MOLECULAR EVOLUTION AMONG SEVERAL ORDERS OF INSECTS BASED ON MITOCHONDRIAL DNA ANALYSIS

by

Hong Liu

B.Sc. East China Institute of Chemical Technology

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in the Department

of

Biological Sciences

SIMON FRASER UNIVERSITY

June 1993

APPROVAL

Name: **HONG LIU**

Degree: **Doctor of Philosophy**

Title of Thesis:

MOLECULAR EVOLUTION AMONG SEVERAL ORDERS OF INSECTS BASED ON MITOCHONDRIAL DNA ANALYSIS

Examining Committee:

Chair: Dr. B.P. Brandhorst, Professor

Dr. A.T. Beckenbach, Associate Professor, Senior Supervisor,
Department of Biological Sciences, SFU

Dr. M.J. Smith, Professor,
Department of Biological Sciences, SFU

Dr. D.L. Baillie, Professor,
Department of Biological Sciences, SFU

Dr. F. Breden, Associate Professor,
Department of Biological Sciences, SFU
Public Examiner

Dr. R.G. Harrison, Professor,
Section of Ecology and Systematics, Cornell University,
Ithaca, New York
External Examiner

Date Approved 19 August 1993

PARTIAL COPYRIGHT LICENSE

I hereby grant to Simon Fraser University the right to lend my thesis, project or extended essay (the title of which is shown below) to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users. I further agree that permission for multiple copying of this work for scholarly purposes may be granted by me or the Dean of Graduate Studies. It is understood that copying or publication of this work for financial gain shall not be allowed without my written permission.

Title of Thesis/Project/Extended Essay

Molecular Evolution Among Several Orders Of Insects Based On Mitochondrial DNA Analysis

Author: _____
(signature)

HONG LIU
(name)

Aug 10, 1993
(date)

Abstract

There are three major parts in my thesis. In the first part, I examined the complete nucleotide sequences for the mitochondrial cytochrome oxidase II gene of 13 species of insects, representing ten orders. The genes range from 673 to 690 bp in length, encoding 226 to 229 amino acids. Several insertion or deletion events can be observed, each involving one or two codons. The 3' end of the gene is extremely variable, both in length and sequence, making alignment of the ends unreliable. Using the first 639 nucleotide positions, for which unambiguous alignments could be obtained, I examined the neighbor-joining trees based on nucleotide divergences, and based on conserved subsets of that data, including transversion, amino acid and second codon position divergences. Each of these subsets produces different trees, none of which can be easily reconciled with trees constructed using morphology and the fossil record. Bootstrap analysis using second codon positions strongly supports affinities between the Order Blatteria (cockroaches) and Order Isoptera (termites) and between a wasp and honeybee (Order Hymenoptera). The divergence of insect orders is very ancient and may have occurred too rapidly for easy resolution using mitochondrial protein sequences. Unambiguous resolution of insect orders will probably require analysis of many additional taxa, using the COII gene and other conserved sequences.

In the second part, I examined the evolution of three mitochondrial tRNA genes. In view of the apparent deep phylogeny of the insects orders, I designed a set of primers to amplify mtDNA fragments encompassing clustered tRNA genes from various insects by the polymerase chain reaction. The region includes three tRNA genes ($tRNA^{trp}$, $tRNA^{cys}$ and $tRNA^{tyr}$) and 5' of the COI gene in *Drosophila yakuba*. For the insects

(holometabolous and hemimetabolous) I examined, they all have the same primitive tRNA gene order as *D. yakuba*. The recently published honeybee sequence has the different gene order compared with the primitive one. Two scenarios has been described to explain this process. By comparing the 5' of COI genes from a broad range of species, I have assigned two putative start codons (TTG and TTA) for the COI genes of some insects. The small tRNA sequence data set cannot give easy resolution for the ancient divergence of the insects studied here at the order level.

The third part of my thesis deals with the mitochondrial gene order of the flea, *Ctenocephalides felis*. I have determined most of the flea mt-genome organization by PCR and direct sequencing of the gene junctions. All together, the locations of 24 genes have been determined, including 15 tRNA genes, 8 protein genes and the large ribosomal subunit. So far, the flea mt-genome order is the same as the fly, *Drosophila yakuba*. The order in which the genes are arranged in mtDNA appears to change at a very slow rate. This may provide us very useful information about the phylogeny of distantly related metazoan groups.

ACKNOWLEDGMENTS

I would like to thank A. Syed for providing some of the insects used in this study. I am indebted to many graduate students and friends for their supports and helpful discussions throughout my time at Simon Fraser University, especially Karen Beckenbach, Dr. Eric Cabot, Tammy Laberge and Cam Muir. I would like to express my gratitude to the members of my committee: Dr. Beckenbach, Dr. Smith, Dr. Baillie, Dr. Breden and Dr. Harrison. Their comments, corrections, suggestions, and discussions have improved my thesis greatly.

Especially, I thank my senior supervisor Dr. Andy Beckenbach for his inspiration, guidance and continuous support.

I dedicate this work to my parents.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION AND REVIEW

Review of Insect Phylogeny

Introduction:

Insects are among the most abundant and successful of terrestrial animals. They include about three-fourths of all the described species of animals, and they have become adapted to a great range of environments, from high latitudes to the equator, from rainforest to desert, from mountains to the shore, and to varied ways of living.

Although insects are very important to the terrestrial ecosystems and are widely used as model systems in genetic and developmental studies, the phylogenetic history of many groups remains controversial (Kristensen, 1989). The higher-order phylogeny of the Class Insecta has become one of the most important unresolved questions in evolutionary biology. Kristensen (1989) notes that "there are no more than two cases where terminal, order-level, sister group relationships seem supported beyond any doubt". Higher-order insect relationships are controversial because of the wide range of morphological diversity found among member taxa, and their long evolutionary history (Matsuda, 1976). The fossil record has been of little use in elucidating insect phylogeny in general, as insects tend to be preserved as wings only and the first fossil insects appear after many of the orders are already extant (Hennig, 1981).

In approaching the study of the evolutionary relationships among insects, it is necessary to mention some of the principles on which phylogenetic inferences should be

based. In the past these principles have differed more or less radically from one group of biologists to another and it is unlikely that general agreement will be reached in the near future. However, there is growing acceptance of the view, first developed by Hennig during the 1950s, that monophyletic groups (i.e. those consisting of all the descendants of a common ancestor) are best recognized by the fact that their members share a number of specialized characteristics (synapomorphies). Shared primitive features on the other hand (symplesiomorphies) are not convincing proof of a common origin, since such characteristics may simply have been retained from more remote ancestral forms. It follows that attempts to infer phylogeny without distinguishing between primitive (plesiomorphic) and specialized (apomorphic) features are unsatisfactory in principle.

How synapomorphic characters can be identified and how they may be distinguished from the resemblances due to convergent evolution is, however, not always easy to decide in practice, with the result that even those who accept Hennig's general principle may differ considerably in the schemes of evolutionary relationship that they devise. Despite the difficulties which arise when one wishes to express a purely phylogenetic scheme in the form of a classification, cladistic arguments, relying on convincing sets of synapomorphic characters, seem destined to play an important part in defining evolutionary relationships. In the case of the insects this role is likely to be especially important, since the fossil record, though of great interest, is far too imperfect to allow direct recognition of ancestor-descendant connections (Carpenter and Burnham, 1985).

Phylogeny of the insects:

Apterygote Relationships: Comparative structural data suggest that the primary evolutionary division among the insects is into a group of Entognathan orders (the Diplura, Protura and Collembola) on the one hand, and all remaining insects on the other (Manton, 1979). Table 1 gives the classification of the Class Insecta with the common names for each orders of the insects. The argument for the monophyly of the Entognatha depends largely on the characteristic way in which the mouthparts are partly or entirely sunk into a gnathal pouch, formed through overgrowth of cephalic folds and from which the tips of the mandibles and maxillae can be protruded for feeding.

The sister-group of the Entognatha consists of all the remaining insects and can be referred to as the Ectognatha. There is little doubt that it is a monophyletic complex, as it shows a large number of synapomorphies. The dominant group of Ectognathan insects is, of course, the Pterygota (winged insects and their secondarily apterous descendants), whose monophyletic character is established by the common possession in their ground-plan of wings and associated thoracic modifications.

Phylogeny of the Pterygota: Fossil remains of winged insects are first known from the lower part of the Upper Carboniferous about 350 million years (Myr) ago (Carpenter and Burnham,1985). It is reasonable to suppose that the Pterygota arose some time before this, perhaps as early as the Devonian. Unfortunately, nothing certain is known of the origin of insect wings and flight. The Permian period was a time of considerable radiation among the winged insects, as illustrated by many species described from important localities of this age. By the end of the Palaeozoic the extant orders Ephemeroptera, Odonata, Orthoptera, Dictyoptera, Plecoptera, Psocoptera and

Table 1. Classification of Class Insecta.

	Orders
Entognatha	
	Protura (Proturans)
	Diplura (Japygids)
	Collembola (Springtails)
Ectognatha	
Apterygota	
	Archaeognatha (Bristle tails)
	Thysanura (Silverfish)
Pterygota	
Palaeoptera	
	Ephemeroptera (Mayflies)
	Odonata (Dragonflies)
Neoptera	
Polyneoptera	
	Dermaptera (Earwigs)
	Grylloblattodea (Grylloblattids)
	Isoptera (Termites)
	Blattodea (Cockroaches)
	Mantodea (Mantids)
	Phasmida (Stick and leaf insects)
	Orthoptera (Crickets and locusts)
	Embioptera (Web spinners)
	Plecoptera (Stoneflies)
Paraneoptera	
	Psocoptera (Booklice)
	Phthiraptera (Chewing lice)
	Hemiptera (True bugs)
	Thysanoptera (Thrips)
	Zoraptera (Zorapterans)
Holometabola	
	Neuroptera (Licewings)
	Coleoptera (Beetles and weevils)
	Mecoptera (Scorpion flies)
	Trichoptera (Caddisflies)
	Lepidoptera (Moths)
	Diptera (Two-winged flies)
	Siphonaptera (Fleas)
	Strepsiptera (Strepsipterans)
	Hymenoptera (Ants and wasps)
	Megaloptera (Alderflies)

Hemiptera had all come into being, as well as the holometabolous Coleoptera, Mecoptera, Trichoptera and perhaps Neuroptera. In addition to these a considerable number of purely fossil orders have been recognized from the Palaeozoic, though any attempt at incorporating them into a general scheme of Pterygote phylogeny is severely restricted by the fact that the fossils usually consist of wing impressions only, with relatively little information available on the structure of other parts of the body.

In the absence of compelling paleontological evidence on the phylogeny of pterygote insects it is necessary to infer phylogeny from compansons of living taxa. This can only be done effectively by examining data for the recent orders, even though one is often limited by the inadequate morphological information available for some groups. One focus of discussion has been relationships among the Ephemeroptera, Odonata and Neoptera. There are three logically possible phylogenetic relationships, and all have been supported by one authority or another (Hennig, 1981; Boudreaux, 1979; Kristensen, 1981).

The identification of major monophyletic groups within the Neoptera also presents some difficulties. Traditionally it has been the custom to recognize three major groups, Polyneoptera, Paraneoptera and Oligoneoptera (Kristensen, 1981). The first of these is the complex of 'Orthopteroid' orders, represented today by the Plecoptera, Grylloblattodea, Orthoptera, Phasmida, Dermaptera, Embioptera, Dictyoptera and Isoptera. The features which they (or at least their more generalized representatives) have in common are the well-developed mandibulate mouthparts, the enlarged anal lobe of the hind wing, and retention of the appendages of the eleventh abdominal segment as cerci. Unfortunately, all of these characteristics are likely to be plesiomorphic

Neopteran features, so there is no convincing proof that the Polyneopteran orders form a natural, monophyletic group.

The Paraneoptera or Hemipteroid orders (Psocoptera, Phthiraptera, Hemiptera and Thysanoptera), on the other hand, are very likely to be a monophyletic group as they share a number of specialized features, e.g. the reduced venation and absence of an enlarged anal lobe in the hind wing; reduction of tarsal subsegments to three or fewer; the absence of cerci. Together they appear to form a sister-group of the Zoraptera and they seem to fall into two monophyletic groups. One of these, consisting of the Psocoptera and the Phthiraptera, has many synapomorphic features and may be further subdivided cladistically. The other, consisting of the Hemiptera and Thysanoptera, is less convincingly monophyletic, though it is difficult to know where else the Thysanoptera could be placed (Boudreaux, 1979).

The third major phyletic branch of the Neopteran insects is composed of the insect orders with a complete metamorphosis, known as the Endopterygota or Holometabola. The characteristic metamorphosis, with larval stages followed by a single pupal instar, is their outstanding synapomorphy and is supported by the specialized larval eyes and the internal development of wings and external genitalia. There is no doubt that this is a monophyletic group, but there is no clear indication of its sister-group and nothing is known of its origin, apart from the fact that it must have occurred early in insect evolution (Kristensen, 1981; Boudreaux, 1979). Permian deposits already contain fossils that may be referred to the Neuroptera, Coleoptera, Mecoptera and Trichoptera. The Trichoptera and Lepidoptera have long been regarded as close relatives, though it is only recently that the full extent of their synapomorphies has been revealed. They are

now known to share over 20 specialized features in their ground-plan. The Mecoptera, and Diptera, together with the Siphonaptera, also seem to constitute a monophyletic group, though with fewer synapomorphies. The Trichoptera + Lepidoptera form a sister-group of the Mecoptera + Diptera + Siphonaptera, the whole assemblage of five orders forming the monophyletic Panorpoid complex (also known as the superorder Panorpida). The affinities of the Coleoptera, however, are far from clear, though they may share a few specializations with the Neuroptera. Whether the Strepsiptera are really close allies of the Coleoptera is also undecided. The Hymenoptera, too, are an isolated group of uncertain affinities (Kristensen, 1981).

Molecular Systematics:

Molecular data have powerful advantages over morphological information, especially in terms of large number of characters, widespread distribution (when nucleic acids are used), and a much better approximation of their evolution to that expected from simple models (Hillis and Moritz, 1990). But these advantages do not mean the end of morphological studies. The explanation of morphological evolution remains a major rationale for phylogenetic studies and the information from fossils remains crucial to understanding evolution.

We can use molecular data to reconstruct phylogenetic trees for all living things. Recent achievements in molecular biology, such as restriction endonuclease fragment analysis, gene cloning, polymerase chain reaction (PCR) and direct sequencing have placed molecular systematics in a new and exciting position (Innis et al., 1990). Traditional fields of evolutionary biology, such as comparative anatomy, morphology and

paleontology restrict themselves mostly to the study of evolutionary relationships among closely related organisms. Using molecular data, we can now build huge family trees connecting vertebrates, insects, plants, fungi and bacteria (Field et al., 1988, Raff et al., 1989).

Review of Animal Mitochondrial DNA

General Features:

Mitochondrial DNA (mtDNA) has been used as a powerful tool for evolutionary studies of animals in the past decade (Moritz et al., 1987). Animal mtDNA has several general features. All the animal mitochondrial genomes (mt-genome) are in the form of a single circular DNA molecule with only one exception. The exception is the mt-genome of the cnidrian *Hydra attenuate*, which comprises two unique 8-kb linear DNA molecules (Warrior and Gall, 1985). The gene content of the animal mtDNA appears to be conserved, although it ranges in size from 14 kb (the nematode, *Caenorhabditis elegans*; Okimoto et al., 1992) to 42 kb (the scallop, *Placeopecten megellanicus*; LaRoche et al., 1990). In animal mtDNA molecule, there are 2 ribosomal RNA (rRNA) genes, 22 transfer RNA (tRNA) genes, and 13 protein genes which code for subunits of enzymes functioning in electron transport or ATP synthesis (Moritz et al., 1987). The protein coding genes are the subunits I, II and III of cytochrome oxidase, subunits 6 and 8 of ATPase, cytochrome b and seven subunits of the NADH reductase system. The ATPase 8 gene has been lost from nematode mtDNA (Wolstenholme et al., 1987) and the Mussel, Mytilis (Hoffmann et al., 1992).

In all metazoan mtDNA molecules sequenced there are very few or no nucleotides between genes. However, there is a single noncoding region that in some vertebrates has been shown to contain sequences essential for the initiation of transcription and replication (Clayton, 1984), and therefore has been designated the control region. Some of the size variation among metazoan mtDNAs results from differences in gene lengths. However, the greatest size variations are attributable to

differences in the length of the control region, some of which contain repeated sequences. In some species, mitochondrial DNA molecules that are of different size or that contain sequence differences are sometimes found in a single individual, a condition known as heteroplasmy (Rand and Harrison, 1989; Brown et al., 1992).

Metazoan mtDNAs have many genetic novelties in contrast to the relative uniformity of gene content. Metazoan mt-genetic codes are highly modified compared to all known genetic codes. At least six unorthodox translation initiation codons are used in transcripts of metazoan mt-protein genes. After primary transcripts of entire mtDNA strands, each transcript is precisely cleaved and have no or few upstream and downstream nontranslated nucleotides. In some cases, gene transcripts end in U or UA that are polyadenylated to provide complete translation termination codons (Okimoto et al., 1992).

The 22 tRNA genes of the animal mitochondrial DNA are generally smaller than their counterparts and lack many of the invariable features (Wilson et al., 1985). Nevertheless, mitochondrial tRNA genes evolve slower than mitochondrial protein coding genes (Bibb et al., 1981; Wolstenholme and Clary, 1985), indicating that they are structurally or functionally more constrained than these protein genes. Ribosomal RNA genes in general have been characterized as highly conserved (Moritz et al., 1987). However, when compared to their nuclear counterparts the mitochondrial ribosomal RNA genes evolve much faster.

The complementary strands of mammalian and some other vertebrate and invertebrate mtDNA molecules differ sufficiently in guanine and thymine content that they can be separated in alkaline cesium chloride gradients. The complementary strands

of these mtDNA molecules thus acquired the designations heavy (H) and light (L) strands that have been used as strand definitions in replication and transcription studies of mammalian mt-genomes (Clayton, 1984, 1991).

Metazoan mitochondrial DNA molecules are mostly maternally inherited, although a few cases of paternal leakage have been reported (Kondo et al., 1990; Hoeh et al., 1991), and there is no evidence for recombination between molecules (Moritz et al., 1987). In vertebrates, and perhaps invertebrates, as well, nucleotide substitutions occur at a much higher rate in mtDNA than in nuclear DNA (Brown, 1985; Vawter and Brown, 1986; DeSalle et al., 1987).

Mitochondrial DNA has some useful characteristics for phylogenetic study (Wilson et al., 1985). These include ease of extraction and manipulation, the simplicity of the molecule, lack of recombination and high mutation rate. But high mutation rate is not necessarily an advantage for studies at high level (e.g. orders of insects). The rate of animal mtDNA evolution varies among lineages, among genes and within genes. The divergence rate of a particular segment of DNA determines the systematic level at which it is appropriately used for phylogenetic study. For example, an investigation of closely related species would be better carried out with highly variable regions such as the control region or sequences from ND6, rather than with the more conserved cytochrome b gene. On the other hand, cytochrome b or COI would be more suitable for looking at differences between orders of insects.

The Rate of mtDNA Evolution Varies Among Gene Regions

There is considerable variation in the rate of evolutionary change between mtDNA regions, depending on the functional role of each. The control region is highly

variable, both within and between species (Brown, 1985). The ribosomal RNA genes, on the other hand, show great conservation between taxa (Hixson and Brown, 1986).

Control Region:

The control region is called A+T rich region in insects (Clary and Wolstenholme, 1985) and D-loop region in vertebrates (Upholt and David, 1977). It contains an origin of replication for the mitochondrial DNA as well as signals for the initiation of transcription. Like some nuclear non-coding regions (e.g. introns, spacer regions, or pseudogenes), the control region does not code for any gene products. It has been shown that sequence variation in the control regions may be useful for phylogenetic studies below the species level (Thomas et al., 1990), because there are few functional constraints (Li et al., 1984; Hillis and Davis, 1986). Most of the size variation that has been detected in mtDNA is contributed by length differences within the control region (Moritz et al., 1987).

Protein Coding Genes:

Animal mtDNA contains 12-13 protein genes (Brown 1985; Wolstenholme et al., 1987; Hoffmann et al., 1992). The protein coding genes are under greater functional constraint than is observed in the control region, because of the need to code for essential components of the electron transport system. Nonetheless, because of the degeneracy of the code, and because not all regions of a polypeptide chain are under stringent functional constraint, protein coding genes provide an intermediate level of variation.

There is a range of evolutionary rates observable among the protein coding genes. In insects, the COI gene is the most conserved, while the ND6 gene is the least

conserved (Clary and Wolstenholme, 1985; Crozier and Crozier, 1993). The rate of evolution is also variable at different codon positions. The third codon position is most variable since transitions do not alter the amino acid sequence of the resulting protein (de Bruijn, 1983). The second codon position shows the slowest rate of evolution, because all substitutions result in chemically important replacements.

tRNA Genes:

The tRNA genes represent a third level of evolutionary rate. The animal mitochondrial genome usually contains 22 tRNA genes. They are generally smaller than nuclear tRNA genes and lack many features that are invariable in nuclear tRNA (Wilson et al., 1985). The tRNA genes are more conserved than the mt-protein genes, indicating stronger structural constraints on them (Bibb et al., 1981; Cantatore et al., 1987). Even within a tRNA gene there are differences in the degree of structural constraints. Each tRNA has as many as four paired stem regions, where substitutions involving only one of the paired bases can disrupt the stability of the molecule. In addition, there are unpaired loop regions which differ in their degree of variability. In comparisons of tRNA genes among vertebrates, the most conserved region is the anticodon loop, followed by the D (dihydrouridine) stem (Gadaleta et al., 1989). The T (TψC) loop and stem are the most variable regions. My result (Chapter III) indicates that the T loop and D loop are the most variable regions for 7 species of insects. Thomas and Beckenbach (1989) examined $tRNA^{Arg}$ and $tRNA^{Gly}$ in closely related salmonoid species and found that variation is concentrated in the D loop. Mitochondrial tRNAs are missing entire loops in nematodes (Wolstenholme et al., 1987).

Ribosomal Genes:

Ribosomal RNA genes in general are highly conserved (Noeller and Woese, 1981; Hixson and Brown, 1986). It is possible to align ribosomal DNA sequences for eukaryotes, archebacteria, eubacteria, plastids and mitochondria (Dams et al. 1988). However, the entire ribosomal gene is not highly conserved; it has some conserved and variable segments as well. In mitochondrial rRNA genes, the structural and functional constraints determine the level of conservation (Simon et al, 1990). For example, the site of tRNA attachment is one of the most conserved segments in the small ribosomal RNA gene.

Wheeler and Honeycutt (1988) discussed the systematic implications of the level of conservation of helical stems versus unpaired loops for 5S and 5.8S rRNA. By analyzing selected arthropods and gastropods, they suggested that trees constructed from unpaired bases were more concordant with trees based on morphological data than those constructed from paired bases.

Genome Organization:

Gene order within phyla:

Although the nucleotide sequence of mtDNA changes at a much higher rate than that of nuclear DNA in vertebrates, the order in which the genes are arranged appears to change at a much slower rate (Brown, 1985). For example, there is one basic gene organization among chordate mtDNA (Anderson et al, 1981,1982; Bibb et al, 1981; Gadaleta et al, 1989; Roe et al, 1985; Johansen, Guddal and Johansen, 1990) with only minor variations (Paabo et al, 1991; Desjardins and Morais, 1990,1991). The mitochondrial gene order is also stable among echinoderm classes (Jacobs et al, 1988;

Cantatore et al, 1989; De Giorgi et al, 1991; Himeno et al, 1987; Smith et al, 1989,1990). However, gene order variation still occurs within phyla. Among vertebrates, birds differ from mammals and amphibians in gene order, suggesting that the transposition of a segment containing one tRNA and one protein-encoding gene occurred in an ancestor of class Aves. Within mammals, marsupials differ from eutherians in the positions of some tRNAs. Smith et al. (1990) have sequenced an 8.0 kb fragment from a sea star (*Pisaster ochraceus*) mitochondrial genome. This sequence confirmed an earlier observation that the sea star mitochondrial genome contains an inversion of approximately 4.6 kb. Within insects, 11 of the 22 tRNA genes are known to vary in position both between the orders Diptera and Hymenoptera (Crozier and Crozier, 1993) and within the order Diptera, with differences observed between *Aedes* and *Drosophila* (Clary and Wolstenholme, 1985).

Gene order among phyla:

The complete gene orders of blue mussel (*Mytilus edulis*) from phylum Mollusca (Hoffmann et al, 1992) and earthworm (*Lumbricus terrestris*) from phylum Annelida (Boore and Brown, personal communication) have been determined. Comparisons of the gene order among *D. yakuba, M. edulis* and *L. terrestris* show very unexpected results. There are extensive gene rearrangements among these three species representing three related phyla.

The conservation of mitochondrial gene organization within a phylum suggests that the arrangements of genes, particularly rRNA and tRNA genes, is important for regulation of gene expression (Montoya, Gaines and Attardi 1983).

Thesis Outline

In this thesis I examine mtDNA sequence variation in members of the Class Insecta. In Chapter II I examine the evolution of the mitochondrial cytochrome oxidase II gene among 10 orders of insects. In Chapter III I examine the evolution of three mitochondrial tRNA genes and the putative start codons of COI gene in five orders of insects. In Chapter IV I extend this method to determine the gene organization of a flea (Order Siphonaptera).

CHAPTER II

EVOLUTION OF THE MITOCHONDRIAL CYTOCHROME OXIDASE II GENE AMONG 10 ORDERS OF INSECTS

Introduction

Insects comprise by far the greatest proportion of described animal species. Fossil evidence concerning their origins is scanty, but offers some insight into the radiation of the major groups. A number of authors have combined cladistic analysis of morphological and developmental characters with the available fossil record to determine a likely sequence of events (Wooten, 1981; Hennig, 1981; Boudreaux, 1979; Kristensen, 1981). The first insect fossils appear in the upper Carboniferous, about 325 million years (Myr) ago (Wooten, 1981). By that time they had already undergone substantial radiation.

It is generally believed that the Paleoptera, represented by order Odonata (dragonflies) in this study, had separated from the Neoptera (which includes the other nine orders examined here) by the end of the Ordovician, more than 400 Myr ago (Hennig, 1981). The most common insect fossils of the Carboniferous belong to the Orthopteroid groups, including the Blatteria (cockroaches) and the Orthoptera (crickets and grasshoppers) (Wooten, 1981). Fossils assignable to the Hemiptera (true bugs) and Coleoptera (beetles) appear during the Permian. The Diptera (true flies) and Hymenoptera (wasps, ants and bees) do not appear until the Triassic (Hennig, 1981). For the Isoptera (termites), Siphonaptera (fleas) and Lepidoptera (butterflies and moths), the fossil record provides few clues concerning their origins. Perhaps the firmest conclusions that can be reached are that the radiation of extant insect orders occurred

over a span of about 70 Myr, and was essentially complete by the start of the Mesozoic, about 280 Myr ago (Hennig, 1981).

Cladistic analysis suggests that the Holometabolous insects (those undergoing complete metamorphosis, including among others, Coleoptera, Hymenoptera, Lepidoptera, Diptera and Siphonaptera) form a monophyletic group (Wooten, 1981; Hennig, 1981; Boudreaux, 1979; Kristensen, 1981). This belief is based primarily on the assumption that the insertion of a pupal stage between the larval and adult stages could not have arisen more than once (Hennig, 1981).

In the absence of a definitive fossil record, molecular data may provide valuable insight into the origins of these groups. It is clear that the branchings are quite deep, that it will be necessary to examine highly conserved sequences or sites, or perhaps gene order in the mitochondrial genome (Moritz et al., 1987). To determine whether sequence comparisons of a conserved mitochondrial gene, cytochrome oxidase II (COII), could clarify some of these relationships, I have sequenced the gene from representative members of ten orders of insects, using PCR amplification of the complete gene and direct sequencing.

Materials and Methods

DNA Sources

A single individual of each species, either from stocks maintained at the Simon Fraser University Insectary or freshly caught, was used as source material for DNA extraction. Species names and the orders to which they belong are shown in Table 2.

DNA Extraction

A small amount of tissue from each insect was ground in 60 μl protease buffer (0.1M Tris-OH, pH 8.0, 0.05 M EDTA, 0.2 M NaCl, 1% SDS, with 0.4 mg/ml protease K), and immediately placed at 65°C for 3 min. The resulting solution was extracted once with phenol (saturated with 10 mM Tris, 1 mM EDTA, pH 8.0) and once with chloroform/isoamyl alcohol (24:1). The supernatant was removed and 2 volumes of 95% ethanol was added to precipitate the DNA. The suspension was pelleted at 40K, 15 min, 4°C, washed twice with 70% ethanol, dried under vacuum, and dissolved in 50μl sterile ddH_2O.

Amplification and Sequencing

The COII gene was amplified using the polymerase chain reaction (PCR) performed in 50 μl reactions using the GeneAmp kit with AmpliTaq DNA polymerase (Perkin Elmer Cetus). I followed the protocol provided with the kit. Thermal cycling was done in an Ericomp Programmable Cyclic Reactor. The PCR consisted of 35 cycles performed as follows: denaturation at 94°C for 1 min, annealing at 50°C for 1 min, and extension at 70°C for 2 min. The primers for PCR amplification and sequencing are given in Table 3. Primers labelled 'A' are oriented 5' → 3' in the direction of transcription of the COII gene. Primers labelled 'B' are oriented in the opposite

Table 2. List of insect species studied in this chapter.

Order	Name
Blattaria	Cockroach (*Periplaneta americana*)
Coleoptera	Ladybird beetle (*Adalia bipunctata*)
Coleoptera	Grain Weevil (*Sitophilus granarius*)
Diptera	Fruit fly (*Drosophila pseudoobscura*)
Hemiptera	Milkweed bug (*Oncopeltus fasciatus*)
Hymenoptera	Wasp (*Excristes roborator*)
Hymenoptera	Ant (*Lasius sp.*)
Isoptera	Termite (*Zootermopsis angusticollis*)
Lepidoptera	Moth (*Galleria mellonella*)
Odonata	Dragonfly (*Sympetrum striolatum*)
Orthoptera	Cricket (*Acheta domesticus*)
Orthoptera	Locust (*Schistocerca gregaria*)
Siphonaptera	Flea (*Ctenocephalides felis*)

Table 3. Primers used in amplification and sequencing of insect COII mtDNA

Name of primer	Sequence	Positions spanned*
A-tLEU	5'-ATGGCAGATTAGTGCAATGG-3'	3018 - 3038
B-tLYS	5'-GTTTAAGAGACCAGTACTTG-3'	3804 - 3784
A-171	5'-GGTCAAACAATTGAGTCTATTTGAAC-3'	3255 - 3281
A-226	5'-ATTGCATTACCATCTCTACGATT-3'	3307 - 3330
A-298	5'-ATTGGACATCAATGATATTGA-3'	3379 - 3400
B-434	5'-GGTAAAACTACTCGATTATCAAC-3'	3516 - 3493
A-470	5'-AGATGTTCTTCACTCATGA-3'	3553 - 3572
A-592	5'-GAAATTTGTGGAGCAAATCATAG-3'	3673 - 3696

*The position of each primer refers to the map position in the *D. yakuba* mitochondrial genome sequence.

direction. The COII gene in *Drosophila* is flanked by $tRNA_{LEU}$ and $tRNA_{LYS}$ (Clary and Wolstenholme, 1985) and the end primers, A-tLEU and B-LYS, were based on the published sequences. The numbers refer to the position of the 5' nucleotide in the *Drosophila yakuba* COII sequence. The internal primers, used for sequencing only, were chosen in regions of sequence conservation between *D. yakuba* and other insects, based on partial sequences as they were obtained.

The double-stranded DNA was purified by electrophoresis in low melting agarose gel, followed by concentration using Amicon 30 microconcentrators. About 7μl was used for double-stranded sequencing using Sequenase kit (United States Biochemical Corporation). I used modifications of the standard Sequenase protocol given by Palumbi et al. (*The Simple Fool's Guide to PCR*, compiled by Palumbi, S., Martin, A., Romano, S., McMillan, W.O., Stice, L., and Grabowski, G.).

Data Analysis

Sequences were aligned and compared with the aid of eyeball sequence editor, ESEE (Cabot and Beckenbach, 1989). The alignment was done using translated sequences and subsequently applied to the nucleotide sequences. Translations to amino acid sequences used the *Drosophila* mitochondrial code (de Bruijn, 1983). I performed the analyses using distance matrices based on the complete nucleotide sequences, the inferred translated sequences, transversion differences, and second codon positions. Distances were corrected using the Jukes-Cantor method (Jukes and Cantor, 1969). Kimura two-parameter corrections (Kimura, 1980) were also calculated for comparisons involving the complete nucleotide sequences and the second codon positions. In the absence of independent estimates of transition and transversion mutation rates, the two-parameter

method uses the observed transition/transversion ratio. In the comparisons presented here, observed transversions exceed the transitions, so this method fails to adequately compensate. Only the Jukes-Cantor corrections are reported here, although they probably represent underestimates of the actual sequence divergences.

The distance trees were constructed by the neighbor-joining method (Saitou and Nei, 1987), using the NEIGHBOR program, version 3.4 of PHYLIP (Felsenstein, 1989). Bootstrap analyses were done on the second codon position data using both parsimony and distance methods. Parsimony bootstrapping was done using the DNABOOT program of PHYLIP, while the distance matrix bootstrapping was carried out using SEQBOOT, DNADIST, NEIGHBOR, and CONSENSE of the PHYLIP package (Felsenstein, 1989).

Availability of sequences.

The sequences have been deposited in the GenBank sequence database under accession numbers, M83959 - M83971.

Results

Nucleotide Sequences.

Fig. 1 shows the aligned nucleotide sequences of the cytochrome oxidase II genes (COII) from 13 species, representing ten orders of insects. The genes range from 673 bp in the wasp to 690 bp in the milkweed bug, and encode 226-229 amino acids. Most of the sequences have ATG for initiation, although ATA occurs at the beginning of the ant and milkweed bug sequences, and ATT appears to serve as initiator in both beetle sequences. In six of the sequences, only the T of the stop codon is coded in the DNA sequence, while seven of the species have TAA at the 3' end, potentially encoding the entire terminator.

Most of the variation in length of the COII genes occurs at or near the 3' end of the gene. Several internal insertion/deletion events can also be seen. In the ant sequence, a deletion of three nucleotides appears at about positions 16-19. In both the wasp and cricket sequences, a deletion of two codons occurs at approximately positions 390-395 in the nucleotide sequences. The regions corresponding to these deletions are subject to considerable variation at the amino acid level, so the alignments of those deletions cannot be made with absolute certainty. In particular, it is possible that the deletions in the wasp and cricket are independent events and may not involve identical codon positions. This two-codon deletion is also present in the honeybee sequence (Crozier et al., 1989), but is absent from the other Hymenopteran sequenced here, the ant. An insertion of three nucleotides occurs in the cricket sequence, at positions 349-351. This insertion is not shared by the locust, a member of the same order, nor by any other of the sequences examined here.

```
Drosophila  ATGTCTACAT GAGCTAATTT AGGTTTACAA GATAGAGCCT CTCCATTAAT GGAACAATTA ATTTTTTTTC ATGACCACGC ATTATTAATT  90
Moth        ...CGA..T. ..T..... TAA..... A....T.... ....TC.T.. A......A.T .......... ....T..TA. T...A.T...
Flea        ...AA...T. ..ATA.... TAA..... A.....AA.. .A..T..... A.......... ..A..... ..A.T..TT. .A.....
Weevil      ..T..... ..AAA..CC. CTT..... .....T. .A...C.... A....TTC.T ..A.GC..C. .....T.. TA..C....C
Lady Beetle ..T.....T. ..AAATC.AG GTTA..TTT. ..C...T.A. ..TTTC..T. A.......... CGA..... ....T..T.. ..........
Wasp        ...AG..... ..T...TAA. .AA.C.T... ...GCCAAT. .A..TA.... A......C.. .....C.... ....T..TA. C...A....C
Ant         ..AAA...T. ..TTA---C. TTCAC.T... A..TC.AAT. .....ACCTA T..TAT.A.. .......... ....TTT.A. TA..A.....
Milkweed Bug ..AG.A.... ..ATA...A. TAA.C.T... ..CGC.AAT. .AT..AC... A.....GC.T .CAA.A.... ....T...A. ....A.....
Cricket     ...G.C.... ..T.C..... .AA..... A.....T.A. ....TC.... A......C.C .....C.... .....TA. T...A.....
Locust      ...G.A.... ..T.A..... .TC.A.T... ...G....T. .A..... ...G..... TCA...C.... .....G.TCA TACTA.GG.C
Cockroach   ...A.A.... ....C...A. .AA..... .....A. .C..TA.T.. A.......... ....AC..C. .....T.. C...A....C
Termite     ...A.A.... ..T.ATG.A. .AA.C..... ..C.....A. .G...A.T.. A.......... .......... ....T...A. CC..A.....
Dragonfly   ...G.A.... .....C.AC. .AAC..T... ...GCG..A. .C...A.... A......C.C CAC.A...C. ....T..TA. .A.GA..G..

Drosophila  TTAGTAATAA TTACAGTATT AGTTGGATAT TTAATATTTA TATTATTTTT TAATAGTTAT GTAAATCGAT TTCTTTTACA TGGACAATTA 180
Moth        ...A.T.... ....TA.T.. ...A..... A.T...A... AC..... .....AA.T. A.T..... ..T..C.TGT .....A..
Flea        A.TT..T... .....A.T.. ...G..... A.T...AGAT CT.....A.A .....AA.TA TAT.....T. A.T.AC.TG. AA.....AAT
Weevil      ...A.CT... ....TA.CA. ...AA..C.A A..C.G..A. G.A..C.... .....AACT. TCTC....T. A...CC.CG. A..T.....
Lady Beetle ...AAT.... ......GTGC T..A.CT... A.T...A... GT...C.... .....AA..C AATC.C.... ..T.A...G. A..T..TAC.
Wasp        ...T..C... ....TA.TAC TA..ATT... A.T..T.C.T C.A.TA...A. A....A..TA ACT...AA.. ..A..A.... AAAT...ACC
Ant         ...A.CT.T. ....TC.TC. TA..CT..T. A.T..... CTA..A..AA .....A..TA A.T..... ..T.AC.... A.....T..C
Milkweed Bug ...ACT.... .....TC.A. T..AACT.T. A.T...G.AT C.A..ACAAC A....CA.TA A.T..... A.T.AC.TG. A..T...ACT
Cricket     C..C..... ....T..T.. .....C.... A.T....CC. .G..... C...CT...C ACT..... ..T.AC.CG. A......ACT
Locust      G..T..T... .....A. T..A..T... GCCC..AG.T ATA....A.. C...GCC... ACT..C..TA A.A.AC.T.. ......T...
Cockroach   A.TA.T.... ..TT.A..G. T..ATC.... A.....A..G C.A..G.... .....AA... A.T..... ..T.A..GG. A..T...A..
Termite     G.TAC..... ..CT.....C ...ATT.... ..T...ACAT C.A..ACCA. ...C.AGA.C .AG..C.... A.A.AC..G. A..C...AC.
Dragonfly   C.G.....T. ....TA.CA. ...A.CT... A.T...GGA. C.A..... .....AGG.. ..T.....G. A.....G. .....AA.

Drosophila  ATTGAAATAA TTTGAACAAT TTTACCGGCA ATTATTCTTT TATTTATTGC TATACCTTCT CTTCGATTAT TATATTTACT AGATGAAATC 270
Moth        ......T... .......TG. .C....A... ....C...AA .T..... AT..... .....T.... .....T. .....T.A
Flea        G.....T. .......... .C....T... T....AT.AA .T..... .T..... ......C.T. .......... ......TTCA
Weevil      .....G.C.. .......... CA.T..T..T .....A .T..A..... AC.T..... T.A.....C .C...A.CT. .....T
Lady Beetle G.....CT. .......... .......... T...C.T.AA .T..... AT....C..A T.AAA....A .T.....A. T.....
Wasp        .....T. .......... .A.C..TAT. ......T.AA .T.A...A.. C..C..C.... ..AAA.A.C. .....AA T.....T
Ant         ......T... .......... .ACT..TAT. ......T.GA .T..A..... A..T..C..A T..AA.A.C. .....AC ......G..A
Milkweed Bug ......T.T. .......T.C AA.T..A..T ....CAT.AA .C..... CC....A... ..A.ATC.T. .....A. T.....T
Cricket     .....T. .C.....T.. ......C... ....CAT.AA .T..C..... AT....C..A T.A..... .....C.... .....TCT
Locust      .....C.. .C......GC ......A... ....CAT.AA .T..... CC.T..A..A T.A...C... ....C..... T.....TTCA
Cockroach   ......T..G CA..... .GCT..A... G.A...T.AA .T..... AG....A... ..A...C... .....A. .....T
Termite     ..C...T... .......... .GC...A..T G.A..C..AG .....CA. ..CC.....C T.A...C..C ....CC.TA. .....
Dragonfly   .....C.G AA.....G.. .G....A.T. T..G..T.GG ..A..... ......C... ..C..G...C ....C..... .....G.A

Drosophila  AATGAACCTT CAATTACACT AAAAAGAATT GGTCATCAAT GATATTGAAG TTATGAATAT TCAGATTTTA ATAACGTA-- -GAATTTGAT 357
Moth        ...A.T...C TT.....CT. .....CT... ..A..... .......... .......... .....C.... ....TA.T-- -.....C
Flea        ...AGG.... T....T.TT. ....GC.... .......... .......... .......... A.T..... ....TA.T-- -TCT.....
Weevil      ...A.T...C A.C.ACT.A. T....TT... ..A..... .....TG .....C ..T....A.. .A..TA.T-- -.....C
Lady Beetle CGGA.T..C. T.G.A..CT. .....CT... ..G..... ....C....C .......... .......... .A..AT..-- -.....
Wasp        ...A.C..C. T...A...A. T...TCC... .......... .......... A.....C ..T..... .A..TA.T-- -..T...A..
Ant         TT.A.TAA.A A.....TA. T...TCTG.G ..C..C.... ....C..... C..C..... ..T.....T TA..TA.T-- -.....
Milkweed Bug ...A.C...G A...A..C.. T...GTT... ..A..... .......... A.....C. ..T.....C. .A..TA.T-- -.....C...
Cricket     .TA..C...C T...C...A. .....C.... ..A..... .....TC A.....AT..... .A..TA.CAT C.....
Locust      GTA..TG.AA T.....A. T....CC... ..A.GA.... ....C..... A.....C..C. TAG.....-- -.....C
Cockroach   ...AC....A ..G.A...T. .....CT... ..A..... .......... A.....CG CA..G...-- -.....
Termite     C.CA.C..AA .T..A...A. C....CC..C ..A..C.... ....C..... .....C..C. T...A...-- -.....C..C
Dragonfly   .....GT. .....C.... ..G..C.... .....C..... .....C. .AC..A.T-- -.....
```

Figure 1. Sequences of the mitochondrial COII gene of 13 species of insects. Dots indicate identity to the *D. pseudoobscura* sequence. Dashes indicate deletion/insertion events.

```
Drosophila      TCTTATATAA TCCCAACAAA TGAATTATCA AATGATGGAT TTCGACTACT AGATGTTGAT AATCGAATTG TTTTACCTAT AAATTCTCAA  447
Moth            .......... .TG...GT.. ....C.TC.T TTAA..AAT. ....T..TT. ......A... .....T...A .......C.. ....AAC...
Flea            ..A....... .T...T.T.. .......AAT TTAA..TC.. ....TT..T. .......... .........A .C........ T.....A...
Weevil          ..A....... .T..T..... A...C.TAAC TCCTT.AAT. .......C.. ...A...... ..C....CAC CAA.T..CTA C..AA.C...
Lady Beetle     .........T .AT.CTATG. .A.T...AAC CCATT.AAT. ....T..TT. ...A...... ..C....CAA .......ATA TTTA...A.T
Wasp            ..A.T..... .TAAT.G... AA.T...AAT C..------. ....TT..T. ......A..C ..C.....AA ..A.C..A.. ....AA....
Ant             ..A.T..... .T..TT.... CC.....AAT CC.A...A.. ....TT..T. ....AC.... .....TTG.A ..C.T...T. T....AC.CT
Milkweed Bug    ..A....... AA........ C......AT. ...A...A.. .........T. ...A..C..C ......G..T .A........ ....AAG....
Cricket         .........T CAG.CCT.G. .A.....AG. TCA------. .C..T..T.. ...C..C... ..C....CCA ..C....... ....A.....
Locust          A......... CA...GA.C. A..CC..GA. .......... .......... ......A... ..C....CAA .CC....A.. ....A.AG..
Cockroach       ..A....... .T..TCA.G. ....A..GAT C..AG.AT.. .C........ .......... ......GCA. .......C.. ....A.ATTT
Termite         ..A..C.... CA...TATG. A...GATAAC ..GA.AAT.. .......... ...AACC..C ....ACG.AA CAC....A.. ...C..ATTT
Dragonfly       .......... .T..TTAC.. ....A..GA. G.GTCA..CC .A...T..T. ...A...... .......CAA CA.....G.. .C.AA.....

Drosophila      ATTCGAATTT TAGTAACAGC AGCTGATGTT ATCCACTCAT GAACAGTCCC CGCTTTAGGA GTTAAGGTAG ACGGAACACC TGGACGATTA  537
Moth            .......... ....T..T.. TA.A.....A ..T.....T. ....TA.... AT........G .....A..T. .T.C..AT.. A..T..T...
Flea            .....T.... ..A.T..T.. TA.C..... C.T..T.... ....CA.T.. .T..C.T... A....AA.T. .T.CT..T.. ......T...
Weevil          ........C. ........TT. ...A.....A ..T.....C. .....A.... AAGAA..A.T A....AA.T. .......... ...T..TC..
Lady Beetle     .....TT.A. ..AC.T.TT. T..A.....G ..T..T.... .G..TA.... TT...C...G .....AA.T. .T.CCT.T.. ...C......
Wasp            ..C..T..AC ..A.T.ATT. ...A.....A .....T..T. ....T..A.. AT.A.....T ..C..AA.T. ..TC.GT... ........A.T
Ant             .......... ..ACT...T. CATA..... ..T.....T. ....T..T.. AT.C.....G A....AA... ..TC...T.. ...G......
Milkweed Bug    ........C. ..A.T..... ...A..C... C....T.... ..G..A.... AT...C..... ..A..AA.T. .T.CT..... A.....TC.T
Cricket         .......CA. .......... T.........A ..T..T..C. ....C..TT. ...CC....T ..G..AACT. .T.CC..C.. A..T......
Locust          G.A...G.A. ..AC..G... .T.A..... C.A..... ..G....T.. T..A.....G .....AA.T. .T.C...G.. A..T..G...
Cockroach       .........A .C........ ...A..C... T.A..... .....A.... .AG.C.C..C ..A..A.CT. .T.C...T.. A..T..CC..
Termite         .........A .C........ C.........A C.A..... .....A.... .AGA..... A....A.CC. ...C..... A.........
Dragonfly       G.......C. ..A.T..... .........A C.T..T.... ....C..A.. AT.A..... A........T. .T.CT..T.. ...G.....G

Drosophila      AATCAAACTA ATTTTTTTAT TAACCGACCT GGTTTATTTT ATGGACAATG TTCAGAAATT TGTGGAGCTA ATCATAGTTT TATACCTATT  627
Moth            .......... ....C..... C..T..T..A ..AA.T.... ....T..... .......... ........A. .......... ..........
Flea            ..C...T.A. .....A.A.. A..T.....A ..A....A.. T.......... .......... ........A. .......A.. ..........
Weevil          ......G.A. ..C.AA..GC C..T...T.. A.AA.C..C. T...T..... C........C .....T..A. ....C..G.. C..........
Lady Beetle     .......TA. GG...ACAC. ...T..CA.A ..AA.T.... .......... .......... ........A. .......A.. ..........
Wasp            .........T TAA.AA.A.. ...T.....A ..AA.T.A.. T......G.. .......... .....TATA. ....C..A.. ..........
Ant             ......T.CT TA..A.A... AT........ ..A....AC. T...C..... ...T..G..C ..C..GA.A. ........A.. ......A...
Milkweed Bug    ......GGA. GAA..AAA.. ......C..A ..AA.T..A. T...C..G.. .......... ........A. ....C..A.. C.........
Cricket         .....G.T.. ....CA.A.. C.....T..A ..A....... .......... C........C .....T..A. ....C..C.. ..........
Locust          ......GGA. CA..CACA.. A..T..... ..A....C. T.......... C........C .....T..A. .C..C..A.. ......A...
Cockroach       ..C...GT.. GA.....A.. ...T..... ..AG.TC.A. ....T..... ...C..G..C ........A. ....C..A.. ......A..C
Termite         ......T.A. GA...A.A.. ......C..C ..AC..C.C. ....T..... C........C ........A. ....C..A.. ......A..C
Dragonfly       ........A. GA........ ...T..T..A ..AA.T..C. T...T..... C.......... ..C..G.... .......A.. ......A..C

Drosophila      GTAATTGAAA GTGTTCCAGT AAATTATTTT ATTAAATGAA TTTCTAATAG AGTAAATTCT T                              688
Moth            ..T....... .AA..T..A. T..AA..... .....T.... ..AA...CTA TTCTT.C.AA
Flea            .......... .AA..TT.A. T....CA... .......... ....ATC..A TTC.T.A
Weevil          ..TT.A.... .AA.CA.TCC C..C.TA..C T.A..T...G ..AT.TC..A ..CTT.A
Lady Beetle     TCT....... .AA..T.TCC TTCAA..... .........G ..AA...ATC CTC.TTA.AA
Wasp            ..T....... .AAC.T.TAA C..CA..... ...TCC...T .AAAATC.TT .T
Ant             ..C......T CAAC.AATT. TTCA..... .AA..T...C .AAAATCATT TT..T.A
Milkweed Bug    ..T....... ....C...A. T..AA.C..C T.A..... ..AAC..ATC .T...G.AGC .AA
Cricket         .....C.... .A..AAATC. ...AA....C .....C.... ..AAA...TA CTC.TCA.
Locust          .......... .AAC.T.... .....TA... .....G...T .A.....G.T .A.TT
Cockroach       .......... ..A..T..AC ....GGC... .....C.... .CCTA..A.T .AAT.TA.AA
Termite         .......... .A..AT..AC ...AA.A... ...G..... ..AAA..CCT .AGCG.A.
Dragonfly       A......... .A...AATA. C..A.CA... .....T.... .CCAA....T .AG.G.AG.A .
```

Figure 1--Continued

Codon Usage

Codon usage for the COII gene in all 13 taxa is given in Table 4. Codon usage in the insects reflects the bias against codons ending in G, as is typical of other mitochondrial protein coding sequences (Brown, 1985). Several codons which are rarely, if ever, used in the sequences obtained from other organisms, appear as apparent sense codons in these insects. AGG, which is evidently a terminator in some mammals (Brown, 1985), and which does not occur in the *Drosophila yakuba* sequence (Clary and Wolstenholme, 1985), appears in these sequences. Codon position 205 (nucleotide positions 613-615) is a conserved serine in the other 12 taxa and is coded by AGG in the weevil sequence. Two other rare codons, CAG (glutamic acid) and CGC (arginine) occur occasionally in the insect COII gene.

Inferred Amino Acid Sequences.

Figure 2 shows the inferred amino acid sequences for the species studied here. Translations used the *Drosophila* mitochondrial code (de Bruijn, 1983), including the codon assignments noted above. Two regions are highly conserved at the amino acid level. The region from amino acid position 101 to 113 has been shown to be involved in the electron transfer system (Millett et al., 1983). The second region, from positions 193-203 in the amino acid sequences includes most of the copper binding site of the protein. The conserved histidine at position 161 also forms part of the copper binding site (Millett et al., 1983).

Other individual residues known to be conserved in COII sequences from vertebrates, yeast and maize are also conserved in all the insects studied here. They include histidine (position 24), glutamic acid (62 and 212), aspartic acid (88, 139, 158 and

```
Drosophila  MSTWANLGLQ DSASPLMEQL IFFHDHALLI LVMITVLVGY LMFMLFFNSY VNRFLLHGQL IEMIWTILPA IILLFIAMPS LRLLYLLDEI  90
Moth        .R..S.FN.. N........I ......T.I. .I...I.... I.IN....KF I...F.V..M ..L...V... .T.I...L.. .........L
Flea        .N..M.FN.. N.N..... M...N.SM.. ILL..I.... I.SS.LY.KL Y..Y..ES.N V.I..... FM.I...L.. ........DS
Weevil      I...K..F.. .........L. MC.....M.. .IL..IM.SQ MLLSML..KL SH.Y..E... ..T....I.. ...IL..L.. .....I....
Lady Beetle I...KSSLFL ..S.F.L... R......... .N...GA.A. I.IS.L..K. NH....E.HT V.T..... FT.I...L.. .K.I..I...
Wasp        ....SMMN.. .AN..M.... ......T.M. .LL..ITII. IISSIIM.NL T.K.IMQN.T ..I....I.M ...IYM.I.. .KI...N...
Ant         .N..L-.S.. N.N..TYDMM .....FTMM. .IF..L.ILF I..TMIN.NL I.....Q.HF ..L....T.M ...IL..I.. FKI...T..M
Milkweed Bug .A..M.IN.. .AN.ST.... TM....T.M. .T...SI.TF I.VSMTT.TL I..Y..E..T ..F...TI.. .T.I...L.. .H....I...
Cricket     .A..S..N.. N.S..... ......T.M. .L......A. I.S.....L. T.....E..T ..I..... .T.I...L.. .........S
Locust      .A..S..SI. .G..... S....DHTMV VLL...I... ALSYML..A. T..NM...H. ..T...A... .T.I...L.. ........DS
Cockroach   .T....MN.. .....I.... .Y......M. II..LMV.S. M.IAMV..K. I.....E..M ..LA...A.. V..I...V.. ......M...
Termite     .T..SCMN.. .....I.... ......T.M. VT..L.S.L. F.TSMTI.KN E..YM.E..T ..L....A.. V..V..TT.. ......M...
Dragonfly   .A...Q.NF. .A...M.... HY....TMMV ..I..IM.A. I.GTM...KD ...Y..D..K ..TE...V.V FV.VI..... .........V

Drosophila  NEPSITLKSI GHQWYWSYEY SDFNNV-EFD SYMIPTNELS NDGFRLLDVD NRIVLPMNSQ IRILVTAADV IHSWTVPALG VKVDGTPGRL 179
Moth        .N.L....T. .......... .....I-... ....AS...P LNN....... ...I....N. .......T.. .....I.S.. ....AN....
Flea        .S.L.S..A. .......... T....I-S.. .....S...N LNS....... ...I..I... ....I..T.. L....I.S.. I.I.A.....
Weevil      .N.QLLI.I. .......... ..YK.I-... ......K..N SFN....E.. ..TPI.YKT. ......S... .....I.SMS I.I.......
Lady Beetle RN.LV...T. ......T... ...KKL-... ...LSYDN.N PFN....E.. ..TI..YL.N ..L.TSS... .....I.SS. ..I.AS....
Wasp        .N.LM.I... .......... ...K.I-D.N .F..NSKN.N H--....... ..MII...N. ..M.INS... .......S.. ..I.SV...I
Ant         FNNK..I..V .......... ...L.I-... .F...S.Q.N PNE.....T. ..CI..F.YP ....T.SM.. .......S.. I.M.S.....
Milkweed Bug .N.EM...V. .......... ...K.I-... ...K.....M .NE....E.. ..VL....K. ....I..... L...AI.S.. ..I.A.....
Cricket     MD.L..M.T. .......... M..K.II... ...SALDK.. S--....... ..TI....T. ..T....... ......S... ..T.A.....
Locust      VDAM..I.T. .R........ ...MD.-... T..T.EQD.E .......... ..TI....TE V.V.TS.S.. L...A..... ..I.A.....
Cockroach   .T.TV...T. .......... ...AK.-... .....QD.MD HSM....... ..A.....TF ...I..... L....I.S.. ..A.A.....
Termite     HN.TM.I.T. .......... ...IK.-... ...T.YE.DN KKM....ET. .HVT.....F ...I..... L....I.S.. I.A.A.....
Dragonfly   ........T. .......... ...KHI-... .....Y..ME ES.L...E.. ..TT...QT. V...I..... L......S.. I...A.....

Drosophila  NQTNFFINRP GLFYGQCSEI CGANHSFMPI VIESVPVNYF IKWISNSVNS                                          229
Moth        .......... .I........ .......... ....ISIKN. .N..N.YSY*
Flea        ..S..MM... ..YF...... .......... ....ILI.S. .....SNS*
Weevil      ..A.LIA..S SI.F...... .......... .L..ITP.L. LN.VISKA*
Lady Beetle ..MS.TL..T .I........ .......... S...ISPSN. ...VNK.SL*
Wasp        ...LMM.... .IYF...... ..M....... ....TSN.N. .S.LKSL
Ant         ..SLLYMY.. ..YF...... ..T....... ....TNFS.. KN.LKSFL*
Milkweed Bug ..GSIK.... .ILF...... .......... ......IKN. L...NK.LS.
Cricket     ..I..M.... .......... .......... .....NLKN. .N..K.YSS
Locust      ..GT.TM... ...F...... .......... ....TS..L. ...L.KMI
Cockroach   ..VS.L.... .VL....... .......... ....IST.G. .N..LKMNM*
Termite     ..SS.M.... ..L....... .......... .....STKK. .E..K.LSE
Dragonfly   ...S...... .I.F...... .......... M....NIKS. .N..Q.MSEA
```

Figure 2. Inferred amino acid sequences for the COII gene of 13 species of insects.

Table 4. Codon usage in COII genes of 13 species of insects.

Phe	TTT	139	Ser	TCT	62	Tyr	TAT	92	Cys	TGT	21
(F)	TTC	31	(S)	TCC	17	(Y)	TAC	21	(C)	TGC	8
Leu	TTA	220		TCA	96	TER	TAA	7	Trp	TGA	77
(L)	TTG	4		TCG	1		TAG	0	(W)	TGG	1
Leu	CTT	44	Pro	CCT	56	His	CAT	52	Arg	CGT	22
(L)	CTC	12	(P)	CCC	17	(H)	CAC	25	(R)	CGC	4
	CTA	69		CCA	52	Gln	CAA	86		CGA	59
	CTG	2		CCG	3	(Q)	CAG	4		CGG	4
Ile	ATT	297	Thr	ACT	62	Asn	AAT	178	Ser	AGT	20
(I)	ATC	62	(T)	ACC	19	(N)	AAC	40	(S)	AGC	4
	ATA	148		ACA	95	Lys	AAA	71		AGA	46
	ATG	14		ACG	3	(K)	AAG	9		AGG	4
Val	GTT	50	Ala	GCT	31	Asp	GAT	95	Gly	GTT	27
(V)	GTC	7	(A)	GCC	18	(D)	GAC	29	(G)	GGC	8
	GTA	65		GCA	59	Glu	GAA	115		GGA	58
	GTG	4		GCG	1	(E)	GAG	7		GGG	10

173). Several positions known to be conserved among bovine, yeast and maize COII sequences show variation in some of the insect groups. Codon 11, which generally codes for aspartic acid, appears to code for asparagine in four of the sequences. This inferred change is from a negative to an uncharged polar amino acid residue. Codon 137, a conserved aspartic acid in other organisms, appears to have conservative replacements of glutamic acid in five of the species. In neither of these cases does the apparent substitution follow obvious phylogenetic patterns.

Divergence.

A total of 468 of the 688 nucleotide sites (229 codons) were variable, including 154 first, 105 second and 209 third codon positions. Corrected nucleotide and amino acid divergences are given in Table 5. Not surprisingly, divergences between different insect orders are quite high, ranging from about 26% (moth - flea) to over 55% (ant - termite). Perhaps more surprising is the observation that the within order comparisons are also high: about 45% divergence between the two beetle sequences, 37% between the wasp and ant and 38% between the cricket and locust. None of these within-order pairs are considered to be closely related pairs, based on traditional taxonomy. Divergences at the amino acid level are also high, ranging from about 23% to over 47%.

Table 5. Corrected interspecific divergences in COII gene sequences in 13 species of insects. Percent DNA sequence divergence using Jukes-Cantor correction and the number of nucleotide differences in terms of transitions/transversions are given above the diagonal. Below the diagonal are the inferred amino acid divergences, and number of substitutions.

	drosop	moth	flea	weevil	l.beetle	wasp	ant	cricket	locust	cockro	termite	mi.bug	dragonfly
drosophila	-	26.63	29.54	40.24	40.75	41.29	45.33	31.02	35.77	31.12	37.22	37.97	31.29
		54/99	55/111	78/134	64/151	66/146	69/161	72/98	73/120	70/104	84/116	74/130	70/106
moth	23.25	-	25.94	37.06	34.04	37.52	40.90	27.97	41.66	32.23	41.42	35.67	35.90
	53		37/112	68/131	62/126	61/136	67/146	70/86	70/147	61/119	86/131	73/121	78/117
flea	30.26	25.44	-	43.30	38.51	38.26	39.89	33.79	41.15	38.02	44.61	35.16	36.58
	69	58		82/143	59/146	55/145	58/152	71/111	83/132	64/139	90/139	62/129	72/125
weevil	35.09	36.40	35.53	-	45.41	44.98	50.32	42.08	47.50	42.52	50.97	43.56	45.14
	80	83	81		54/147	82/144	86/163	81/135	88/151	77/144	98/154	83/142	93/138
l.beetle	37.28	34.21	38.60	36.84	-	43.38	48.90	35.83	48.34	41.51	48.04	43.30	45.67
	85	78	88	84		65/155	75/168	77/114	85/157	78/141	93/149	74/151	99/135
wasp	38.94	35.84	37.61	40.26	43.36	-	37.18	40.01	47.73	43.38	50.01	35.12	43.38
	88	81	85	91	98		65/129	78/129	88/148	81/139	95/149	63/124	80/140

ant	39.91	38.16	39.04	42.98	47.81	34.51	-	50.16	51.99	51.18	55.05	45.59	50.32
	91	87	89	98	109	78		93/151	77/176	92/159	96/168	76/155	81/167
cricket	23.24	20.61	28.95	37.72	33.77	34.51	38.77	-	38.71	37.75	43.56	40.72	39.72
	53	47	66	86	77	78	88		72/130	84/115	95/128	80/131	92/115
locust	32.90	37.72	36.84	42.98	42.11	43.36	45.37	31.58	-	39.64	43.47	43.47	42.69
	75	86	84	98	96	98	103	72		83/126	99/125	92/132	82/139
cockroach	28.51	28.51	35.53	37.72	37.72	40.71	44.05	31.14	38.77	-	31.24	36.61	36.37
	65	65	81	86	86	92	100	71	88		96/78	86/112	83/114
termite	33.33	32.90	36.40	41.67	40.35	39.82	43.17	31.58	40.53	23.68	-	41.16	41.16
	76	75	83	95	92	90	98	72	92	54		95/121	103/113
milkweed bug	32.02	29.83	32.46	36.40	37.28	35.84	40.53	30.70	38.77	32.90	32.02	-	42.44
	73	68	74	83	85	81	92	70	88	75	73		94/128
dragonfly	30.26	31.14	35.09	38.16	41.67	41.59	44.05	31.14	38.33	32.75	34.21	33.33	-
	69	71	80	87	95	94	100	71	87	75	78	76	

Discussion

Comparative analysis of aligned homologous DNA sequences from related taxa provides a powerful way of studying evolution at the DNA sequence level. Some of the taxa examined here diverged more than 300 million years ago (Wooten, 1981). These long divergence times are reflected in the nucleotide and inferred amino acid sequence divergences, which range from 26 - 55% and 23 - 47%, respectively. Nonetheless, alignment using the translated sequences was unambiguous, except at the site(s) of the two-codon deletions in the wasp and cricket sequences, and about 14 amino acids at the carboxy terminus of the translated sequence (Fig. 2). Because of the difficulties in alignment at the 3' end of the gene, in the analysis that follows I ignore the 3' end, from position 640 to the ends of the genes.

COII sequence evolution.

Before any attempt can be made to apply sequence data to phylogenetic analysis, it is necessary to examine the details of sequence evolution for the gene. It is well known that transitions occur at a much higher frequency than transversions in mitochondrial DNA, but this excess is only evident in comparisons involving very closely related species (Brown, 1985). For more distant comparisons, the occurrence of multiple transition mutations at variable sites and the accumulation of transversions, which erase transition differences (DeSalle et al., 198; Brown et al., 1982), can eliminate the transition bias.

Figure 3 shows the proportion of transitions in pairwise sequence comparisons of the COII gene for a broad spectrum of relationships. Data points for divergences of less than 12% are within-group comparisons of the *Drosophila obscura* group, and

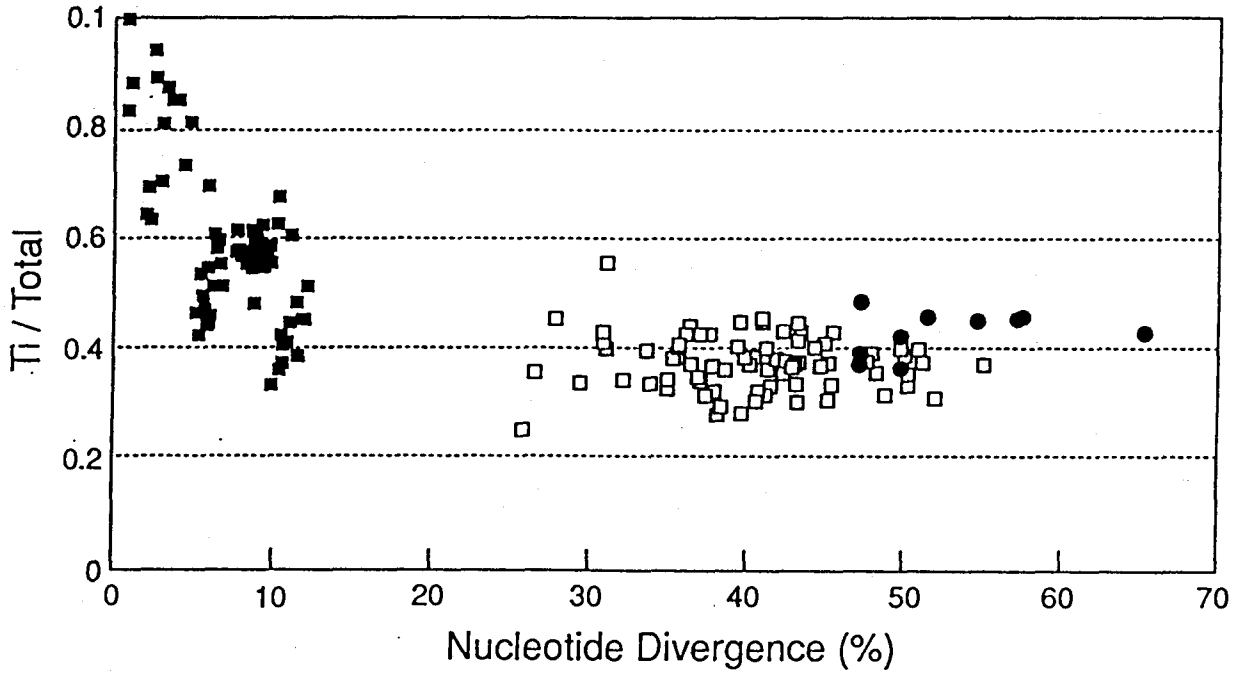

Figure 3. The proportion of transitions in pairwise comparisons of the COII gene for a broad spectrum of relationships. The percent nucleotide divergences are corrected using Jukes-Cantor's method. Solid squares indicate the within-group comparisons of the *Drosophila obscura* group and comparisons between *D. yakuba* and members of the *D. obscura* species group (Beckenbach et al, 1993). Open squares indicate the comparisons between all the insects studied here. Solid cycles indicate the extended comparisons between the *D. pseudoobscura* and published urchin, frog, mouse, cow and human COII sequences, and comparisons of the urchin to the mammal sequences.

comparisons between *D. yakuba* and members of the *D. obscura* species group (Beckenbach et al., 1993). It appears that the transition excess is essentially eliminated in comparisons of sequences that are only 12% diverged. Pairwise comparisons of sequences presented here (Fig. 1 and Table 5) range from 25 to 55% divergent. The proportion of transitions in these comparisons average 36% and appear to be stable throughout this range. To determine if this value is a limiting value, I have extended the comparisons to include those between *D. pseudoobscura* and the published COII sequences of sea urchin (Jacobs et al., 1988), frog (Roe et al., 1985), cow (Anderson et al., 1982), mouse (Bibb et al., 1981), and human (Anderson et al., 1981). These values (solid dots, Fig. 3), having divergences between 47 and 65%, show no further reduction in the proportion of transitions observed.

Holmquist (1983) has analyzed the limiting ratios for transversions to transitions, R_{∞}, based on observed base composition. Using his formula 11, we calculate R_{∞} of 2.50, corresponding to a proportion of transitions of 28.6%. The main conclusions from this analysis, are that the proportion of observed transitions stabilizes at a value somewhat higher than that predicted solely on the basis of base composition, presumably because of functional constraints at the protein level, and that this limiting value holds for all divergences greater than 25%. Thus, all comparisons in this study appear to have reached this limiting value. It is clear that transitions cannot contribute phylogenetic information in these comparisons, and it is possible that even transversions have reached, or are approaching, a level of saturation.

Greater resolution may be obtained by analyzing the changes according to codon position. Transitions and transversions can have very different effects, depending on

codon position. Table 6 gives the base composition and bias by codon position for each of the insect COII sequences examined here. Figure 4 summarizes these results. The A + T bias observed in other insect mtDNA sequences occurs at all positions, but is most extreme at the third position. If one assumes that the A + T bias is maintained selectively, then we would expect less constraint against synonymous transversions (A ↔ T) than transitions at the third codon position. If we recalculate R_∞ by codon position, weighted by the number of observed variable sites, the value increases to 3.13, corresponding to a transition frequency of 24.2%. This value is even farther from the observed value (36%) than we obtained by ignoring codon positions. I suggest the transition excess relative to random expectation, based on base composition, occurs mostly at the first and second positions, indicating that amino acid substitutions generated by transitions are less severe than those generated by transversions.

Phylogenetic inferences.

Computer programs will generate trees from sequence data, regardless of whether phylogenetic information is present in the sequences. Trees can be generated using the complete data set, or by using restricted subsets of the data. DeSalle, et al. (1987) and Beckenbach et al. (1993) found that by confining attention to transversion differences, reasonable resolution of relationships for time spans of tens of millions of years can be obtained. Edwards et al. (1991) and Irwin et al. (1991) found that the inclusion of third position changes can obscure relationships in deep phylogenies. Moritz et al. (1987) suggested that resolution of deep phylogenies may be achieved by confining attention to the second codon position differences. They reasoned that second position substitutions result in non-conservative amino acid substitutions, while those arising from first or third

Table 6. Base composition at first, second, and third codon positions

Species	First				Second				Third			
	G	A	T	C	G	A	T	C	G	A	T	C
Dragonfly	25.65	33.04	22.61	18.70	12.66	29.69	37.55	20.09	7.42	45.41	37.55	9.61
Milkweed Bug	20.00	40.43	22.17	17.39	12.17	29.13	37.83	20.87	1.30	49.57	38.26	10.87
Locust	27.19	29.82	25.88	17.11	14.10	25.99	37.00	22.91	3.96	52.86	31.72	11.45
Cricket	19.30	35.09	27.19	18.42	11.45	26.87	40.09	21.59	1.76	44.05	36.56	17.62
Cockroach	25.33	35.81	22.27	16.59	13.10	26.64	40.61	19.65	1.75	49.78	37.12	11.35
Termite	21.40	38.43	20.96	19.21	12.72	28.95	36.40	21.93	2.19	54.39	21.93	21.49
Drosophila	23.91	30.00	29.57	16.52	14.41	26.64	41.05	17.90	1.75	45.85	47.16	5.24
Moth	17.47	37.99	27.95	16.59	12.66	29.26	41.48	16.59	0.87	34.50	58.95	5.68
Flea	14.47	36.40	33.33	15.79	12.72	28.51	39.91	18.86	1.32	42.54	53.51	2.63
Weevil	17.11	36.40	23.68	22.81	13.16	27.63	39.91	19.30	1.32	41.67	40.35	16.67
Ladybird Beetle	17.90	31.44	36.68	13.97	13.10	27.51	36.68	22.71	3.49	41.92	47.60	6.99
Wasp	13.33	46.67	25.33	14.67	10.71	29.91	41.96	17.41	0.89	40.62	47.32	11.16
Ant	12.33	37.89	33.92	15.86	10.57	27.75	41.85	19.82	3.08	37.00	49.34	10.57
Mean	19.65	36.11	27.04	17.20	12.58	28.04	39.41	19.97	2.39	44.63	42.11	10.87
Bias*		0.175				0.233				0.490		

* *Bias in base composition is calculated as* $C=(2/3)\sum_{i=1}^{4}|c_i - 0.25|$, *where C is the compositional bias and* c_i *is the frequency of the ith base*

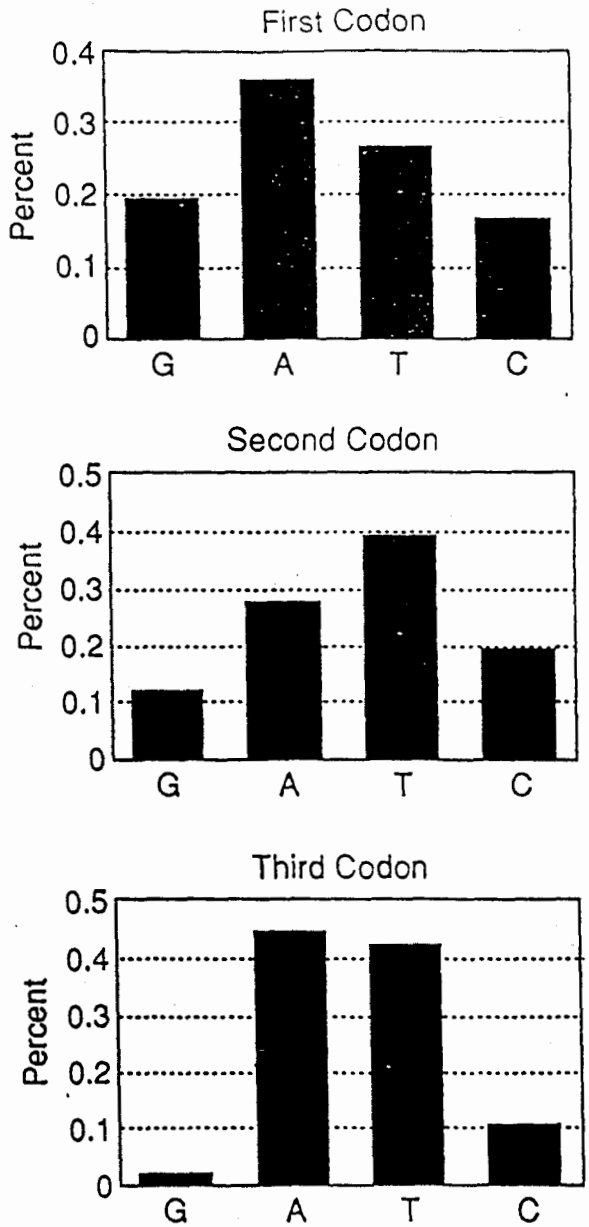

Figure 4. Average base composition of 13 insect mitochondrial COII genes at first, second and third codon positions.

position substitutions are often chemically conservative.

In view of the apparent saturation in this data set, not only of transitions, but perhaps of transversions as well, I examined the neighbor-joining trees produced by using each of the following series of subsets of the data: (1) all substitutions; (2) transversions only; (3) inferred amino acid substitutions; and (4) second codon position differences. Edwards et al. (1991) and Irwin et al. (1991) used first and second codon positions to analyse relatively deep phylogenies. This procedure includes some silent first position substitutions, and ignores some non-silent substitutions at third codon positions. By comparing the inferred amino acid sequences, I include all non-silent differences, and ignore all silent nucleotide substitutions. Some amino acid substitutions can arise from a single non-silent nucleotide substitution, while others require two or even three nucleotide changes. To partially account for these differences, I weight the amino acid substitutions by the minimum number of nucleotide substitutions required, based on the insect mitochondrial code.

I compared the resulting trees to those based on consideration of morphology and the fossil record. There are a number of relationships that are widely accepted. Each of the orders included in this study is believed to represent a monophyletic grouping. The orders of holometabolous insects (including Coleoptera, Diptera, Hymenoptera, Lepidoptera and Siphonaptera, studied here) are also believed to comprise a monophyletic grouping. The Blatteria (cockroaches) are closely related to the Isoptera (termites). The Odonata (dragonflies) should represent an outgroup to the other nine orders included here.

Figure 5 gives the neighbor-joining trees resulting from the complete data set (Fig. 5A) or conserved subsets of the data. The published honeybee COII sequence (Crozier et al., 1989) is included in this analysis. All of the trees are different. The only relationships supported by all trees is the close affinities of the cockroaches and termites, and the monophyly of the Hymenoptera. Some of the trees fail to support monophyly of some orders. The two Coleoptera, a ladybird beetle and a weevil, are not clustered when all data are used (Fig. 5A) or when only transversions are considered (Fig. 5B). The two members of Orthoptera, a cricket and locust, are separated when only transversions (Fig. 5B) or amino acid sequences (Fig. 5C) are used. In all trees, there is a curious mix of the holometabolous orders with the hemimetobolous orders, believed to have diverged prior to the radiation of the holometabolous groups.

The presumed outgroup, Order Odonata, appears in a variety of places in the trees, but nowhere as an outgroup. This result is not surprising for networks that include a number of highly diverged taxa. In such cases, the network will tend to be rooted by the longest branch (Wheeler, 1990), which depends not only on time, but on evolutionary rates as well. I re-examined all of the trees shown in Fig. 5, specifying Odonata as an outgroup. The results showed no obvious improvement in discerning relationships.

It appears from this analysis that these data provide little basis for confirming or rejecting any particular relationships among the orders, or even for confirming monophyly of the orders themselves. The only exceptions are the close relationships of cockroaches with termites, and within the order Hymenoptera. The tree supporting the greatest proportion of our preconceived ideas of insect relationships is that based on the most conserved subset of the data, the second codon positions (Fig. 5D).

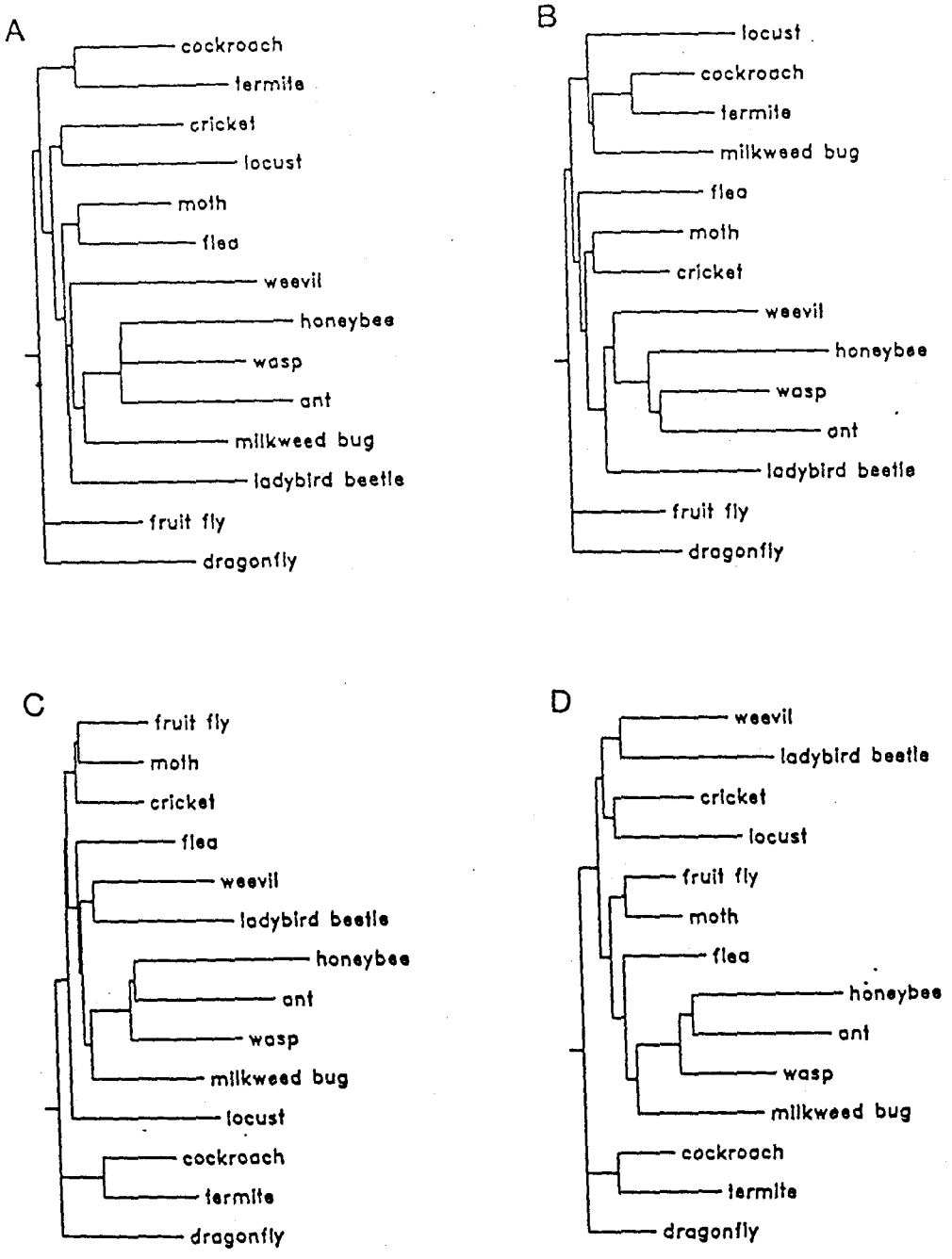

Figure 5. Neighbor-joining trees constructed from the nucleotide and inferred amino acid sequences. The published honeybee COII sequence is included in the analysis. All trees were obtained by NEIGHBOR program in PHYLIP 3.4 (13). The trees are based on: (A) all nucleotide sites; (B) transversions differences; (C) amino acid sequence divergences; and (D) second codon position divergences. All are unrooted trees.

Fig. 6 gives bootstrap trees based on second codon positions. Fig. 6A is based on neighbor-joining distance tree, while 6B is a bootstrap using the maximum parsimony tree. In both trees, the dragonfly was used as the outgroup. In both trees, the termite-cockroach cluster is strongly supported. In addition, monophyly of the Hymenoptera is supported in the distance tree, but not in the tree based on parsimony. The distance tree weakly supports (bootstrapping number is 25 out of 100) monophyly of the Orthoptera (cricket and locust) and Coleoptera (ladybird beetle and weevil), but only the latter is supported by the parsimony tree.

Conclusions

The COII gene is a highly conserved protein coding gene in the mitochondrial genome. I initially hoped that sequence comparisons, using the translated sequences or second codon position variation, would provide a basis for inferring the phylogenetic relationships of the holometabolous orders of insects, and their positions relative to several hemimetabolous orders. It appears, however, that the divergence of insect orders is so ancient that it not easily resolved using this gene. More extensive analyses involving many additional taxa, using COII or other conserved protein coding genes, may provide the necessary resolution to answer these questions, and will certainly provide a greater understanding of the evolution of the genes.

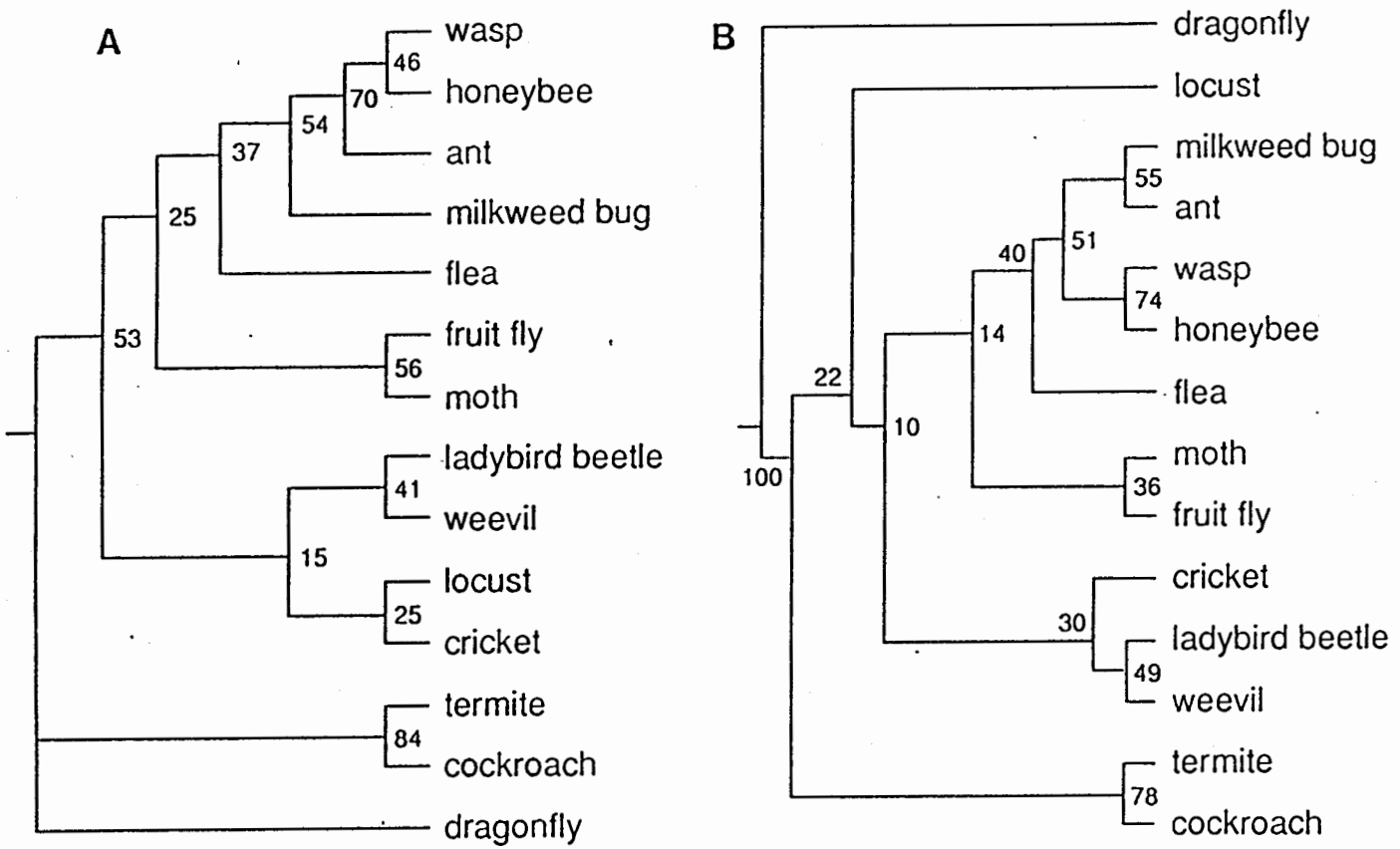

Figure 6. Bootstrap trees based on the second codon positions. Both trees were done using 100 replicates with the dragonfly as the outgroup. A. Bootstrap of distance matrices, using the neighbor-joining method. B. Parsimony bootstrap using DNABOOT of PHYLIP 3.4.

CHAPTER III

EVOLUTION OF THREE MITOCHONDRIAL tRNA GENES AND THE PUTATIVE START CODONS OF COI GENE IN FIVE ORDERS OF INSECTS

Introduction

The animal mitochondrial genome has proven to be one of the most interesting and useful molecules for studies of evolution (Moritz et al, 1987). It encodes 12-13 of the proteins necessary for mitochondrial function. It also has minimal machinery (2 rRNA genes and only 22 tRNA genes) for translation of those protein genes. These tRNAs are sufficient to decode the metazoan mt-protein genes, because a single tRNA can apparently translate all codons of a four-codon family (Anderson et al, 1981).

mtDNA sequences are less useful for reconstructing phylogenies with deep branches, e.g. animals that diverged more than 100 million years ago. This is because the rapid rates of sequence evolution of mtDNAs compared to those of nuclear DNAs produce an apparent saturation of sequence differences among distantly related species due to multiple sequence changes at the same sites (Brown et al, 1982; Moritz et al, 1987).

Recently, Kumazawa and Nishida (in press) used all mt-tRNAs of several vertebrates and sea urchin that diverged 20-600 million years ago to examine the usefulness of mtDNA sequences for deep-branch phylogenies. Examination of tempo and mode of evolution of all tRNA genes and for two protein coding genes revealed that stem-forming nucleotides of mitochondrial tRNA genes accumulate variation over long periods of time. They concluded that tRNA genes give more efficient phylogenetic

performance than Cytb and COI genes in terms of an index for randomness of sequence data, as well as bootstrap probabilities for some well established branches.

Although mitochondrial genome organization in general appears to change at a very slow rate (Brown, 1985), the rearrangements involving tRNA genes occur more frequently than rearrangements involving protein and rRNA genes (Dubin et al., 1986; Crozier and Crozier, 1993; Paabo et al., 1991; Desjardins and Morais, 1991; Smith et al., 1989,1990). Differences in relative locations of two tRNA genes have been noted between the mtDNA molecules of *Drosophila yakuba* and the mosquito *Aedes albopictus* (Dubin et al., 1986), and for 11 tRNA genes between the mtDNA molecules of *D. yakuba* and the honeybee, *Apis mellifera* (Crozier and Crozier, 1993). The marsupials have a rearranged tRNA gene cluster (Paabo et al., 1991). Among echinoderms, there is a major inversion in the mtDNA of sea stars in contrast to the sea urchins (Smith et al., 1989,1990; Jacobs et al., 1989; Asakawa et al., 1991). The inversion contains portions of the tRNA cluster, the 16S RNA gene and the ND1 and ND2 genes.

In view of the apparent deep phylogeny for the insects at order level (Liu and Beckenbach, 1992), I designed a set of primers based on *D. yakuba* mitochondrial sequence to amplify an mtDNA fragment encompassing clustered tRNA genes from various insects by the polymerase chain reaction. I expected the region to include three tRNA genes ($tRNA^{Trp}$, $tRNA^{Cys}$ and $tRNA^{Tyr}$) and 5' of COI gene. There are several reasons to look at this region. First, it is a region where known rearrangements exist between orders of holometabolous insects, honeybee compared to Drosophila (Clary and Wolstenholme, 1985; Crozier and Crozier, 1993). Second, the start of COI gene is very enigmatic. A four base codon with a TAA terminator in frame, ATAA, has been

hypothesized for Drosophila (de Bruijin, 1983; Clary and Wolstenholme, 1985). Honeybee, however, has consecutive ATA (Met) codons which are commonly used for initiation in mtDNA. The COI gene start codon position is also unclear in blue mussel, *Mytilus edulis* (Hoffmann et al., 1992). The inferred starts of COI genes in both blue mussel and nematodes are 9-10 amino acids upstream from the putative start in Drosophila (Okimoto et al., 1990). Also, Smith et al., (1993) have reported a short unassigned sequence in ophiuroid, *Ophiopholus aculeata* preceeding the COI start codon. Third, there is the potential of using conserved stem regions of the tRNA genes for resolving deep phylogenies (Kumazawa and Nishida, 1993).

Materials and Methods

DNA Sources. A single individual of each species, either from stocks maintained at the Simon Fraser University Insectary or freshly caught, was used as source material for DNA extraction. Species names and the orders to which they belong are shown in Table 7.

DNA Extraction. A small amount of tissue from each insect was ground in 60 μl protease buffer (0.1M Tris-OH, pH 8.0, 0.05 M EDTA, 0.2 M NaCl, 1% SDS, with 0.4 mg/ml protease K), and immediately placed at 65°C for 3 min. The resulting solution was extracted once with phenol (saturated with 10 mM Tris, 1 mM EDTA, pH 8.0) and once with chloroform/isoamyl alcohol (24:1). The supernatant was removed and 2 volumes of 95% ethanol was added to precipitate the DNA. The suspension was pelleted at 40K, 15 min, 4°C, washed twice with 70% ethanol, dried under vacuum, and dissolved in 50μl sterile ddH_2O.

Amplification and Sequencing. The tRNA genes were amplified using the polymerase chain reaction (PCR) performed in 50μl reactions using the GeneAmp kit with AmpliTaq DNA polymerase (Perkin Elmer Cetus). I followed the protocol provided with the kit. Thermal cycling was done in an Ericomp Programmable Cyclic Reactor. The PCR consisted of 35 cycles performed as follows: denaturation at 94°C for 1 min, annealing at 50°C for 1 min, and extension at 70°C for 2 min. The primers for PCR amplification and sequencing are primer F and G:

Primer F: $^{5'}$-TGTTCCTACTATTCCGG(AC)TCA-$^{3'}$

Primer G: $^{5'}$-TAGGTGGACTACCTCCATTTT(CT)AGG-$^{3'}$

Figure 7 gives the location and the direction of the primers based on the *D. yakuba* map.

Table 7. List of insect species studied in this chapter.

Order	Name
Blattaria	Cockroach (*Periplaneta americana*)
Lepidoptera	Moth (*Galleria mellonella*)
Odonata	Dragonfly (*Sympetrum striolatum*)
Orthoptera	Cricket (*Acheta domesticus*)
Orthoptera	Locust (*Schistocerca gregaria*)
Siphonaptera	Flea (*Ctenocephalides felis*)

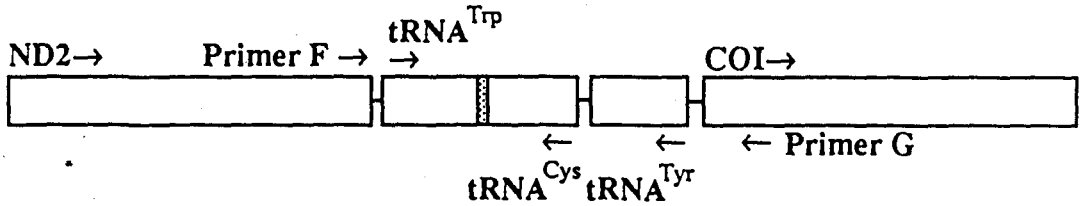

Fig. 7. Map of location of primers used in amplifying the tRNAs.

The double-stranded DNA (about 500bp) was purified by electrophoresis in low melting agarose gel and followed by concentration using Amicon 30 microconcentrators. About 7μl was used for double-stranded sequencing using Sequenase kit (United States Biochemical Corporation). I used modifications of the standard Sequenase protocol given by Palumbi et al. (*The Simple Fool's Guide to PCR*, compiled by Palumbi, S., Martin, A., Romano, S., McMillan, W.O., Stice, L., and Grabowski, G.).

Data Analysis. All tRNA structures were hand-drawn and the sequences were aligned based on the derived structures. Sequences were compared with the aid of eyeball sequence editor, ESEE (Cabot and Beckenbach, 1989). I performed the phylogenetic analysis based on two sets of data. One set was the three tRNA sequences ($tRNA^{trp}$, $tRNA^{cys}$ and $tRNA^{tyr}$) without the D loop and the TψC loop, two highly variable loops in terms of sequences and sites. Another data set was the second codon sequences of the cytochrome oxidase II gene (Liu and Beckenbach, 1992). The distance trees were constructed by the neighbor-joining method (Saitou and Nei, 1987), using NEIGHBOR program, version 3.4 of PHYLIP (Felsenstein, 1989). Parsimony analysis was done using DNAPARS program of PHYLIP.

Results and Discussion

Mode of tRNA Sequence Evolution.

Figure 8 shows the aligned nucleotide sequences of three tRNAs ($tRNA^{Trp}$, $tRNA^{Cys}$ and $tRNA^{Tyr}$) from 7 species, representing 6 orders of insects. The sequences were compared to *Drosophila yakuba* sequence (Clary and Wolstenholme, 1985) and all the tRNAs from other insects studied here have the same organization and orientation as *D. yakuba*. Figure 8 also contains the 5' portion of the cytochrome oxidase I genes (COI).

All the alignments were based on the hand drawn tRNA structures, that is, the alignments were made separately based on accepter stems, Dihydrouridine loops and stems, anticodon loops and stems and TψC loops and stems. Figure 9 to figure 11 give the structures for the three tRNAs: $tRNA^{Trp}$, $tRNA^{Cys}$ and $tRNA^{Tyr}$. The majority of base pairs in stem regions are Watson-Crick pairs, but G-U pairs and mismatched pairs were also observed. There is an 8 nucleotide overlap between the AC stem at the 3' end of the $tRNA^{Trp}$ and the AC stem at 3' of the $tRNA^{Cys}$ on the other strand in all insect species studied. This might indicate a strong three way constraint against any change in the AC stems of $tRNA^{Trp}$ and $tRNA^{Cys}$ for these insects.

Recently, the complete mitochondrial DNA sequence of honeybee *Apis mellifera* has been determined (Crozier and Crozier, 1993). In the honeybee sequence, the ND2 and COI junction also contains the same 3 tRNAs as studied here, but the gene order is different. In the honeybee, the gene order is ND2, $tRNA^{Cys}$, $tRNA^{Tyr}$, $tRNA^{Trp}$ followed by COI, with the same orientation as *D. yakuba*. The overlap I found in all the insects studied here between $tRNA^{Trp}$ and $tRNA^{Cys}$ does not exist in honeybee sequence. In that species these two tRNAs are separated by the $tRNA^{Tyr}$. It is reasonable to suggest that

```
                AA stem   D stem      D stem  AC stem      AC stem    T stem        T stem  AA stem  T stem
                *******   ****        ****    *****        *****      *****         *****   ********  ****
fruit fly   acttt|AAGGCTT TA AGTT AACTA- AACT A ATAGC CTTCAAA GCTGT AAAT AAAGG --GTATT CCTTT |AAGTCTTA| GTAA --
cockroach   cttta|AAGGCTT TA AGTT AATTTA AACT A ATATC CTTCAAA GCTAT AAAT AAAGA TTAAATT TCTTT |AAGCCTTA| GTAA --
dragonfly   atatt|AAGATTT TA TGTT AAATA- AACA A TTAAC CTTCAAA GTTAA AAAT ACAAG TAATAGT CTTGT |AAGTCTTA| GTAA --
locust      atact|AAGGACT TA AGTT AATCA- AACT A ATAAC CTTCAAA GTTAT AATT AAAAG -AATAAT CTTTT |AGGCCTTA| GTAA --
cricket     aattt|AAGGCCT TA AGTT AAAAA- AACT A TTAGC CTTCAAA GCTGA AAAT GAAA- TTCATAT -TTTT |AGGCCTTA| GTA- --
moth        tttat|AAGGTTT TA GGTT AATAAA AACT A ATAAT CTTCAAA ATTAT TCAT AAAGA --AAATT TCTTT |AAGCCTTA| ATAA AT
flea        attta|AAGGTTT TA AGTT AAAAA- AACT A ATAAC CTTCAAA GTTAT AAAT AAAGA- --ATT- TCTTT |AAACCTTA| GTA- --
                 |tRNAtrp-->                                                                |<--  -->|

                T stem    AC stem      AC stem  D stem         D stem   AA stem                          AA
                ****      *****        *****    ****           ****     *******                         ***
fruit fly   -AAAT TTAC TCCT TCAAA ATTGCAG TTTGA T ATCA -----TTAT TGAC TA TAAGACC|tagatttaatttat----------|T GAT
cockroach   ATAAT TTAC TCCT ATAGA ATTGCAT TCTAT T ATCA TACAAAATA TGAA TA TAAGACC|tttaacaaatttattcattcttga|T AGT
dragonfly   AATTT TTAC ACCT TCAGA ATTGCAG TCTAA A ATCA ----TAATT TGAA TA TAAGACC|----------------------|C GGT
locust      -AATT TTAC ACCT CTAGA ATTGCAG TCTAG A ATCA -----TAAT TGAA TA TAAGACC|taaata----------------|T GAT
cricket     -AAAT -TAC ACCT TTAGA ATTGCAG TCTAA A ATCA -----TTAT TGAA TA TAAGGCC|a----------------------|T GAT
moth        TTTCT TTAT ACCA TGAAA TTTGCAA TTTCA T ATCA ----TTCTT TGAC TA TAAAACT|ta---------------------|T AAT
flea        -TTAA -TAC TCCT TTAAA TTTGCAA TTTAA T ATCA -----TTTT TGAC TA TAAGATT|c----------------------|T GAT
                                                                  <--tRNAcys|                          |<--

            stem T stem     T stem    AC stem      AC stem  D stem       D stem    AA stem
            **** ****       ****      *****        *****    *****        *****     *******
fruit fly   TAAG AAGA -ATAAT TCTT ATAA ATAGA TTTACAA TCTAT C GCCTA AA---CT TCAGC CA CTTAATC|------------------
cockroach   AGAA GAGA AATAAT TCTC CTTA ATAGA TTTACAA TCTAT C ACCTA TTCAACC TCAGC CA TTCTACT|------------------
dragonfly   AAAA GAGG -ATACC CCTC CTAT ATAGA TTTACAG TCTAT C GCCTA AAGT--C TCGAC CA TCTTACC|taagttatattaagggtca
locust      AAGA GAGA -AAACA TCTC ATTA GTAGA TTTACAG TCTAC C ACCTA AAA---T TCAGC CA TCTTACC|------------------
cricket     AAAG GAAA -TTAT- ATTC ATGA AAAGA TTTACAA TCTAT C GCCTA T----CT TCAGC CA CTTTATC|------------------
moth        AAAA GAGA --AATT TCTC GTTA ATAAA TTTACAA TTTAT C GCTTA TAAA--C TCAGC CA TTTTATT|------------------
flea        AAGA GAGT --TTTT ACTC ATAA ATAAA TTTACAA TTTAT T ACCTA AA----T TCAGC CA TCTTATC|------------------
                                                                          <--tRNAtyr|

fruit fly   --------------cataatcgcgacaatggttattttctacaaatcataaaga
cockroach   --------------aactttgcaacgatgaatattttcaacaaatcataaaga
dragonfly   taatataatatttaagatattgcgacgatggctatttttctacgaatcataagga
locust      --------------------gcaaaaatgattattctcaacaaaccataagga
cricket     --------------------gcaacgatgattattctctactaatcataaaga
moth        --------------ttttttgcgaaaatgacttttttcaacaaatcataaaga
flea        ---------------------gtaaaatgattattttctactaatcataaaaa
```

Figure 8. Sequences of the 3 mitochondrial tRNA genes ($tRNA^{Trp}$, $tRNA^{Cys}$ and $tRNA^{Tyr}$) of 7 species of insects. Mitochondrial tRNA genes were aligned based on the loop and stem regions separately. Solid bars indicate the beginnings and ends of the genes. Dashes were used to make the alignments among these genes. Open boxes indicate the anticodons of the tRNAs. Nucleotide positions in stem regions were highlighted by asterisks.

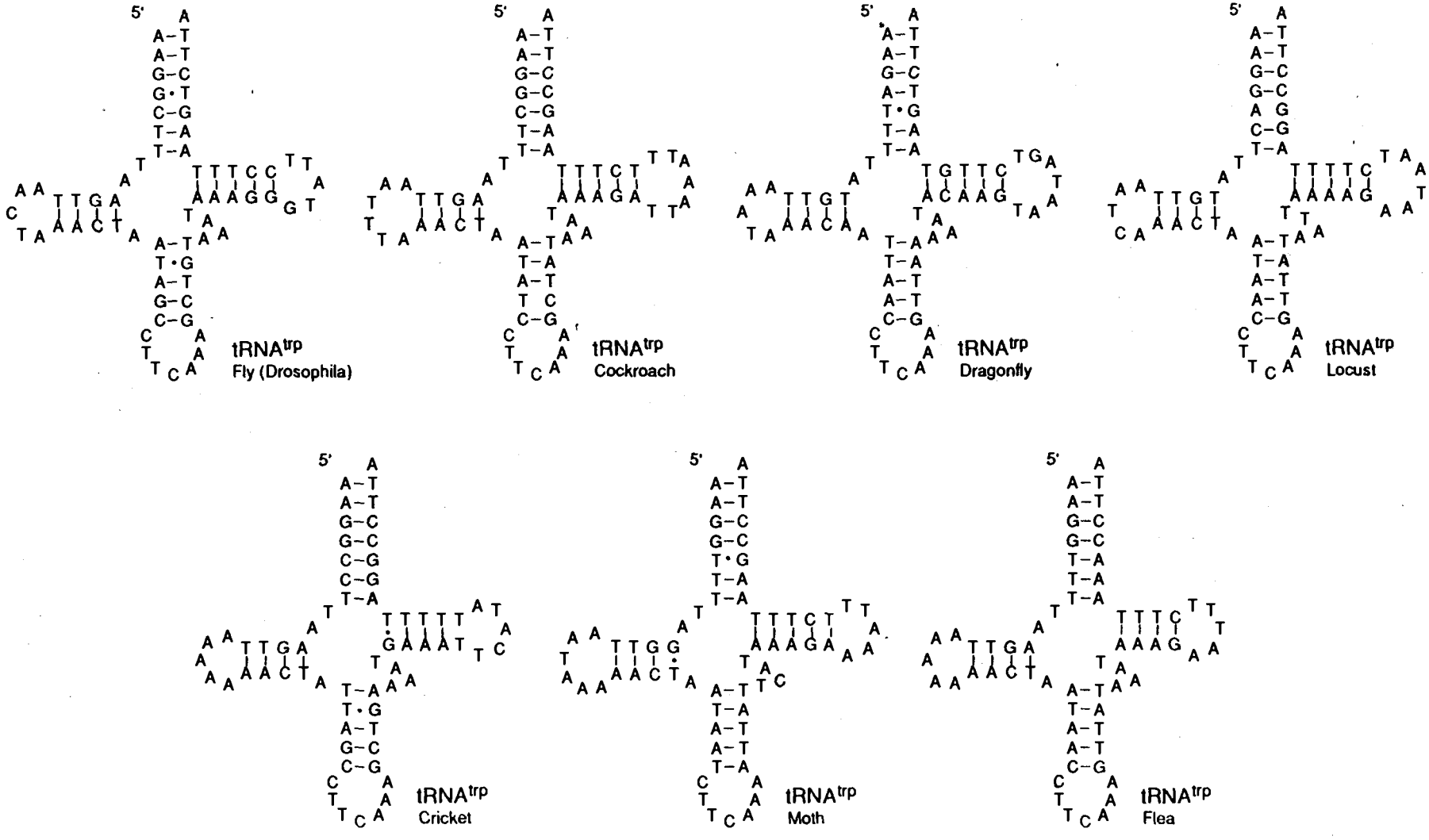

Figure 9. Mitochondrial tRNA^{trp} structures of 7 species of insects.

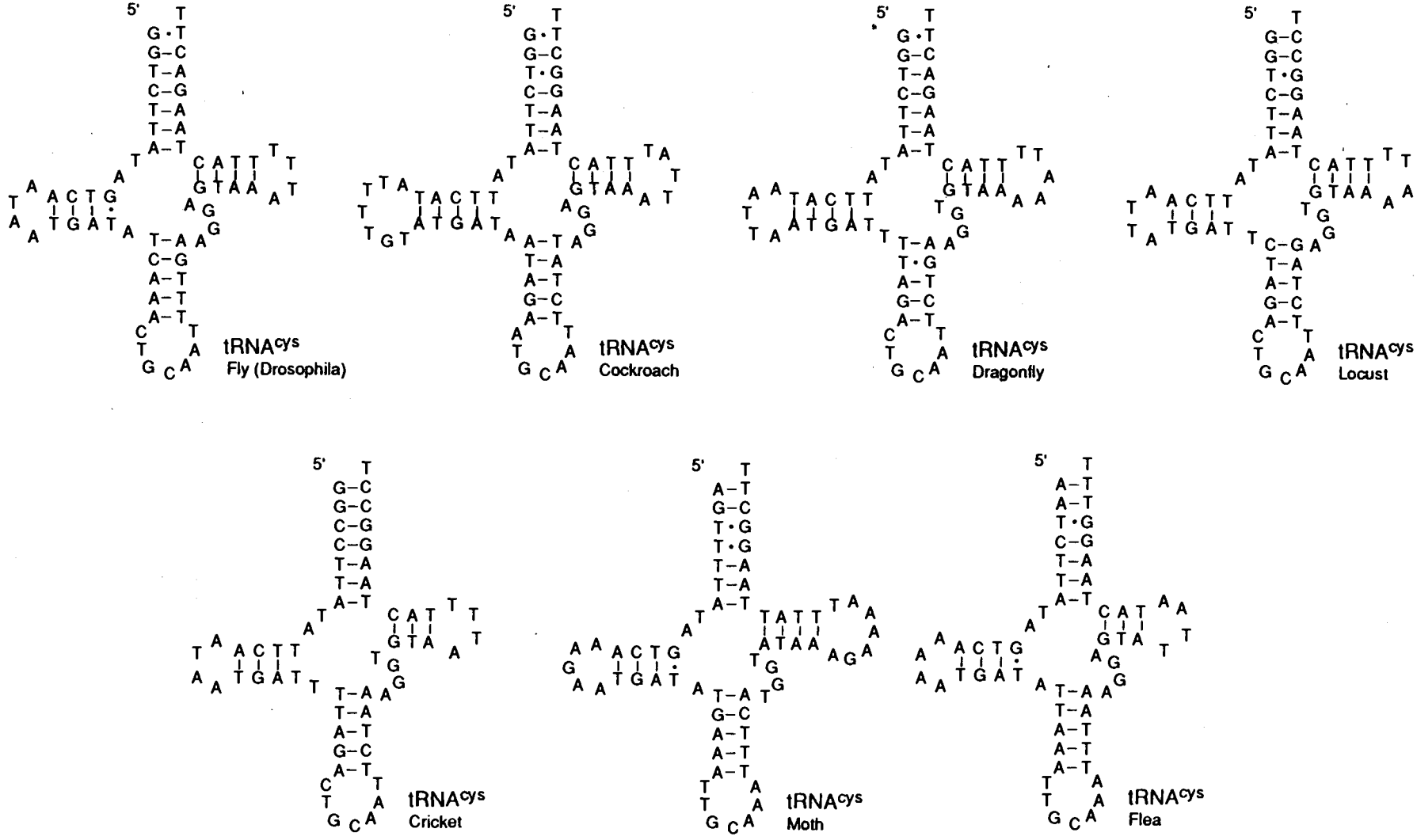

Figure 10. Mitochondrial tRNA^{cys} structures of 7 species of insects.

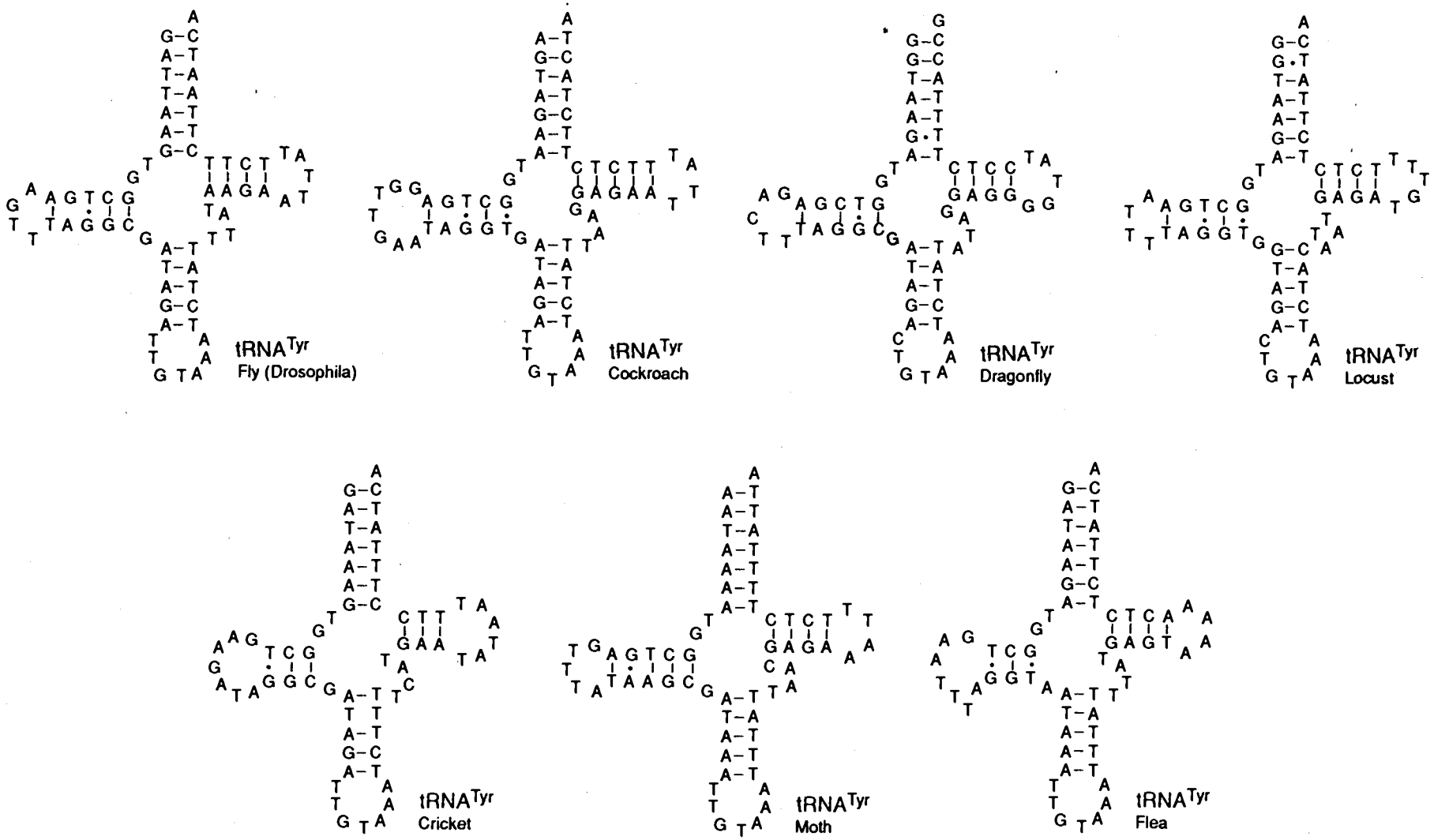

Figure 11. Mitochondrial tRNA^{Tyr} structures of 7 species of insects.

Drosophila and other insects studied have the primitive gene order and that the honeybee organization represents a derived state. The honeybee organization must have arisen as a direct duplication of the three tRNA genes followed by two deletions or arisen as a direct duplication of the overlap region followed by a simple transposition. Figure 12 shows both scenarios require duplications.

Tempo of tRNA Sequence Evolution

Table 8 shows the interspecific divergences in COII genes (Liu and Beckenbach, 1992) and three tRNA genes of 7 species of insects. The average sequence divergence using the Kimura two parameter correction between species are 0.30 and 0.38 for the three tRNAs and the COII sequence, respectively. In 19 of the 21 pairwise comparisons, the divergences observed for tRNA genes is less than was observed for COII genes. The average difference is about 0.08. That indicates that the tRNAs are structurally or functionally more constrained than the COII gene. This observation also agrees with the finding that mitochondrial tRNA sequences generally evolve more slowly than most protein coding sequences (Cann et al., 1982; Brown et al., 1982).

Figure 13 shows the proportion of transitions in pairwise sequence comparisons of three tRNA genes and the COII genes (Liu and Beckenbach, 1992) from these 6 species of insects. The percentage nucleotide divergences were corrected using the Kimura's method (Kimura, 1980). Pairwise comparisons of sequences presented here range from 20 to 41% divergent for three tRNAs and from 25 to 42% divergent for the COII gene. It appears that the transition excesses are essentially eliminated for both COII and the tRNA genes. The proportions of transitions in these comparisons average 53% for three tRNA genes and 38% for COII gene. To determine if these values are limiting values,

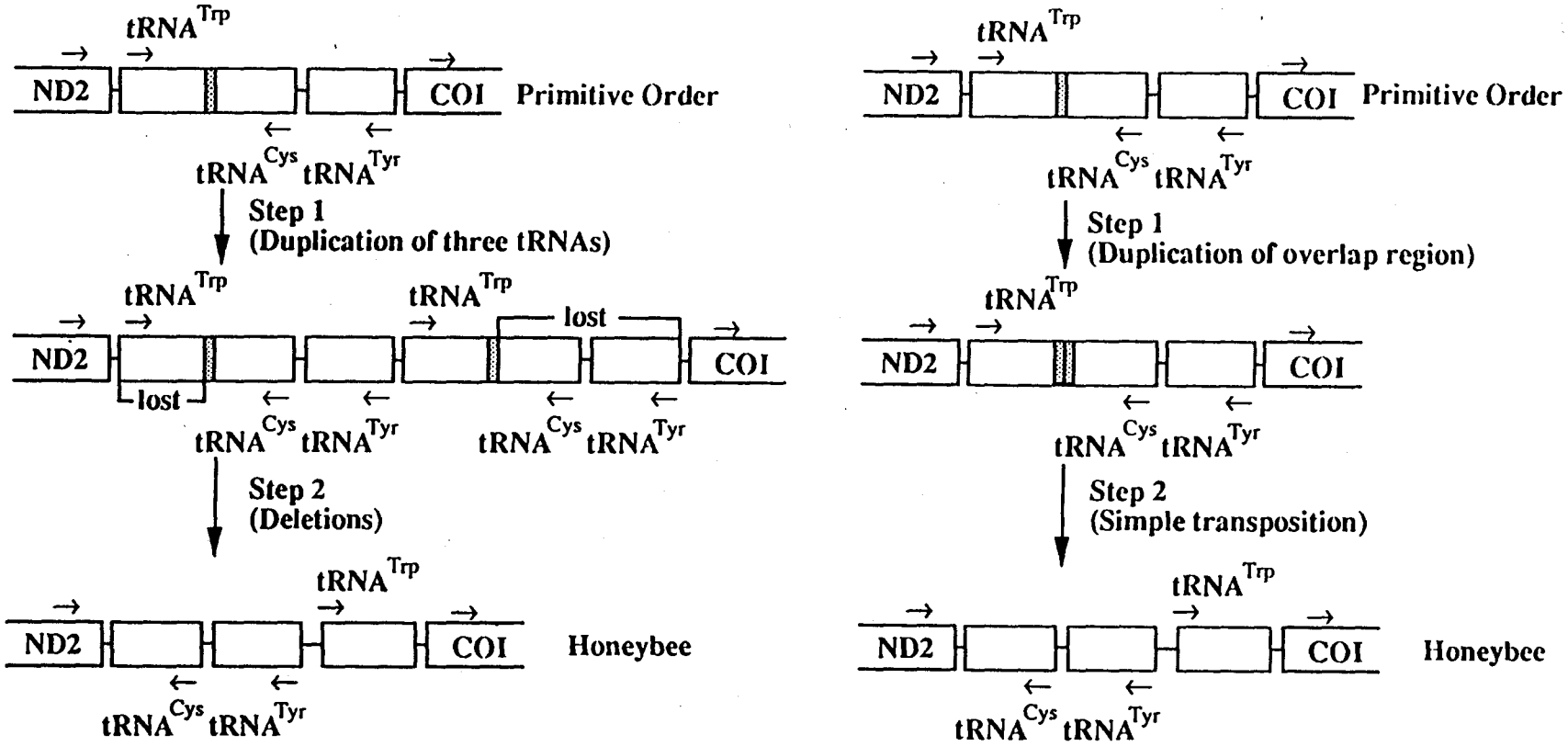

Fig. 12. Two scenarios to explain the honeybee gene organization. Both require duplications.

Table 8. Interspecific divergences in COII gene and three tRNA genes of 7 species of insects. Percent DNA sequence divergence using Kimura correction and the number of nucleotide differences in terms of transition/transversion are given as well. Above the diagonal are the COII gene comparisons. Below the diagonal are the three tRNA gene comparisons.

	dros	cock	drag	locu	cric	moth	flea
drosophila	-	77/98 33.17	62/115 34.08	83/125 43.24	82/94 33.84	54/95 27.29	55/119 33.84
cockroach	22/18 25.95	-	83/114 38.85	83/126 43.56	84/115 40.57	61/119 35.27	64/139 42.50
dragonfly	26/21 31.95	22/29 35.10	-	82/139 47.73	92/115 42.24	78/117 38.52	72/125 40.40
locust	28/14 27.15	19/24 28.79	21/17 24.21	-	72/130 42.53	70/147 47.22	83/132 45.57
cricket	16/18 22.59	19/25 30.75	26/20 32.09	19/17 23.99	-	70/86 28.76	71/111 36.08
moth	29/19 32.67	19/26 30.23	33/28 44.48	34/19 36.53	32/22 40.20	-	37/112 28.10
flea	22/15 24.37	19/23 29.25	25/24 35.29	20/23 30.07	27/16 29.70	18/16 22.24	-

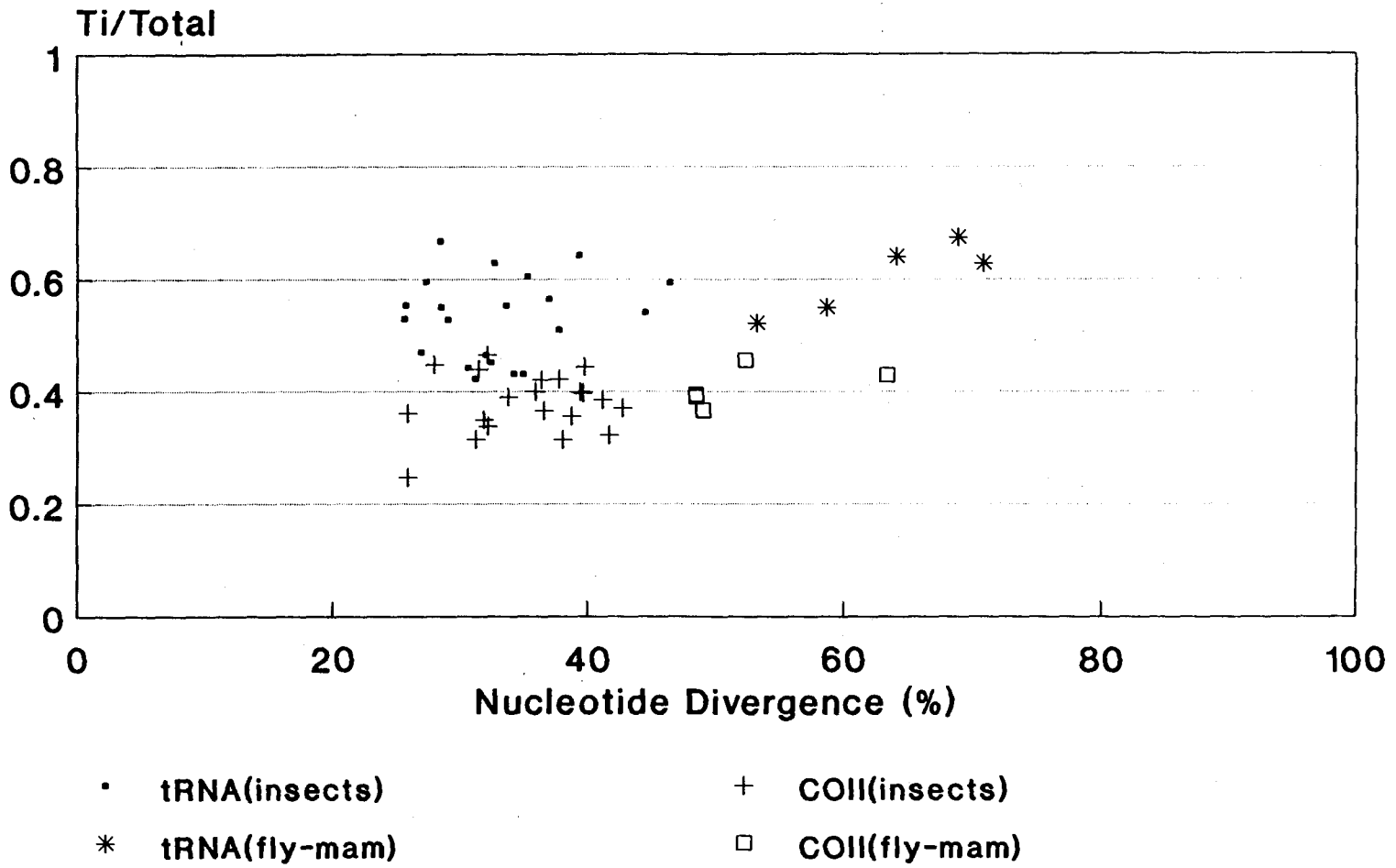

Figure 13. The proportion of transitions in pairwise comparisons of the three tRNA genes and the COII gene for a broad spectrum of relationships. The percentage nucleotide divergences are corrected using the Kimura's method.

we have extended the comparisons to include those between *D. yakuba* and the published tRNA and COII sequences of sea urchin (Jacobs *et al.,* 1988), frog (Roe *et al.*, 1985), cow (Anderson *et al.*, 1982), mouse (Bibb *et al.*, 1981), and human (Anderson *et al.*, 1981). These values, having divergences between 39 and 56% for three tRNAs and between 48 and 63% for COII gene, show no further reduction in the proportion of transitions observed. That might indicate that these tRNA genes are structurally or functionally more constrained than the COII gene.

Phylogenetic Analysis

In view of the apparent saturation in the tRNA data set, I obtained another tRNA sequence data set (118 bp) by combining the three tRNA sequences without the two highly variable loops, D loop and TψC loop. That has been proved to be an effective way to avoid the noise in the tree analysis for tRNA data (Kumazawa and Nishida, in press). The COII sequence data set derived from previous data (Liu and Beckenbach, 1992) is also used to compare the results. The COII data set only contains the second codon position sequences.

Two data sets were analyzed using DNAPARS and NEIGHBOR programs of PHYLIP, version 3.4. The results are shown in Fig 14. In all trees, the dragonfly is treated as a outgroup.

For each data set, distance and parsimony analysis gave very similar trees. In all trees, the three holometabolous insects are clustered together. The cricket and locust are close to each other, especially in the COII trees. One major difference between the trees from two data sets is the position of the cockroach. In the COII trees, the cockroach is close to the dragonfly, but in tRNA trees, the cockroach is more closely

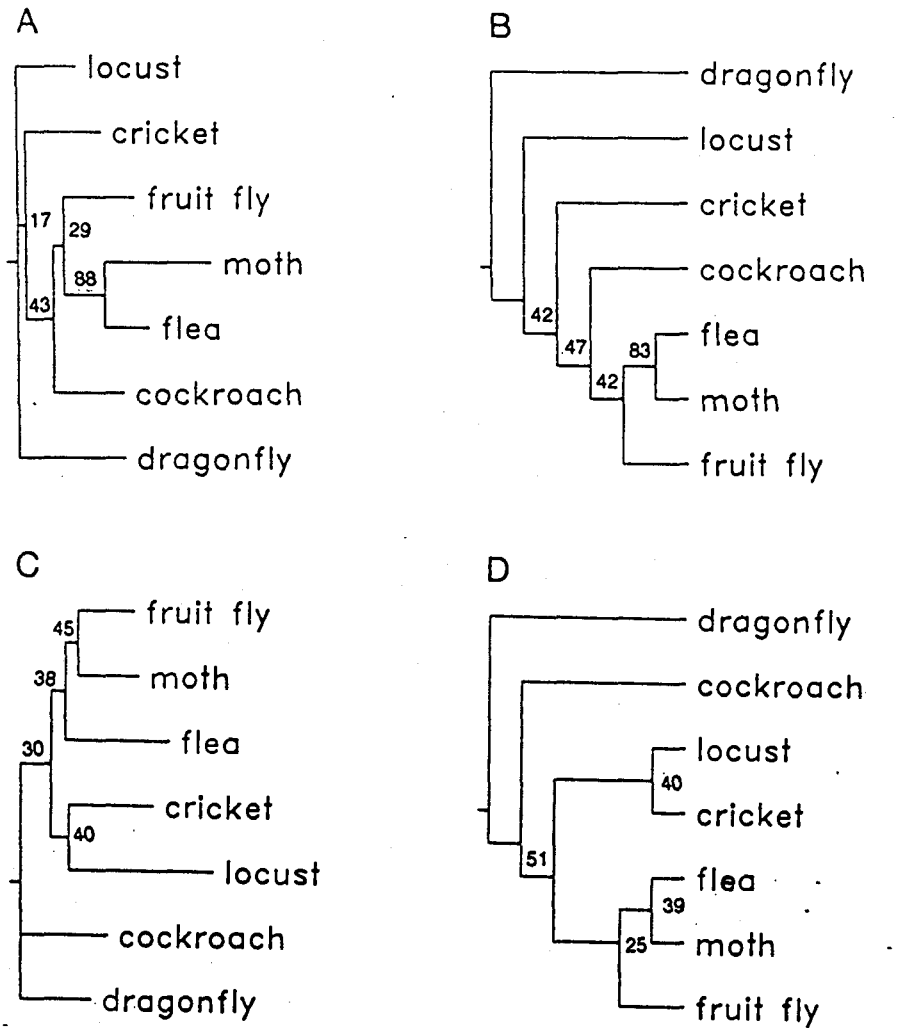

Figure 14. Neighbor-joining and parsimony trees constructed from the tRNA data set and the COII data set. All trees were obtained by the NEIGHBOR and DNAPARS programs in PHYLIP 3.4. (A) neighbor-joining tree based on tRNA data, (B) parsimony tree based on tRNA data, (C) neighbor-joining tree based on COII data, (D) parsimony tree based on COII data. In all trees, the dragonfly is specified as an outgroup.

associated with the flea, moth and fly cluster. The trees based on COII second codon positions support the traditional view of the relationships of these taxa.

Unusual Start Codons of COI Gene.

Figure 15 shows the alignments of the beginning part of the COI genes for 8 species of insects, including the honeybee sequence (Crozier and Crozier, 1992), as well as the human, frog and urchin sequences. The translations were based on the insect mitochondrial code (de Bruijn, 1983; Liu and Beckenbach, 1992) for all the insect species. The inferred insect COI protein sequences were confirmed by comparing vertebrates corresponding protein sequences.

Among many metazoan mtDNAs, triplets other than ATG are often used as translation initiation codons. Some or all ATN codons are used in this way among mammals, Drosophila and echinoderms (Anderson *et al.*, 1981, 1982; Bibb *et al.*, 1981; Clary and Wolstenholme, 1985; Jacobs *et al.*, 1988). Also, ATAA has been suggested as the translation initiation codon of the COI gene of *Drosophila melanogaster* and *D. yakuba* mtDNAs (de Bruijin, 1983; Clary and Wolstenholme, 1985). GTG appears to be used as the translation initiation codon of one gene in the mtDNAs of mouse (ND1, Clary and Wolstenholme, 1985); rat (Gadaleta et al, 1989); chicken (Desjardins and Morais, 1990); two echinoderms, *Paracentrotus lividus* and *Stronglyocentrotus purpuratus* (ATPase 8) (Jacobs et al, 1988; Cantatore et al, 1989); and *D. yakuba* (ND5) (Clary and Wolstenholme, 1985).

Among the completely sequenced mtDNAs, only *Xenopus laevis* contains a full set of protein genes that all appear to begin with ATG (Roe et al, 1985). The mtDNAs of *Caenorhabditis elegans* and *Ascaris suum* are the only ones that totally lack protein genes

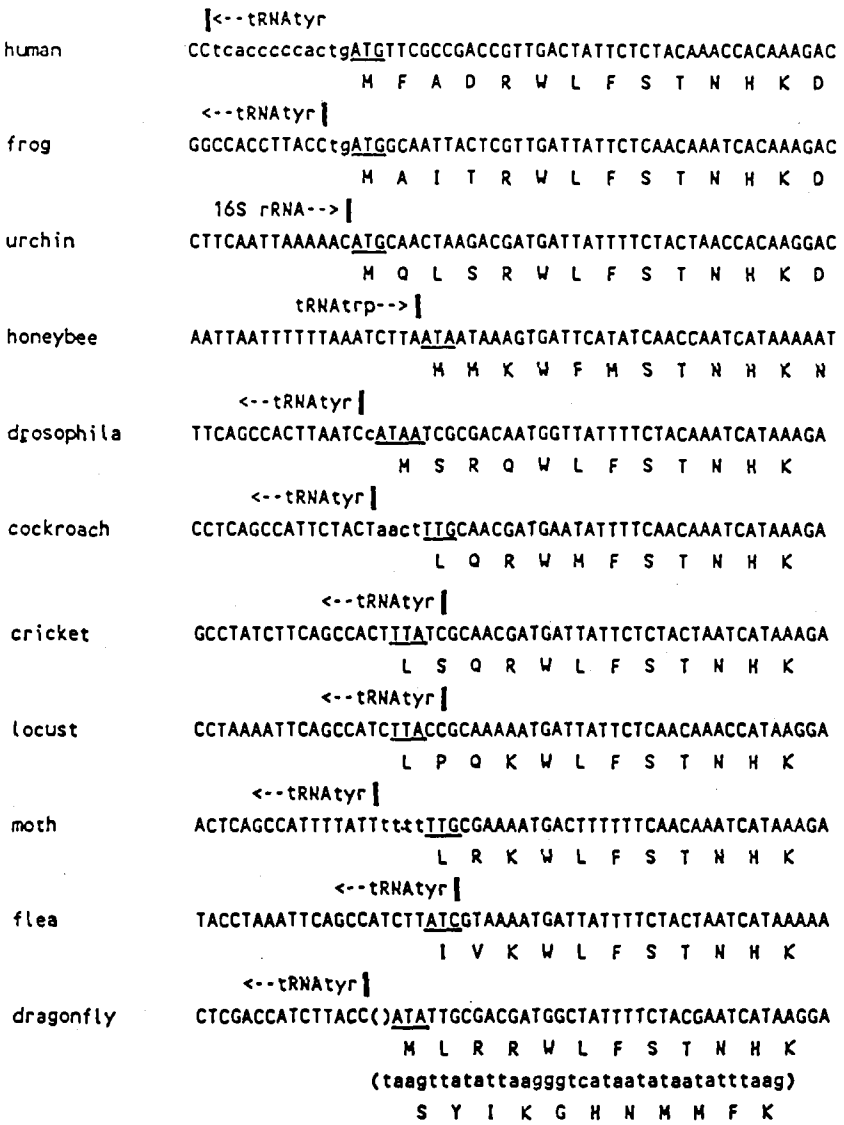

Figure 15. Alignments of the beginnings of COI genes for 8 species of insects as well as the human, frog and urchin sequences. Solid bars indicate the beginnings or the ends of the tRNA genes. Start codons and putative start codons for some insects were underlined. Small letters were used for the gene junction sequences. The unassigned sequence between the $tRNA^{tyr}$ and COI gene in dragonfly was marked with brackets.

with an ATG initiation codon (Okimoto et al, 1990). These mtDNAs are unusual in that 6 of the 12 mt-protein genes of *A. suum*, and 3 of the mt-protein genes of *C. elegans* appear to begin with TTG. Five of the remaining 6 *A. suum* mt-protein genes have a putative ATT initiation codon, and one, the COIII gene, has a putative GTT initiation codon. The translation initiation codons of the remaining *C. elegans* mt-protein genes appear to be ATT (6 genes) or ATA (3 genes). As homologous genes in mtDNAs of different species within a class or phylum sometimes begin with different putative translation initiation codons, it seems unlikely that the unorthodox codons play any role in regulating mt-protein gene expression, as has been suggested for the TTG initiation codon of the *E. coli* adenylate cyclase gene (Reddy et al, 1985).

Comparison of the junction between $tRNA^{Tyr}$ and COI for these insect species does not provide any easy resolution of the problem of locating the COI start codon. The data presented in Fig 15 indicates that TTG may be used by two insect species, moth and cockroach. I believe that the triplet TTA, which normally specifies leucine, is the most plausible translation initiation codon of the two orthopteran insects, the cricket and locust. In the dragonfly, there is a unique 35 nucleotide sequence between the $tRNA^{Tyr}$ and the putative start codon of COI. This sequence represents a continuation of the COI open reading frame in the 3' direction. In fact, the reading frame is open through part of the $tRNA^{Tyr}$ sequence as well. In the resently published blue mussel (Hoffmann *et al.*, 1992) and nematode (Okimoto *et al.*, 1992) COI sequences are both 9-10 amino acids longer at the upstream of this protein gene. When the amino acids from this region of the dragonfly and the 9 amino acids from blue mussel and nematodes are compared, no significant similar region can be found among them. I am unwilling to

assign these nucleotides as part of the COI gene, because of the relatively consistent size of the COI gene for the other insects studied here.

Some insect species studied here (cricket, locust and flea) apparently have a COI region overlapping with the $tRNA^{Tyr}$, on the complementary strand.

Conclusion

By amplifying the junction regions between the ND2 and COI genes, I obtained sequence data for 3 tRNA genes from 5 orders of insects. All the insects I examined, including holometabolous and hemimetabolous, have the same tRNA gene order as *Drosophila yakuba*. I suggest, therefore, that this organization is primitive. The recently published honeybee sequence has a different gene order from that found in other insects. Two scenarios have been described to explain the process. By comparing the 5' end of COI genes from a broad range of species, I have assigned two putative start codons (TTG and TTA) for the COI genes of some insects. The small tRNA data set cannot provide resolution of the ancient divergence of the insect orders.

CHAPTER IV

THE MITOCHONDRIAL GENE ORDER OF FLEA *Ctenocephalides felis*

Introduction

Animal mtDNA has been a very useful tool for molecular studies of evolution because of several advantageous characteristics of this genome. Recent advances in the PCR-aided direct DNA sequencing technique (Gyllensten and Erlich, 1988; Kocher et al., 1989; Liu and Beckenbach, 1992) and methods for analyzing sequence relationships (Felsenstein, 1988; Miyamoto and Cracraft, 1991) have made possible the rapid and accurate sequence determination of mtDNAs and the subsequent construction of phylogenetic trees. The noncoding region of mtDNA, which changes its sequence more rapidly than the gene-coding regions, is thought to be suited for the analysis of populational relationships within species, whereas gene-coding regions are usually used to infer phylogenetic relationships above the species level (Moritz et al., 1987; Liu and Beckenbach, 1992).

However, mtDNA sequences have not been proved as useful in reconstructing phylogenies for deep branches as they are for shallow ones (Liu and Beckenbach, 1992; Beckenbach et al., 1993). That is because the rapid rates of sequence evolution of mtDNAs compared to those of nuclear DNAs produce an apparent saturation of sequence differences among distantly related species due to multiple substitutions at single sites (Brown, 1985).

The order in which the genes are arranged in mtDNA appears to change at a very slow rate (Brown, 1985; Jacobs et al., 1988). This may provide very useful information

about the phylogeny of distantly related metazoan groups. Recently, Smith et al. (1993) have analyzed the mitochondrial gene order in extant echinoderm classes. Previous analyses have demonstrated that the sea star mitochondrial genome contains a large inversion in comparison to the mitochondrial DNA of sea urchins (Smith et al., 1989,1990; Jacobs et al., 1989; Asakawa et al., 1991). Using PCR and direct sequencing, Smith et al. have been able to demonstrate that brittle stars and sea stars have similar mitochondrial gene arrangements. This result clarifies the relationships among these classes.

Complete nucleotide sequences and gene content have been determined for five mammals, human, cow, mouse, rat and fin whale (Anderson., 1981, 1982; Bibb et al., 1981; Gadaleta et al., 1989; Arnason et al, 1991); a bird, *Gallus domesticus* (Desjardins and Morais, 1990); an amphibian, *Xenopus laevis* (Roe et al., 1985); two sea urchins, *Strongylocentrotus purpuratus*, and *Paracentrotus lividus* (Jacobs et al., 1988; Cantatore et al., 1989); a fish, *Gadus morhua* (Johansen et al, 1990), three nematodes, *Caenorhabditis elegans*, *Ascaris suum* and *Meloidogyne javanica* (Okimoto et al., 1991,1992) and two insects, *Drosophila yakuba* and *Apis mellifera* (Clary and Wolstenholme, 1985; Crozier and Crozier, 1993). Also, sufficient nucleotide sequences have been obtained from mtDNA molecules of a mollusk, *Mytilus edulis* (Hoffman et al., 1992), to establish gene orders in these molecules.

Partial mtDNA sequences have also been obtained for a wide range of species including the following: chimpanzee, *Pan troglodytes*, pygmy chimpanzee, *P. paniscus*, lowland gorilla, *Gorilla gorilla* (Foran et al., 1988); japanese monkey, *Macaca fucata* (Hayasaka et al., 1991); two dolphins, *Cephalorhynchus commersonii* and *Delphinus*

delphis (Southern et al., 1988); a bird, japanese quail, *Coturnix japonica* (Desjardins and Morais, 1991); six species of salmonid fish, four of the genus *Orcorhynchus* and two of the genus *Salma* (Thomas and Beckenbach, 1989); a sea urchin, *Arabacia lixula* (De Giorgi et al., 1991); starfish and sea stars, *Pisaster ochraceus*, *Asterias forbessi*, *Asterias amureusis* (Smith et al., 1989, 1990; Jacobs et al., 1989) and *Asternia pectinifera* (Asakawa et al., 1991); a brittle star, *Ophiopholus aculeata* and a sea cucumber, *Parastichopus californicus* (Smith et al., 1993); the insects *Drosophila melanogaster* (de Bruijn, 1983; Garesse, 1988), *D. virilis* (Clary and Wolstenholme, 1987), and mosquito, *Aedes albopictus* (HsuChén et al., 1984; Dubin et al., 1986), and locust, *Locusta migratoria* (McCracken et al., 1987; Uhlenbusch et al., 1987; Hancke and Gellissen, 1988); the brine shrimp, *Artemia* sp. (Batuecas et al., 1988); a platyhelminth, the liver fluke, *Fasciola hepatica* (Garey and Wolstenholme, 1989).

Although the methods for analyzing the evolutionary relationships among gene rearrangements are being developed (Sankoff et al., 1990), more data on mitochondrial gene order from all taxa are needed for assessment of its value in clarifying deeply branched phylogenies.

The evolutionary position for the flea, Order Siphonaptera, has been an unsolved question in biology (Hristensen, 1981). Is Siphonaptera the sister group of the Diptera, Mecoptera, or even Lepidoptera? Recent complete mtDNA sequence data for the honeybee (Crozier and Crozier, 1993) indicates that 11 of the 22 tRNA genes are in altered positions relative to the mitochondrial genome of *Drosophila yakuba*. Does the flea mitochondrial gene order reflect the same extent of rearrangement compared to *D. yakuba*? To answer these questions, I have determined most of the flea mt-genome

organization by PCR and direct sequencing of the gene junctions, including 15 tRNA genes.

Materials and Methods

DNA Source

A single individual of the flea (*Ctenocephalides felis*) freshly caught was used as source material for DNA extraction.

DNA Extraction

The whole flea was ground in 60μl protease buffer (0.1M Tris-OH, pH 8.0, 0.05 M EDTA, 0.2 M NaCl, 1% SDS, with 0.4 mg/ml protease K), and immediately placed at 65°C for 3 min. The resulting solution was extracted once with phenol (saturated with 10 mM Tris, 1 mM EDTA, pH 8.0) and once with chloroform/isoamyl alcohol (24:1). The supernatant was removed and 2 volumes of 95% ethanol was added to precipitate the DNA. The suspension was pelleted at 40K, 15 min, 4°C, washed twice with 70% ethanol, dried under vacuum, and dissolved in 50μl sterile ddH_2O.

Primers

Fourteen sets of primers were designed to amplify all the tRNA genes. By comparing all the protein and rRNA sequences from mammals, sea urchin, frog and *Drosophila yakuba*, the primers were placed within relatively conservative portions in protein and rRNA genes surrounding gene junctions containing single or clusters of tRNA genes. Table 9 lists the primers used for PCR and sequencing. Table 10 shows the 12 most successful primers for PCR and sequencing the flea mitochondrial genome. Table 10 also shows the primer sequence alignments among the fly, bee, sea urchin, cow, frog and the nematode, C. elegans. Since mismatched base pairs in proximity to the 3' end of PCR primers are known to have a critical effect on the efficiency of amplification (Sommer and Tautz, 1989), more than one nucleotide was included in some primers at

Table 9. Primers used in amplification and sequencing of the flea mtDNA

Name of primer	Sequence	Positions spanned*
Primer A	5'-ATGGCAGATTAGTGCAATGG-3'	3018 - 3038
Primer B	5'-GTTTAAGAGACCAGTACTTG-3'	3804 - 3784
Primer C	5'-TCATCTTATAGGTACTATTTGAGG-3'	3935 - 3915
Primer D	5'-GAATGAAATCAAGG(GT)ATATTA-3'	5931 - 5950
Primer E	5'-GA(CT)CAAGGTTGGTCAGAA-3'	6580 - 6561
Primer F	5'-TGTTCCTACTATTCCGG(AC)TCA-3'	1560 - 1540
Primer G	5'-TAGGTGGACTACCTCCATTTT(CT)AGG-3'	982 - 1005
Primer H	5'-TGTACACATCGCCCGTC-3'	14237 - 14220
Primer I	5'-GAAATTTGTGGAGCAAATCATAG-3'	3673 - 3696
Primer J	5'-TTAGGGTCAAATCCACA-3'	5750 - 5733
Primer K	5'-GTATAACCGCGACTGCTGGCA-3'	14744 - 14765
Primer M	5'-TCTAAACCTATTCAAGCT-3'	346 - 329
Primer N	5'-ATCCAACATCGAGGTCG-3'	12947 - 12964
Primer O	5'-GCTGTTCGATTAACAGCTA-3'	4536 - 4555
Primer P	5'-GCTAATATAGCAGCTCCTCC-3'	8503 - 8483
Primer Q	5'-CCAACTTTATTAGGTCC-3'	12546 - 12563
Primer R	5'-ACATGATCTGAGTTCAAACCGG-3'	12888 - 12864
Primer S	5'-ACATGAATTGGAGCTCGACCAGT-3'	11522 - 11545
Primer T	5'-GATTTTGCTGAAGGTGAATCAGA-3'	12074 - 12051
Primer U	5'-TCTGCTGTGTAGTGTATAGCTA-3'	10690 - 10669

Name of primer	Sequence	Positions spanned*
Primer V	$^{5'}$-CGTTCCGGTTGATAACCTCATC-$^{3'}$	9152 - 9174
Primer W	$^{5'}$-TTCTGACCAACCTTGATC-$^{3'}$	6561 - 6580
Primer X	$^{5'}$-AGGAGCTTCAACATGAGCT-$^{3'}$	8923 - 8942
Primer Y	$^{5'}$-AAATACAACAAGTATTCCTCC-$^{3'}$	10184 - 10163

*The position of each primer refers to the complete *D. yakuba* sequence.

Table 10. 12 most successful primers for PCR and sequencing the flea mitochondrial genome. Dots indicate identity to the primer sequences. Dashes indicate delesion/insertion events. The positon of each primer refers to the original sequence.

Primer A

(20mer)	5'- ATGGCA-GATTAGTGCAATGG -3'	Position*
Fly	......-.....	3018
Bee	......GA..A......T..A	3361
Urchin	G.A...-A.G.G..TA--AT.	2093
Bovine	G.....-..GCCCG.T...T.	3030
Xenopus	G.....-..GCCTG..T.AT.	4730
C.elegans	..A...TA.GA......T.T.	3313

Primer B

(20mer)	5'- GTTTAAGAGACCAGTACTTG -3'	Position
Fly	.....	3804
Bee	A.....A...T..ACT...A	4407
Urchin	AA.......T.T.AAG...T	8440
Bovine	AC....A..GTT...G..AT	8094
Xenopus	.C....A..G.TGTCG..AA	9838
C.elegans	AA....A..TTTG.C....T	3283

Primer C

(24mer)	5'- TCATCTTATAGGTACTATTTGAGG -3'	Position
Fly	......A..T...G.......T..	3937
Bee	....T.........T.........	4470
Urchin	.....AAGC.AA.T...G...T..	8502
Bovine	....G..GAC.TGT...G...C..	8155
Xenopus	....GGGCCT..GTT..AC..T..	9899

Primer D

(21mer)	5'- GAATGAAATCAAGGTATATTA -3'	Position
Fly	.....	5932
Bee	.......TAA.TAA.TAT...	6503
Urchin	..G....TAA....GGGCC..	10516
Bovine	..G....C....AAAGG.C..	10139
Xenopus	......CT......AGGCC..	11871
C.elegans	..G...TGAT.T....A....	11656

Table 10--Continued

Primer E

(18mer)	5'- GATCAAGGTTGGTCAGAA -3'	Position
Fly	.....	6579
Bee	T.....AAAAT.T.T.T	7019
Urchin	..C.G...A...CA....	13828
Bovine	..C.T.ATC..ACT....	13768
Xenopus	....T.TCC..A.AT...	15517
C.elegans	T..TT.TAAAATAA....	13104

Primer F

(21mer)	5'- TGTTCCTACTATTCCGGCTCA -3'	Position
Fly	.....	1560
Bee	..A.....G.....A.A...	1880
Urchin	...G.....C..G..A.....	5877
Bovine	.....A.....C..	5779
Xenopus	G.....G..G.GC..T.....	7489
C.elegans	A..A..A..C..A..A.A...	7961

Primer G

(21mer)	5'- TAGGTGGACTACCTCCATTTTTAGG -3'	Position
Fly	.....T.....	982
Bee	..CA.TCTTA..A.TA..A....AT	1194
Urchin	....A..C..T..C.....AACC..	3952
Bovine	....A.....C.....CC.A.CT..	5017
Xenopus	.....C..T..A..TC...C...	6729
C.elegans	.T.AG...AATTT.TA.A.A.G.TA	4082

Primer P

(20mer)	5'- GCTAATATAGCAGCTCCTCC -3'	Position
Fly	.....	8503
Bee	T.A.....G.T....AGT	8913
Urchin	..C..AT.G.G.TTA.....	11664
Bovine	A.C..CT....TCTA..C..	11621
Xenopus	.....CCTA..A..	13362
C.elegans	T.A....GG.GT.TA..A..	7481

Table 10--Continued

Primer R

(22mer)	5'- ACATGATCTGAGTTCAAACCGG -3'	Position
Fly	.....	12887
Bee	.....T.....T..A	13415
Urchin	..G.....G.....	5661
Bovine	..G.....G.....	2853
Xenopus	..G.....G.....	4551

Primer S

(23mer)	5'- ACATGAATTGGAGCTCGACCAGT -3'	Position
Fly	.....	11523
Bee	......T.A..TAAA.A.TT.A.	12012
Urchin	.....G..A..CAGA.A...C..	15517
Bovine	.....GA.A.....	15519
Xenopus	..C.....G..A.....	17254
C.elegans	.G....C.A..TCAAT.CA.T..	5488

Primer T

(23mer)	5'- GATTTTGCTGAAGGTGAATCAGA -3'	Position
Fly	.....	12074
Bee	.....AAT.....A.....T..	12628
Urchin	..CC.AA.A.....A.....T..	2767
Bovine	.....AA.....A.....	3695
Xenopus	..CC..A.A.....A.....T..	5408
C.elegans	......T.....AGG..	2339

Primer W

(18mer)	5'- TTCTGACCAACCTTGATC -3'	Position
Fly	.....	6562
Bee	A.A.A.ATTTT.....A	7002
Urchin	....TG...T...C.G..	13845
Bovine	....AGT..GAT.A.G..	13785
Xenopus	...AT.T..GGA.A....	15534
C.elegans	....TTATTTTA.AA..A	13121

*The position of each primer refers to the complete sequence of *D. yakuba* (Clary and Wolstenholme, 1985), honeybee (Crozier and Crozier, 1993), sea urchin (Jacobs et al., 1988), cow (Anderson et al., 1982), frog (Roe et al., 1985) and *C. elegans* (Okimoto et al., 1992).

their 3' end (Table 10 only use one nucleotide for the aligments), so that an exact match could take place with an ambiguous position of template protein genes.

Amplification and Sequencing.

The gene junctions were amplified using the polymerase chain reaction (PCR) performed in 50μl reactions using the GeneAmp kit with AmpliTaq DNA polymerase (Perkin Elmer Cetus). I followed the protocol provided with the kit. Thermal cycling was done in an Ericomp Programmable Cyclic Reactor. The PCR consisted of 35 cycles performed as follows: denaturation at 94°C for 1 min, annealing at 50°C for 1 min, and extension at 70°C for 2 min.

The double-stranded DNA was purified by electrophoresis in low melting agarose gel and followed by concentration using Amicon 30 microconcentrators. About 7μl was used for double-stranded sequencing using Sequenase kit (United States Biochemical Corporation). I used modifications of the standard Sequenase protocol given by Palumbi et al. (*The Simple Fool's Guide to PCR*, compiled by Palumbi, S., Martin, A., Romano, S., McMillan, W.O., Stice, L., and Grabowski, G.).

Data Analysis

Sequences were entered into the computer with the aid of eyeball sequence editor, ESEE (Cabot and Beckenbach, 1989). All sequences were then compared and aligned with the whole *Drosophila yakuba* sequence using FASTA programs (Pearson, 1991). The putative tRNA sequences, which aligned well with the *D. yakuba* tRNA sequences, were folded by hand into the configurations inferred for the corresponding tRNAs.

Results and Discussion

The flea mitochondrial genome organization

Figure 16 shows the mitochondrial genome organization map for the fly (*Drosophila yakuba*) and the flea (*Ctenocephalides felis*). The flea mitochondrial gene order is partially determined (the dark region) by PCR and direct sequencing of gene junctions. So far, the flea mitochondrial gene order is the same as the fly. All together, the locations of 24 genes have been determined, including 15 tRNA genes, 8 protein genes and one large subunit of the ribosomal gene. The determined tRNA genes are: $tRNA^{trp}$,$tRNA^{cys}$, $tRNA^{tyr}$, $tRNA^{leu}_{UUR}$, $tRNA^{lys}$, $tRNA^{asp}$, $tRNA^{ala}$, $tRNA^{arg}$, $tRNA^{asn}$, $tRNA^{ser}_{AGY}$, $tRNA^{glu}$, $tRNA^{phe}$, $tRNA^{his}$, $tRNA^{ser}_{UCN}$ and $tRNA^{leu}_{CUN}$.

Recently the honeybee (*Apis mellifera*) mitochondrial genome organization has been determined (Crozier and Crozier, 1993). Comparison of the flea and the honeybee gene orders indicates 8 tRNA genes are in altered position. These tRNAs are: $tRNA^{cys}$, $tRNA^{tyr}$, $tRNA^{trp}$, $tRNA^{asp}$, $tRNA^{lys}$, $tRNA^{glu}$, $tRNA^{ser}_{AGN}$ and $tRNA^{ala}$. All the protein genes and the lrRNA gene are in the same order.

Cytb, $tRNA^{Ser}_{UCN}$ and ND1 genes

Primer S and primer T are based on sequence comparisions of the Cytb and ND1 genes across a wide range of species. They can prime from the conservative portions at the 3' end of the Cytb gene and 3' end of the ND1 gene, and amplify the junction between the Cytb and ND1 genes of the flea.

Figure 17 shows the aligned nucleotide sequences of some portions of Cytb and ND1 genes and also the $tRNA^{Ser}_{UCN}$ gene for the flea and *D. yakuba*. Unlike *D. yakuba*, the flea uses a single T as the incomplete stop codon for the Cytb gene. Five of the stop

Drosophila yakuba / Ctenocephalides felis

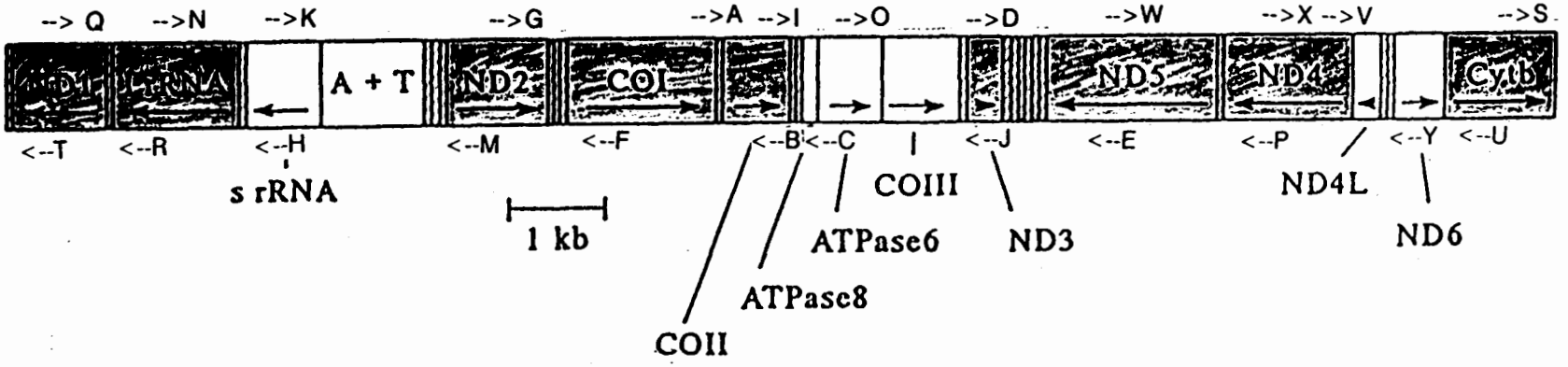

Figure 16. Mitochondrial genome organization map for the fly (Drosophila yakuba) and the flea (Ctenocephalides felis). The dark region indicates the determined mtDNA gene order for the flea. The location and direction for each primer used in this Chapter is also indicated.

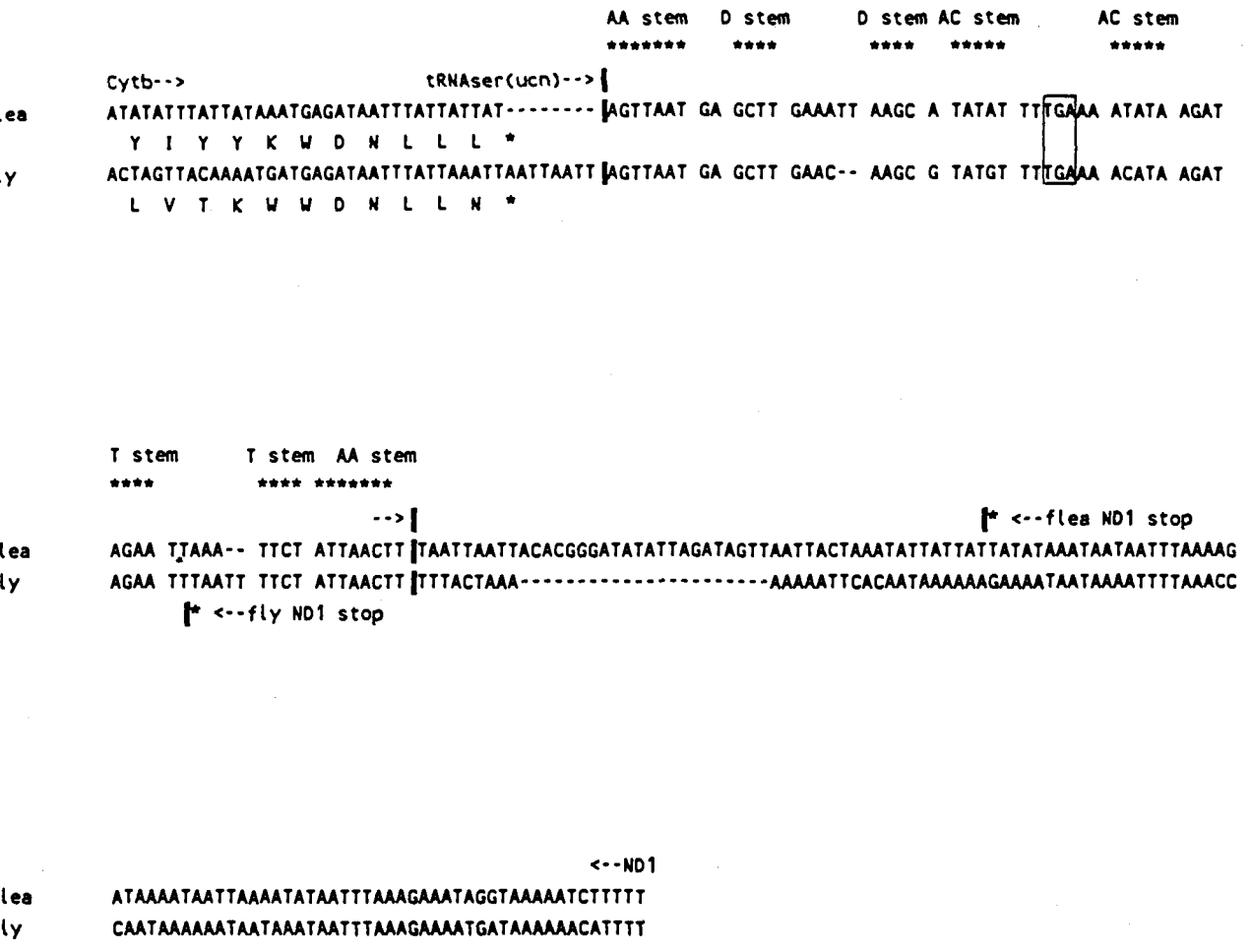

Figure 17. Sequences of portions of Cytb and ND1 genes and the $tRNA^{ser}_{UCN}$ gene for the flea and the fly. The gap was introduced to align the sequences. Solid bars indicate the beginnings and the ends of the genes. Open box indicates the anticodon of the tRNA. Nucleotide positions in stem regions were highlighted by asterisks.

codons of *D. yakuba* mtDNA are reported to be incomplete, either T or TA (Clary and Wolstenholme, 1985). There is no spacer between Cytb and $tRNA^{ser}_{UCN}$ for the flea and there is a 6 bp spacer between these two genes in *D. yakuba*. Also, the flea has a 49 bp long spacer between the $tRNA^{ser}_{UCN}$ gene and the ND1 gene. For the *D. yakuba*, the ND1 gene extends into the complementary strand of the $tRNA^{ser}_{UCN}$ gene.

Figure 18 shows the aligned nucleotide and amino acid sequences for the 3' end of ND1 gene for the flea, fly and honeybee. The 3' end of the flea ND1 gene codes for three successive TAA terminators and the protein is 15 amino acids shorter than the fly. The protein sequence of the honeybee is also shorter (by 13 amino acids) than that of the fly. Figure 19 shows that the tRNA sequence can be folded into the configurations inferred for the corresponding $tRNA^{Ser}_{UCN}$.

ND1, $tRNA^{Leu}_{CUN}$ and lrRNA genes

Primer R was constructed based on a conserved region at the 3' end of lrRNA gene. Primer S and primer R were used to amplify the sequence between the Cytb and lrRNA genes of the flea. The amplified fragment is 1.25 kb, and matches very well with the corresponding fragment from *D. yakuba*.

Figure 20 shows the aligned nucleotide sequences of some portions of ND1 and lrRNA genes as well as the $tRNA^{Leu}_{CUN}$ gene for the flea and the fly. In the flea, there is a 2 bp spacer between the ND1 and the $tRNA^{leu}_{CUN}$. It's shorter than the 10 bp fly spacer. There is no spacer between the $tRNA^{leu}_{CUN}$ gene and the lrRNA gene in the flea. Figure 21 shows that the tRNA sequence can be folded into the configuration inferred for the corresponding $tRNA^{Leu}_{CUN}$.

```
fly         ATTTTTATAGCAGAATATGCTAGAATTTTATTTATAAGAATATTGTTTTGTGTAATTTTTTTAGGTTGTGATGTATTTAATTTATTATTTTATGTAAAA
flea        ......T.G..T.....................T.TA.A......T....G.....G..AG.G...TAT..ATC......T....TCT.....
honeybee    ......T..T.T......ATA.AT...A..........G.TA.T..AA.AT....A....AT..A.T.A.ATAT.GATCAA.TAA....AT.T...TT
fly          I  F  M  A  E  Y  A  S  I  L  F  M  S  M  L  F  C  V  I  F  L  G  C  D  V  F  N  L  L  F  Y  V  K
flea         .  .  L  .  .  .  .  .  .  .  .  .  .  F  M  .  .  L  M  .  .  .  G  .  Y  L  S  .  F  .  F  L  .
honeybee     .  .  L  S  .  .  M  N  .  M  .  .  .  V  I  L  S  L  M  .  Y  .  F  K  Y  W  S  I  K  .  I  L  I

fly         TTAACTTTTATTTCATTTGTATTTATTTGAGCTCGAGGTACATTACCTCGGTTTCGTTATGATAAATTAATATATTTAGCTTGAAAATGTTTTTTATCA
flea        ...GTA...T.A..T...A............T......A...........A.A...A........G.....G..............A.A......C.T
honeybee    .ATTTA...CA.ATT.G.T..A........AT........TT........AA....A..............A..A..TG.....CTGAAA.A...AT.
fly          L  T  F  I  S  F  V  F  I  W  A  R  G  T  L  P  R  F  R  Y  D  K  L  M  Y  L  A  W  K  C  F  L  S
flea         .  V  .  L  .  .  M  .  .  .  V  .  .  .  .  .  .  Y  .  .  .  .  .  .  .  .  .  .  .  S  .  .  P
honeybee     Y  L  .  H  I  C  L  I  .  .  I  .  .  I  .  .  .  I  .  .  .  .  .  .  N  M  C  .  T  E  M  .  M

fly         TTTTCTTTAAATTATTTATTATTTTTTATTGGGTTTAAAATTTTATTATTTTCTTTTTTATTGTGAATTTTTTTTTAGTAAAAAGTTAATAGAAAATTAA
flea        A..............A.T...A..A..T.ATCT......T.A.....TA.A.AA.AA.A.
honeybee    ..AGT.A...T.......A...A...ATA.TTTA.A...GAA..T....G.AT..AA
fly          F  S  L  N  Y  L  L  F  F  I  G  F  K  I  L  L  F  S  F  L  L  W  I  F  F  S  K  K  L  M  E  N  *
flea         I  .  .  .  .  I  .  I  I  L  S  .  .  L  .  F  M  *  *  *
honeybee     L  V  M  I  .  .  M  Y  L  Y  F  M  .  E  F  .  C  I  *
```

Figure 18. Aligned nucleotide and amino acid sequences for the 3' end of the ND1 genes for the flea, fly and honeybee. Dots indicate identity to the fly sequence. Asterisks indicate the stop codons.

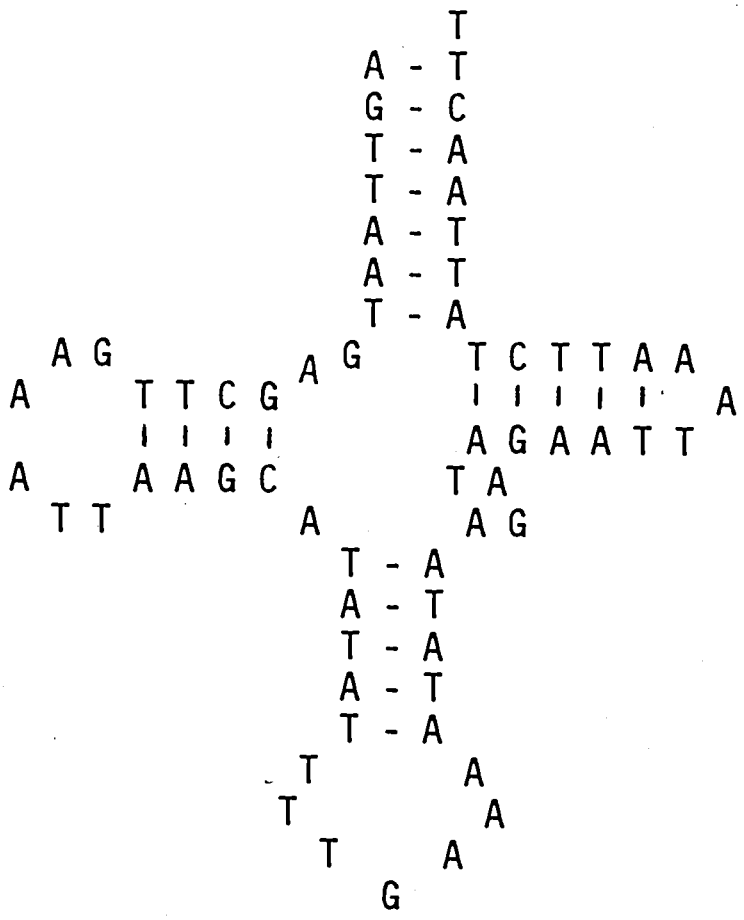

Figure 19. Mitochondrial $tRNA^{ser}_{UCN}$ structure of the flea.

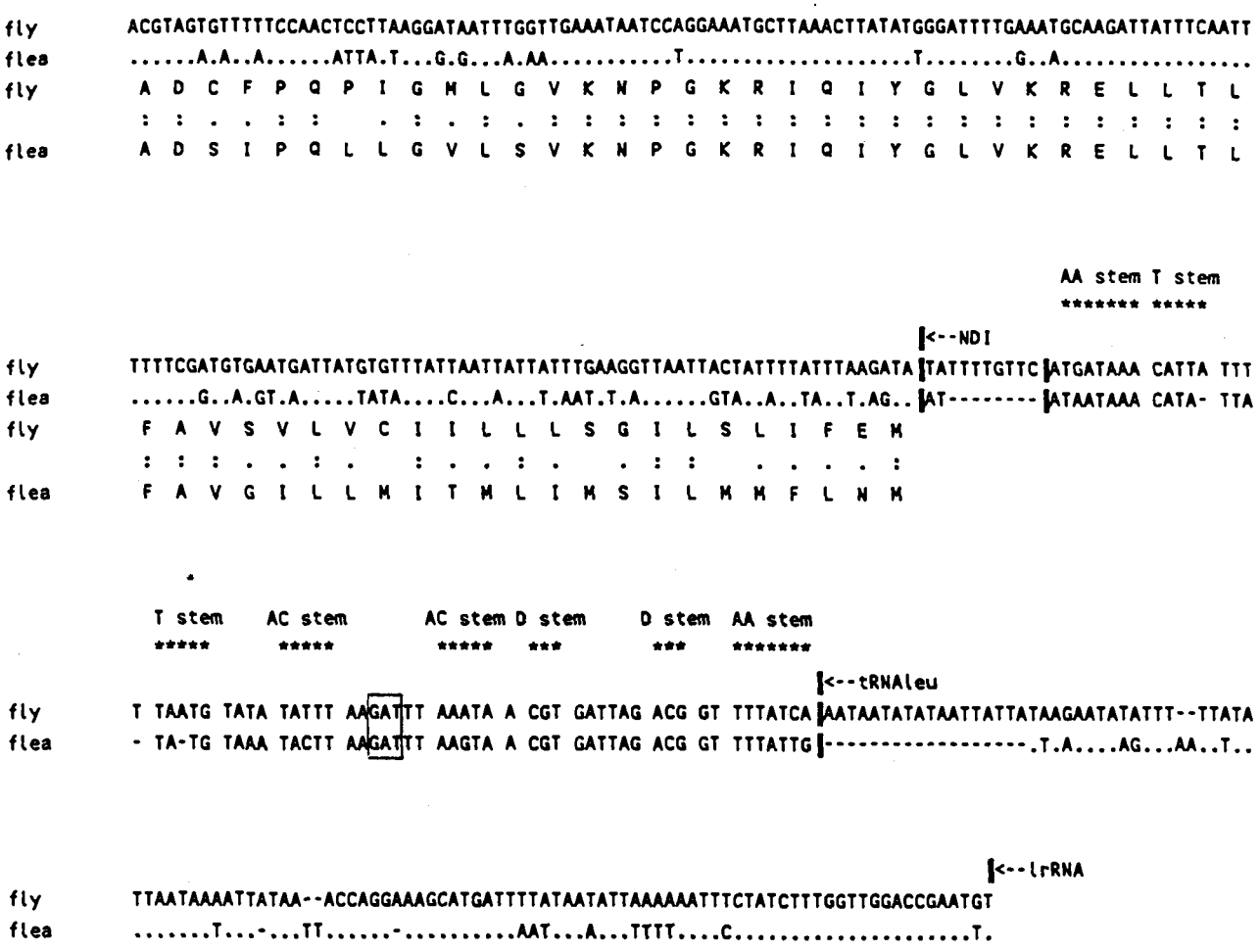

Figure 20. Sequences of portions of mitochondrial ND1 and lrRNA genes and the $tRNA^{leu}_{CUN}$ gene for the flea and the fly. Dashes were used to align the sequences. Solid bars indicate the beginnings and the ends of the genes. Open box indicates the anticodon of the tRNA gene. Nucleotide positions in stem regions were highlighted by asterisks.

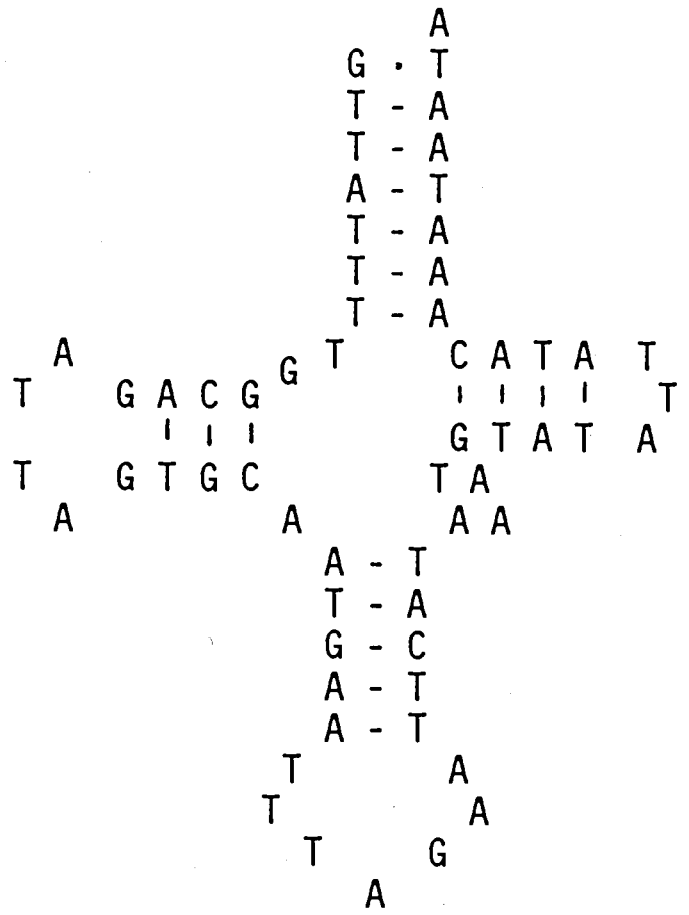

Figure 21. Mitochondrial $tRNA^{leu}_{CUN}$ structure of the flea.

ND2, tRNA^{Trp}, tRNA^{Cys}, tRNA^{Tyr} and COI genes

Primer F and G are based on the conservative portions at the 5' end of COI gene and 3' end of ND2 gene of *D. yakuba* respectively. They were used to amplify the sequences between the COI and ND2 genes. Figure 22 shows the aligned nucleotide sequences of some portions of COI genes as well as the tRNA^{Trp}, tRNA^{Cys} and tRNA^{Tyr} genes for the flea and the fly (see Chapter III for details). Figure 9, 10 and 11 show that the tRNA sequence can be folded into the configurations inferred for the corresponding tRNAs.

tRNA$^{leu}_{UUR}$, COII, tRNA^{lys} and tRNA^{asp} genes

Primer A and C are based on the conservative portions of the tRNA$^{leu}_{UUR}$ gene and ATPase 8 gene of *D. yakuba* respectively. They were used to amplify the COII gene as well as the two tRNA genes, tRNA^{lys} and tRNA^{asp} (see Chapter II for details).

Figure 23 shows the aligned nucleotide sequences of the 3' end of COII genes as well as the tRNA^{lys} and tRNA^{asp} genes for the flea and the fly. Figure 24 shows that part of this sequence can be folded into the configuration inferred for the tRNA^{lys}. The tRNA^{asp} gene sequence for the flea is not complete because primer C is too close to the tRNA^{asp} sequence. But that is enough to determine that it is the tRNA^{asp} gene for the flea.

ND3, tRNA^{Ala}, tRNA^{Arg}, tRNA^{Asn}, tRNA$^{Ser}_{AGN}$, tRNA^{Glu}, tRNA^{Phe} and ND5 genes

Primer D and E are based on the conserved portions at the 3' ends of *D. yakuba's* ND3 and ND5 genes respectively. They amplify the junction between the ND3 and ND5 genes. Figure 25 shows the aligned nucleotide sequences of some portions of ND5 gene as well as 6 tRNA genes. In the flea, there is no spacer between any tRNAs. In

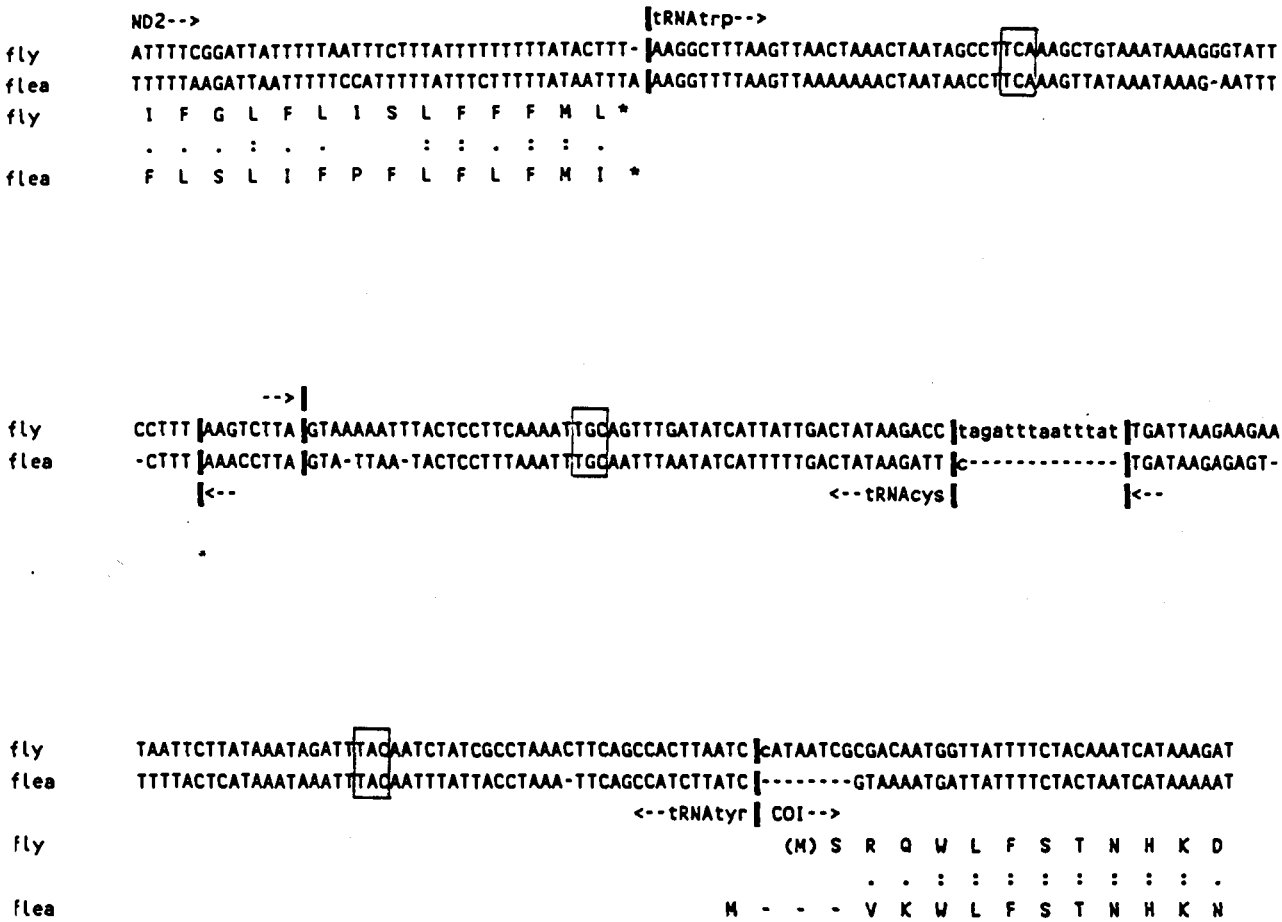

Figure 22. Sequences of portions of the mitochondrial ND2 and COI genes as well as the complete $tRNA^{trp}$, $tRNA^{cys}$ and $tRNA^{tyr}$ genes for the flea and the fly. Dashes were used to align the sequences. Solid bars indicate the beginnings and the ends of the genes. Open boxes indicate the anticodons for the tRNA genes. See Fig. 15 about the start codons of the COI gene.

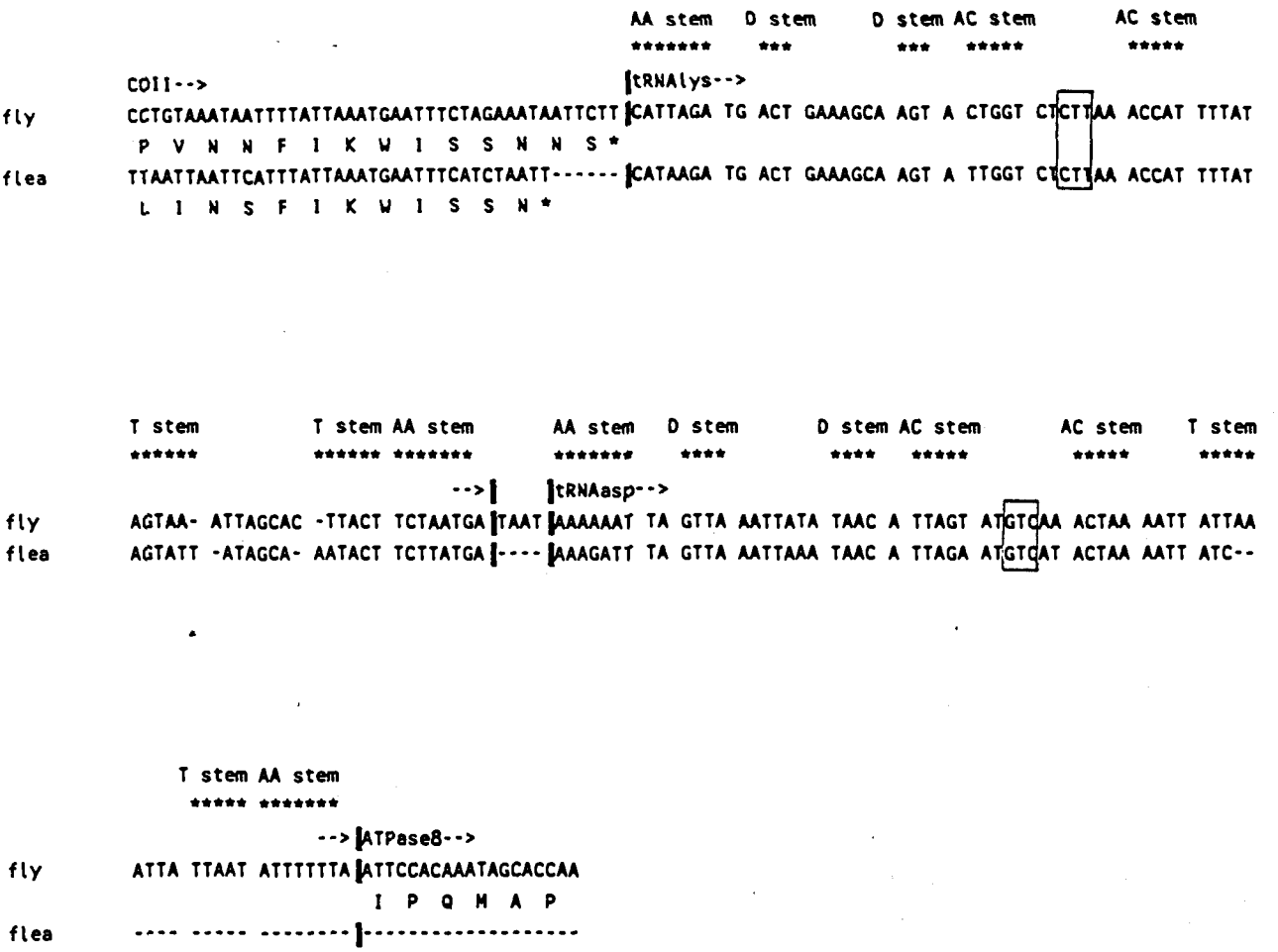

Figure 23. Sequences of portions of mitochondrial COII and $tRNA^{asp}$ genes as well as the $tRNA^{lys}$ gene for the flea and the fly. Dashes were used to align the sequences. Solid bars indicate the beginnings and the ends of the genes. Open box indicates the anticodon of the tRNA gene. Nucleotide positions in stem regions were highlighted by asterisks.

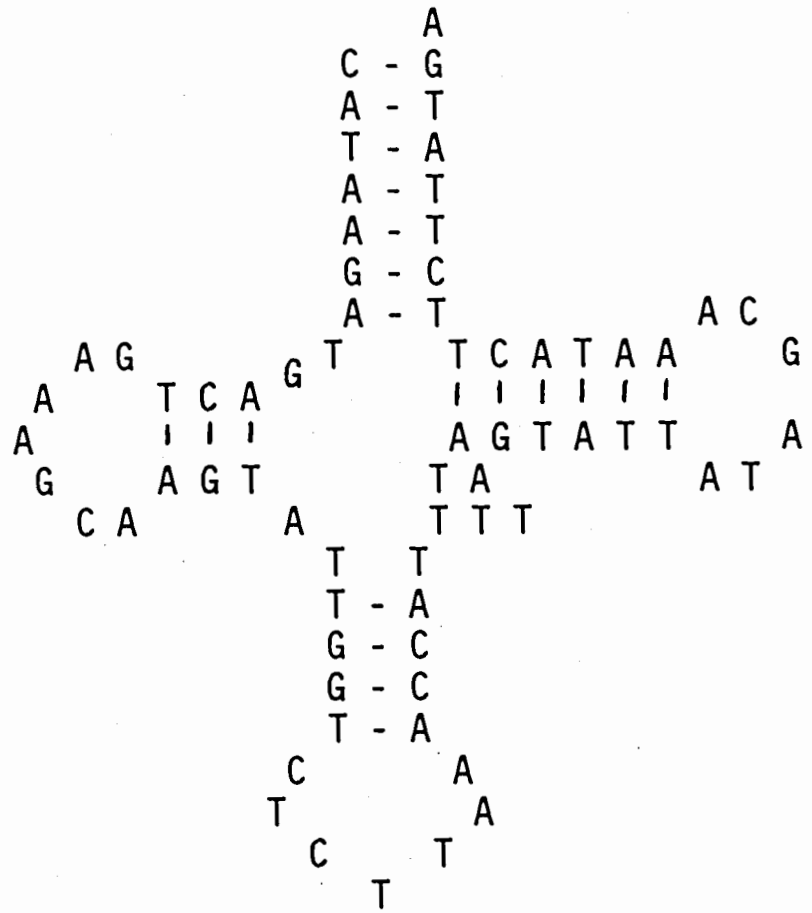

Figure 24. Mitochondrial $tRNA^{lys}$ structure of the flea.

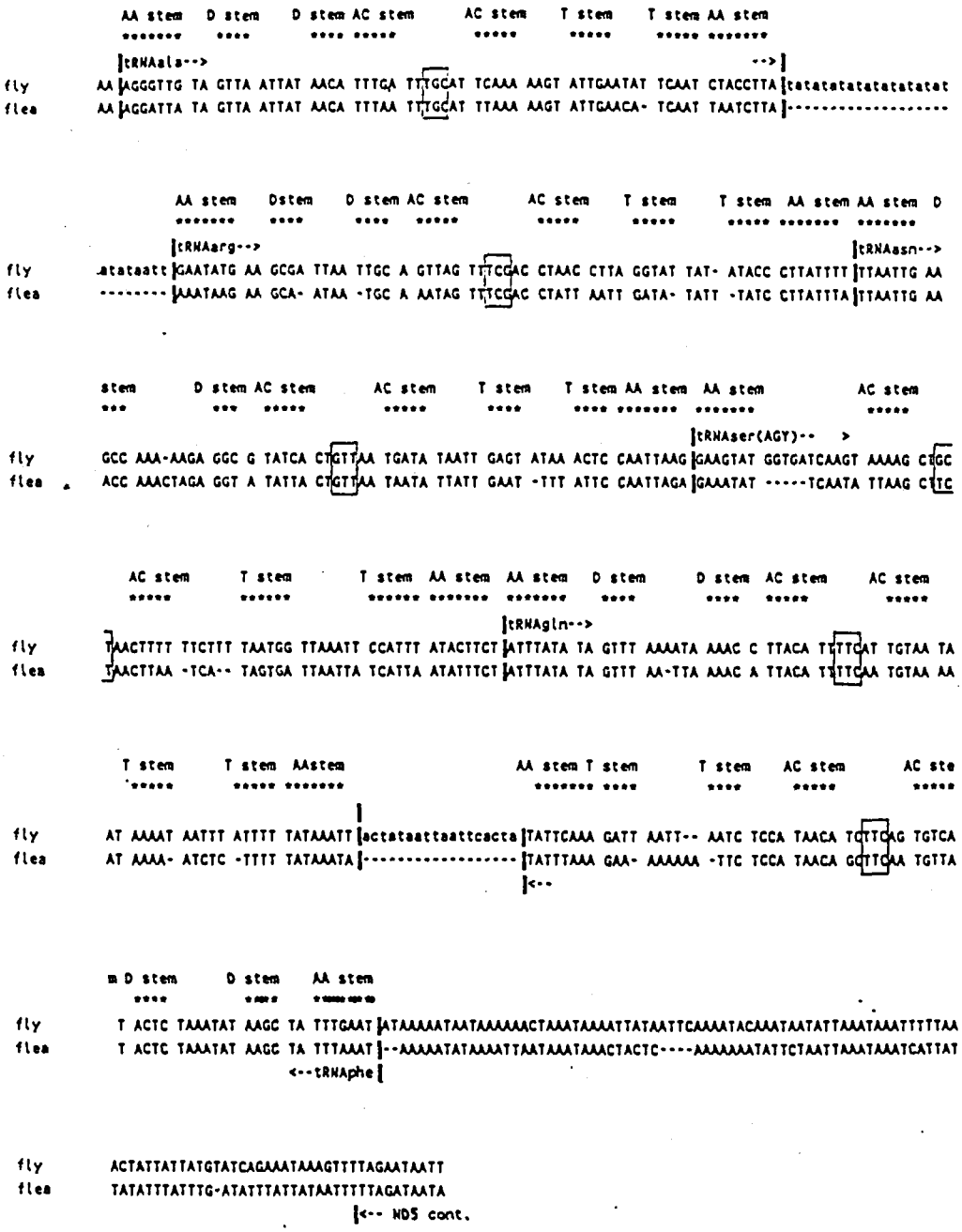

Figure 25. Sequences of portions of the mitochondrial ND5 genes as well as the complete $tRNA^{ala}$, $tRNA^{arg}$, $tRNA^{asn}$, $tRNA^{ser}_{AGY}$, $tRNA^{gln}$ and $tRNA^{phe}$ genes for the flea and the fly. Dashes were used to align the sequences. Solid bars indicate the beginnings and the ends of the genes. Open boxes indicate the anticodons for the tRNA genes. Nucleotide positions in stem regions were highlighted by asterisks.

D. yakuba, there is a 27 bp of AT repeats between the $tRNA^{ala}$ and the $tRNA^{arg}$ genes and a 18 bp spacer between the $tRNA^{gln}$ and the $tRNA^{phe}$ genes. Figure 26 shows that the tRNA sequences can be folded into the configurations inferred for the corresponding tRNAs: $tRNA^{Ala}$, $tRNA^{Arg}$, $tRNA^{Asn}$, $tRNA^{Ser}_{AGN}$, $tRNA^{Glu}$ and $tRNA^{Phe}$.

ND5, $tRNA^{His}$ and ND4 genes

Primer P and W are based on the conserved portions at the 3' ends of *D. yakuba*'s ND4 and ND5 genes. Figure 27 shows the aligned nucleotide sequences of some portions of ND5 and ND4 genes as well as the $tRNA^{His}$ gene. In the flea, the spacer is 6 bp long between the ND5 gene and the $tRNA^{his}$ gene. In *D. yakuba*, the spacer is 15 bp long. Figure 27 also shows the 3' end of the ND4 genes. Both the flea and the fly use single T as their ND4 stop codon and there is no spacer between the ND4 and the $tRNA^{his}$ genes. Figure 28 shows that the tRNA sequence can be folded into the configurations inferred for the corresponding tRNA: $tRNA^{His}$.

Conclusion

In Chapters II and III, I have examined the use of sequence data (COII and tRNA genes) for establishing the relationships among orders of insects. Mitochondrial gene organization provides the alternative route. Gene rearrangements represent unique events. They require multiple and rare events, for example, gene duplications followed by deletions. Parallel or reverse mutations are very unlikely for gene rearrangements.

In this Chapter, I presented an efficient way to determine the mitochondrial gene order. Mt-genome order is commonly determined from complete DNA sequence. That usually requires cloning, screening of clones and a lot of sequencing. I have developed

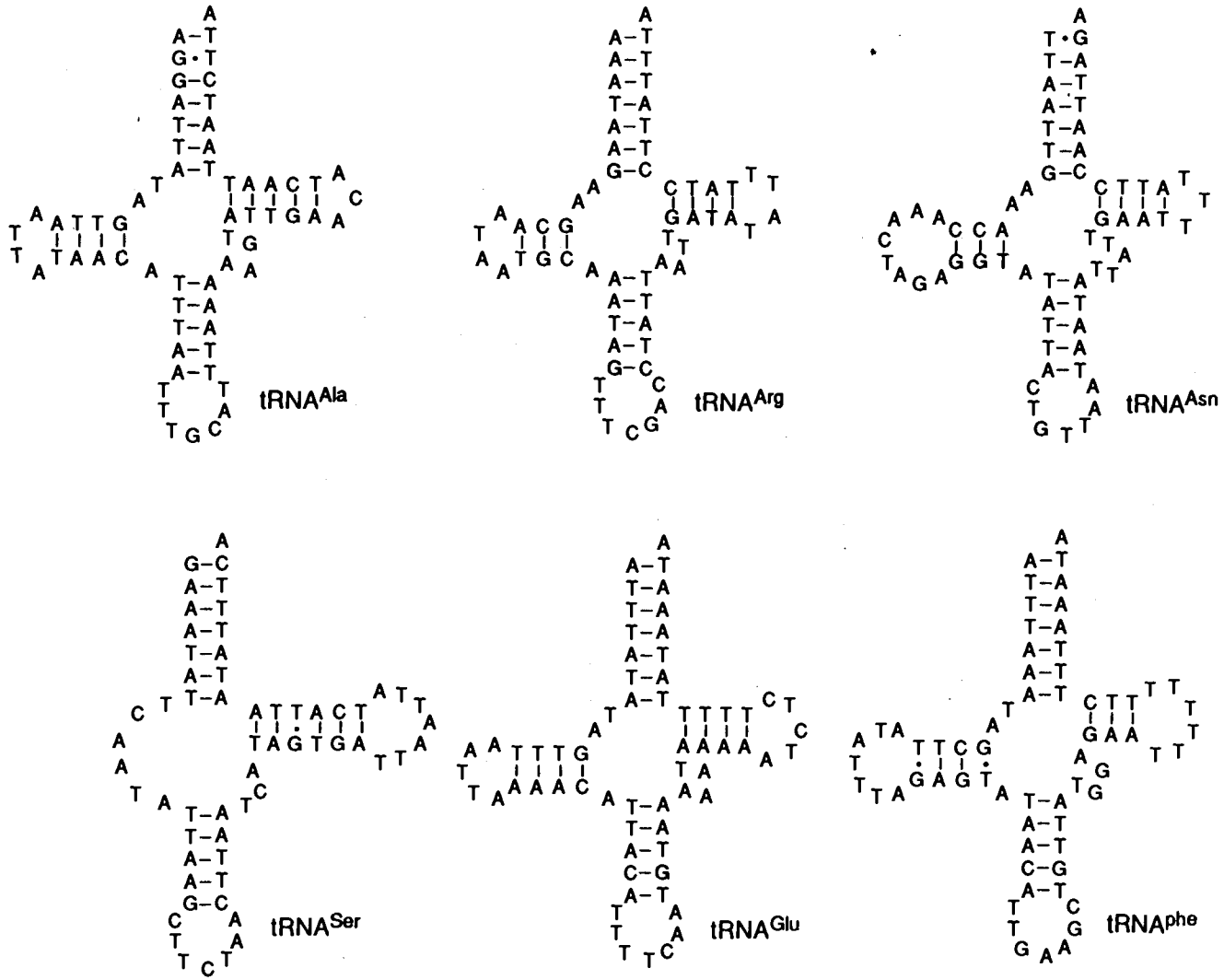

Figure 26. The 6 tRNA genes of flea mtDNA, folded into the configurations inferred for the resulting tRNAs.

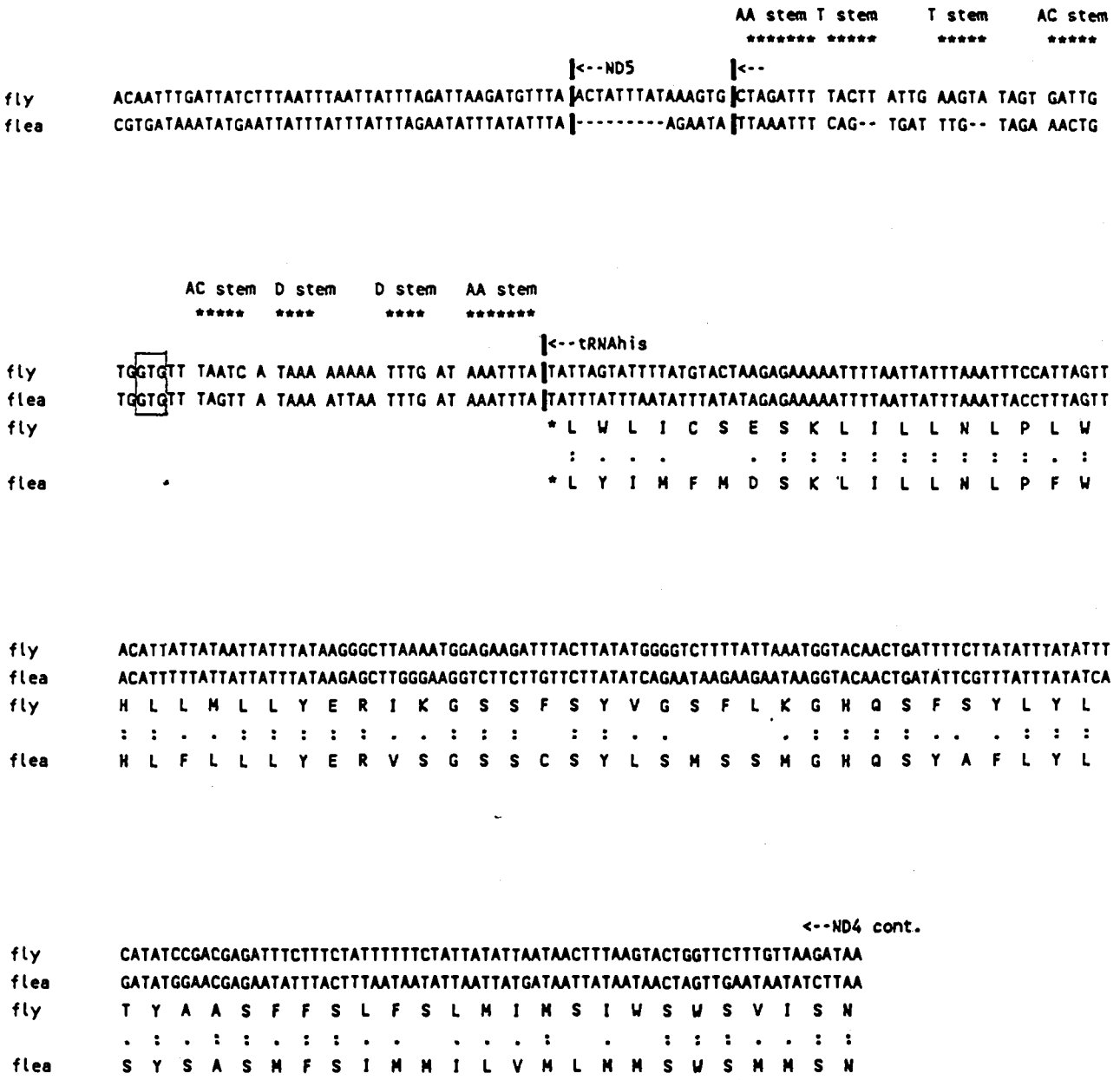

Figure 27. Sequences of portions of the mitochondrial ND4 and ND5 genes as well as the complete $tRNA^{his}$ genes for the flea and the fly. Dashes were used to align the sequences. Solid bars indicate the beginnings and the ends of the genes. Open box indicates the anticodon for the tRNA gene. Nucleotide positions in stem regions were highlighted by asterisks.

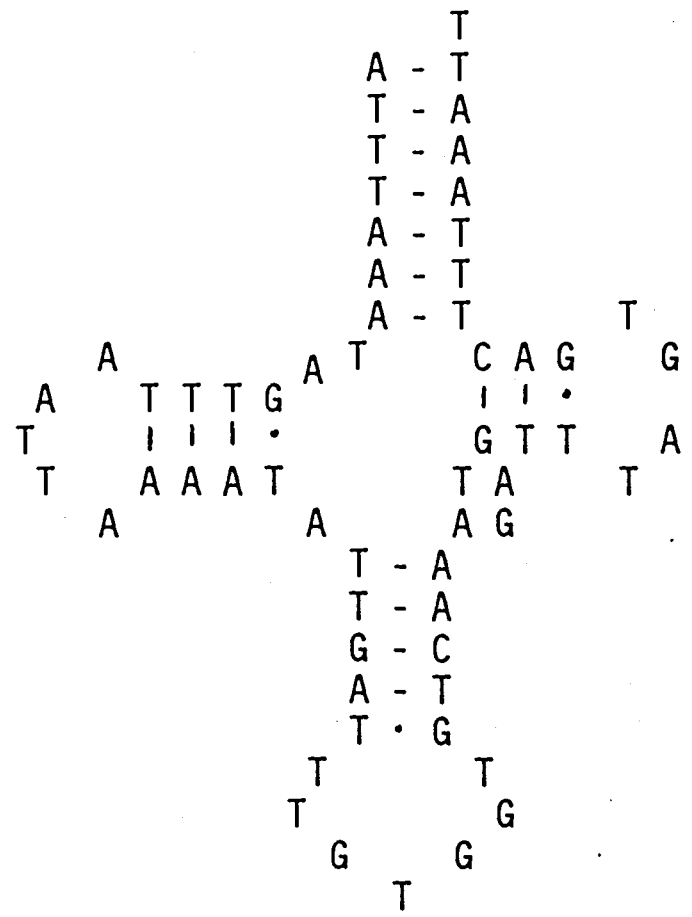

Figure 28. Mitochondrial $tRNA^{his}$ structure of the flea.

13 pairs of primers based on comparison of sequences of protein coding regions. By amplifying and sequencing gene junctions, only 22 tRNA genes plus minimal flanking sequences need to be determined in order to obtain the entire mtDNA genome organization.

CHAPTER V

SYNTHESIS OF RESULTS

In this thesis, I examined the evolution of mtDNA of insects at three different levels. The first is at protein level. I used the fairly conserved mitochondrial cytochrome oxidase II genes to examine the evolution among 10 orders of insects. The conclusion I can draw from this study is the divergence of insect orders is very ancient and may have occurred too rapidly for easy resolution using mitochondrial protein sequences. The second codon position data may hold more promise but still they are not conserved enough for this kind of deep phylogeny.

The second is at tRNA level. By amplifying the junction region between the ND2 and COI genes, I obtained sequence data for 3 tRNA genes from 5 orders of insects. The conclusion that can be drawn for resolving the phylogeny of these insects at the order level is basically the same as for the COII genes. But that could be partially because of the small data set. By looking at all 22 tRNA sequences, especially the stem regions, we may have a better resolution about the phylogeny of insects. Neither the COII nor the tRNA data analysis support the monophyly of the Holometabola which has been assumed based on morphological data.

Mitochondrial gene organization data may hold more promise for resolving relationships at this level simply because gene order changes at much slower rate. Only a few groups have been investigated and data from many more are needed for even a preliminary assessment. For about 30 orders of insects, only two complete mt-genome orders have been published. These are *Drosophila yakuba* (order Diptera) and *Apis*

mellifera (order Hymenoptera). We need more information on insect gene organization to make use of it. I think the technique I used with the flea provides an efficient way of getting this information. By amplifying and sequencing gene junctions, only 22 tRNA genes need to be determined in order to obtain the entire mtDNA genome organization.

REFERENCES

Anderson, S., M.H.L. de Bruijn, A.R. Coulson, I.C. Eperon, F. Sanger and I.G. Young. (1982). Complete sequence of bovine mitochondrial DNA. *J. Mol. Biol.* **156**: 683-717.

Anderson, S., A.T. Bankier, B.G. Barrell, M.H.L. de Bruijn, A.R. Coulson, J. Drouin, I.C. Eperon, 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 mitochondrial genome. *Nature* Vol. **290**: 457-465.

Arnason, U., A. Gullberg and B. Widegren (1991) The complete nucleotide sequence of the mitochondrial DNA of the fin whale, *Balaenoptera physalus*. *J. Mol. Evol.* **34**:493-505.

Asakawa, S., Y. Kumazawa, T. Araki, H. Himeno, K. Miura and K. Watanabe (1991) Strand-specific nucleotide composition bias in echinoderm and vertebrate mitochondrial genomes. *J. Mol. Evol.* **32**:511-520.

Beckenbach, A.T., Y.W. Wei, and H. Liu (1993) Evolution of the mitochondrial cytochrome oxidase ii gene in the *Drosophila obscura* species group. *Mol. Biol. Evol.* **10**(3):619-634.

Bibb M.j., R.A. Van Etten, C.T. Wright, M.W. Walberg, and D.A. Clayton (1981) Sequence and gene organization of mouse mitochondrial DNA. *Cell* **26**:167-180.

Boudreaux, H.B. (1979) *Arthropod phylogeny with special reference to insects*. New York: John Wiley and Sons.

Brown, J.R., A.T. Beckenbach and M.J. Smith (1992) Mitochondrial DNA length variation and heteroplasmy in populations of white sturgeon (*Acipenser*

transmontanus). *Genetics* **132**:221-228.

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

Brown, W.M., E.M. Prager, A. Wang and A.C. Wilson. (1982) Mitochondrial DNA sequences of primates: Tempo and mode of evolution. *J. Mol. Evol.* **18**: 225-239.

Cabot, E.L. and A.T. Beckenbach. (1989) Simultaneous editing of multiple nucleic acid and protein sequences with ESEE. *Computer Applications in Biosciences* **5**: 233-234.

Cantatore P., M.N. Gadaleta, M. Roberti, C. Saccone and A.C. Wilson (1987) Duplication and remoulding of tRNA genes during the evolutionary rearrangement of mitochondrial genome. *Nature* **329**:853-855.

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

Carpenter, F.M. and L. Burnham (1985) The geological record of insects. *Ann. Rev. Earth Planet. Sci.* **13**: 297-314.

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

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.

Crozier, R.H. and Crozier, Y.C. (1993) The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. *Genetics* **133**:97-117.

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

Dams E, L. Hendriks, Van de Peer Y, J-M Neefs, G. Smits, I. Vandenbempt, and R. DeWachter (1988) Compilation of small ribosomal subunit RNA sequences. *Nucl. Acids Res.* **16** suppl:r87-r173.

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

De Giorgi, C., C. Lanave, M.D. Musci and C. Saccone (1991) Mitochondrial DNA in the sea urchin *Arbacia lixula*: evolutionary inferences from nucleotide sequence analysis. *Mol. Biol. Evol.* **8**:515-529.

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.

Desjardins, 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.

Desjardins, P., and R. Morais (1991) Nucleotide sequence and evolution of coding and noncoding regions of a quail mitochondrial genome. *J. Mol. Evol.* **32**:153-161.

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

Edwards, S.V., P. Arctander and A.C. Wilson (1991) Mitochondrial resolution of a deep branch in the genealogical tree for perching birds. *Proc. R. Soc. Lond.* **B 243**: 99-107.

Felsenstein, J. (1988) Phylogenies from molecular sequences: inference and reliability. *Annu. Rev. Genet.* **22**:521-565.

Felsenstein, J. (1989) PHYLIP -- Phylogeny inference package (version 3.2). *Cladistics* **5**: 164-166.

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

Foran, D.R., J.E. Hixson and W.M. Brown (1988) Comparisons of ape and human sequences that regulate mitochondrial DNA transcription and D-loop DNA synthesis. *Nucleic Acids Res.* **16**:5841-5861.

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

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

Garey, J.R., and D.R. Wolstenholme (1989) Platyhelminth mitochondrial DNA: evidence for early evolutionary origin of a tRNA^{ser}AGN that contains a dihydrouridine arm replacement loop, and of serine-specifying AGA and AGG codons *J. Mol. Evol.* **28**:374-387.

Gyllensten U.B. and H.A. Erlich (1988) Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus. *Proc. Natl. Acad. Sci. U.S.A.* **85**:7652-7656.

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

Hayasaka, K., T. Ishida and S. Horai (1991) Heteroplasmy and polymorphism in the major noncoding region of mitochondrial DNA in Japanese Monkeys: association with tandemly repeated sequences. *Mol. Biol. Evol.* **8**:399-415.

Hennig, W. (1981) Insect Phylogeny. John Wiley and Sons.

Hillis, D.M. and S.K. Davis (1986) Evolution of ribosomal DNA: Fifty million years of recorded history in the frog genus *Rana*. *Evolution* **40**:1275-1288.

Hillis, D.M. and C. Moritz (1990) Molecular systematics: context and controversies. In *Molecular Systematics* (ed. D.M. Hillis and C. Moritz) pp. 1-10, Sinauer publishers.

Himeno, H., H. Masaki, T. Kawai, T. Ohta, I. Kumagai, K. Miura and K. Watanabe (1987) Unusual genetic codes and a novel gene structure for tRNA$^{ser}_{AGY}$ in starfish mitochondrial DNA. *Gene* **56**:219-230.

Hixson J.E. and W.M. Brown (1986) A comparison of the small ribosomal RNA genes from the mitochondrial DNA of the great apes and humans: sequence, structure,

evolution, and phylogenetic implications. *Mol. Biol. Evol.* **3**:1-18.

Hoeh, W.R., K.H. Blakley and W.M. Brown (1991) Heteroplasmy suggests limited biparental inheritance of *Mytilus* mitochondrial DNA. *Science* **251**:1488-1490.

Hoffmann R.J., J.L. Boore, and W.M. Brown (1992) A novel mitochondrial genome organization for the blue mussel, *Mytilus edulis*. Genetics **131**:397-412.

Holmquist, R. (1983) Transition and transversions in evolutionary descent: an approach to understanding. *J. Mol. Evol.* **19**:134-144.

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

Innis, M.A., D.H. Gelfand, J.J. Sninsky and T.J. White (1990) *PCR Protocols: A Guide to Methods and Applications*. Academic Press, San Diego.

Irwin, D.M., T.D. Kocher, and A.C. Wilson. (1991) Evolution of the cytochrome B gene of mammals. *J. Mol. Evol.* **32**:128-144.

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

Johansen, J., P.H. Guddal and T. Johansen. (1990) Organization of the mitochondrial genome of Atlantic cod, *Gadus morhua*. *Nucleic Acids Res.* **18**:411-419.

Jukes, T.H. and C.R. Cantor. (1969) Evolution of protein molecules. Pp. 21-126 *in* H.N. Munro, ed. Mammalian protein metabolism. Vol. 3. Academic Press, New York.

Kimura, M. (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.*

16:111-120.

Kocher T.D., W.K. Thomas, A. Meyer, S.V. Edwards, S. Paabo, F.X. Yillablanca, and A.C. Wilson (1989) Dynamics of mitochondrial DNA evolution in animals: Amplification and sequencing with conserved primers. *Proc. Natnl. Acad. Sci. USA* **86**:6196-6200.

Kondo, R., Y. Satta, E.T. Matsuura, H. Ishiwa, N. Takahato and S.I. Chigusa (1990) Incomplete maternal transmission of mitochondrial DNA in Drosophila. *Genetics* **126**:657-663.

Kristensen, N.P. (1981) Phylogeny of insect orders. *Ann. Rev. Entomol.* **26**: 135-157.

Kristensen, N.P. (1989) Insect phylogeny based on morphological evidence. In *Hierarchy of life* (ed. B. Fernholm et al) pp. 295-306, Elsevier science publishers.

Kumazawa, Y. and M. Nishida (1993) Sequence evolution of mitochondrial tRNA genes and deep-branch animal phylogenetics. *J. Mol. Evol.* (in press).

LaRoche, J., M. Snyder, D.I. Cook, K. Fuller and E. Zouros (1990) Molecular characterization of a repeat element causing large-scale size variation in the mitochondrial DNA of the sea scallop *Placopecten magellanicus*. *Mol. Biol. Evol.* **7**:45-64.

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

Liu, H. and A.T. Beckenbach (1992) Evolution of the mitochondrial cytochrome oxidase II gene among 10 orders of insects. *Mol. Phylogenet. Evol.* **1**:41-52.

Manton, S.M. (1979) Functional morphology and evolution of the hexapod classes in Arthropod phylogeny. (ed. A.P. Gupta) pp. 387-465. New York: Van Nostrand-Reinhold.

Matsuda, R. (1976) *Morphology and evolution of the insect abdomen.* Pergamon, New York.

McCracken, A., I. Uhlenbusch and G. Gellissen (1987) Structure of the cloned *Locust migratoria* mitochondrial genome: restriction mapping and sequences of its ND1 gene. *Curr. Genet.* **11**:631-638.

Millett, F., de Jong, C, Paulson, L. and A. Capaldi. (1983) Identification of specific carboxylate groups on cytochrome *c* oxidase that are involved in binding cytochrome *c*. *Biochemistry* **22**: 546-552.

Miyamoto, M.M. and J. Cracraft (1991) *Phylogenetic analysis of DNA sequences.* (Oxford University Press).

Montoya J, G.L. Gaines, and G. Attardi (1983) The pattern of transcription of the human mitochondrial rRNA genes reveals two overlapping transcription units. *Cell* **34**:151-159.

Moritz C, T.E. Dowling, and W.M. Brown (1987) Evolution of animal mitochondrial DNA: Relevance for population biology and systematics. *Ann. Rev. Ecol. Syst.* **18**:269-92.

Noller, H.F. and C.R. Woese (1981) Secondary structure of 16S ribosomal RNA. *Science* **212**:403-411.

Okimoto, R., H. Chamberlin, J. Macfarlane and D. Wolstenholme (1990) Repeated sequence sets in mitochondrial DNA molecules of root-knot nematodes

(*Meloidogyne*): nucleotide sequences, genome location and potential for host-race

identification. *Nucleic Acids Res* **19**:1619-1626.

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**:471-497.

Paabo, S., W.K. Thomas, K.M. Whitefield, Y. Kumazawa and A.C. Wilson. (1991) Rearrangements of mitochondrial transfer RNA genes in marsupials. *J. Mol. Evol.* **33**:426-430.

Pearson, W.R. (1990) Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods in Enzymology* Vol. 183, pp 63-97. Academic Press.

Raff, R.A., K.G. Field, G.J. Olsen, S.J. Giovannoni, D.J. Lane, M.T. Ghiselin, N.R. Pace and E.C. Raff (1989) *The Hierarchy of Life* pp 247-260 (Elsevier Science Publishers B.V.).

Rand, D.M., and R.G. Harrison (1989) Molecular population genetics of mtDNA size variation in crickets. *Genetics* **121**:551-569.

Reddy, P., A. Peterkofsky and K. McKenney (1985) Translationl efficiency of the *Escherichia coli* adenylate cyclase gene: mutating the UUG initiation codon to GUG or AUG results in increased gene expression. *Proc. Natl. Acad. Sci. U.S.A.* **82**:5656-5660.

Roe, B.A., D. Ma, R.K. Wilson and J.H. Wong. (1985). The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome. *J. Biol. Chem.* Vol. 260 No. **17**: 9759-9774.

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

Sankoff, D., R. Cedergren and Y. Abel. (1990) Genomic divergence through gene rearrangement. *Methods Enzymol.* **183**:426-438.

Simon C, S. Paabo, T. Kocher, and A.C. Wilson (1990) Evolution of the mitochondrial ribosomal RNA in insects as shown by the polymerase chain reaction. In Clegg M (ed). Molecular Evolution. *UCLA Sym. Mol. Cellu. Bio.* Vol.**122** p235.

Smith, M.J., D.K. Banfield, K. Doteval, S. Gorski and D.J. Kowbel (1989) Gene arrangement in sea star mitochondrial DNA demonstrates a major inversion event during echinoderm evolution. *Gene* **76**:181-185.

Smith, M.J., D.K. Banfield, K. Doteval, S. Gorski and D.J. Kowbel (1990) Nucleotide sequence of nine protein-coding genes and 22 tRNAs in the mitochondrial DNA of the sea star *Pisaster ochraceus*. *J. Mol. Evol.* **31**:195-204.

Smith, M.J., A. Arndt, S. Gorski and E. Fajber (1993) The phylogeny of echinoderm classes based on mitochondrial gene arrangements. *J. Mol. Evol.* **36**:545-554.

Sommer, R. and D. Tautz (1989) Minimal homology requirements for PCR primers. *Nucleic Acids Res* **17**:6749.

Southern, S.O., P.J. Southern and A.E. Dizon (1988) Molecular characterization of a cloned Dolphin mitochondrial genome. *J. Mol. Evol.* **28**:32-42.

Thomas, W.K. and A.T. Beckenbach. (1989) Variation in salmonid mitochondrial DNA: Evolutionary constraints and mechanisms of substitution. *J. Mol. Evol.* **29**: 233-245.

Thomas W.K., S. Paabo, F.X. Villablanca, and A.C. Wilson (1990) Spatial and temporal continuity of kangaroo rat populations shown by sequencing

mitochondrial DNA from museum specimens. *J. Mol. Evol.* **31**:101-112.

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

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

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

Warrior, R., and J. Gall (1985) The mitochondrial DNA of *Hydra attenuate* and *Hydra littoralis* consists of two linear molecules. *Arch. Sci. Geneva* **38**:439-445.

Wheeler, W.C. (1990). Nucleic acid sequence phylogeny and random outgroups. *Cladistics* **6**: 363-367.

Wheeler W, and R. Honeycut (1988) Paired sequence difference in ribosomal RNAs: Evolutionary and phylogenetic implications. *Mol. Biol. Evol.* **5**:90-96.

Wilson A.C., R.L. Cann, S.M. Carr, M. George, U.B. Gyllensten, K.M. Helm-Bychowski, R.G. Higuchi, S.R. Palumbi, E.M. Prager, R.D. Sage, and M. Stoneking (1985). Mitochondrial DNA and two perspectives on evolutionary genetics. *Biol. J. Linn. Soc.* **26**:375-400.

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, J.A. Wahleithner (1987) Bizarre tRNAs inferred from DNA sequences of mitochondrial genomes of

nematode worms. *Proc. Natl. Acad. Sci. USA* **84**:1324-1328.

Wooten, R.J. (1981) Paleozoic insects. *Annu. Rev. Entomol.* **26**:319-344.