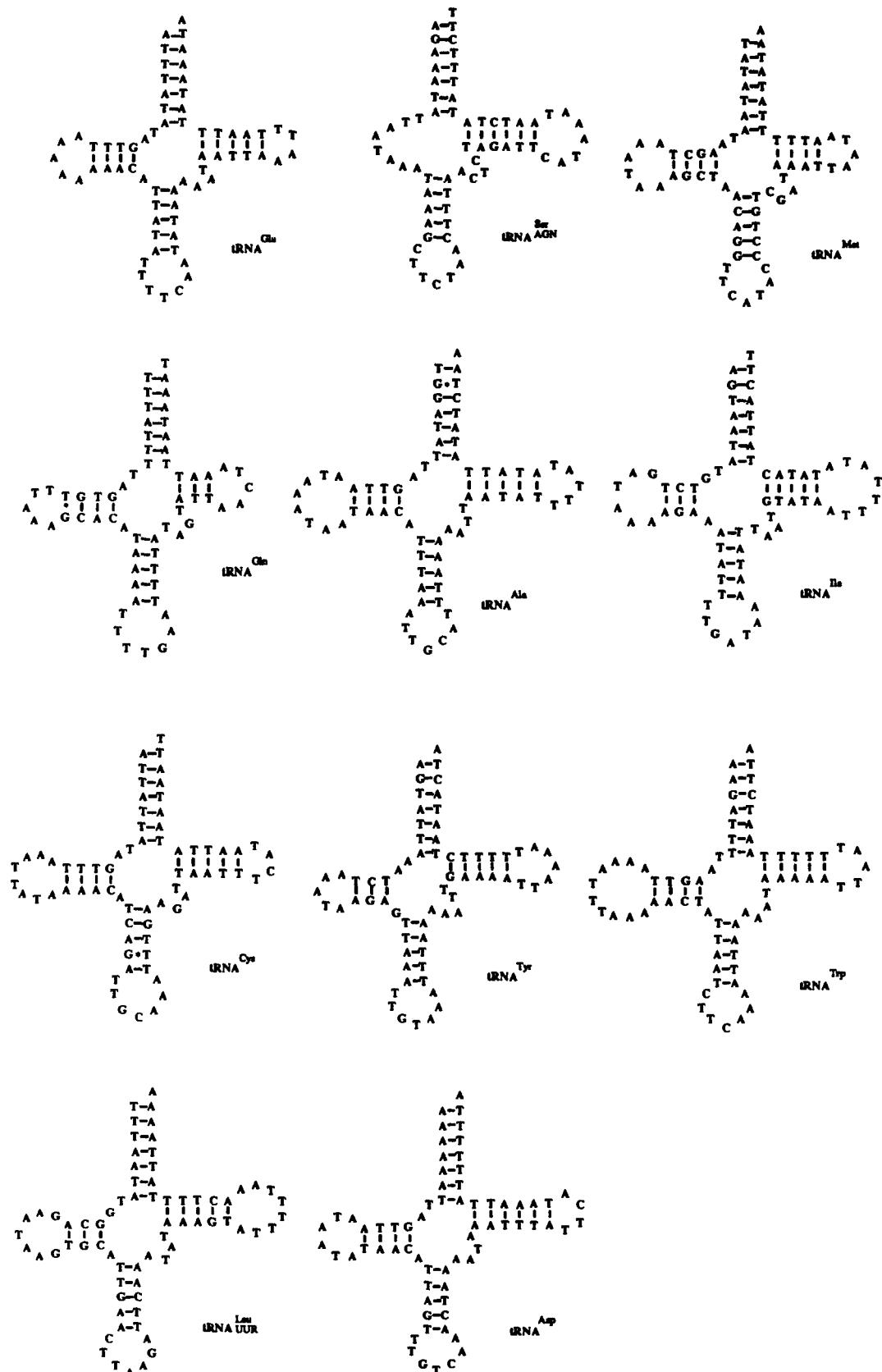


FIGURE 4.—The 22 tRNA genes of honeybee mtDNA, folded into the configurations inferred for the resulting tRNAs.

1992), the size of the codon family was found to be the most important predictor for the fly and the two vertebrates, but to be relatively unimportant for the

bee, for which only the %T in the codon family is a significant predictor of amino acid use. The significant X variables for the four species are given in Table 4.

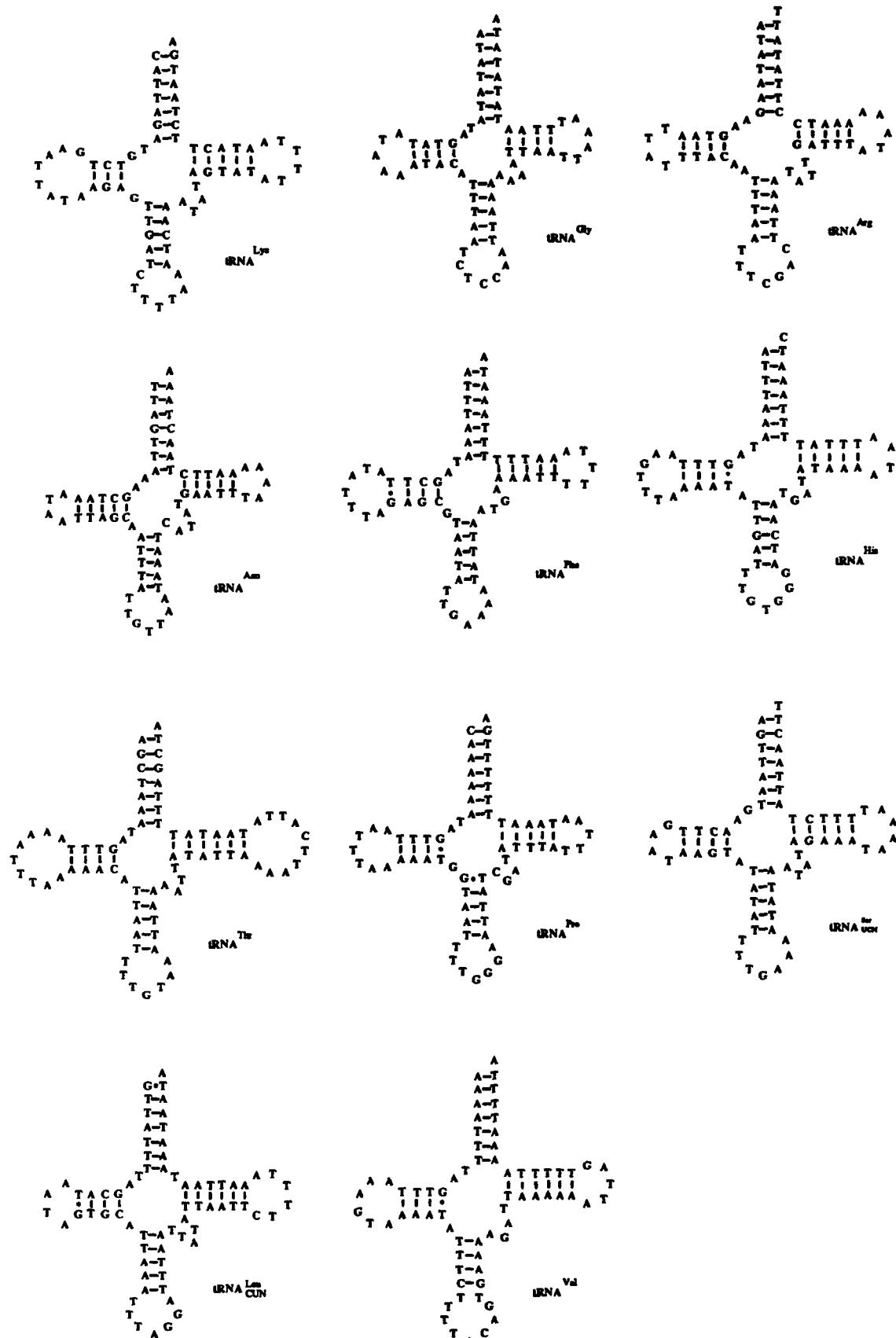


Fig. 4 part 2.

The relationship between amino acid usage and %T in codon family is shown in Figure 7. A strong positive relationship is apparent.

The greater base composition bias in *Apis* than in *Drosophila* raises the possibility that this bias will affect amino acid use in the bee relative to the fly. This was

FIGURE 5.—The two ribosomal RNA genes of *A. mellifera* aligned with their *D. yakuba* counterparts; in each case the bee sequence is the upper one. Hyphens denote gaps inserted to improve the alignment.

TABLE 1
Codon usage and inferred genetic code for the honey bee, *Apis mellifera*, compared to *Drosophila yakuba*

	Apis	Dros									
TTT F	354	313	TCT S	53	120	TAT Y	209	142	TGT C	25	40
TTC F	26	17	TCC S	11	4	TAC Y	10	28	TGC C	0	2
TTA L	472	542	TCA S	166	102	TAA *	11	7	TGA W	78	96
TTG L	24	25	TCG S	1	3	TAG *	0	0	TGG W	5	6
CTT L	35	36	CCT P	36	79	CAT H	57	65	CGT R	9	8
CTC L	1	2	CCC P	3	3	CAC H	3	12	CGC R	0	0
CTA L	36	19	CCA P	64	45	CAA Q	39	70	CGA R	29	45
CTG L	0	2	CCG P	0	3	CAG Q	2	0	CGG R	1	6
ATT I	476	345	ACT T	53	97	AAT N	238	193	AGT S	18	34
ATC I	24	15	ACC T	5	3	AAC N	11	13	AGC S	2	1
ATA M	312	195	ACA T	72	85	AAA K	152	76	AGA S	81	73
ATG M	22	18	ACG T	1	2	AAG K	8	9	AGG S	2	0
GTT V	67	90	GCT A	20	125	GAT D	52	54	GGT G	47	67
GTC V	1	3	GCC A	1	9	GAC D	5	10	GGC G	0	2
GTA V	53	93	GCA A	36	37	GAA E	74	82	GGA G	85	129
GTG V	0	8	GCG A	0	2	GAG E	5	1	GGG G	3	22

The frequency of codon usage is calculated from all codons in the protein-encoding genes (3686 in *Apis*, 3735 in *Drosophila*). Amino acids are indicated using the standard one-letter code. Data for *D. yakuba* are from CLARY and WOLSTENHOLME (1985).

TABLE 2
Base composition of honeybee protein-encoding genes

Codon position	Base			
	A	T	G	C
1	40.05	39.24	12.17	8.54
2	23.77	51.63	10.44	14.16
3	47.75	47.45	2.01	2.79
Overall	37.19	46.10	8.21	8.50

Percentage composition is determined from all codons in the protein-encoding genes of *A. mellifera*.

investigated by using the same multiple regression *X* variables as before as predictors for the quantity abundance in *Apis*/abundance in *Drosophila*. Only %AT was selected as a significantly associated variable ($P < 0.001$). This relationship is shown in Figure 8. The differences in amino acid occurrences between bee and fly are in some cases extreme. For example, the proportion of residues which are lysine in the inferred honeybee proteins is almost twice that of *Drosophila*, whereas the proportion of alanines in honeybee proteins is only about one third that in the fly.

The differences in amino acid composition between honeybee and fly proteins are in two cases associated with a markedly skewed pattern of replacements. Thus, alanines in drosophilid mitochondrial proteins are as often replaced by serines in honeybee proteins (43 occurrences) as they are conserved (42 occurrences), and a similar result holds for valine, which is replaced by isoleucine in *Drosophila* proteins (51 occurrences) almost as often as it is conserved (53 occurrences). In each case the replacement is conserva-

tive according to the criteria of FRENCH and ROBSON (1983) (see Figure 6). The most conserved amino acid overall is arginine: 95% of the occurrences in the honeybee proteins are conserved in those of *Drosophila*, and 63% of those of *Drosophila* are conserved in the honeybee.

The association between base composition of codon family and the occurrence of amino acids in honeybee mitochondrial proteins parallels findings reported elsewhere for other systems. For example, the amino acid composition of mammalian nuclear genes tends to reflect the GC content of the chromosomal segments in which the genes occur (e.g., BERNARDI and BERNARDI 1986; BERNARDI 1989). The effect reported here, of amino acid occurrence varying according to base composition, can also be expressed in terms of the ratio of the total occurrences of "A + T" (F, I, M, Y, N and K in the standard one-letter code) and "G + C" (P, A, R and G) amino acids: for *D. yakuba* the ratio of "GC" to "AT" amino acid occurrences is 0.43 (JUKES and BHUSHAN 1986) whereas in the honeybee it is 0.18.

Concluding discussion: Honeybee genes and proteins have diverged much more from those of the common ancestor of *Apis* and *Drosophila* than have those of this fly (CROZIER, CROZIER and MACKINLAY 1989; CROZIER and CROZIER 1992). Possible explanations for this divergence include the smaller effective population size of honeybees compared to *Drosophila* (CROZIER 1980), poorer mutation repair in bee than in fly mitochondria, and different stationary states due to the differing base compositions (*i.e.*, a role for AT pressure). The population size explana-

ATPase 6		
MKLILMMNLFEMFDPTSNNLSMNWLMMPLIIIFPSIFWLIQSRIMFIMKTLMNFMYNEFK-VVSKSKYQSNIIFISLMLYIMITNIFSILPYVFTLT ---m.T...Sv....aI...l...STF.glmI...y..mP..YNIFWNsillT1HK...T11GP.GHnGSTF.....FSL.lFN.FMG.F..i..S.	99	96
SHLLNMLSLT1LWFSLFLIY-LIYNNYIMFLSHLVPLNSPVFLMNFVIELISLIIRPWTLSIRLSANLISGHILTLGIFISNFISILPINLMIQNM ...T.T1s.a.P..LC.m1.GW.NHTQH..-a....QGt.AI..P..C..T..N....G..av..t..m.a..1....NTGPSMSY1.VTF.lvAqi	198	195
-L LTLEIFMSM IQSYVF SILLI LYFSESN A..V..SAvt.....av.ST..S..V.	226	224
ATPase 8		
IPQMMPMKWF LIYFI YLLI-FYLFI MLI NSMLIKTKINKET-LK-IKLKKWNWFWA.iS.L.1..VFS.T.I..CS..YYSYMPTSP.SNE..N.N.NSM..K.	52	53
Cytochrome b		
MKKFMNFFSSNEFLKMIMSTIYLPTPVNINYMWNFGSILGIFLMIQIISGFILSMHYCPNIDIAFW SITNIMKDMNSGWLFR LIHMNGASFYFLMMYIHI .h.PLRN-.HPL.KIANNaLvD..a..SSW....1..1c.i...lt.LF.a...TAdvnl..y.vNH.Cr.v.Y...L.T1.A....f.iCi.l..	100	99
SRNLFYCSYKLNNVWGIGIMILLMSMAAFMGYVLPWGQMSYWGATVITNL SAIPYIGDTIVLWIWGGFSINNATLNRF FSLHFILPLLILF MVILHLF G.Giy..G..LFTPT.Lv.vi..F1V.Gt.....f.....1.MD1.Q.1....avd....T...tF....Fiv.A.Tmi..L	200	199
ALHTGSSNPLGSNFNNYKISFH PYFSIKD L LGFYI I LFIMF FINQF QPYH LGDPDNFKIANPMNTPTHIKPEWYFLF AYSILRAIPNKLGGVIGLVMSI F..Q...N..I..L.S.ID..P....tF..iv..Iv mi..LiS1VLIS.NL.....IP...1V..a..Q.....a..s.....A..1..	300	299
LILYIMIFYNN-KMMNNKF NMLNKI YYWMPIINN FILLTWLGKOLIEYFP TNINM LFTTYFLYFFLNFYLSKLWDNLIWNSPLN A..M..1P...LS.FRGQI.YPi.Q.Lf.SM1VT.i.AR Pv.E.yVL.GQ1L.II....yLi.PLvt.W.....	383	378
Cytochrome oxidase I		
M-MKWF M STNHKNIGI LYI I ALW SGMLGSSMSL IIRME LSSPGS WISNDQIYNTIVTSHAFLM IFFMVMPFLIGGFGNWL I PLMLGSPDMAFPRMNNIS .SRQ.LF....d..T..F.FGA.a..v.t.l.il..A..GH..aL.Gd....V..a..i.....Im.....v.....a.....m.	99	100
FWLLPPSLF MLLLSNLFYPSPGTGVYPPLSAYLYHSSPSVDFAI FSLHMSGI SSIMGSLN LMVTIMMMK NFSM NYDQISLFPWSV FITAILLIMSLPVa.S1..v.SmVENGA.....sGia.GGA...L.....la.....1.av.FiT.viN.rSTG1L.RmP..V...V...1..11....	199	200
LAGAITMLLFDRNFNTSFFDPMGGDP ILYQHFLWFGHPEVYI I LPGF GLISHIVMNE SKKE IFGNLS MIYAMLGIGFLGFIVWAHMFTVGLDVDTT...L.....A.....m.....m..iSq.....T..S.G.....A..L.....m....	299	300
RAYFTSATMIIAVPTGIKVFSW LATYHGSKLKL NISI LWSLG FIMLFTIGGLTGIMLNSNSIDIILHD TYYV GHFHYVLSMGAVFAI ISSFIHWYPLITi.....L..tQ.SYSpa...a..vF...v.....vv.a..v.....A.....maG.....F.	399	400
GLLLNIKWLKIQFIMMF1GVNL TFPQHFLGLMSM PRRYSDY PD SYCWN SISI MMGSM ISLN SMI FLIFI ILES LISK RMLLF KFN-QSSLEWLNFLPPL ..T..N....S..i.....AG.....a.TT..Vv.ti..T..LG1l.FFy..W...v.Q.QviyPi qLn..i..YqNT..A	498	500
DHS HLEIPLLIK NLN LKS ILIK e..YS.1...TN-----	520	512
Cytochrome oxidase II		
ISTWF MF MFQESNSYYADNLISFHN MVM MIIIMISLT VYI I LDLF MNKFSN LFL KNHNIEII WTI IPI I ILLIICFPSLK ILYL IDEIVNPFFSIKSI m...ANLGL.d.A.PLMeq..F..dHAll.lv..tV.VG.lmFM..F.NyV.R...hGQL..m....l.A....F.AL...r1..1...Ne.SVt1...	100	100
GHQWYWSYEYPEFNNIEFD SYMLN YNNLN--QFRLLET DNR MVIPMKIPLRLITTSTDVIHSWTVPSLGIKVDAVGRINQLN LISKRP GIFFGQCSEICSd.....iPt.e.AIDG....dv...v i l..NSQi.ilv.aa.....a..v....GT...l..T.FFIN...l.y.....	198	200
GMNHSFMP IMIESTSFQYFLN WVN KQI- .A.....v...VPVnN.iK.iSSnNS	225	228
Cytochrome oxidase III		
MKK--NFPFHMVTNSPWI ILSFS FMNT LIST VIWIYSS-ISM FMI LN FINS I LIMMLW FRDI I REST FQGMHSMF ITNFLKFSMILF I SELMFFISFF .STHS.H...1.DY....1TGaIGA.T.vSGM.K.FHQYD..1.11G.I.T...TvYQ.w..vS..G.y..1.tYAv.IG.rwG.....vl..v...	97	99
WTFFPHSSISP NIE INMTWPPKNIKFFNPMEIPLNSF ILVSSGFTV TL SHYLLI INNLKLSKSYLLTILLGIYFTI LQTE YSNSFFCFNDSIYGSIFF .a.....1..A..1GAs...MG.IS...Fq.....tA..la..V...Wa.Hs.mES.HSQtTQG.FF.v.....ay..IeaP.TIA..v...T.y	197	199
MATGFHGLHV LIGSIFLLISLYRMMN I HFSNMHN MNFELAIWYWHFVDVIWLF LYTFIYLLI-v.....tT...vc.L.H1.N...KN.HFG..A.A.....v.....IT..WWGG	259	262
NADH dehydrogenase 1		

IWVLINLLILMIMV L ISVAFL T L LERK IGYI QDRKGPNK IMLFGMFQPFSDALKL LSKEWFFFN YSNLFYI --SPMLMFFL S LVMW ILYPWFGF MY MEF.LS..GS.1.1.C..v.....v....I.....vg.M.iP...C..i..Ft..QTyPLL..YLS.YI..ifSL...Fv.mCM.f.VK1.	95	100
YIEFSILFMLLVLGLSVYPVLFVGWISNCNYA I LGSM RL VSTMISFEI NLF LVFS LMMV E SFNE FFFQNNIK FAILLYPLYLM MFTSMLI ELNRT SFnLGG..F.CCTS.G..T.mVA..S..S...1..G1.A.aQT..y..vs.ALimL.FiFliG.yNMIY..yy.IYmW.L.i.f.mS.vWL.IS.A.T..	195	200
PFDLIEGESELVSGF NI EYHSSMFV LIFLSE YMNIMFMSV ILSLMF YGFKYWSIKF I L YLFHICLII WIRGIL PRIRYDKLMN MCWTEMLMLVMIYLMY ...FA.....v..S.GG.A...ma..AS.1...mlFCv1..L.CDVF N1L.YvKLT.ISFvF..A..T...F.....Y1A.KCF.SFS1N..lf	295	300
LYFMKEFLCI----- FIGF.IL.FSFL WI FF SKKL MEN	305	324

FIGURE 6.—The inferred amino acid sequences for the proteins encoded by honeybee (*A. mellifera*) mtDNA, aligned with their *D. yakuba* counterparts. In each case the honeybee sequence is the upper one. Amino acids identical to those in the honeybee are indicated by dots; conservative replacements are indicated by lowercase letters, and nonconservative replacements relative to the honeybee are indicated by uppercase letters. Replacements were treated as conservative if, according to the protocol of FRENCH and ROBSON (1983), they involved amino acids of the same group (KHR, DENQ, GP, AST, ILMV, or FWY, with C ungrouped).

NADH dehydrogenase 2		
IFFFMNFKYHWFIYFLITIFVLMNNNIFIQWMLMEFGTIISISLINIKST--NKTPSLIYY-SVSVISSIPLFFMIIIVYL-SSISFTKTDTFNFMVQM-..yNSS.I-L.TTIm.IGTl1tVt..SWLGA..G1.INL1SF.P.1SDNNNLSTEA..K.fLtQalA.tvL..SS.11M.ANN1NNEINes.TS.1I.S	95	99
MFFLKIGTFFHFWMIYSYEMMNWKQIFLMSTLIKFIPIYMMVSMTKINSWTLYFLITNSLYISFYANKFYTLLKLLACSTIFNSFYFIFILELNKNMFI ALL..S.aA....FPNMM.G1T.MnALm1M.WQ.IA.1..11.YLN.KNLL.ISv.LSVIIGaIGGLNQTs.r..m.F.s.NHLGwMISS.MISESiwl	195	198
AMIILYSPNYFLLISFLNKPNIQNPNFMF--YNKYQMYTF-LTLMF-NYSMPPIFLSPVVIWNLIFMMVSVKAYNWILFLLMISSMLIWNYIIIKRVF IYF.F...LS.v.TFMP.I.K1FH1.Q1.SWFVN SKILK.S.FmN.LSLGG1.P..G.1P..Lv.QQ1TMCNQ.-f1.T.m.m.t1t1ff.1R.CYS.A.	291	297
LKMNFYKNNFIDDKD-NKMYHHSYFALTLLS-FNISPFITLNFL-- m-1.yFE..w.MeMnM.SNNNTNL.L1m.FF.I.G1..L.s.F.FML	333	341
NADH dehydrogenase 3		
MKFIFMYFIFIILISSLLNKFISIYKKKDYEKSSPFECGFNPITKANLPFSLPFFLMTMMFLIFDVEIILFLPIIFYLKSSSTMISYLMISIF-LIL iFS.IiiASV.l..ttvvmF.ASiL.KKALI.R.....d.KsSsR.....R..i.ii.....A.I..m.II..Y.NI..WTITSI..I...	99	100
LITTLILEWMNNYLNWL G-.YH..NqGM...SN	117	117
NADH dehydrogenase 4		
MYLLLLLIMLNMLMMMSM1YL FML-FMNKMKNNNL1IIGNLIIINLLNLFNLNWIDWIYIFCNLSFNMYSYGLIMLTLWIFGLIFI---SLNN-NSLNCL -----.Ki1f-11f.TPvC.i.N.YMvqim1FFSF.F..m.N.MNY.SeIS.--.Gcd.L....v1.s...CS.mLlASE.i.KY.NYKN.	95	88
F-MN-LLLIMISLLLVLFSMNLFLYFLYEFGLLIFLVVKWGYSENRLSGFYLMFYTMIIFSLPMLYIYYYIYLIDSYLFNFMLMEMLNLNLNMMFLIYL .L1.Ivi.11L.v.T.S..S.Fm....f.SS.iPTLf.11G...QPe.LQa.V..1...11V....IG.f.vMNKTG.m..Y..nNMF.Yd-1.yFC.	193	187
LMSFLVKIP1YLFHGWLKAHVEAPYYGSMILASIMLKGGYGMRLMIYKNEFILIQKIL-VMINSFGVLILSLMCLSQFDMSKIIAISSIVHMGLMI .Ca....m.mf.V.L..P.....VS.....G.....1..-v.NfLqlMN.KYSFvWIS.SLV.Gv1m..v..R.T.1.al..Y..vA...ivl	292	286
MSMMTFLKISLIGGYLMMISHGLSSGLFFLVNV1Y5QTSRSLMFINKGMINFMPMSM1LLWFMLCSSNMGPVSLN1SEVMLLIGMISWLKFMMILMM AG11.MTYWG.C.S.T1..a...C....C.A..S.ERLG..S.L....11....a.t.W..1.S.a..Aa.Pt...1G.is..NSiv..SWIS.im.SF	392	386
YCLFSFIY1YLFMF1NHGKIFI-MFKIKNG1LVEYFVLLHW1PLNLMPLKLYFI---- LSF..AA.t1..yS.Sq...1.SGvySFSS.KiR..L1m....1....1I..SESC1LWL	447	446
NADH dehydrogenase 4L		
-IKLLFV--MMLLFFFMLWYY--N-VNFLSFLILMELFLVITVLFF1IGY---EINSWLFLI-FLVFSVCELVGLSLLVSMNYELGHQKLSVMDLIY M.Mi.yWSLP.i..11g.fcFvS.RKHL..M.1s1..1.1M1F.M1FI.LNMLNYENY.SmM..T....GA....1....IRTH.NdYFQSFSim-	87	96
NADH dehydrogenase 5		
I1KMMVCGILLFEFSFLMMLMSLYLLYLNKEFFFEWN1YTFNSMKFNLLL1DYKSLMF1FLVSMIFSMII1Y5IS1SYM1DSELKMDRFLYLMFLISMY --ICS1SF.N.ISI.LTCF.1...Y.LN.MVY.I..evVsL..SIVMTF.F.wM..1.mSF.L..A.lv.F..KE..e..dENin..iM.v1m.v1..M	100	97
MLILSPNMLS1ILGWDGLGLISYCLVIYMMKMSFTSGMVTT1LNRLGD1GL1L1GLMTYYGSWNLSFY---KMNEF---MMI--YILLMAFTKSAQ1PF 1..i..1v..1.....v.....fqn1..yNa..1.A.S..1..va...AiAW.LN.....YI..LEVMQ..Sm1..GSLvM.A.M.....	194	198
STWLPMMAMPTPVSSLVHSSTLVTAGIYLLIRYVNLLDFNY-KNYIML1ASLTMLFAGLVANFELDLKKVVAYSTLSQLGFMMMSMLSIGSTELVFLHLF .s....A..A.....a.....v.....fnIV.STSwLGq111.lsG...FM...G....F....ii.L.....L....1..m.FYK.AMF..L	293	298
IHAMFKSLIMFCVGSYMHYMYSNQDIRMYYGMMYIYPMKSMIL1F1S1LSCGFPLVGGYSKDL1IEMFFSKM1YFSMINLI1G1T1F1TVSYSFRMI-LV T..1..a.1...A.all.N.NNS....1MG.1SIHM.1T.ACfvN.N.a...M..A.f....m.1.1vSI.N1N..FFLYFFS.GL.....lvYYS	392	398
LTSKFLMMNV-IYSKEDK1MC1S1MM1F1S1IY1SKL1F1NLM-NFNLLGIN1LLMIYKLMVFKM1MVG1IMGF--NFYKL1LNNK1GYFKMSFLFMNLIY m.GDLNCGS1NmLND.SWv.LRG.1G11FM.i.GGS1NW.iFP.PYm-.C.PGyl.m1TFvC1..G1F.yL1S1N.YS..KS1LNyNlt..1GSmw	488	496
KIYKK1I1MMFTYEVYIEKS1IE1L1---SKFMSVTLN1YELK1S1N1M1Y1L1T1IY1L-IY1L1Y1LINF----- FmP.ISTYG.i..-P1NYGQLvvKSFDQGW.EyFG-GQH1..Y..1..YSKT1f.MHNNS.K...m1fvFW1M1LFSPLLFL	554	573
NADH dehydrogenase 6		
IMLT1I1M1-SK1FM1S1L1M1T1Y1L1N1F1N1P1S1M1L1--L1Y1L1S1Y1M1S1M1F1T1M1C1S1M1N1S1L1M1L1I1V1F1L1S1G1M1F1S1Y1F1S1L1-NEPL1K1K1K1---P --i1Q1..Y.....1TT1I.FF.M.HPLALG.TL..Q.T.FV1L1SG..T1K1F1W1S1Y1---.F1..i..G...v1.I.VT..AS..MFN.S1.LT1F	93	87
FIQTLFL1I1TMK1Y1N1K1L1S1Q1N1E1H1F1N1Y1--F1K1N1D1-MY1Y1M1K1M1N1S1--T1F1M1L1M1I1T1L1L1M1T1K1Y1IE--K1T1R1K1K1K1 --- SmF1..FmF.LSM.Ld.T.ITLFLM.NEMQSI.emNS.FTENSL.LNK.yNFPTNFvT.L.mNYLL..L.VVv.I.KLF.GP1RMMS	167	174

Fig. 6 part 2.

tion is weakened by the fact that the two lineages have had different life patterns to the current ones for most of their evolutionary history (CROZIER, CROZIER and MACKINLAY 1989), and the second explanation is entirely conjectural (although testable in principle). The third explanation is rendered plausible by an association between the amount of divergence of insect mtDNA and AT% (L. S. JERMIIN and R. H. CROZIER in preparation).

The honeybee and Drosophila COI and COII genes remain more similar than expected by chance at all codon positions when the chance expectations are determined as the products of base proportions (CROZIER 1992). The sequences reported here extend this result to all of the mitochondrial protein-encoding genes (analyses not shown). One possible explanation for the persistence of greater than chance similarity is that changes have not yet saturated; another is that

TABLE 3

Comparison of protein-encoding genes of *Apis* and *Drosophila*

Gene	No. of amino acids		Percent similarity	
	Apis	Drosophila	Nucleotides	Amino acids
ATPase6	226	224	64	47
ATPase8	52	53	62	46
COI	520	512	74	70
COII	225	228	68	55
COIII	259	262	66	53
Cyt b	383	378	66	53
ND1	305	324	64	47
ND2	333	341	51	27
ND3	117	117	63	49
ND4	447	446	64	45
ND4L	87	96	57	37
ND5	554	573	61	42
ND6	167	174	53	31

Data for *D. yakuba* are from CLARY and WOLSTENHOLME (1985).

TABLE 4

Variables selected by stepwise multiple regression analysis as significantly associated with the number of occurrences of amino acids

Species	Variables	F-to-remove	r^2
<i>Apis mellifera</i>	%T in codon family	9.89	0.35
<i>Caenorhabditis elegans</i>	%T in codon family	43.02	0.75
	Codon family size	26.83	
	%A in codon family	6.49	
<i>Drosophila yakuba</i>	Codon family size	20.81	0.63
	%T in codon family	4.87	
	%AT in codon family	4.74	
<i>Mus musculus</i>	Codon family size	28.60	0.59
	%AT in codon family	12.92	
<i>Xenopus laevis</i>	Codon family size	30.95	0.61
	%AT in codon family	9.14	

Total suite of variables: %A, %T, %G, %C, and %AT in codon family, and codon family size (number of codons). Variables in table are listed in order of importance for each species according to F-test (threshold value for both F-to-enter and F-to-remove set at 4.41). The coefficient of determination, r^2 , indicates the proportion of the variance explained by the significant variables. Data sources: *A. mellifera* (this paper), *C. elegans* (OKIMOTO *et al.* 1992), *D. yakuba* (CLARY and WOLSTENHOLME 1985), *M. musculus* (BIBB *et al.* 1981), *X. laevis* (ROE *et al.* 1985).

DNA-level processes (ALMAGOR 1985) constrain variation even at the third positions of codons. The first hypothesis seems unlikely given the 280 million years (My) of separation between these species (CARPENTER and BURNHAM 1985). The second hypothesis at first appears plausible because analysis of the association of nucleotides at adjoining positions across all protein genes shows a significant association for each of the three possible classes (first codon position with second; second with third; and third with the first position of the following codon; in all cases $P < 0.001$ by G-test); such associations have been reported for coding se-

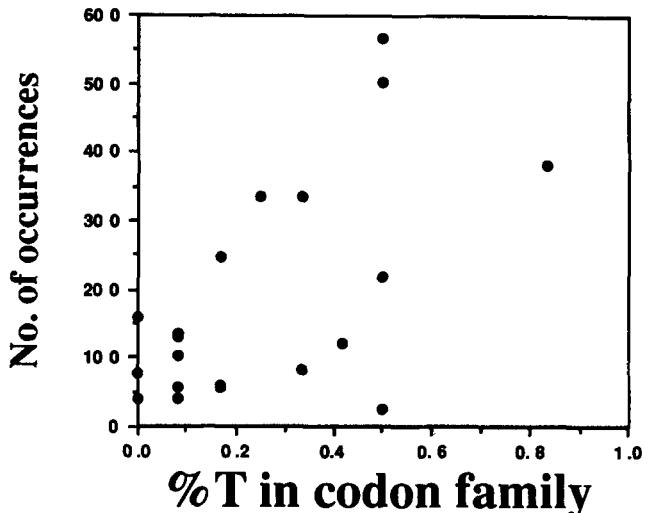


FIGURE 7.—Relationship between use of amino acids in honeybee mtDNA proteins and %T in codon family.

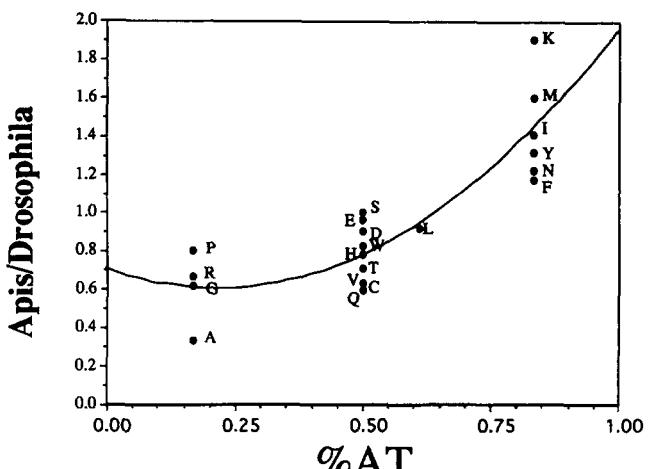


FIGURE 8.—The ratio of proportions of amino acid occurrences in mitochondrially encoded honeybee (*A. mellifera*) proteins to those of *D. yakuba* as a function of the %AT in codon family.

quences generally (reviewed by CROZIER 1992; MADDOX 1992), suggesting a general occurrence of DNA-level processes. However, there is no association between doublet type and the likelihood of a position being conserved between honeybee and *Drosophila* (analyses not shown), indicating that an explanation based on DNA-level processes is inadequate.

Transitions have often been suggested to occur much more frequently during the evolution of mitochondrial genomes than transversions. Comparisons of mtDNA haplotypes in Hawaiian *Drosophila* indicate a ratio of transitions to transversions of 16:1 (DESALLE *et al.* 1987), and THOMAS and WILSON (1991) report a ratio of 12:1 for intraspecific comparisons in the nematode genus *Caenorhabditis*. Such a high ratio is expected to occur only over short stretches of evolutionary time, because once a transversion has occurred all subsequent transitions appear to be transversions when the sequences are compared.

TABLE 5

Transitions and transversions observed between *Apis* and *Drosophila* for the complete suite of protein-encoding genes

	Codon position			
	1	2	3	All
Transitions	400	357	243	1000
Transversions	1000	643	1318	2961
Observed (S/V)	0.400	0.555	0.184	0.334
Expected (S/V)	0.382	0.487	0.104	0.324

Data are from this paper for *A. mellifera* and from CLARY and WOLSTENHOLME (1985) for *D. yakuba*. The observed and expected transition/transversion ratios are also shown (Observed (S/V) and Expected (S/V)). Transitions at the third codon position are significantly more common than expected by chance ($\chi^2 = 68.8, P < 0.001$); transitions at the second codon position are marginally more common than expected by chance ($\chi^2 = 3.98, P < 0.05$).

TABLE 6

Transitions and transversions between the small and large subunit of *Apis* and *Drosophila* mitochondrial ribosomal RNA genes

	Observed	Expected	Observed (S/V)	Expected (S/V)
s-rRNA				
Transitions	81	56.96	0.526	0.320
Transversions	154	178.04		
l-rRNA				
Transitions	79	77.76	0.273	0.268
Transversions	289	290.24		

Data are from this paper for *A. mellifera* and from CLARY and WOLSTENHOLME (1985) for *D. yakuba*. The observed and expected transition/transversion ratios (S/V) are also shown. There is an excess of transitions over chance expectation for the small subunit gene ($\chi^2 = 13.4, P < 0.001$) but not for the large subunit gene.

As might be expected under this reasoning for a pair of species separated for 280 My (CARPENTER and BURNHAM 1985), most of the transition/transversion ratios determinable in the comparison with *Drosophila* are close to chance expectation based on the nucleotide proportions for both protein-encoding genes (taken as a whole) and the ribosomal RNA genes (Tables 5 and 6). The deviations from chance expectation are in the direction of an excess of transitions, despite the long period of separation.

The similarity between honeybee and *D. yakuba* at the third codon position differs markedly between genes relative to expectations based on the products of the base proportions (Figure 9) and the observed third-position similarities are significantly associated with those at the second codon positions (Figure 10).

The persistent similarity between *Drosophila* and *Apis* at the third codon position, the excess of transitions over expectation at the third codon positions when these insects are compared, and the association between second and third codon similarities can be simply explained by considering conservation of the amino acid sequence. Twelve of the codon families

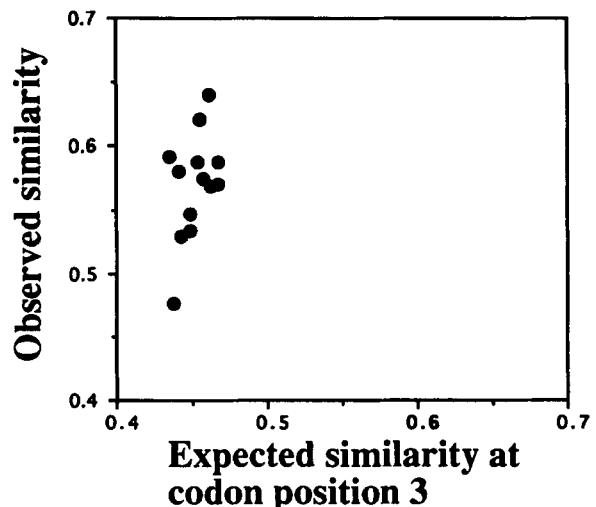


FIGURE 9.—Relationship between observed and expected similarities at codon position 3 between *A. mellifera* and *D. yakuba* protein-encoding mitochondrial genes. Expectations were calculated from the products of the base proportions in the two sets of genes.

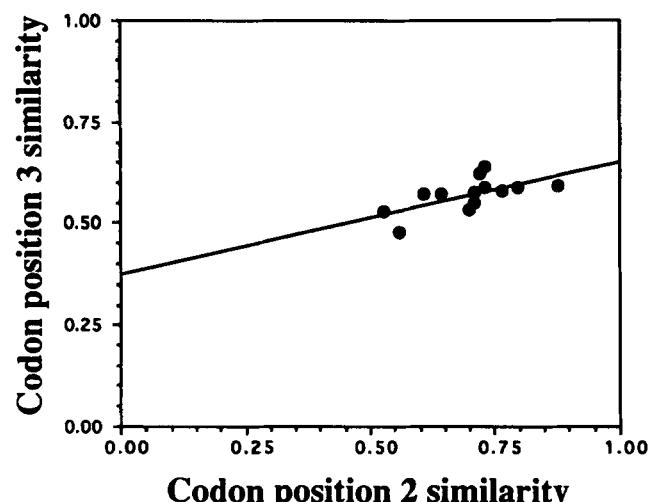


FIGURE 10.—Relationship between similarities between *A. mellifera* and *D. yakuba* at codon position 3 and codon position 2 for protein-encoding genes. The regression line has a coefficient of determination of $r^2 = 0.39$.

consist of two codons, greatly limiting the possible bases for conserved amino acids for these families. For example, for phenylalanine, only T and C are possible at the third position, and most such codons for both insects will be TTT. Hence, there will be a relationship between conservation at the amino acid level (and hence at the first and second codon positions as well) and similarity at the third codon position. When expectations are determined in terms of codons for conserved or not conserved amino acids using the base proportions for each codon family in each insect, there is a good fit of observation to expectation (e.g., Figure 11).

Similarly, for conserved amino acids with two-codon

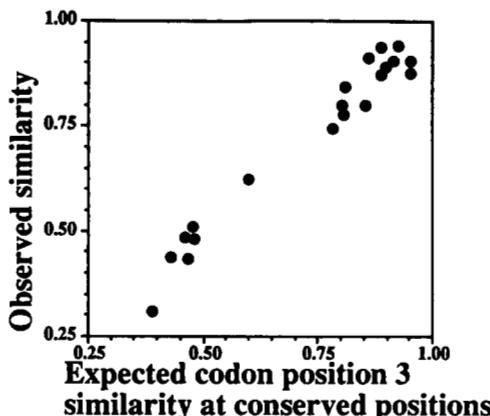


FIGURE 11.—Relationship between observed and expected similarities between *A. mellifera* and *D. yakuba* for third codon positions for codons for conserved amino acids. Expectations were calculated as the product of base proportions between the sequences for each codon family.

families, the only detectable substitutions will be transitions at the third codon position.

The excess of transitions at the second codon position is of borderline statistical significance, and should be regarded with caution.

The excess of transitions over expectation for the srRNA gene, but not for the lrRNA gene, clearly cannot be due to the same factor as that advanced for the protein-encoding genes, but may indicate that the folding pattern of the small ribosomal subunit RNA is relatively sensitive to nucleotide size, favoring transitions because these substitutions will not change nucleotide sizes whereas transversions will.

The likelihood that transitions will greatly exceed transversions in the evolution of honeybee mtDNA is uncertain. Given the predominance of adenine and thymine, most substitutions must be A → T or T → A, which are transversions. If it were otherwise, then the high AT content actually seen would have been eroded, unless each transition tends to be reversed by a further transition at the same site (W. K. THOMAS, personal communication). Such an explanation accords with the observed excess of transitions over transversions seen in the evolution of nematode and *Drosophila* mtDNA, both of which are AT-rich genomes. Data are now being collected to test this latter possibility.

Although it is clear that a bias of substitutions toward A and T has affected the nucleotide composition of honeybee mtDNA, and appears to have markedly affected the amino acid composition of honeybee mitochondrial proteins, the source of this bias is uncertain. It is, however, known that some natural mutational processes favor the generation of an AT-bias. Thus, guanine is prone via alkylation to conversion to O⁶-methylguanine which often mispairs with thymine, leading to replacement of GC pairs with AT pairs (WATSON *et al.* 1987). Alkylating agents capable of

producing O⁶-methylguanine are endogenously produced in prokaryotic cells (REEBECK and SAMSON 1991), and it is reasonable to suppose that this is true for eukaryotes too. Endopterygote insects may then be characterized either by a higher endogenous production of alkylating agents in mitochondria, or by relatively inefficient import into mitochondria of DNA repair methyltransferases.

This work was supported by grants from the Australian Research Council, the Ian Potter Foundation, and La Trobe and New South Wales universities. We thank B. CRESPI, B. G. HALL, L. S. JERMIN, A. G. MACKINLAY, C. SIMON, M. WESTERMAN, R. G. WHITE and an anonymous reviewer for constructive comments on the manuscript. The early stages of this work benefitted from aid by C. HAWKINS and the hospitable laboratory of A. G. MACKINLEY.

LITERATURE CITED

- ALMAGOR, H., 1983 A Markov analysis of DNA sequences. *J. Theor. Biol.* **104**: 633–645.
- BERNARDI, G., 1989 The isochore organization of the human genome. *Annu. Rev. Genet.* **23**: 637–661.
- BERNARDI, G., and G. BERNARDI, 1986 Compositional constraints and genome evolution. *J. Mol. Evol.* **24**: 1–11.
- BIBB, M. J., R. A. VAN ETEN, C. T. WRIGHT, M. W. WALBERG and D. A. CLAYTON, 1981 Sequence and gene organization of mouse mitochondrial DNA. *Cell* **26**: 167–180.
- BROWN, W. M., 1985 The mitochondrial genome of animals, pp 95–130 in *Molecular Evolutionary Genetics*, edited by R. J. MACINTYRE. Plenum Press, New York.
- CARPENTER, F. M., and L. BURNHAM, 1985 The geological record of insects. *Annu. Rev. Earth Planet. Sci.* **13**: 297–314.
- CLARY, D. O., J. M. GODDARD, S. C. MARTIN, C. M.-R. FAURON and D. R. WOLSTENHOLME, 1982 *Drosophila* mitochondrial DNA: a novel gene order. *Nucleic Acids Res.* **10**: 6619–6637.
- CLARY, D. O., J. A. WAHLSTEINER and D. R. WOLSTENHOLME, 1984 Sequence and arrangement of the genes for cytochrome b, URF1, URF4, URF5, URF6 and five tRNAs in *Drosophila* mitochondrial DNA. *Nucleic Acids Res.* **12**: 3747–3762.
- CLARY, D. O., and D. R. WOLSTENHOLME, 1983a Genes for cytochrome c oxidase subunit I, URF2, and three tRNAs in *Drosophila* mitochondrial DNA. *Nucleic Acids Res.* **11**: 6859–6872.
- CLARY, D. O., and D. R. WOLSTENHOLME, 1983b Nucleotide sequence of a segment of *Drosophila* mitochondrial DNA that contains the genes for cytochrome c oxidase subunits II and III and ATPase subunit 6. *Nucleic Acids Res.* **11**: 4211–4277.
- CLARY, D. O., and D. R. WOLSTENHOLME, 1984 A cluster of six tRNA genes in *Drosophila* mitochondrial DNA that includes a gene for an unusual tRNA^{Ser}. *Nucleic Acids Res.* **12**: 2367–2379.
- CLARY, D. O., and D. R. WOLSTENHOLME, 1985 The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. *J. Mol. Evol.* **22**: 252–271.
- CLARY, D. O., and D. R. WOLSTENHOLME, 1987 *Drosophila* mitochondrial DNA: conserved sequences 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.
- CORNUEL, J.-M., L. GARNERY and 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.
- CROZIER, R. H., 1980 Genetical structure of social insect populations, pp. 129–146 in *Evolution of Social Behavior: Hypotheses*

- and Empirical Tests, edited by H. MARKL. Verlag Chemie, Weinheim.
- CROZIER, R. H., 1992 Molecular methods for insect phylogenetics, in *Molecular Approaches to Pure and Applied Entomology*, edited by M. J. WHITTEN and J. G. OAKESHOTT. Springer-Verlag, New York (in press).
- CROZIER, R. H., and Y. C. CROZIER, 1992 The cytochrome *b* and ATPase genes of honeybee mitochondrial DNA. *Mol. Biol. Evol.* **9**: 474–482.
- CROZIER, R. H., Y. C. CROZIER and A. G. MACKINLAY, 1989 The COI and COII region of honey-bee mitochondrial DNA: evidence for variation in insect mitochondrial rates. *Mol. Biol. Evol.* **6**: 399–411.
- DE BRUIJN, M. H. L., 1983 *Drosophila melanogaster* mitochondrial DNA, a novel organization and genetic code. *Nature* **304**: 234–241.
- DESALLE, R., T. FREEDMAN, E. M. PRAGER and A. C. WILSON, 1987 Tempo and mode of sequence evolution in mitochondrial DNA of Hawaiian Drosophila. *J. Mol. Evol.* **26**: 157–164.
- DESGARDINS, P., and R. MORAIS, 1990 Sequence and gene organization of the chicken mitochondrial genome. A novel gene order in higher vertebrates. *J. Mol. Biol.* **212**: 599–634.
- FRENCH, S., and B. ROBSON, 1983 What is a conservative substitution? *J. Mol. Evol.* **19**: 171–175.
- GILBERT, D. G., 1989 IUBio archive of molecular and general biology software and data. An Internet resource available via anonymous ftp to ftp.bio.indiana.edu.
- GODDARD, J. M., and D. R. WOLSTENHOLME, 1978 Origin and direction of replication in mitochondrial DNA molecules from *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **75**: 3886–3890.
- GODDARD, J. M., and D. R. WOLSTENHOLME, 1980 Origin and direction of replication in mitochondrial DNA molecules from the genus *Drosophila*. *Nucleic Acids Res.* **8**: 741–757.
- HARRISON, R. G., 1989 Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. *Trends Ecol. Evol.* **4**: 6–11.
- HSUCHEN, C.-C., and 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 and 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.
- JUKES, T. H., and V. BHUSHAN, 1986 Silent nucleotide substitutions and G + C content of some mitochondrial and bacterial genes. *J. Mol. Evol.* **24**: 39–44.
- KOULIANOS, S., and R. H. CROZIER, 1991 Two ancient alleles in Australian honeybees. *Apidologie* **22**: 621–626.
- MADDOX, J., 1992 Long-range correlations within DNA. *Nature* **358**: 103.
- OJALA, D., J. MONTOYA and G. ATTARDI, 1981 tRNA punctuation model of RNA processing in human mitochondria. *Nature* **290**: 470–474.
- OKIMOTO, R., J. L. MACFARLANE, D. O. CLARY and D. R. WOLSTENHOLME, 1992 The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. *Genetics* **130**: 471–498.
- OSAWA, S., T. H. JUKES, K. WATANABE and A. MUTO, 1992 Recent evidence for evolution of the genetic code. *Microbiol. Rev.* **56**: 229–264.
- PÄÄBO, S., W. K. THOMAS, K. M. WHITFIELD, Y. KUMAZAWA and A. C. WILSON, 1991 Rearrangements of mitochondrial transfer RNA genes in marsupials. *J. Mol. Evol.* **33**: 426–430.
- REEBECK, G. W., and L. SAMSON, 1991 Increased spontaneous mutation and alkylation sensitivity of *Escherichia coli* strains lacking the *ogg* O⁶-methylguanine DNA repair methyltransferase. *J. Bacteriol.* **173**: 2068–2076.
- ROE, B. A., D. P. MA, R. K. WILSON and J. F. H. WONG, 1985 The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome. *J. Biol. Chem.* **260**: 9759–9774.
- SACCONE, S., G. PESOLE and E. SBISÁ, 1991 The main regulatory region of mammalian mitochondrial DNA: structure-function model and evolutionary pattern. *J. Mol. Evol.* **33**: 83–91.
- SANGER, F., A. R. COULSEN, B. G. BARRELL, A. J. H. SMITH and A. ROE, 1980 Cloning in single-stranded bacteriophage as an aid to DNA sequencing. *J. Mol. Biol.* **143**: 161–178.
- SATTA, Y., H. ISHIWA and S. I. CHIGUSA, 1987 Analysis of nucleotide substitutions of mitochondrial DNAs in *Drosophila melanogaster* and its sibling species. *Mol. Biol. Evol.* **4**: 638–650.
- SMITH, D. R., and W. M. BROWN, 1988 Polymorphisms in mitochondrial DNA of European and Africanized honeybees (*Apis mellifera*). *Experientia* **44**: 257–260.
- SOKAL, R. R., and F. J. ROHLF, 1981 pp. 661–671 in *Biometry*, Ed. 2. Freeman, San Francisco.
- THOMAS, W. K., and A. C. WILSON, 1991 Mode and tempo of molecular evolution in the nematode *Caenorhabditis*: cytochrome oxidase II and calmodulin sequences. *Genetics* **128**: 269–279.
- UHLENBUSCH, I., R. M. RIPPE and G. GELLISEN, 1987 Mitochondrial tRNA genes in *Locusta migratoria*. *Naturwissenschaften* **74**: 142–143.
- VLASAK, I., S. BURGSCHWAIGER and G. KREIL, 1987 Nucleotide sequence of the large ribosomal RNA of honeybee mitochondria. *Nucleic Acids Res.* **15**: 2388.
- WATSON, J. D., N. H. HOPKINS, J. W. ROBERTS, J. A. STEITZ and A. M. WEINER, 1987 pp. 346–347 in *Molecular Biology of the Gene*, Ed. 4, Vol. 1. Benjamin/Cummings, Menlo Park, Calif.
- WOLSTENHOLME, D. R., 1992 Animal mitochondrial DNA: structure and evolution. *Int. Rev. Cytol.* **141**: (in press).
- WOLSTENHOLME, D. R., and 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 and J. A. WAHLEITHNER, 1987 Bizarre tRNAs inferred from DNA sequences of mitochondrial genomes of nematode worms. *Proc. Natl. Acad. Sci. USA* **84**: 1324–1328.

Communicating editor: A. H. D. BROWN