Int. J. Indust. Entomol. Vol. 2, No. 1, 2001, pp. 1~6

[Review]

International Journal of *Industrial Entomology*

Comparative Analysis of Completely Sequenced Insect Mitochondrial Genomes

Jin Sung Lee, Ki Hwan Kim[1], Dong Sang Suh[1], Jae Heung Park, Ji Yoeun Suh[2], Kyu Hoi Chung[3] and Jae Sam Hwang[4,*]

CoreBio Research Institute of Life Science & Biotechnology, CoreBio System Co., Ltd., Seoul 603-23, Korea.
[1]Department of Genetic Engineering, Sungkyunkwan University, Suwon 440-746, Korea.
[2]Infectious Diseases Research Institute, Samsung Biomedical Research Institute, Seoul 135-710, Korea.
[3]Department of Biology, Kyonggi University, Suwon 442-760, Korea.
[4]Department of Sericulture and Entomology, National Institute of Agricultural Science and Technology, RDA, Suwon 441-744, Korea.

(Received 12 January 2001; Accepted 12 February 2001)

This paper reports a few characteristics of seven insect mitochondrial genomes sequenced completely (*Bombyx mori*, *Drosophila melanogaster*, *D. yakuba*, *Apis mellifera*, *Anopheles gambiae*, *A. quadrimaculatus*, and *Locusta migratoria*). Comparative analysis of complete mt genome sequences from several species revealed a number of interesting features (base composition, gene content, A+T-rich region, and gene arrangement, etc) of insect mitochondrial genome. The properties revealed by our work shed new light on the organization and evolution of the insect mitochondrial genome and more importantly open up the way to clearly aimed experimental studies for understanding critical roles of the regulatory mechanisms (transcription and translation) in mitochondrial gene expression.

Key words : Mitochondrial genome, A+T-rich region, Mt tRNA, Mt rRNA, Gene arrangement, Ribosomal transcriptional signal, Base composition

Introduction

Mitochondrial (mt) genome is smaller than nuclear DNA with a simpler composition and has the genetic characteristic of being transmitted maternally. For example, mt genome size of human being is about 1/8000 of an average DNA, and this takes up 0.5% in the total chromosome (somatic cell). Intervening sequence between transcribed genes and spacer sequence between genes are not generally detected in animal mt genome. Also, unlikely various nuclear DNAs that are rearranged in a complex mode, rearrangement among genes is not occurred usually in mt genome in higher eukaryotes. Mt genome has a simple composition and has a very simple genetic mode excluded from the influence of nuclear DNA (Cann *et al.*, 1987; Nei and Kohe *et al.*, 1983; Strachan *et al.*, 1976). Thus, mt genome is a useful mateial in studying molecular-evolution of living things. Moreover, mt genome is transmitted through cytoplasm so that it is also helpful in studies of inter-evolution between nucleus and cytoplasm (Brown *et al.*, 1985; Clayton *et al.*, 1975; Wolstenholmn *et al.*, 1992).

Table 1 shows the comparison of the characteristics between nuclear DNA and mt genome. Until now, seven species belonging to four orders in insect have been reported in 4 orders of 6 species [Hymenoptera; *Apis mellifera* (L06178), Diptera; *Anopheles gambiae* (L20934), *A. quadrimaculatus* (L04272), *Drosophila melanogaster* (U37541), *D. yakuba* (X03240), Orthoptera; *Locusta migratoria* (X80245), Lepidoptera; *Bombyx mori* (AF 149768) (Clary *et al.*, 1985; Garesse *et al.*, 1985; Crozier *et al.*, 1993; Michell *et al.*, 1993; Beard *et al.*, 1993; Flook *et al.*, 1995, Lee *et al.*, 1999)] (Table 2). These reported cases take 4.4% of 135 types of mt genomes registered with GOBASE (http://megasun.bch.umontreal, ca), only seven completed mt genome sequences in insects suggest lack of extensive mt genome analysis in

*To whom correspondence should be addressed.
Department of Sericulture and Entomology, National Institute of Agricultural Science and Technology, RDA, Suwon 441-744, Korea. Tel: 82-031-290-8432; Fax : 82-031-295-2176;
E-mail : jhwangjs@rda.go.kr

Table 1. Comparison of human nuclear and mitochondrial genome

	Nuclear genome	Mitochondrial genome
Total size	3,000 Mbp	16.6 kbp
Total no. of DNA molecule per one cell	23 in haploid cell; 46 in diploid cell	Several$\times 10^3$
Associated protein	Several classes of histon and non-histone protein	Largely free of protein
Repetitive DNA	Large fraction	Very little
Intron	Found in most genes	Absent
% of coding DNA	About 3%	About 93%
Recombination	At least once for each pair of homologs at meiosis	None
Inheritance	Mendelian mode	Exclusively maternal mode

Table 2. Comparative analysis of A+T-rich region, two rRNA genes and protein-coding genes of insect mitochondrial genomes

Organism	Accession no.	Total size (bp)	Total A+T (%)	No. of codons	A+T content of PCG* (%)	LrRNA gene		SrRNA gene		A+T-rich region	
						Size (bp)	A+T (%)	Size (bp)	A+T (%)	Size (bp)	A+T (%)
B. mori	AF149768	15,643	81.3	3,714	79.5	1,375	84.4	783	85.6	499	95.4
A. gambiae	L20934	15,363	77.6	3,733	75.9	1,325	82.5	800	79.6	519	94.2
A. qadrima.	L04272	15,455	77.4	3,728	75.4	1,321	82.2	794	80.5	625	93.5
D. yakuba	X03240	16,019	78.6	3,727	76.7	1,326	83.4	789	79.3	1,077	92.9
L. migratoria	X80245	15,722	75.3	3,713	74.1	1,314	78.9	829	76.0	875	86.2
A. mellifera	L06178	16,343	84.9	3,675	83.2	1,371	85.3	786	81.4	827	96.1

*PCG indicates thirteen protein-coding gene of mitochondrial genome.

insect species compared with other taxonomic groups. In the present review, we discussed a few characteristics of insect mt genomes using seven completely sequenced mt genomes.

Genome size

Insect mt genome is reported to be between 15 kb and 20 kb which is the similar size as that of eukaryotic organisms including mammals. No evolutionary relationship could be present between the classification group and the size of mt genome. This presumption could be confirmed dramatically in the *Drosophila* species that shows the genome with various sizes compared to other species. Also, this similarity in sizes could suggest the smallest size for mt genome to maintain its function in multicellular organisms. Slight differences in the sizes of insect mt genomes usually occur due to deletion and addition, and especially, the change in size occurs according to the presence or absence of tandomly repeated sequences in the A+T-rich region, which plays similar roles as displacement-loop (D-loop) that is termed as replication origin in mammals (Clary *et al.*, 1987; Moritz and Brown *et al.*, 1987). As shown in Fig. 1, the A+T-rich region in *D. melanogaster* is composed of 4,584 bp with 10 tandom repeat sequences whereas that in *D. yakuba* is composed of 1,077 bp with the presence of 2 tandom repeat sequences.

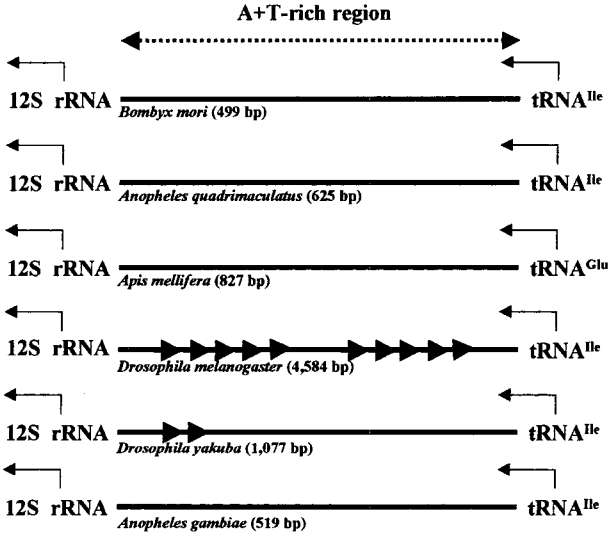

Fig. 1. Comparison of A+T rich region of insect mitochondrial genome. The arrow heads indicate tandom repetitive sequences. The bent arrows indicate transcriptional direction of flanking mitochondrial genes of A+T-rich region.

Thus, the change of about 3 kb in size exists in the A+T-rich region. However, the remaining four insect species including *B. mori* did not possessed tandomly repeated sequences.

Base composition and control region

One of the characteristics of insect mt genome is that the content of A and T is significantly higher than other classes of organisms. Among those studies reported, *A. mellifera* (84.9%) and *B. mori* (81.3%) are the heavily biased in A and T contents, which are known to be extreme compared to 55% to 60% in mammals. When the A and T content within mt genome according to gene was analyzed, A and T content is shown in the descending order of A+T-rich region, rRNA gene, and protein-coding gene. Thus, we could determine that the A+T-rich region that is the region responsible for insect mt genome replication contains the highest content of A and T. In the case of *B. mori*, the A+T-rich region showed the A and T content of 95.4%, which is highly disproportionate compared to that in the completely mt genome (Table 2). Thus, due to these reasons, the replication region of insect mt genome is not called D-loop but is called the A+T- rich region (Debruijn *et al.*, 1983; Fauron *et al.*, 1980). Unlikely insects, the highly conserved content of A and T in the A+T-rich region is not detected in mammals, making motif analysis difficult for replication origin through similarity analysis (Clayton *et al.*, 1984, 1991; Dairaghi *et al.*, 1995; Fisher *et al.*, 1992; Kruse *et al.*, 1989; Parisi *et al.*, 1991). Thus, studies related to replication mechanism of insect mt genome and related regions are not well understood in many areas.

Gene content and its arrangement

Mt genome in eukaryotic organisms has five necessary enzyme complexes in the oxidative phosphorylation and when adenine nucleotide translocator (ANT) that exists as a homodimer expressed in nuclear DNA is included, about 90 proteins form the five protein complexes to accomplish ATP synthesis and respiratory process (Clayton., 1991). Among 39 subunits of complex I, 7 subunits exist on mt genome, and rest on nuclear DNA. Four subunits of complex II exist on nuclear DNA, and complex III is composed of a total of 10 subunits, and mt genome encodes only cytochrome B. Complex IV is composed of a total of subunits, and only 3 of the genes exist on mt genome. Lastly, complex V is composed of 12 subunits among which only ATPase6 and ATPase8 are detected on mt genome. Also, mt genome encodes 22 tRNA genes and 2 rRNA (12S and 16S rRNA) necessary for the translating procedure of mitochondrial protein synthesis (Clayton, 1975; Stachan *et al.*, 1996).

Insect mt genome includes the above 37 genes (Fig. 2). Thus, the gene contents and function of mt genome in higher eukaryotes are discovered commonly throughout various species but has many differences in gene arrangement. The arrangement of insect mt genome can be divided into two types, homologous and non-homologous arrangements. With 13 protein-encoding genes, homologous arrangement has the relatively same arrangement in all insects. The main source of non-homologous arrangement stems from tRNA genes and rRNA genes. For example, except for the inversion of one $tRNA^{Ser}$ gene in *A. gambiae* and *A. quadrimaculatus*, all the gene arrangements are same throughout insect species (Fig. 2). However, this is not true among different classes. Especially, compared to mt gene arrangements reported until now, the fact is noticeable

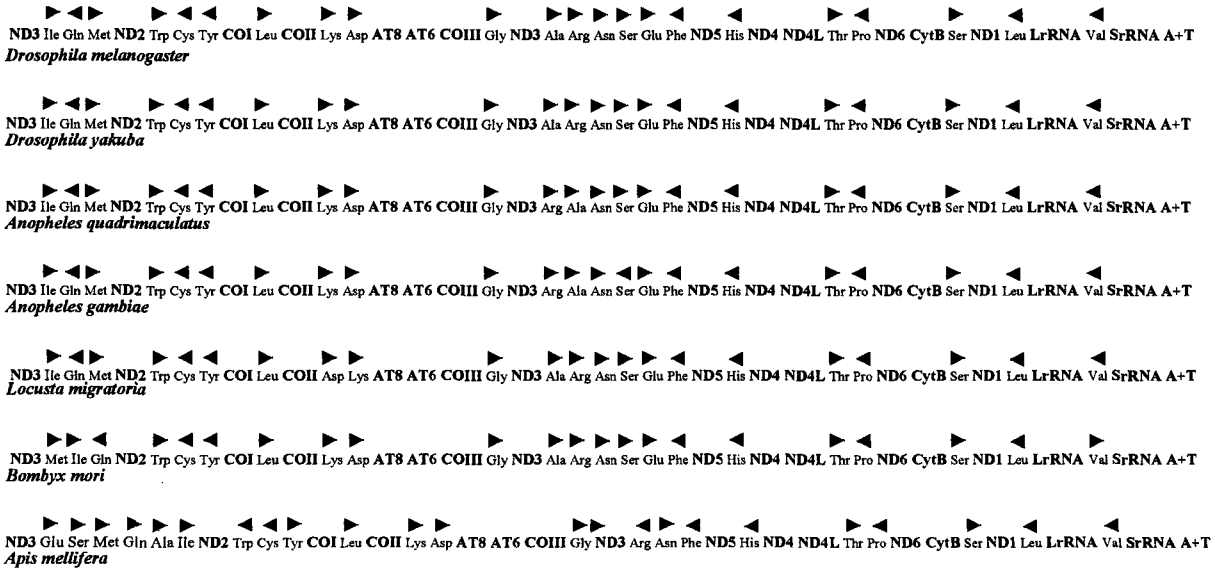

Fig. 2. Comparative analysis of arrangement of insect mitochondrial genes. The arrow heads indicate transcriptional direction of 22 mitochondrial tRNA genes.

Table 3. Comparison of initiation and termination signals in protein-coding genes of insect mitochondrial genomes

PCG	Initiation					Termination				
	Bmo	Aga	Aqu	Dya	Ame	Bmo	Aga	Aqu	Dya	Ame
ATPase6	ATG	ATG	ATG	ATG	ATG	TAA	TAA	TAA	TAA	TAA
ATPase8	ATA	ATC	ATC	ATT	ATT	TAA	TAA	TAA	TAA	TAA
COI	CGA*	TCG*	TCG*	ATAA*	ATA	Taa	Taa	Taa	TAA	Taa
COII	ATG	ATG	ATG	ATG	ATT	Taa	Taa	Taa	Taa	Taa
COIII	ATG	ATG	ATG	ATG	ATG	TAA	Taa	Taa	TAA	TAA
CytB	ATA	ATG	ATG	ATG	ATG	TAA	Taa	TAA	TAA	TAA
ND1	ATT	ATA	ATT	ATA	ATT	TAA	TAA	TAA	TAA	TAA
ND2	ACA	ATC	ATT	ATT	ATC	TAA	Taa	Taa	Taa	TAa
ND3	ATA	ATA	ATA	ATT	ATA	TAA	Taa	Taa	TAA	TAA
ND4	ATG	ATG	ATG	ATG	ATA	TAA	Taa	Taa	Taa	TAA
ND4L	ATG	ATA	ATG	ATG	ATT	Taa	TAA	TAA	TAA	TAA
ND5	ATT	GTG*	ATT	ATT	ATT	TAA	TAA	Taa	Taa	TAA
ND6	ATT	ATT	ATT	ATT	ATT	TAA	TAA	TAA	TAA	TAA

TAa and Taa signifies incomplete termination codons. Bmo, Aga, Aqu, Dya and Ame indicate *Bombyx mori*, *Anopheles gambiae*, *Anophele quadrimaculatus*, Drosophila yakuba and *Apis mellifera*, respectively. *indicates incomplete or unusual termination codon.

that many of the mt gene arrangement in *A. mellifera* and *B. mori* are translocational arrangement rather than inversion. For example, except for *A. mellifera*, the arrangement of $tRNA^{Ala}$-$tRNA^{Arg}$-$tRNA^{Asn}$-$tRNA^{Ser}$-$tRNA^{Glu}$-$tRNA^{Phe}$ that is composed with the largest gene group among the group of tRNA gene is almost the same gene arrangement (Fig. 2). Also, $tRNA^{Val}$ gene located between 12S rRNA and 16S rRNA genes has the same direction in six insect species, but is inversed in *B. mori* (Fig. 2).

Translation initiation and termination codons

As mt gene in other animals, triplet ATN (N = A or T or C or G) is used as insect translation initiation codon. Among these, not only ATN but also TTG and GTG are used as initiation codon in *Drosophila*, and, especially, COI of *D. yakuba* uses a substitute translation initiation codon that encodes quadruplet codon called ATAA. Also, in COI gene, *B. mori* uses CGA as the translation initiation codon, specifically, and ND5 in *A. gambiae* uses GTG as the translation initiation codon. Thus, among those protein-coding genes of insect mt genome, COI and ND5 use special codons among insect species (Table 3). However, molecular biology studies have not been reported on these genes. In the case of termination codon, all of the genes in other taxonomic groups use TAA codon as mt genome. However, it has been reported that some unstable termination codons are rearranged due to the post-transcriptional polyadenylation process, forming complete TAA (Clayton, 1984, 1991; Shadel, 1993).

Transcriptional-termination signal of 16S rRNA gene

Generally, the expression of mt gene goes through the polycistronic process where several genes in L-strand promoter (LSP) and H-strand promoter (HSP) located on D-loop near the replication origin transcribed at once (Chang and Clayton, 1987; Doda *et al.*, 1981). The transcription and translation of mt genes seem to be accomplished through the similar process also in the case of insects. However, based on the transcription termination signal of rRNA gene in human mt genome, the rRNA gene termination signal was confirmed in insect mt genome. As shown in Fig. 3, the transcription termination signal in insect mt 16S rRNA gene is highly conserved within $tRNA^{Leu}$ gene in insects, which is located in downstream of 16S rRNA genes as in vertebrates, although other taxonomic groups are conserved within COI, ND3, and non-coding region, respectively. Thus, the transcription termination signal of 16S rRNA gene is suggested to be related more with the conservation of transcription termination signal rather than the location of tRNA located in the downstream of rRNA and structural conservation.

Summary

Other than what has been discussed above, many characteristics in insect mt genomes can be analyzed and found. However, not many studies are at present dealing with interesting aspects of insect mt genomes in depth due maybe to the fact that the mt genome study has maily been focused on molecular evoulation, and phylogenetics, species identification. For an example, the total nucleotide sequence of mt genome of *A. gambia* and *A. quadrimaculatus* was determined in order to analyze the genetic relationship of the mosquito that is the host insect transmitting malaria. Thus, from the aspect of understanding molecular

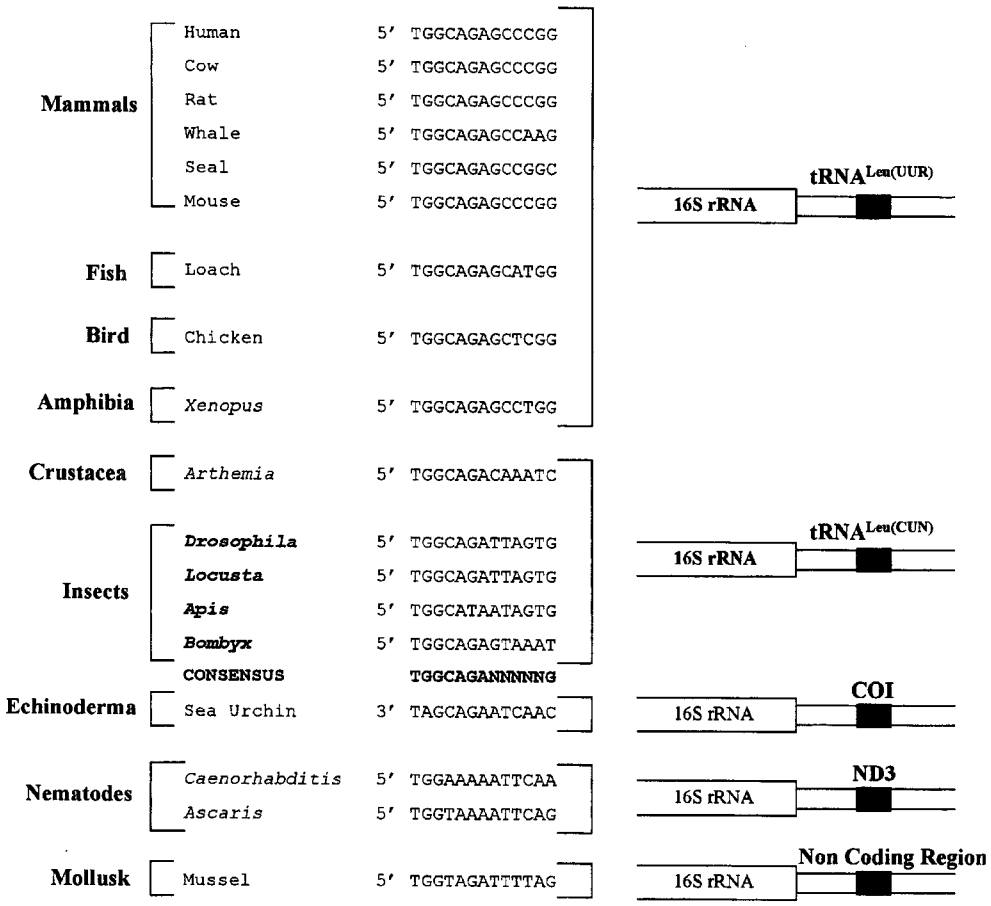

Fig. 3. Conservation of the termination signal of 16S rRNA genes in insect mitochondrial genomes. A heptanucleotide sequence with high similarity to the first seven nucleotide of the insect transcriptional signal can be found in downstream of 16S rRNA genes in a variety of mitochondrial genomes representing different phyla. The gene or region of the mitochondrial genome adjacent to the large rRNA (16S rRNA) is schematically shown at right; the position of the conserved sequence is closed box. A consensus sequence has been deduced for vertebrates, arthropod, and echinoderm; the first seven nucleotides are maintained in other phyla with one to three mismatches.

evolution, extensive analysis on overall mt genome would give many phylogenetic data since the extent of understanding molecular-evolution in living organisms could be broadened from the fact that mt genome in overall could be analyzed in depth beyond the dimension of understanding with each gene and the advantage of analyzing from various view points. However, considering the many insect species, which takes up most of existing living species, the usefulness and application of mt genome in the 7 species that have been analyzed until now are still insufficient.

References

Anderson, S., M. H. L. Bruijn, A. R. Clulson, I. C. Eperon, F. Sanfer and I. G. Young (1982) Complete sequence of bovine mitochondrial DNA: conserved features of the mammalian mitochondrial genome. *J. Mol. Biol.* **156**, 683-717.

Anderson, S., A. T. Bankier, B. G. Barrell, M. H. L. Brujin, A. R. Coulson, J. Drouin, I. C. Eperson, D. P. Nierlich, B. A. Roe, F. Sanger, P. H. Schreier, A. J. H. Smith, R. Staden and I. G. Young (1981) Sequence and organization of the human mitchondrial genome. *Nature* **290**, 457-465.

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

Brown, W. M. (1985) The mitochondrial genome of animals; in *Molecular evolutionary genetics*. Maclntyre, R. J. (ed), pp 95-130, Plenum Press, New York.

Cann, R. L., M. Stoneking and A. C. Wilson (1987) Mitochondrial DNA and human evolution. *Nature* **325**, 31-36.

Chang, D. D. and D. A. Clayton (1987) A mamalian mitochon-

drial RNA processing activity contains nucleus encoded RNA. *Science* **235**, 1178-1184.

Clayton, D. A. (1984) Transcription of the mammalian mitochondrial genome. *Annu. Rev. Biochem.* **53**, 573-594.

Clayton, D. A. (1991) Replication and transcription of vertebrate mitochondrial DNA. *Annu. Rev. Cell Biol.* **7**, 453-478.

Clayton, D. A. and C. A. Smith (1975) Complex mitochondrial DNA. *Int. Rev. Exp. Pathol.* **14**, 1-67.

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

Dairaghi, D. J., G. S. Shadel and D. A. Clayton (1995) Addition of a 29 residue carboxyl terminal tail converts a simple HMG box containing protein into a transcriptional activator. *J. Mol. Biol.* **249**, 11-28.

Clary, D. O. and D. R. Wolstenholmn (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.* **22**, 116-125.

Cornuet, 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. and Y. C. Crozier (1993) The mitochondrial genome of the honey bee *Apis mellifera* complete sequence and genome organization. *Genetics* **133**, 97-117.

Debruijn, M. H. L. (1983) *Drosophila* melanogaster mitochondrial DNA, a novel organization and genetic code. *Nature* **304**, 234-241.

Doda, J. N., C. T. Wright and D. A. Clayton (1981) Elongation of displacement loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. *Proc. Natl. Acad. Sci. USA* **78**, 6116-6120.

Fauron, C. M. R. and D. R. Wolstenholmn (1980) Extensive diversity among *Drosophila* species with respect to nucleotide sequences within the adenine + thymine rich region of mitochondrial DNA molecules. *Nucleic Acids Res.* **8**, 2439-2452.

Fisher, R. P., T. Lisowsky, M. A. Parisi and D. A. Clayton (1992) DNA wrapping and bending by a mitochondrial high mobility group like transcriptional activator protein. *J. Biol. Chem.* **267**, 3358-3367.

Flook, P. I., C. H. Rowell and G. Gellissen (1995) The sequence, organization, and evolution of the *Locusta migratoria* mitochondrial genome. *J. Mol. Evol.* **41**, 928-941.

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

Kruse, B., N. Narasimhan and G. Attardi (1989) Termination of transcription in human mitochondria : identification and purification of a DNA binding protein factor that promotes termination. *Cell* **58**, 391-97.

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

Moritz, C. and W. M. Brown (1987) Tandem duplication in animal mitochondrial DNAs: variation in incidence and gene content among lizards. *Proc. Natl. Acad. Sci. USA* **84**, 7183-7187.

Nei, M. and R. K. Kohen (1983) Evolution of animal mitochondrial DNA; in *Evolution of genes and proteins*. Sinauer Associates, Sunderland, MA, USA.

Ojala, D. C. Merkel, R. Gelfand and G. Attardi (1980) The tRNA genes punctuate the reading of genetic information in human mitochondrial DNA. *Cell* **22**, 393-403.

Ojala, D., J. Montoya and G. Attardi (1981) tRNA punctuation model of RNA processing in human mitochondria. *Nature* **290**, 470-474.

Parisi, M. A. and D. A. Clayton (1991) Similarity of human mitochondrial transcription factor 1 to high mobility group proteins. *Science* **252**, 965-969.

Rand, D. M. (1993) Endoderms, ectotherms, and mitochondrial genome size variation. *J. Mol. Evol.* **37**, 281-295.

Shadel, G. S. and D. A. Clayton (1993) Mitochondrial transcription initiation. Variation and conservation. *J. Biol. Chem.* **268**, 16083-16086.

Strachan, O. and A. P. Read (1996) Human Molecular Genetics. BIOS Scientific Publisher Ltd., Magdalen Road, Oxford, UK.

Wolstenholmn, D. R. (1992) Animal mitochondrial DNA: structure and evolution. *Int. Rev. Cytol.* **141**, 173-216.