THE EVOLUTION OF MITOCHONDRIAL GENOME STRUCTURE AND FUNCTION IN INSECTS

by
James Bruce Stewart
B.Sc.H, University of Waterloo, Waterloo, Ontario, 1997
M.Sc. Simon Fraser University, Burnaby, British Columbia, 2001

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in the
Department of Molecular Biology and Biochemistry

© James Bruce Stewart, 2005
SIMON FRASER UNIVERSITY
Summer 2005

All rights reserved. This work may not be
reproduced in whole or in part, by photocopy
or other means, without permission of the author.

Approval

Name: **James Bruce Stewart**
Degree: Doctor of Philosophy

Title of thesis:

The Evolution of Mitochondrial Genome Structure and Function in Insects

Examining Committee:
Chair: **Dr. Nicholas Harden**, Associate Professor, Department of Molecular Biology and Biochemistry

---

Dr. Andrew T. Beckenbach, Professor
Senior Supervisor
Department of Biological Sciences

Dr. David L. Baillie, Professor
Supervisory Committee Member
Department of Molecular Biology and Biochemistry

Dr. Peter J. Unrau, Assistant Professor
Supervisory Committee Member
Department of Molecular Biology and Biochemistry

Dr. Esther M. Verheyen, Associate Professor
Public Examiner
Department of Molecular Biology and Biochemstry

Dr. Laurie S. Kaguni, Professor
External Examiner
Department of Biochemistry and Molecular Biology
Michigan State University

Date Approved: August 19, 2005

SIMON FRASER UNIVERSITY

PARTIAL COPYRIGHT LICENCE

The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the right to lend this thesis, project or extended essay 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.

The author has further granted permission to Simon Fraser University to keep or make a digital copy for use in its circulating collection.

The author has further agreed that permission for multiple copying of this work for scholarly purposes may be granted by either the author or the Dean of Graduate Studies.

It is understood that copying or publication of this work for financial gain shall not be allowed without the author's written permission.

Permission for public performance, or limited permission for private scholarly use, of any multimedia materials forming part of this work, may have been granted by the author. This information may be found on the separately catalogued multimedia material and in the signed Partial Copyright Licence.

The original Partial Copyright Licence attesting to these terms, and signed by this author, may be found in the original bound copy of this work, retained in the Simon Fraser University Archive.

W. A. C. Bennett Library
Simon Fraser University
Burnaby, BC, Canada

Abstract

The mitochondrial genomes (mt-genomes) of animals are very compact in structure, encoding thirteen protein genes involved in the production of ATP, and the key components of the translation system to express these proteins. The mitochondrial expression system, which functions separately from that of the nucleus, shows characteristics of both prokaryotic and eukaryotic expression systems, and has diverged greatly from that currently observed in the closest living relatives of mitochondria, the α-proteobacteria. Current understanding of transcript maturation is that large multi-gene transcripts are processed by the removal of intervening tRNA genes, leaving behind RNA templates to be matured into the functional mRNAs and rRNAs. One of the most striking features of insect mt-genomes has been the apparent replacement of a start codon with a stop codon for the essential mitochondrial gene *cytochrome c oxidase subunit 1* (*cox1*). When first observed in *Drosophila*, Clary and Wolstenholme proposed a highly unusual four-base "ATAA" start codon. With the expanded sampling of mitochondrial sequence across the various insect orders, the data does not support the use of this aberrant initiation for *cox1.* At the initiation of this study, the diversity of insect groups represented by complete mt-genome sequence was very poor. To address this deficit, I undertook sequencing projects to increase the number of insect orders represented in the mitochondrial sequence databases. I report the complete mt-genome sequences for two insects, the spittlebug *Philaenus spumarius*, and the giant stonefly *Pteronarcys princeps.* The sequences are annotated and compared to other insect mt-genomes in the sequence databases. I report the cDNA sequences of *Drosophila melanogaster* mitochondrial mRNAs, rRNA subunits, and a population of pre-mRNA molecules that are intermediates of the RNA processing system. Models to explain mitochondrial transcript maturation in light of these new observations are proposed. Comparative analyses were undertaken to apply the information gained from the mitochondrial transcripts of *D. melanogaster* to the mitochondrial structure and annotations of mt-genomes from the other insects. These analyses suggest a 5' specific modification to the tRNA punctuation model for insect mitochondria. This modification may represent a further evolutionary simplification of the mitochondrial expression system.

To Laila, for everything,

And to You who we patiently await.

In memory:

Grandpa Bruce Stewart
December 6, 1922 - February 27, 2001

and

Great-Grandpa Ted Baxter
October 20, 1897 – May 5, 1990

Acknowledgements

To my senior supervisor Andy Beckenbach, for allowing me the chance to explore this project and believing that I could get it done. The last few years have been a lot of fun, and I will miss our lab chats.

A special thanks to my former lab-mate Russ Watkins for keeping in touch after moving on, and for supplying a key tip for my early RNA struggles. To my lab-mates Andrea Scouras, Dave Russell, Alison Cloutier, Karen Beckenbach, Denise Reichow, Cam Muir and Mike Vankoeveringe. Thanks for all the tips, advice and silliness.

Thanks to the RNA-work troubleshooters Alex Ebhardt and Don Sinclair, and to the computer guys Jeff Bryer and Duncan Napier. To Jillian Smith, Darrell Bessette and Theresa Kitos for the help with cell culture.

To my committee members Peter Unrau and Dave Baillie for their time, effort and advice over the course of my degree work. I would also like to thank Mike Smith for his role as unofficial committee member. Thank-you to Jamie Cates for the very important discussion of translation at a key time.

To Laurie Kaguni and Esther Verheyen, my external examining committee members, for the useful, constructive and kind comments regarding my thesis.

To my fellow graduate students in Biology and the MBB for all the fun and for the shop talk. I especially want to thank John Taylor, Ian Hamilton, Jeremy Mitchell, Tom Chapman, Lynnette Kuervers, Nigel O'Neil, Greg Vatcher, Steve Springer, Michelle Harrison, Susanne Kühnholz, Richard Fahlman, Anat Feldman, Carrie Simms and Daniela Ginta for important discussions over the course of my study. My personal thanks to the Biergarten Entrepreneurs, past and present. Where else would the discussions take place? To Carlo Artieri, Leslie Mitchell and Connie Roth for seminar practice sessions.

To our Vancouver friends; Annick, Kevin and Giselle; Cam; Cathy and Rasmus; David; Kym, Dana and Tayme; Kat and Greg; Tina and Olie; Jennifer and Stevo; Markus, Lisa and Anna; Cris and Diana; Anna, Jurgen, Max and Sophie; Nat, Dirk, Steffen and Gabby; Leah, Omar and Amira; Melissa. From Ontario, to Pete, Tina, Mason and Rowan; Jilly, Bomina, Susan and Carole; Scott, Kurt and Karl and Christine for continued support and friendship.

Thanks to Cathy, Don, Mike, Kelly Joel, Ardis, Fred, and to Norm, Joe, Vicky, Rob, Tracy, Mary and all the rest of my family, for your patience, support, and well wishes over the years. Thank-you to Agneta and Zubin for welcoming me.

Table of Contents

Approval ii
Abstract iii
Dedication iv
Acknowledgements v
Table of Contents vi
List of Tables ix
List of Figures x

CHAPTER 1. **Introduction** 1

1.1. Mitochondria 1

1.2. Mitochondrial Evolution 2

1.3. Animal Mitochondrial Genomes 5
1.3.1. Animal mitochondrial gene content 5
1.3.2. Animal mitochondrial genome structure 9
1.3.3. Mitochondrial genome sequencing 10

1.4. Mitochondrial Transcription And Replication 14
1.4.1. Mitochondrial replication in animals 14
1.4.2. Variation in transcriptional systems between eukaryotic groups 17
1.4.3. Mammalian mitochondrial transcription 18
1.4.4. Transcript processing in mammalian mitochondria 19
1.4.5. Transcription in *Drosophila* mitochondria 20
1.4.6. RNA editing and programmed frameshifting in animal mitochondria 23

1.5. Thesis Outline 24

CHAPTER 2. The Complete Mitochondrial Genome Sequence of the Meadow Spittlebug *Philaenus spumarius* (Hemiptera: Auchenorrhyncha: Cercopoidae) 26

2.1. Introduction 26

2.2. Materials and Methods 28
2.2.1. DNA extraction and PCR amplification 28
2.2.2. Sequence assembly, annotation and analysis 32

2.3. Results and Discussion 33
2.3.1. Genome content 33
2.3.2. Protein coding genes 39
2.3.3. Genomic comparisons 43

2.4. Conclusions 44

CHAPTER 3. The Complete Mitochondrial Genome Sequence of a Giant Stonefly, *Pteronarcys princeps* (Plecoptera; Pteronarcyidae)..........47

3.1. Introduction..........47

3.2. Materials and Methods..........51
3.2.1. Specimen collection and identification..........51
3.2.2. DNA isolation..........51
3.2.3. PCR and Long-PCR amplification..........52
3.2.4. Purification of PCR products..........54
3.2.5. DNA sequencing..........56
3.2.6. Sequence assembly, annotation and analysis..........56
3.2.7. Sequence comparisons..........57

3.3. Results and Discussion..........61
3.3.1. Genome annotation..........61
3.3.2. Genome architecture..........65
3.3.3. Nucleotide composition..........65
3.3.4. Stonefly A+T rich regions..........74

3.4 Conclusions..........80

CHAPTER 4. Characterization Of Mitochondrial Transcripts In *Drosophila melanogaster*..........82

4.1. Introduction..........82
4.1.1. The *cox1* initiation mystery in insect mitochondria..........82
4.1.2. Mitochondrial transcription and RNA processing in *Drosophila*..........84
4.1.3. 5' amplification and sequencing..........86

4.2. Materials and methods..........89
4.2.1. Blast searches for nuclear copies of the *cox1* gene in *Drosophila*..........89
4.2.2. Maintenance of *Drosophila* fly stocks..........89
4.2.3. Maintenence of *Drosophila* Schneider-2 cell lines..........90
4.2.4. Mitochondrial isolation protocols..........91
4.2.5. RNA extraction protocol..........93
4.2.6. Reverse-transcription pcr protocols..........96
4.2.7. Protocol for 5'-RACE amplification..........100
4.2.8. Protocol for 3'-RACE amplicication..........101
4.2.9. RNA circularization for RT-PCR..........101
4.2.10. Cloning of amplification products..........102
4.2.11. Sequencing of cDNAs and sequence alignment..........104

4.3. Results..........104
4.3.1. Search for a nuclear copy of *cox1*..........104
4.3.2. RNA extraction from mitochondria..........106
4.3.3. Detection of unprocessed pre-mRNA molecules in flies..........107
4.3.4. Mapping of mRNA and rRNA boundaries by RACE..........107
4.3.5. Detection of RACE amplified RNAs from Schneider-2 cells..........111
4.3.6. Characterization of RNAs by CR-RT-PCR..........111

4.4. Discussion **117**
4.4.1. Schnieder-2 cells versus adult fly data 117
4.4.2. Detection of pre-mRNA molecules 118
4.4.3. Description of mitochondrial mRNAs 119
4.4.4. Intitiation of the *cox1* gene 127
4.4.5. Characterization of the ribosomal RNAs 129
4.4.6. Polyadenylated pre-mRNAs 130

CHAPTER 5. Conclusions And Discussion **136**

5.1. Transcript Characterization in *Drosophila* Mitochondria **136**
5.1.1. Characterization of mt-rRNAs 136
5.1.2. Characterization of mt-mRNAs 138
5.1.3. Processing of tRNAs versus mRNAs in insect mitochondria 141

5.2. Insect Mitochondrial Genomics **143**
5.2.1. Two insect mitochondrial genomes 143
5.2.2. Annotation of mitochondrial genes 143
5.2.3. Genome structure and rearrangements 145

5.3. Future Directions **148**
5.3.1. Modification of the CR-RT-PCR protocol 148
5.3.2. Transcrption of A+T rich region sequence 149
5.3.3. Further characterization of nuclear transcription and processing factors 151

APPENDIX A. Genbank Flatfile for *Philaenus spumarius* **155**

APPENDIX B. Genbank Flatfile for *Pteronarcys princeps* **165**

REFERENCES **174**

List of Tables

Table 2.1. List of Primer Pairs Utilized in the Amplification of the Mitochondrial Genome of *Philaenus spumarius* 30

Table 2.2. Insect Mitochondrial Genomes Utilized in Alignments and Comparative Analyses 35

Table 2.3. Nucleotide Composition of Meadow Spittlebug (*P. spumarius*) Genome Features 40

Table 2.4. Relative Synonymous Codon Usage for Meadow Spittlebug, the Kissing Bug and the Migratory Locust Mitochondrial Genomes 45

Table 3.1 Primers Used in the Amplification of the Mitochondrial Genome of the Giant Stonefly *Pteronarcys princeps* 53

Table 3.2. Complete and Near-Complete Hexapod Mitochondrial Genome Sequences Utilized in this Study, with their GenBank Accession Numbers 59

Table 3.3. Nucleotide Composition of Stonefly Mitochondrial Genome Features 66

Table 3.4. Comparison of Relative Synonymous Codon Usage for the Giant Stonefly, a Silverfish and a Locust 68

Table 3.5. Positional Nucleotide Bias in the Mt-Genome of the Giant Stonefly and Other Representative Insects 73

Table 4.1. Primers Used in 5' and 3' RACE Amplification 97

Table 4.2. Primers Utilized in CR-RT-PCR Reactions 98

Table 4.3. Results of tBlastn Searches for a Nuclear *cox1* Gene 105

Table 4.4. Positional Annotation of Transcribed RNA Nucleotides as Determined Through cDNA Sequencing 115

List of Figures

Figure 1.1. Summary of Major Changes in the Evolution of Animal Mitochondrial Genome Content....7

Figure 1.2. An Insect Phylogeny....13

Figure 1.3. Schematic Representation of Asymmetric Unidirectional Versus Bi-directional, Symmetric Semidiscontinious Replication in Animal Mitochondria....15

Figure 1.4. Summary of Results from Berthier et al. (1986) Study of *Drosophila* Mitochondrial RNAs....22

Figure 2.1. One view of Evolutionary Relationships of the Order Hemiptera....27

Figure 2.2. Schematic Representation of Amplification Strategy Employed for the *Philaenus spumarius* Mitochondrial Genome....31

Figure 2.3. Putative secondary structure folds for the tRNA genes of the *Philaenus spumarius* mt-genome....36

Figure 2.4. Putative stem-loop structures for two repeat units of Repeat Region 1.... 41

Figure 2.5. Map of the Region Between the *srRNA* and the *nad2* Genes....42

Figure 3.1. Differing Phylogenetic Views of the Placement of Plecoptera within the Insects.... 50

Figure 3.2. Proposed Secondary Structure for the tRNA Genes of the Giant Stonefly.... 63

Figure 3.3. Illustration of the Directional Mutational Pressure Observed in the Mitochondrial genome of the Giant Stonefly....71

Figure 3.4. Dot Plot Analyses of the A+T Rich Region of *Pteronarcys princeps.*....76

Figure 3.5. Predicted Secondary Structure Folds for Potential Stem-Loop Regions Identified within the A+T Rich Region of the Giant Stonefly............................ 77

Figure 3.6. The A+T Rich Regions of *Pteronarcys princeps* and Two *Peltoperla* Species..78

Figure 4.1. Alignment of the 5' End of the cox1 Gene for Various Insect, Springtail and Crustacean Representatives...85

Figure 4.2. Summary of the Analysis of Mitochondrial Transcripts from *Drosophila*, as Described by Berthier *et al.* (1986)..87

Figure 4.3. A Schematic Representation of the Circularization RT-PCR Reaction............103

Figure 4.4. Alignment of 3' cDNA Sequence to the Mitochondrial DNA Sense Strand for *Drosophila melanogaster*..109

Figure 4.5. Alignment of 5' cDNA Sequence to the Mitochondrial DNA Sense Strand for *Drosophila melanogaster*..110

Figure 4.6. Example of Sequencing Results Obtained from CR-RT-PCR Generated cDNAs...113

Figure 4.7. Characterization of Polyadenylated mRNA and pre-mRNA Molecules from *Drosophila* mitochondria..116

Figure 4.8. A Model for Processing of the *nad3* - $tRNA^{Glu}$ RNA Transcript......................132

Figure 4.9. A Model for the Processing of the *cox2* - *atp8* Transcript..............................134

Figure 5.1. Transcription Map for the *D. melanogaster* Mitochondrial Chromosome......137

Figure 5.2. Proposed Annotation of Insect *cox1* 5' ends for Representative Insect and Hexapod Taxa...146

Chapter 1. Introduction

1.1. Mitochondria

Mitochondria are organelles found within most eukaryotic cells. They are observed to be small (0.5-1.0 μm), typically rod shaped, with two distinct membrane bilayers surrounding them. They are responsible for the generation of cellular ATP through the oxidative phosphorylation enzyme pathway. A proton gradient across the mitochondrial inner membrane, maintained by the oxidative phosphorylation enzyme complexes, is utilized to drive the production of ATP from ADP and phosphate within the mitochondrial inner matrix.

The term "mitochondria" was originally coined by C. Benda in 1898, formally naming these entities that had been observed within cells for the preceding 60 years (Ernster and Schatz, 1981). During the early stages of molecular technology, DNA was isolated from the mitochondrial organelles, showing that the mitochondria had a genetic system independent of the nuclear genome of the cell (Nass and Nass, 1963a; Nass and Nass, 1963b; Nass, *et al.*, 1965). This discovery lead to a renewed interest in the "Serial Endosymbiotic Theory", first proposed by Altmann in 1890, that the mitochondrion was the result of a bacterium that entered into a symbiotic relationship with the ancestor of the eukaryotic cell (Ernster and Schatz, 1981). Widespread acceptance of an endosymbiotic origin of mitochondria and chloroplasts appears to have been achieved after clear arguments that would support the theory were described by Gray and Doolittle (1982). Subsequent studies have supported endosymbiotic theory, and remaining controversies focusing on the Serial Endosymbiotic Theory versus a simultaneous origin of the

mitochondria and nucleus in eukaryotic cells (Gray, 1992; Gray, *et al.*, 1999; Dyall, *et al.*, 2004a; Gray, 2005).

1.2. Mitochondrial Evolution

Current opinions of mitochondrial origin are based on molecular phylogenetic analyses of various molecular markers, which consistently describe the mitochondria as branching with extant α-proteobacteria (Olsen, *et al.*, 1994; Viale and Arakaki, 1994; Andersson, *et al.*, 1998). This endosymbiotic association is thought to have occurred at a very early stage of the evolution of eukaryotes. Eukaryotic cells exist that lack mitochondria, but most groups have now been investigated and appear to either encode genes that are clearly of mitochondrial origin, or contain organelles that may be modified mitochondria (Gray, *et al.*, 1999). Currently there are claims that organelles known as hydrogenosomes may be highly modified mitochondria that have evolved in place of the aerobically respiring mitochondria in anaerobic eukaryotes (Hrdy, *et al.*, 2004; Boxma, *et al.*, 2005), however the evidence supporting this assertion is still controversial (Dyall, *et al.*, 2004b; Gray, 2005). Mitosomes are also considered to be degenerate mitochondrial organelles (Gray, *et al.*, 2004), minimizing the potential number of protists that may have maintained a primitive amitochondriate state (Gray, 2005).

The diversity of structural forms that extant mt-genomes have adopted is often not appreciated by those familiar with only animal, plant and yeast mitochondrial systems. The circular chromosome presumed to be ancestral state has been modified to single linear molecules in some protist groups, split into multiple linear chromosomes in *Amoebidium* protists, or developed into complex groups of gene encoding maxicircles and minicircles that direct RNA editing in trypanosome mitochondria (Gray, *et al.*, 1999;

Burger, *et al.*, 2003a; Burger, *et al.*, 2003b; Bullerwell and Gray, 2004; Gray, *et al.*, 2004).

Great functional diversity has also been observed, including numerous examples of RNA editing systems (Gray, 2003). There have been descriptions of rRNA genes expressed as pieces that must self-associate to function (Boer and Gray, 1988; Gillespie, *et al.*, 1999). Protein genes have been divided into independently expressed protein subunits (Edqvist, *et al.*, 2000), with one portion expressed in the mitochondria and the other expressed in the cytoplasm (Nedelcu, *et al.*, 2000; Perez-Martinez, *et al.*, 2001).

The mt-genomes for which complete sequence exists show that extensive gene loss by the mitochondrial chromosome has occurred in all eukaryotic lineages. The hypothetical ancestor of the mitochondria would be expected to lose genes required by the free living organism as purifying selection on these genes was diminished within the new host. Further gene loss has since occurred as genes were moved from the mitochondrial genome to the nuclear genome, presumably to enable nuclear coordination of expression of the mitochondrial genes and thereby increase nuclear control over mitochondrial function (Muller and Martin, 1999; de Grey, 2005). The gene content within mt-genomes varies greatly between the different extant groups, from 98 genes in the freshwater protozoon *Reclinomonas americana* (Lang, *et al.*, 1997) to only three genes in *Plasmodium* species (Conway, *et al.*, 2000).

Given that from 600 to 1000 other genes have successfully transferred from the mt-genome to the nuclear genome, considerable effort has gone into developing a theory to describe why mitochondria have actually maintained any genes, and not allowed for complete transfer of genes to the nucleus. Three main theories are currently cited to

explain the persistence of organellar DNA; the hydrophobicity hypothesis, the codon disparity hypothesis, and location based expressional control (Adams and Palmer, 2003; de Grey, 2005). The hydrophobicity hypothesis demonstrates that mitochondrial gene products are among the most hydrophobic proteins, are difficult to import across the outer and inner mitochondrial membranes, and argue that these properties of the proteins have prevented gene transfer to the nucleus (von Heijne, 1986; Popot and de Vitry, 1990; Claros, *et al.*, 1995).

The second theory notes that the changes in genetic code between nuclear and mitochondrial genomes, especially the common change of UGA to code for tryptophan in mitochondria instead of a stop codon in the nuclear code, would lead to proteins being translated with severe truncations or amino acid substitutions and would be selected against gene transfer (Andersson and Kurland, 1991; Jacobs, 1991; Leblanc, *et al.*, 1997). The codon changes appear to have occurred after the loss of the majority of genes, so may be involved in maintenance of the now heavily reduced mt-chromosome (de Grey, 2005).

The third hypothesis proposes that key genes involved in oxidative phosphorylation are expressed in the mitochondria so that their proximity to the enzyme complexes may regulate their expression (Allen, 1993). Location based expressional control theory is gaining support within the community (Race, *et al.*, 1999; Adams and Palmer, 2003; Allen, 2003; Gaspari, *et al.*, 2004b). Experimental elucidation of translational-based control of mt-mRNAs in yeast (Naithani, *et al.*, 2003; Barrientos, *et al.*, 2004) and potentially in humans (Mili and Pinol-Roma, 2003; Mootha, *et al.*, 2003; Xu, *et al.*, 2004) are adding considerable support to this hypothesis.

1.3. Animal Mitochondrial Genomes

1.3.1. Animal Mitochondrial Gene Content

In terms of structure and gene content, animal mt-genomes are far more stable and predictable than mt-genomes in other eukaryotic groups (Wolstenholme, 1992; Boore, 1999). The standard animal mt-genome encodes thirteen protein-coding genes that are subunits of four of the five mitochondrial membrane associated protein complexes involved in oxidative phosphorylation. The genes are *NADH dehydrogenase subunit 1* (*nad1*), *nad2, nad3, nad4, nad4L, nad5* and *nad6* from complex I (NADH-ubiquinol oxidoreductase), *cytochrome b* (*cytb* or *cob*) from complex III, (ubiquinone-cytochrome-c-oxidoreductase), *cytochrome c oxidase subunit I* (*cox1*), *cox2,* and *cox3* from complex IV (cytochrome c oxidase), and ATP synthase FO subunit 6 (*atp6*) and *atp8* from complex V (ATP synthase). In addition to the protein coding genes, the mitochondria encode a minimal translation system, including the large and small mitochondrial ribosomal subunits (*lrRNA* and *srRNA,* respectively), as well as 22 tRNA genes[1]. The 22 tRNA genes are the entire complement found within the mitochondria, and interact with the two or four degenerate codons for that amino acid through a "super-wobble" codon pairing interaction (Taanman, 1999; Lavrov, *et al.*, 2005). The final structural feature is a large non-coding region recognized for its involvement in replication and transcript initiation, and variously designated the major control region, displacement loop (D-loop), major non-coding region or A+T rich region (in insects and other hexapods).

---

[1] The tRNA genes will be referred to by the superscripting of their three-letter amino acid codes, for example $tRNA^{Ile}$ for the isoleucine. The serine and leucine tRNAs will be specified by their three letter code, and the codons they decode (for example $tRNA^{Ser\text{-}AGN}$).

Variations from this standard genome content are rare, but known from a variety of different animal groups (see Figure 1.1). The most basal of the metazoa, the sponges (phylum Porifera), are represented by three complete mitochondrial genome sequences and appear to have retained a small number of genes that are found in the mitochondria of fungi, protists and plants (Gray, *et al.*, 1999; Burger, *et al.*, 2003a). In the three sponge mt-genomes, the *atp9* subunit of complex V has been retained within the mt-genome, whereas it has been transferred to the nuclear genome in other animal groups (Lavrov, *et al.*, 2005; Lavrov and Lang, 2005). A *Tethya* sponge mt-genome has retained three additional tRNA genes not seen in other animal mitochondrial genomes. A second methionine tRNA (the additional elongator $tRNA^{Met}$ as well as the f-met initiator $tRNA^{Met}$), the $tRNA^{Ile\text{-}ATA}$, and the $tRNA^{\text{-}Arg\text{-}AGY}$ are found in addition to the 22 tRNAs normally found in the animals (Lavrov, *et al.*, 2005). Two additional sponge mt-genome sequences are available and appear to have subsequently lost the elongator $tRNA^{Met}$ but retain $tRNA^{Ile\text{-}ATA}$ and $tRNA^{Arg\text{-}AGY}$ (Lavrov, *et al.*, 2005; Lavrov and Lang, 2005).

Addition of non-ancestral genes to the mitochondrial chromosome has only been reported in the Cnidaria (corals, jellyfish and hydras). The first is a gene originally identified as a member of the MutS mismatch repair system (Pont-Kingdon, *et al.*, 1995; Pont-Kingdon, *et al.*, 1998) and is currently proposed to be fusion of a MutS2 gene with a mismatch recognition-nicking domain (Helfenbein, *et al.*, 2004).

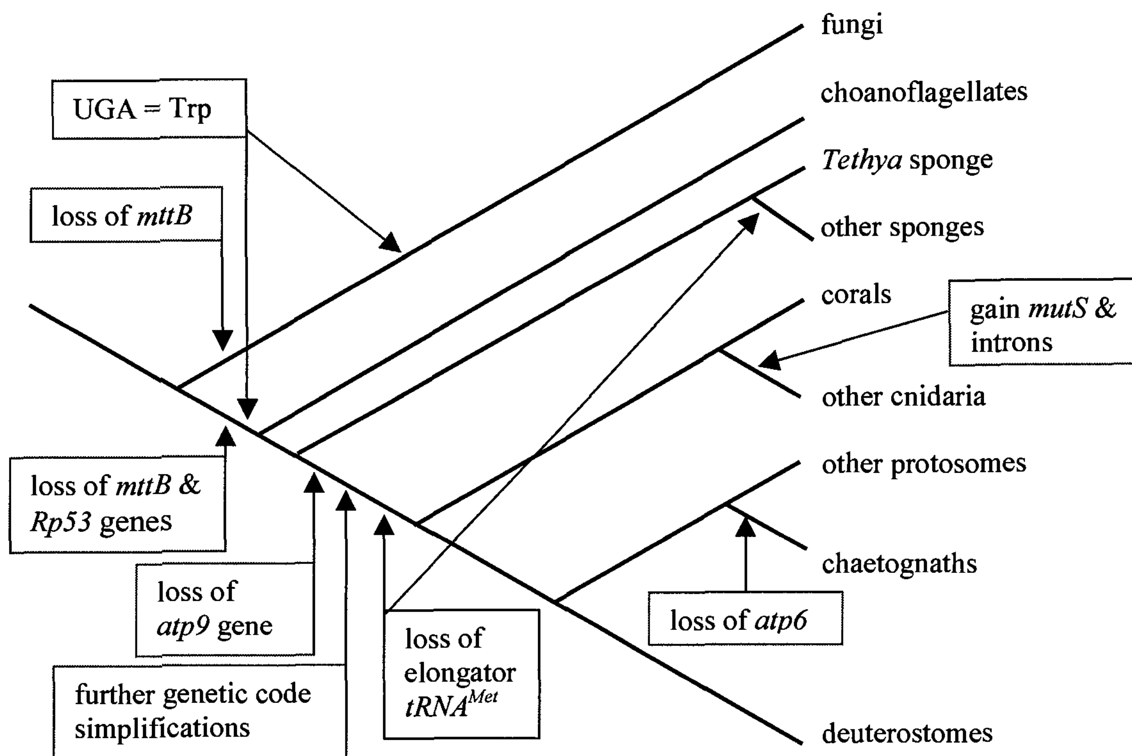


Figure 1.1. Summary of Major Changes in the Evolution of Animal Mitochondrial Genome Content. Adapted from de Grey (2005) with updates from Helfenbein, *et al.*(2004), Papillon, *et al.* (2004), Watkins and Beckenbach (1999) and Lavrov *et al.* (2005). Arrows point to proposed timing of major genome additions, losses or changes in animal evolution. The independent losses of *atp8* in mollusks, nematodes and chaetognaths are not represented.

A number of independent animal lineages have lost the *atp8* gene from the mitochondrial chromosome. The first examples of this were found in the nematodes, including the genetic model organism *Caenorhabditis elegans* and the pig intestinal roundworm *Ascaris suum* (Okimoto, *et al.*, 1992). A second gene has been acquired in the Cnidaria, which appears to be a homing endonuclease gene that may be involved in the removal of intron sequences that have evolved in the cnidarian mitochondria (Beagley, *et al.*, 1998).

Gene loss in animal mitochondria is rare, but more frequent than the acquisition of new genes. In the chaetognath *Paraspadella gotoi,* the mt-genome appears to have lost all but one mitochondrial tRNA, as well as the *atp6* and *atp8* genes normally found in animal mitochondria (Helfenbein, *et al.*, 2004). A second chaetognath mt-genome has been completely sequenced, and also appears to lack the same protein coding genes and may have lost the remaining tRNA gene (Papillon, *et al.*, 2004). Bivalves, mollusks and platyhelminth flat worm mt-genomes have independently lost the *atp8* gene as well (Hoffmann, *et al.*, 1992; Le, *et al.*, 2000). Whether the *atp8* gene has transferred to the nucleus or had its function replaced by other proteins has been of interest to those studying the molecular evolution of the mitochondria. The *atp8* protein is very small (51 to 65 aa in length) with very little sequence conservation. At present, it has not yet been identified through sequence similarity to small open reading frames within the complete nuclear genome of *C. elegans* (Helfenbein, *et al.*, 2004).

Change in tRNA content varies widely across the animals, but generally it appears that early animals shared the 25 tRNA content of the sponges, where the only genetic code change is UGA to coding tryptophan, rather than a terminator (Lavrov, *et al.*, 2005).

The number of tRNAs was reduced to the current 22 by the loss of the elongator $tRNA^{Met-CAU}$, the $tRNA^{Arg-UCY}$, and the $tRNA^{Ile-AUA}$. This altered the genetic codes further to encode only two and four-fold degenerate tRNAs, so that the 22 tRNAs are sufficient to recognize the 62 or 60 coding codons, depending on the particular genetic code (Boore, 1999; Lavrov and Lang, 2005).

1.3.2. Animal Mitochondrial Genome Structure

Typically the animal mitochondrial genome is a circular DNA molecule. Early reports of linear animal mitochondrial genomes have been published, but no sequence confirmation of these molecules has yet been reported (Bridge, *et al.*, 1992). The animal mt-genome is typically very small and compact, at 15-17 kb in size with very few non-coding nucleotides (Wolstenholme, 1992). Variations in the size of a mitochondrial genome are quite large, with the largest reported mt-genome belonging to the deep-sea scallop, *Placopecten magellanicus,* at 42 kb (Snyder, *et al.*, 1987), with the smallest being the mt-genome of the chaetognath (or arrow worm) *Paraspadella gotoi* at 11423bp (Helfenbein, *et al.*, 2004). Larger than average mitochondrial genomes tend to be the result of repeat unit expansions (Boyce, *et al.*, 1989; La Roche, *et al.*, 1990; Fuller and Zouros, 1993) , duplication of genome regions (Arndt and Smith, 1998), or the development of very large non-coding regions, not through increase in gene content. In contrast, the smallest mt-genome has been observed to be the result of loss of protein coding and tRNA genes, and the contraction of the major non-coding region (Helfenbein, *et al.*, 2004; Papillon, *et al.*, 2004).

1.3.3. Mitochondrial Genome Sequencing

Complete mitochondrial DNA sequences are being generated and released in the sequence databases at a very rapid rate. As of June 2005, complete (or nearly complete) NCBI Reference Sequences (RefSeq, Pruitt, *et al.*, 2005) mt-genome sequences were available in GenBank for 721 different species of plants, animals, fungi and protists.

Animal sequences are strongly over represented with 649 (90%) of the mt-genome sequences. Besides our own phylum Chordata (with 484 mt-genome sequences), Arthropoda is the best represented phylum, with 95 mt-genome sequences.

Within the arthropods, the insects and other hexapods are the best represented with mt-genome sequences, with 49 species represented (45 from class insecta and four collembolan springtails). The remaining major divisions of arthropods show less mt-genome sampling at this time: 4 Myriapoda (centipedes and millipedes), 17 Chelicerata (spiders, ticks, scorpions and horseshoe crabs) and 25 crustaceans.

The first insect mitochondrial genome sequences were obtained for *Drosophila yakuba* and *D. melanogaster*. The mt-genome of *D. yakuba* was finished in advance of *D. melanogaster* primarily due to the large (4.6 kb) and complex non-coding region found in *D. melanogaster* that made for difficult cloning and sequencing (Goddard and Wolstenholme, 1980; Clary and Wolstenholme, 1985a; Lewis, *et al.*, 1994). The mitochondrial gene order originally observed in *Drosophila* has been demonstrated to be the ancestral gene order of the hexapods and crustaceans, as representatives with this gene arrangement are observed in each of these groups (Clary and Wolstenholme, 1985a; Crease, 1999; Hwang, *et al.*, 2001; Nardi, *et al.*, 2003b). This gene order is generally conserved, with occasional tRNA rearrangements via a duplication / random loss model

(Moritz, *et al.*, 1987) that seem to have occurred independently (Flook, *et al.*, 1995; Yukuhiro, *et al.*, 2002). Mosquitoes and hymenoptera appear to have undergone tRNA translocations relative to the ancestral gene order (Moritz, *et al.*, 1987; Crozier and Crozier, 1993; Flook, *et al.*, 1995; Dowton and Austin, 1999; Dowton, *et al.*, 2003). Protein gene and rRNA rearrangements are rare within the insects, and currently appear to be confined to the group known as the Paraneoptera (Shao, *et al.*, 2001; Shao and Barker, 2003; Shao, *et al.*, 2003; Thao, *et al.*, 2004).

A unique feature was noted in these earliest insect mt-genomes that set insect mitochondria apart from the standard mitochondrial peculiarities. The most unexpected of these was the lack of an in-frame start codon for the otherwise highly conserved *cox1* gene (Clary and Wolstenholme, 1983a; de Bruijn, 1983). In response to this unusual situation, a four base start "ATAA" start codon was described, incorporating the in-frame TAA stop plus the 5' encoded A nucleotide at positions 1470-1473 in both *Drosophila* species. Subsequent analyses of other *Drosophila* species did not support the conservation of a 4-base ATAA start codons (Satta, *et al.*, 1987). As more sequences accumulated, it became apparent that a multi-base start codon was not a reasonable model for translation. Such an unusual mechanism would be expected to be well conserved, but the proposed multi-base initiation codon was not conserved between the insects studied (Beard, *et al.*, 1993). The uncertainty has persisted, and annotation of *cox1* has become a difficult problem for insect mt-genome researchers (discussed in Mitchell, *et al.*, 1993; Flook, *et al.*, 1995; Lessinger, *et al.*, 2000; Spanos, *et al.*, 2000; Nardi, *et al.*, 2001; Yukuhiro, *et al.*, 2002; Friedrich and Muqim, 2003; Nardi, *et al.*, 2003a; Stewart and Beckenbach, 2003; Junqueira, *et al.*, 2004; Yamauchi, *et al.*, 2004; Coates, *et al.*, 2005).

The *nad4* gene of the plague thrips *Thrips imaginis* encodes a similarly cryptic initiation codon (Shao and Barker, 2003). Aside from annotation questions, these unusual genes raise interesting questions regarding transcription and translation of these genes.

Much of the subsequent interest in complete mt-genome sequences of insects has been promoted by molecular evolution and molecular phylogenetic studies. The insects are a very diverse class of animals, with more living species than all other animal groups (Kristensen, 1991). This rich diversity has proven to be very problematic in deciphering the interrelationships between the insect orders through more classical taxonomic methods, with disputes regarding the between-order relationships being very common (see Boudreaux, 1979; Hennig, 1981; Kristensen, 1991). Expectations were that molecular evolutionary studies might discriminate among the phylogenetic hypotheses and thereby resolve some of these relationships. Analyses using nuclear genes have so far not been successful in the resolution of order-level phylogenetics, and have led to some very high profile controversies in the literature (Whiting and Wheeler, 1994; Carmean and Crespi, 1995; Huelsenbeck, 1997; Whiting, *et al.*, 1997; Huelsenbeck, 1998; Hwang, *et al.*, 1998; Whiting, 1998; Wheeler, *et al.*, 2001). It has been suggested that more complete sampling of insect mt-genomes would create a data set sufficient to resolve the order level relationships of insects and help to resolve these unclear relationships (Friedrich and Muqim, 2003; Nardi, *et al.*, 2003b; Stewart and Beckenbach, 2003; Bae, *et al.*, 2004; Cameron, *et al.*, 2004; Castro and Dowton, 2005; Kim, *et al.*, 2005). The phylogeny proposed by Kristensen (1991) is presented in Figure 1.2.

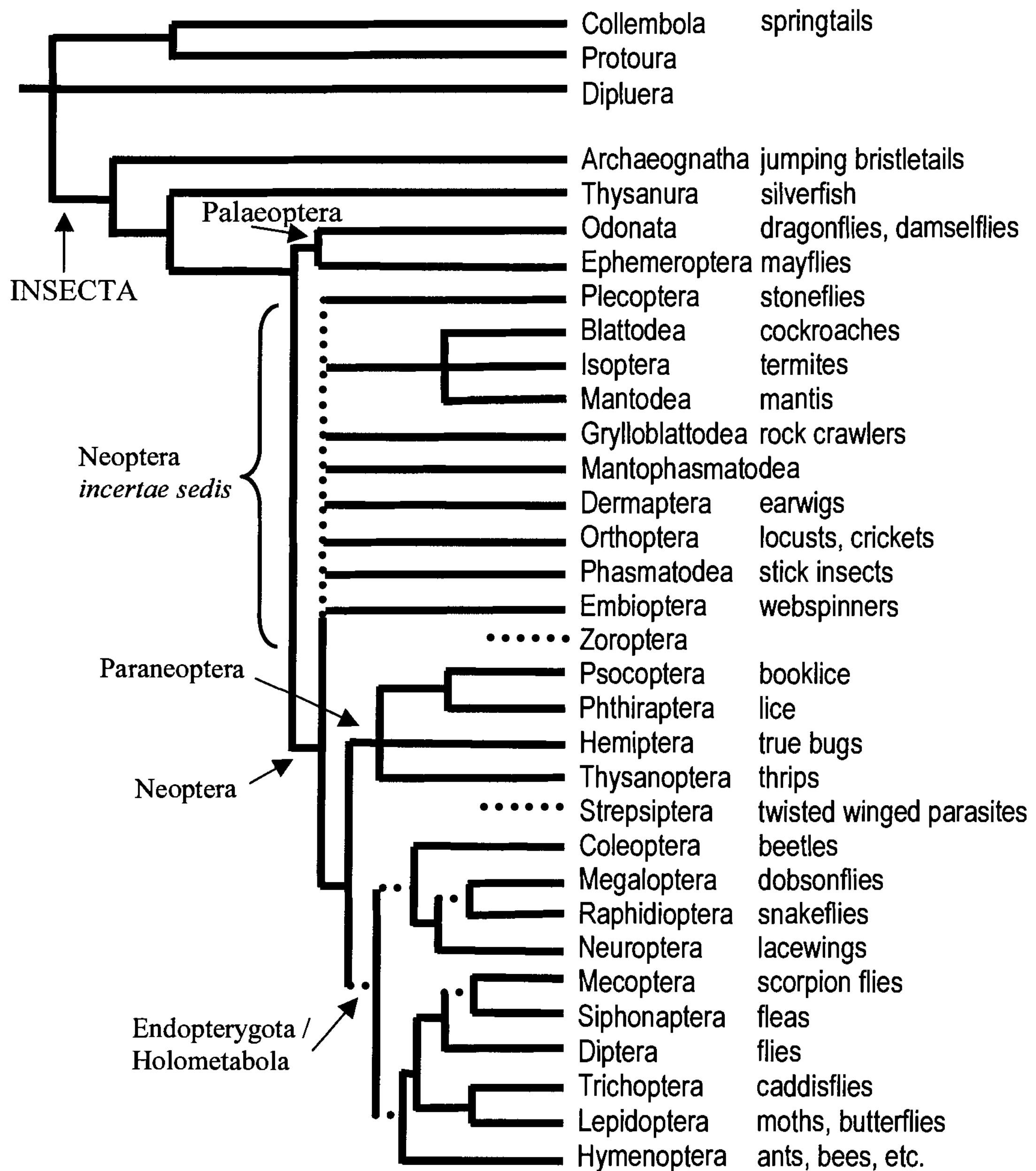


Figure 1.2. An Insect Phylogeny. A reconstruction of the phylogeny of insect orders proposed by Kristensen (1991). Names of orders are followed by a common name of representatives of that order. Dotted lines reprsent branches that are not well supported in the view of Kristensen (1991).

1.4. Mitochondrial Transcription and Replication

1.4.1. Mitochondrial Replication in Animals

Mitochondrial replication in animals has been studied for over 30 years in vertebrate and invertebrate systems (Wolstenholme, 1973; Goddard and Wolstenholme, 1978; Crews, *et al.*, 1979; Goddard and Wolstenholme, 1980; Montoya, *et al.*, 1982 and others). Replication in mammalian mitochondria initiates at the origin of heavy strand replication (O_H) which is found within the D-loop region of the mt-DNA. Replication of the heavy strand proceeds asymmetrically and unidirectionally for approximately two thirds the way around the mt-genome until the origin of light strand replication (O_L) is reached (see figure 1.3A). Once heavy strand replication uncovers the O_L, lagging strand replication initiates and proceeds unidirectionally in the opposite direction. Until O_L is reached, the displaced light strand stays single-stranded and exposed to the environment of the mitochondrial inner matrix (reviewed in Clayton, 2000).

At present, the known and characterized enzymes involved in mitochondrial replication are the gamma polymerase (POLγ) (Kaguni, 2004), a T7-phage like helicase (TWINKLE) (Spelbrink, *et al.*, 2001), and the mitochondrial single-strand binding protein (mtSSB) (Korhonen, *et al.*, 2004). Insect homologues for POLγ (Wernette and Kaguni, 1986; Wernette, *et al.*, 1988) and mtSSB (Stroumbakis, *et al.*, 1994; Farr, *et al.*, 1999) are known for *D. melanogaster.*

Recently, controversy surrounding the model of mt-DNA replication has appeared in the literature (Bogenhagen and Clayton, 2003; Holt and Jacobs, 2003; Yasukawa, *et*

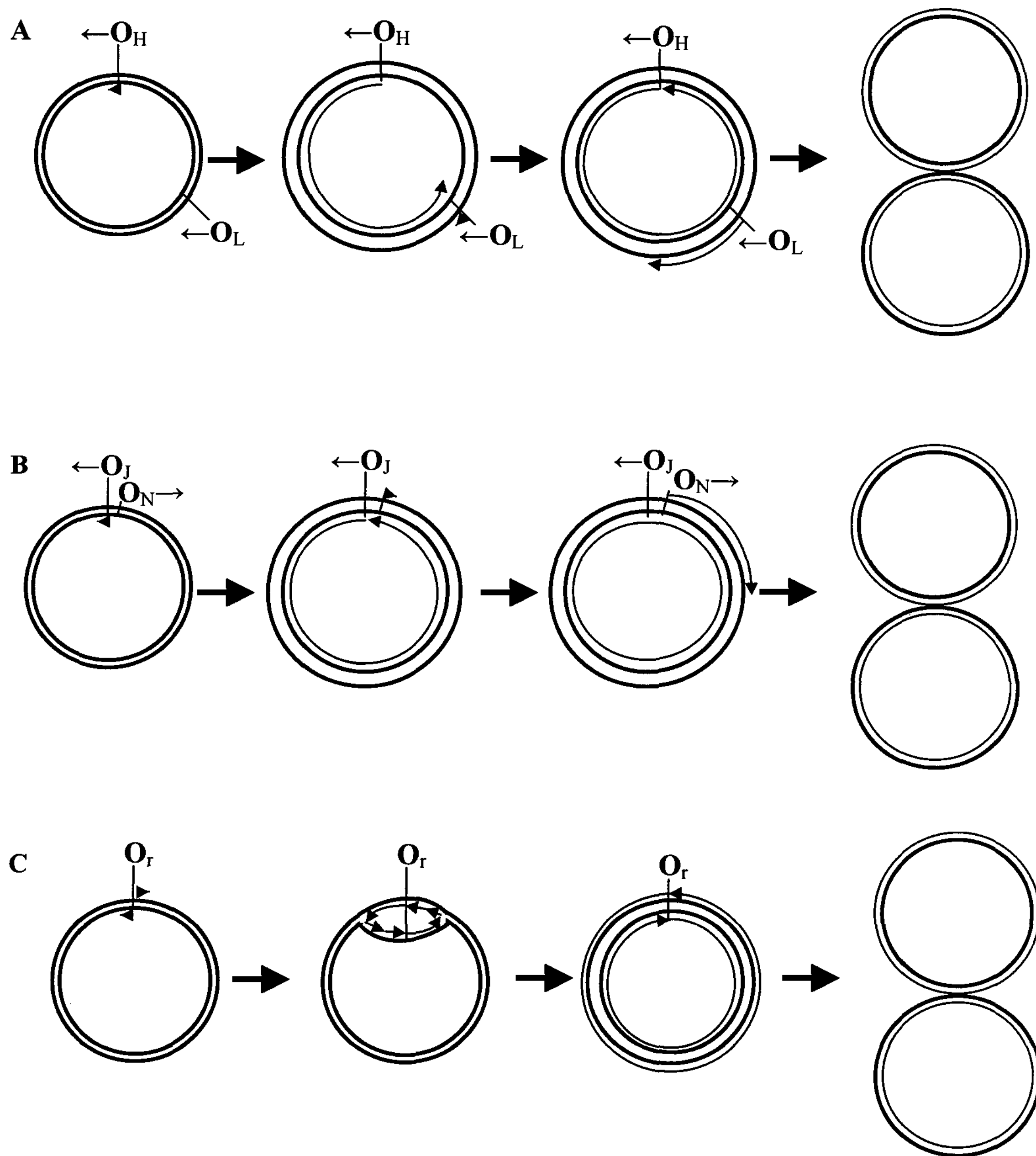


Figure 1.3. Schematic Representation of Asymmetric Unidirectional Versus Bi-directional, Symmetric Semidiscontinious Replication in Animal Mitochondria. Origins of replication are denoted and labeled. Thick lines represent the parental molecule and thin represents the newly replicated molecule, with the arrowhead representing the polymerase. A – Replication in mammal mitochondria, with O_L initiation occurring after approximately two-thirds of the H-strand replication (Clayton, 2000). B - Replication in *Drosophila*, where N-strand replication initiates after >90% of the replication of the J-Strand (Goddard and Wolstenholme, 1978, 1980). C – Bidirectional replication of a mitochondrial chromosome (Bowmaker *et al.*, 2003).

al., 2005). Evidence for a bi-directional, symmetric and semidiscontinuous replication model has been described in mammalian mitochondria (Yang, *et al.*, 2002; Bowmaker, *et al.*, 2003) and bird mitochondria (L'Abbe, *et al.*, 1991; Reyes, *et al.*, 2005)(outlined in figure 1.3C). In response to these reports, strong defense of the asymmetric model has been published (Clayton, 2003; Fish, *et al.*, 2004). In support of asymmetric replication models, a mutational pattern has been observed in mt-DNA that is consistent with the displaced single strand in replication being exposed to stronger mutational pressures in regions of the mt-DNA that are maintained in the single stranded state longer. In these regions, mt-DNA on the lagging strand appears to select against the deamination-sensitive A and C nucleotides, resulting in characteristic AT and GC skews in nucleotide composition (Reyes, *et al.*, 1998; Saccone, *et al.*, 1999).

A possible explanation of these conflicting results was published much earlier based on direct observation of replicating mt-DNA in *Drosophila.* Electron microscopy of restriction endonuclease digested mitochondria undergoing replication have shown that a majority of mitochondria undergo extreme asymmetric replication (figure 1.3B), yet a minority of molecules were observed that suggest a bi-directional or only slightly asymmetric model of replication (figure 1.3C) (Goddard and Wolstenholme, 1978). The two *Drosophila* replication origins appear to be encoded in the central region of the major non-coding region (A+T rich region) in a number of different *Drosophila* species (Goddard and Wolstenholme, 1980). Thus the use of two modes of replication within animal mitochondria had already been suggested.

1.4.2. Variation in Transcriptional Systems Between Eukaryotic Groups

The transcriptional initiation, regulation and processing machinery of mt-genomes has varied over the course of eukaryote evolution. Three systems have received the majority of the research interest, the mammal mitochondria, plant mitochondria and yeast mitochondria. The study of yeast and plant mitochondrial systems has a number of advantages over animal mitochondrial systems. Both yeast and plant mitochondria can be manipulated by the addition of DNA into the mitochondria (Johnston, *et al.*, 1988; Farre and Araya, 2001; Koulintchenko, *et al.*, 2003), a technique not yet possible in animal mitochondria. Yeast's ability to grown via fermentation when their aerobic respiration system is disrupted has allowed for the isolation of mitochondrial mutants that would have been lethal to animal model organisms (Gagliardi, *et al.*, 2004).

Important differences between animal and other eukaryotic systems persist and necessitate the study of animal systems. Both plant and yeast transcriptional systems encode important functional information within the 3' untranslated region (UTR) of the mRNAs, with animal mitochondrial transcripts lacking UTRs completely, or encoding only very few non-coding nucleotides (Gagliardi, *et al.*, 2004). Animals require polyadenylation of their mature mt-transcripts for function (Ojala, *et al.*, 1980a; Ojala, *et al.*, 1981), while plants utilized polyadenylation as a signal to promote RNA degradation, and polyadenylation of RNA is absent in yeast (Binder and Brennicke, 2003; Gagliardi, *et al.*, 2004). Animal transcription is also unusual in that it proceeds as a small number of very large polycistronic messages that are post-transcriptionally processed into mature RNA molecules (see section 1.4.4) (Ojala, *et al.*, 1980a; Ojala, *et al.*, 1981). Also, apoptosis is absent in the unicellular yeast (Foury and Kucej, 2002), yet studies show that

mitochondria play a pivotal role in the apoptotic cascade in animals (Adrain and Martin, 2001).

1.4.3. Mammalian Mitochondrial Transcription

All animal mitochondrial genes involved in transcription of the mt-genome are encoded within the nuclear genome, and are imported into the mitochondria. *In vitro* transcription of cloned mitochondrial sequence has been accomplished for mammalian systems (Falkenberg, *et al.*, 2002; Gaspari, *et al.*, 2004a). Transcription *in vitro* requires mitochondrial RNA polymerase (POLMT), the transcription factor alpha (TFAM), and either of the mammalian transcription factor beta proteins (TFB1 or TRB2)(Falkenberg, *et al.*, 2002; Rantanen, *et al.*, 2003). The transcription activities of TFB1 and TFB2 have not yet been confirmed in cell culture or in whole organism studies. The TFB genes are interesting in that they show strong sequence similarity to bacterial RNA methyl transferase genes (Falkenberg, *et al.*, 2002). TFB1 has been shown to have methylase activity in addition to a transcription factor role (McCulloch and Shadel, 2003; Seidel-Rogol, *et al.*, 2003). These genes appear to have evolved as methyl transferase genes and were recruited to their current role in transcription initiation (Falkenberg, *et al.*, 2002; Shoubridge, 2002; Rantanen, *et al.*, 2003).

Mammalian mitochondria transcription initiates at either the light strand promoter (Lsp) or one of two heavy strand promoters (Hsp1 or Hsp2), which are located within the D-loop region of the mammalian mt-genome (Fernandez-Silva, *et al.*, 2003). The Lsp and Hsp2 promoters appear to be involved in the generation of the large polycistronic messages that produce mRNAs, tRNAs and rRNAs (Montoya, *et al.*, 1982; Montoya, *et al.*, 1983), while the Hsp2 promoter is more active (approximately 20x more active) and

may be involved in transcribing the rRNA cassette (Montoya, *et al.*, 1983; Kruse, *et al.*, 1989; Fernandez-Silva, *et al.*, 2003). The promoters require binding of TFAM and the associated POLMT / TFB2 complex to initiate transcription, and appear to interact in a nucleotide sequence-specific manner with the proteins (Gaspari, *et al.*, 2004a). The Lsp and Hsp2 initiated mt-transcripts are then extended to transcribe nearly the complete mt-genome before termination, while Hsp1 initiated transcripts are terminated just downstream of the *lrRNA* gene (Montoya, *et al.*, 1982; Montoya, *et al.*, 1983; Tracy and Stern, 1995; Nakamichi, *et al.*, 1998; Fernandez-Silva, *et al.*, 2003). These large polycistronic messages are then processed into mature RNA molecules (see section 1.4.4)

Other nuclear-encoded mammalian genes involved in the mitochondrial transcription and transcript maturation have been characterized, such as a termination factor (Fernandez-Silva, *et al.*, 1997), the RNase P (Doersen, *et al.*, 1985; Puranam, *et al.*, 2001) a poly(A) polymerase (Tomecki, *et al.*, 2004; Nagaike, *et al.*, 2005) and a polynucleotide phosphorylase (Nagaike, *et al.*, 2005).

1.4.4. Transcript Processing in Mammalian Mitochondria

The processing of animal mitochondria has been most thoroughly described from human cell line mitochondria. The large polycistronic pre-RNA messages generated by the transcriptional machinery are then recognized by a currently undescribed RNA processing machinery. These large transcripts are then thought to be processed through a "tRNA punctuation" model of processing, whereby the removal and maturation of the intervening tRNA genes liberates molecules that are then processed into mature rRNA and mRNA molecules (Ojala, *et al.*, 1980a; Ojala, *et al.*, 1981). Polyadenylation of the mRNAs was shown to be required to complete some of the stop codons, which were

predicted to encode only in-frame U or UA nucleotides after the removal of the 3' abutting tRNA genes (Ojala, *et al.*, 1980a; Anderson, *et al.*, 1981), and is thought to impart stability to the mRNAs similar to mechanisms in the nucleus (Fernandez-Silva, *et al.*, 2003; Gagliardi, *et al.*, 2004). It has been recognized that this could not be the complete description of the processing machinery, as *cox3* gene in mammal mitochondria directly abuts the *atp6* gene, but both are expressed as separate transcripts (Ojala, *et al.*, 1980a; Montoya, *et al.*, 1981; Ojala, *et al.*, 1981; Chrzanowska-Lightowlers, *et al.*, 2004). Due to the interspersing of tRNAs between the rRNA and protein genes of other animal mitochondria, this model for transcript processing has been broadly applied to all animal genomes (Wolstenholme, 1992; Taanman, 1999; Fernandez-Silva, *et al.*, 2003).

1.4.5. Transcription in *Drosophila* Mitochondria

The homologues of a number of the mammalian mitochondrial transcription genes are known in *Drosophila*, including the TFAM (Goto, *et al.*, 2001; Takata, *et al.*, 2001), TFB2 (Matsushima, *et al.*, 2004), a termination factor DmTTF (Roberti, *et al.*, 2003) and the RNA polymerase (Goldenthal and Nishiura, 1987). Interestingly, the TFB1 homologue in *Drosophila* shows RNA methyl transferase activity, and does not appear to function in transcription, but RNAi knockdown of the TFB1 has an impact on protein synthesis, implying a role in translation (Matsushima, *et al.*, 2005).

Detailed analyses of mt-RNA pools were conducted in the late 1970s and early 1980s. Mitochondrial RNAs were first identified in *Drosophila* as thirteen molecules whose expression profiles were uninfluenced by heat shock or camptothecin treatment of *Drosophila* Schneider-2 cell lines, but were sensitive to actinomycin treatment (Spradling, *et al.*, 1977). Merten and Pardue (1981) characterized the mitochondrial RNA

fraction and described two rRNA genes and "nine or ten" mRNA molecules. It was not until after the complete *D. yakuba* mitochondrial DNA sequence was published (Clary and Wolstenholme, 1985a) that a study by Berthier *et al.* (1986) was able to identify the genetic identities of each of the mRNA and rRNA genes.

Berthier *et al.* (1986) determined the relative abundance of RNA in a LiCl insoluble fraction of RNAs smaller than 200bp (30%), the poly-A fraction (5%) and the non-poly-A fraction (65%) of the RNA pool. They were also able to identify eleven poly-A mRNAs (nine corresponding to single gene mRNAs and two bicistronic messages) based on hybridization of the RNAs to restriction digested *Drosophila* mt-DNA. The *srRNA* was found in the non-poly-A fraction, but the *lrRNA* was identified in the poly-A fraction. The poly-A state of the *lrRNA* has been reported elsewhere (Spradling, *et al.*, 1977; Benkel, *et al.*, 1988). Twenty-eight large RNA species were also identified that were interpreted as processing intermediates of the tRNA punctuation mediated maturation of the transcripts. By hybridizing these large RNAs to digested mt-DNA they were able to identify five clusters of polycistronic transcriptional products which led them to infer five transcriptional cassettes for *Drosophila*, in contrast to the two observed in mammals. The results of this study are summarized in Figure 1.4.

Unfortunately the only RNA sequencing technologies available at the time were through the digestion of radiolabeled RNA with selective RNases (Ojala, *et al.*, 1980a; Montoya, *et al.*, 1981; Ojala, *et al.*, 1981), nuclease protection assays and primer extension assays (Thompson, *et al.*, 1979; Rohrbaugh and Hardison, 1983). Nucleotide sequence information was therefore not available for the RNAs.

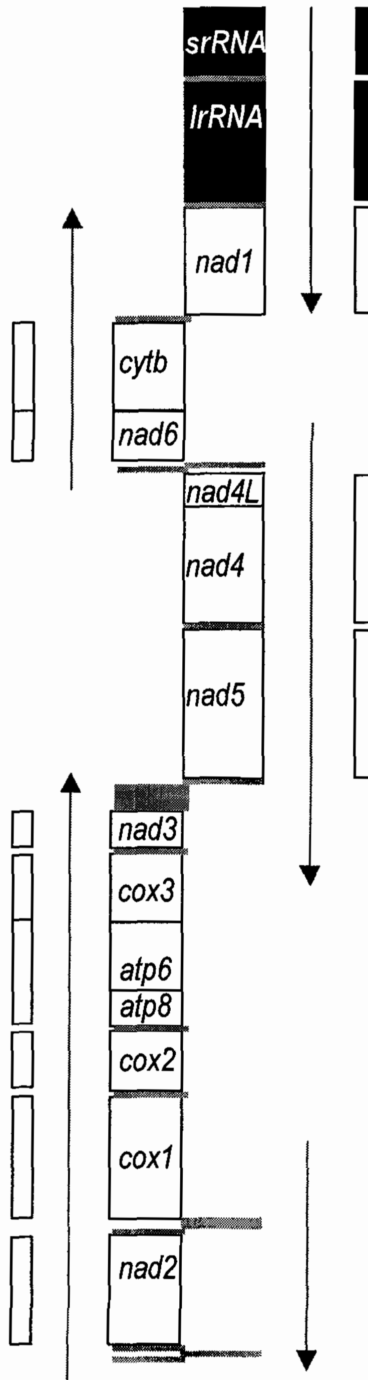


Figure 1.4. Summary of Results from Berthier *et al.*, (1986) Study of Drosophila Mitochondrial RNAs. The genes are labeled in the center of the figure, with genes on one strand offset from the other. Arrows depict transcription units, with the start of the line at the predicted transcript initiation site and the arrowhead at the predicted transcript termination site. White rectangles to the outside represent mRNAs, black the rRNAs.

1.4.6. RNA Editing and Programmed Frameshifting in Animal Mitochondria

Though some of the more exotic transcriptional mechanisms seen in other eukaryotic mitochondria are absent in animals, two unusual mechanisms are described that cause variation between the inferred transcribed information and the final gene function. RNA editing refers to the modification of RNA from the original transcript, whether by the insertion or deletion of nucleotides or the specific C to U modification of bases to alter the nucleotide sequence of the mature RNA. This phenomenon is distinct from RNA modification that is commonly observed in rRNA and tRNA sequences, where non-standard nucleotides are created by post-transcriptional modification of coded bases. RNA editing has been extensively studied in mitochondrial systems of plants and specific protist lineages, such as the trypanosomes and dinoflagellates (Lin, *et al.*, 2002; Gray, 2003; Knoop, 2004; Simpson, *et al.*, 2004; Stuart, *et al.*, 2005).

In animal mitochondria, numerous mitochondrial tRNA genes have been determined to undergo editing in diverse groups such as marsupials (Janke and Paabo, 1993; Morl, *et al.*, 1995; Borner, *et al.*, 1996), land snails (Yokobori and Paabo, 1995), a squid (Tomita, *et al.*, 1996), centipedes (Lavrov, *et al.*, 2000), and potentially in spiders (Masta, 2000; Masta and Boore, 2004; Qiu, *et al.*, 2005). Editing of mRNAs has not been yet been observed in animal mitochondria (Orr, *et al.*, 1997).

Programmed translational frameshifting is a second mechanism by which the mRNA molecule is decoded in a way other than by the reading of the simple open reading frame. Bird and turtle mitochondrial *nad3* have been observed to encode a single C insertion that disrupts the open reading frame and would result in an unusually short *NAD3* protein (Desjardins and Morais, 1990; Harlid, *et al.*, 1997; Zardoya and Meyer,

1998). Mindell, *et al* (1998) investigated this extra nucleotide by sampling a large number of bird taxa, as well as some turtles. Determining that the presence of the extra nucleotide varied between different bird taxa, they proposed a frameshifting model in an attempt to explain the highly variable *nad3* gene structure within birds. A similar phenomenon has now been observed in the *cytb* gene of ant mitochondria, and a more detailed description of a possible frameshifting mechanism has been published (Beckenbach, *et al.*, 2005).

1.5. Thesis Outline

In this study I combine a comparative mitochondrial genomic approach with molecular characterization of RNAs found in the mitochondria of *Drosophila melanogaster* in an attempt to understand the molecular evolution of the insect mitochondrial genome. Chapters 2 and 3 describe the sequencing, annotation, characterization and analyses of complete mitochondrial genomes from two insect species, the meadow spittlebug *Philaenus spumarius* (Hemiptera; Auchenorryncha; Cercopoidae) and the giant stonefly *Pteronarcys princeps* (Plecoptera; Pteronarcyidae), respectively.

Chapter 4 describes the sequencing and characterization of the mitochondrial mRNA and rRNA pools from isolated mitochondria from adult *Drosophila melanogaster*. A discussion of models proposed to describe the expression and transcript processing of mt-RNAs based on these sequences is provided.

Chapter 5 discusses the comparative analyses of the available complete mitochondrial genome sequences, and what comparative analyses can tell us of the evolution of insect mitochondria. It also discusses the implication of the observed

mitochondrial transcripts on the annotation of insect mitochondrial genomes. Finally, a description of what the transcript analysis has added to our understanding of insect mitochondrial function, and a discussion of potential research to follow this study is included.

Chapter 2. The Complete Mitochondrial Genome Sequence of the Meadow Spittlebug *Philaenus spumarius* (Hemiptera: Auchenorrhyncha: Cercopoidae)[2]

2.1. Introduction

The order Hemiptera is the largest group of insects outside of the holometabolous orders. Three suborders are recognized within the Hemiptera: the Heteroptera (true bugs), Stremorrhyncha (aphids, scale bugs, whiteflies and psyllids) and the Auchenorrhyncha (planthoppers, leafhoppers, spittlebugs and cicadas) (Carver, *et al.*, 1991). The interrelationships of these three suborders have traditionally been controversial. At one time the Stremorrhyncha and the Auchenorrhyncha were considered to form the order Homoptera, with Heteroptera being classified as an independent order. More recently, the three suborders were united under the single order Hemiptera, as a result of morphological comparisons and nuclear DNA sequence data arguing against the monophyly of the groups constituting the Homoptera (see Kristensen (1991) and von Dohlen and Moran (1995)). A current phylogenetic view of the Hemiptera is summarized in Figure 2.1.

There has been a recent interest in mitochondrial genome (mt-genome) sequencing within the Paraneoptera, which includes the orders Hemiptera, Psocoptera, Phthiraptera and Thysanura. To date, the heteropteran *Triatoma dimidiata* (the kissing

---

[2] This chapter was published in the journal *Genome*, volume 48(1): pg. 46-54, February 2005, under the co-authorship of James Bruce Stewart and Andrew T. Beckenbach.

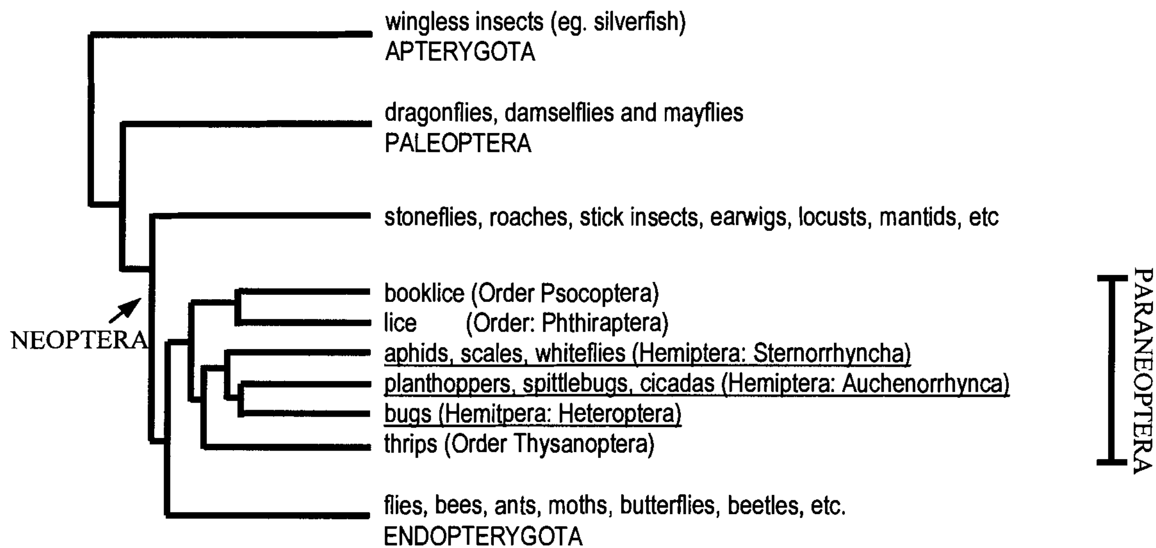


Figure 2.1. One view of evolutionary relationships of the Order Hemiptera. Adapted from (Kristensen, 1991) and (von Dohlen and Moran, 1995). Names in all capital letters are informal name designations of larger insect groups. Underlined names represent the three recognized groups within the Order Hemiptera.

bug), the phthirapteran *Heterodoxus macropus* (the wallaby louse), the thysanuran *Thrips imaginis* (the plague thrips,) and a psocopteran (a booklouse identified as lepidopsocid RS-2001) have their complete mitochondrial DNA sequences available in the sequence databases (Dotson and Beard, 2001; Shao, *et al.*, 2001; Shao and Barker, 2003; Shao, *et al.*, 2003). The thrips, louse and booklouse mt-genomes include the only known examples of rearrangements of relative position of protein coding genes and ribosomal RNA genes within the hexapods. This unusual trait has made this group an interesting one in the study of mitochondrial sequence evolution.

In this chapter I describe the mt-genome of the meadow spittlebug, *Philaenus spumarius*. The spittlebug is a member of the Auchenorrhyncha (family Cercopoidea), and represents an individual from a second suborder to represent the large order Hemiptera.

2.2. Materials and Methods

2.2.1. DNA extraction and PCR amplification

An adult specimen of *P. spumarius* (the meadow spittlebug) was collected in Coquitlam, British Columbia. The specimen was keyed using Hamilton (1982). The insect was quartered, with each quarter ground and digested in 100 µl protease buffer at 65°C (0.01M Tris pH 7.8, 5mM EDTA, 5% w/v SDS, 50ng/µl proteinase K). The digestion was extracted with 1 volume of Tris-buffer saturated phenol, then 1 volume of SEVAG (24:1 chloroform: isoamyl alcohol). DNA was precipitated over night in 2.5 volumes of 95% ethanol at -20°C, then pelleted through centrifugation. The pellets were washed once with 70% ethanol, and stored at -20°C. Prior to use, the pellets were

dissolved in 100 μl ddH_2O. One-tenth dilutions were utilized as template in PCR reactions.

The genome was amplified in overlapping PCR fragments (see Table 2.1 for list of primer pairs, and Figure 2.2 for a map of the amplification fragments). Initial amplifications were conducted using sets of heterologous primers we have developed, based on aligned insect and hexapod sequence. As more conserved genes were sequenced, primers were chosen based on obtained sequence, and occasionally were designed specifically from the spittlebug sequence.

PCR amplifications were carried out in an Eppendorf Mastercycler 5333 thermocycler. QIAGEN *Taq* DNA polymerase was utilized in 50 μl reactions, with magnesium concentration adjusted to 2.0 mM. The PCR cycling consisted of a 2 minute denaturation step at 94°C, followed by five 30 second cycles of denaturation at 94°C, annealing for 30 seconds at 45°C, then 120 second elongation at 72°C. Thirty additional cycles were carried out as above, with the annealing temperature increased to 50°C.

Amplification products for the region between the *nad2* and *srRNA* genes, including the major non-coding (or A+T rich) region, were produced using the GeneAmp XL PCR kit (Applied Biosystems), utilizing the reaction concentration suggested by the manufacturer, scaled to 50μl reactions. Cycling conditions were followed as per the product information, with a 60°C or 65°C anneal/elongation step for 8 minutes per cycle. For the primer pair TI-N18 and SR-J14766, lower annealing temperatures gave two fragments of 5.5 and 6.3 kb in size, which were digested with *HincII* (Gibco BRL). The fragments (2.0 kb, 3.5 kb and 2.5kb, 3.8 kb, respectively) were A-tailed with *Taq* polymerase in PCR reaction buffer containing only dATP at 72°C for 10 minutes, then

Table 2.1. List of Primer Pairs Utilized in the Amplification of the Mitochondrial Genome of *Philaenus spumarius*.

	Primer 1	Sequence (5'-3')	Primer 2	Sequence (5'-3')
1	TI-J34	GCCTGATAAAAAGGRTTAYYTTGATA	N2-N993	GGTAAAAATCCTAAAAATGGNGG
2	N2-J574	ATAGGAGCCGCTCCNTTYCA	C1-N1828	GTTCATCCTGTWCCWGCWCC
3	C1-J1751	GGAGCCCCTGATATAGCWTTYCC	C1-N2326	AATATATGATGAGCTCANAC
4	C1-J2222	TTAATTTTACCAGGATTYGG	C2-N3695	TACAATAGGTATAAATCT
5	TL-3034	TAATATGGCAGATTAGTGCA	A8-N4067	CTGAGAATAAGTTTGTTATCA
6	A8-J3927	TTTTATTCCTCAAATAGCWCC	A6-N4552	ATGTCCWGCAATYATATTWGC
7	A6-J4103	TCAAYATTTGAYCCNTCAAC	C3-N5346	ACATGTAGCCCATGYAANCC
8	C3-J4948	CGAGAAGGAACTTTTCARGG	N3-N5731	TAAGGGTCAAATCCRCAYTC
9	C3-5524	TTTCTATCTAYTGATGAGGTAG	N5-N6551 #	TTGAGGGGTATATTGAGGACC
10	TS-J6246#	GCTGCTAACTATATTTATAAGCGG	N5-N7208	GCTTTATTATTTATGTGYGCWGG
11	N5-J7077	AAATCCTTTGARTAAAANCC	N5-N7832	CTATAATATTTTTAATTATTAGTCC
12	N5-J7518	AAAGCAGANACWGGRGTWGG	N5-N8622	TCATGGTCTATGTTCWTCTGG
13	N5-J7806	ACTAATCCTAACCCATCTCA	N4-N8727	GCTTTAATTGCATAYTCNTC
14	N4-J8644 #	AAAGCAGANACWGGRGTWGG	N4-N9153	TGAGGTTATCAACCNGARCG
15	N4-J8941	ACAGGAGCTTCAACRTGNGC	N4L-9629	GTTTGTGAGGGWGYTTTRGG
16	N4-J9360	ATAACCATTAAGCATCTAATCC	N6-N10484 #	GTTTTTTATACGTAGAGGACCATC
17	TT-J9885	AAAAAAAACATTGGTCTTGTA	CB-N10917	CAAAATGATAATTGWCCTCANGG
18	CB-J10747	TGTCGAGATGTAAATTATGGNTG	CB-N11526	TTCAACTGGTCGRGCTCCAATYCA
19	CB-J11389#	CCCAACAAATTGGGGGGG	N1-N12067	AATCGWACTCCWTTTGATTTTGC
20	CB-J11858	TTATCATAACGAAANCGNGG	LR-N13003	TTACCTTAGGGATAACAGCG
21	LR-J12967	TTAATCCAACATCGAGGTCGCAA	SR-N14220	TTGTACACATGCCCCGTC
22	LR-J13900	TGATAAACCCTGATACAMAAGG	SR-N14588	CAAACCAGGATTAGATACCCTGTTAT
23	SR-J14247	GGCGATGTGTGCATAATTAAGAGC	DL-N #	GAAATAACCATTAAAGTGAATTTG
24	SR-J14766	TAACCGCAACTGCTGGCAC	TI-N18	TCCTATCAARRAAYCCTTT
25	SR-J14766	(listed above)	C1-N1828	(listed above)
26	DL-J #	AGATAAAAATATAATTATGGTTTC	N2-N558 #	TCATATATGGAATGGTGCGACTCC

Specific primer derived from *P. spumarius* sequence.

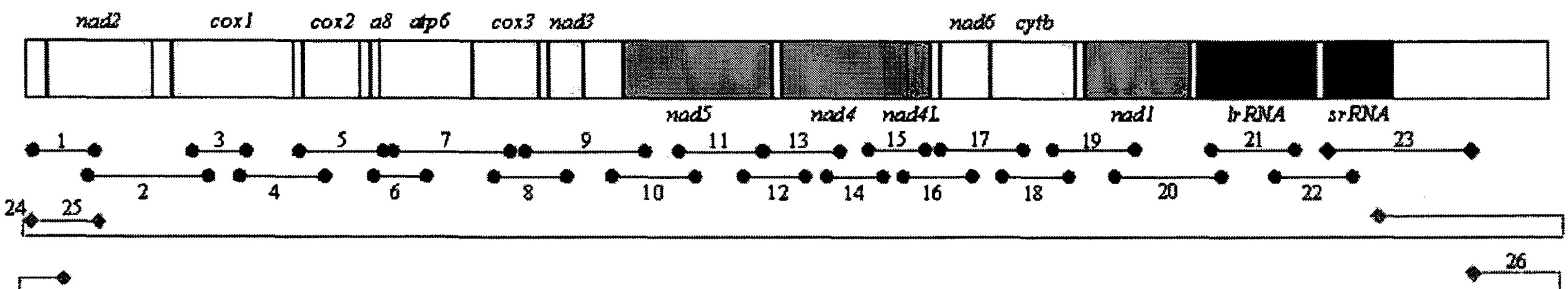


Figure 2.2. Schematic Representation of Amplification Strategy Employed for the *Philaenus spumarius* Mitochondrial Genome. Lines below the linearized genome map represent the amplification products. Numbers identify the primer pair, listed in Table 2.1. Black lines represent amplification products obtained with standard PCR protocols, while grey lines represent those obtained through long-PCR.

cloned using the Original TA Cloning® Kit (Invitrogen). The 65°C annealing resulted in a 2 kb fragment that was A-tailed as above and cloned using the Original TA Cloning® Kit. Multiple copies of all cloned fragments were independently sequenced to confirm their sequence identity.

Sequencing was conducted by the University of Calgary DNA Sequencing Center, using an ABI Prism model 377 sequencing machine and the ABI BigDye Terminators Version 3.0 or Version 3.1 sequencing kits.

2.2.2. Sequence Assembly, Annotation and Analysis

Sequences were aligned, and the genome assembled using BioEdit v. 5.0.9 (Hall, 1999). Nucleotide composition calculations were also conducted with BioEdit features. Protein coding genes and ribosomal RNA genes were identified through alignment to the other insect mt-genomes available in GenBank. The 22 tRNA genes were identified using the tRNAScan-SE server, or through alignment to other insect mt-tRNA genes (Lowe and Eddy, 1997) <http://www.genetics.wustl.edu/eddy/tRNAscan-SE/> . The $tRNA^{Ser\text{-}AGN}$ secondary structure was developed using constraints proposed by Steinberg *et al.* (1994). Relative synonymous codon usage (RSCU) was calculated using MEGA version 2.1 (Kumar, *et al.*, 2001), but the calculated 8-fold serine and 6-fold leucine families generated by the software were re-calculated as two 4-fold families and a 4-fold and 2-fold family, respectively.

Potential secondary structure folds of non-coding and A+T rich regions of the genome were calculated using the DNA mfold web server, using default settings and a folding temperature of 22°C (Zuker, 2003) <http://www.bioinfo.rpi.edu/applications/mfold/>. Dot-plot comparison of sequences was

conducted using the Windows version of DOTTER version 1.0 (Sonnhammer and Durbin, 1995). Repeat regions were identified using dot plots.

The genome *P. spumarius* sequence is available from GenBank under the accession number AY630340. Comparative analysis utilized the 20 hexapod mitochondrial genomes found in Table 2.2. These organisms were selected to cover the broadest phylogenetic range possible at the time, and limited each genus to a single representative. Near-complete genomes found in GenBank that were missing sequences for tRNA genes or portions of protein coding or ribosomal RNA genes were not utilized.

2.3. RESULTS AND DISCUSSION

2.3.1. Genome Content

The typical gene content for an animal mt-genome was observed for the spittlebug genome, with the genes arranged in the same order as that observed in *Drosophila yakuba* (Clary and Wolstenholme, 1985a). The majority-coding strand (J-strand) encodes the sense strand of 23 genes (*nad2, cox1, cox2, atp8, atp6, cox3, nad3, nad6, cob,* plus the tRNA genes $tRNA^{Ile}$, $tRNA^{Met}$, $tRNA^{Trp}$, $tRNA^{Leu\text{-CUN}}$, $tRNA^{Lys}$, $tRNA^{Asp}$, $tRNA^{Gly}$, $tRNA^{Ala}$, $tRNA^{Arg}$, $tRNA^{Asn}$, $tRNA^{Ser\text{-AGN}}$, $tRNA^{Glu}$, $tRNA^{Thr}$, $tRNA^{Ser\text{-TCN}}$). The minority-strand (N-strand) encodes the remaining 14 genes (*nad5, nad4, nad4L, nad1, lrRNA, srRNA,* plus the tRNAs $tRNA^{Gln}$, $tRNA^{Cys}$, $tRNA^{Tyr}$, $tRNA^{Phe}$, $tRNA^{His}$, $tRNA^{Pro}$, $tRNA^{Leu\text{-TTR}}$, $tRNA^{Val}$).

Table 2.2. Insect Mitochondrial Genomes Utilized in Alignments and Comparative Analyses.

Accession Number	RefSeq Number	Species	Common Name	Reference
Collembola: Arthropleona				
AY191995.1	NC_005438	Gomphiocephalus hodgsoni	springtail	(Nardi, *et al.*, 2003b)
AF272824.1	NC_002735	*Tetrodontophora bielanensis*	giant springtail	(Nardi, *et al.*, 2001)
Insecta: Coleoptera				
AF467886.1	NC_003372	*Crioceris duodecimpunctata*	asparagus beetle	(Stewart and Beckenbach, 2003)
AF452048.1	NC_003970	*Pyrocoelia rufa*	firefly	(Bae, *et al.*, 2004)
AJ312413.2	NC_003081	Tribolium castaneum	flour beetle	(Friedrich and Muqim, 2003)
Insects: Diptera				
L04272.1	NC_000875	*Anopheles quadrimaculatus*	mosquito	(Mitchell, *et al.*, 1993)
AY210702.1	NC_005333	*Bactrocera oleae*	olivefly	(Nardi, *et al.*, 2003a)
AJ242872.1	NC_000857	*Ceratitis capitata*	Mediterranean fruitfly	(Spanos, *et al.*, 2000)
AF352790.1	NC_002697	*Chrysomya putoria*	latrinefly	Junqueira *et al.*, 2003, unpublished
AF260826.1	NC_002660	*Cochliomyia hominivorax*	screwworm fly	(Lessinger, *et al.*, 2000)
X03240.1	NC_001322	*Drosophila yakuba*	-	(Clary and Wolstenholme, 1985a)
Insecta: Hemiptera				
AY630340.1	NC_005944	*Philaenus spumarius*	meadow spittlebug	This Study, (Stewart and Beckenbach, 2005)
AF301594.1	NC_002609	*Triatoma dimidiata*	kissing bug	(Dotson and Beard, 2001)
Insecta: Hymenoptera				
L06178.1	NC_001566	*Apis mellifera ligustica*	honeybee	(Crozier and Crozier, 1993)
Insecta: Lepidoptera				
AY242996.1	NC_004622	*Antheraea pernyi*	Chinese oak silkmoth	Lui *et al.*, 2003, unpublished
AB070263.1	NC_003395	*Bombyx mandarina*	wild silkmoth	(Yukuhiro, *et al.*, 2002)
Insecta: Orthoptera				
X80245.1	NC_001712	*Locusta migratoria*	migratory locust	(Flook, *et al.*, 1995)
Insecta: Phthiraptera				
AF270939.1	NC_002651	*Heterodoxus macropus*	wallaby louse	(Shao, *et al.*, 2001)
Insecta: Psocoptera				
AF335994.2	NC_004816	Lepidopsocid RS-2001	scaly-winged barklouse	(Shao, *et al.*, 2003)
Insecta: Thysanoptera				
AF335993.2	NC_004371	*Thrips imaginis*	plague thrips	(Shao and Barker, 2003)
Insecta: Thysanura				
AY191994.1	NC_005437	*Tricholepidion gertschi*	silverfish	(Nardi, *et al.*, 2003b)

The standard 22 tRNA genes were identified in the same relative genomic positions as observed for the *Drosophila yakuba* genome. The tRNAScane-SE was unable to detect $tRNA^{Arg}$ and $tRNA^{Ser\text{-AGN}}$ (settings: Organellar search, invertebrate mitochondrial code, cove cut-off score = 10). $tRNA^{Arg}$ and $tRNA^{Ser\text{-AGN}}$ were identified through alignment of other insect tRNA genes to unassigned nucleotide gaps corresponding to the conserved relative position of these missing tRNA genes. Predicted cloverleaf folds are given in Figure 2.3.

The A+T composition is very close to the mean observed for other insect mt-genomes, at 77.0% for the entire genome, and 76.7% for the coding sequences only (mean 0.759, standard deviation 0.037 for the 20 selected species).

At 16324 bp, the spittlebug mt-genome is the sixth largest complete insect mt-genome sequence reported to date, behind *Drosophila melanogaster*, the firefly, the kissing bug, the booklouse, and the honeybee (Crozier and Crozier, 1993; Lewis, *et al.*, 1995; Dotson and Beard, 2001; Shao, *et al.*, 2003; Bae, *et al.*, 2004). The majority of difference in the length of the insect mt-genomes observed to date has been in the size of the A+T rich region (the inferred control region) or the insertion of other large non-coding regions within the genome. The A+T rich region varies greatly in insects, from 4599bp in *Drosophila melanogaster* to 73bp in the louse *Heterodoxus* (Garesse, 1988; Shao, *et al.*, 2001). The control region of the spittlebug was observed to be 1834 bp in length, and includes two repeat regions. The first repeat region is situated to the *rrnS* gene-side of the A+T rich region. This repeat region consists of a 46 basepair repeat unit

A.

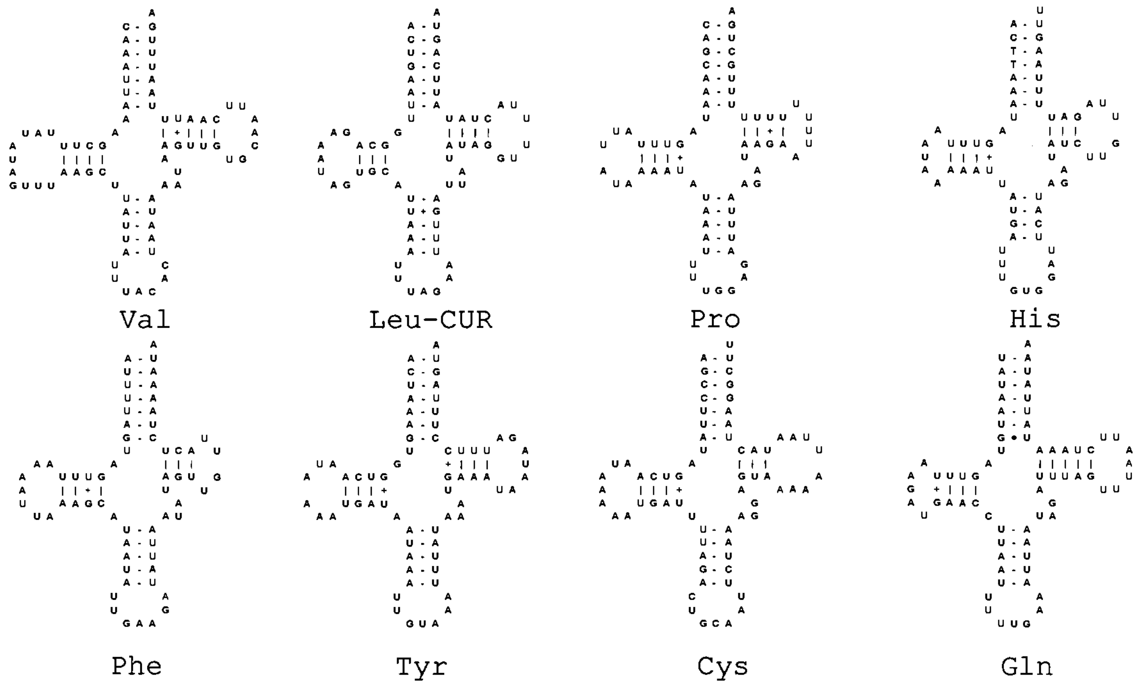
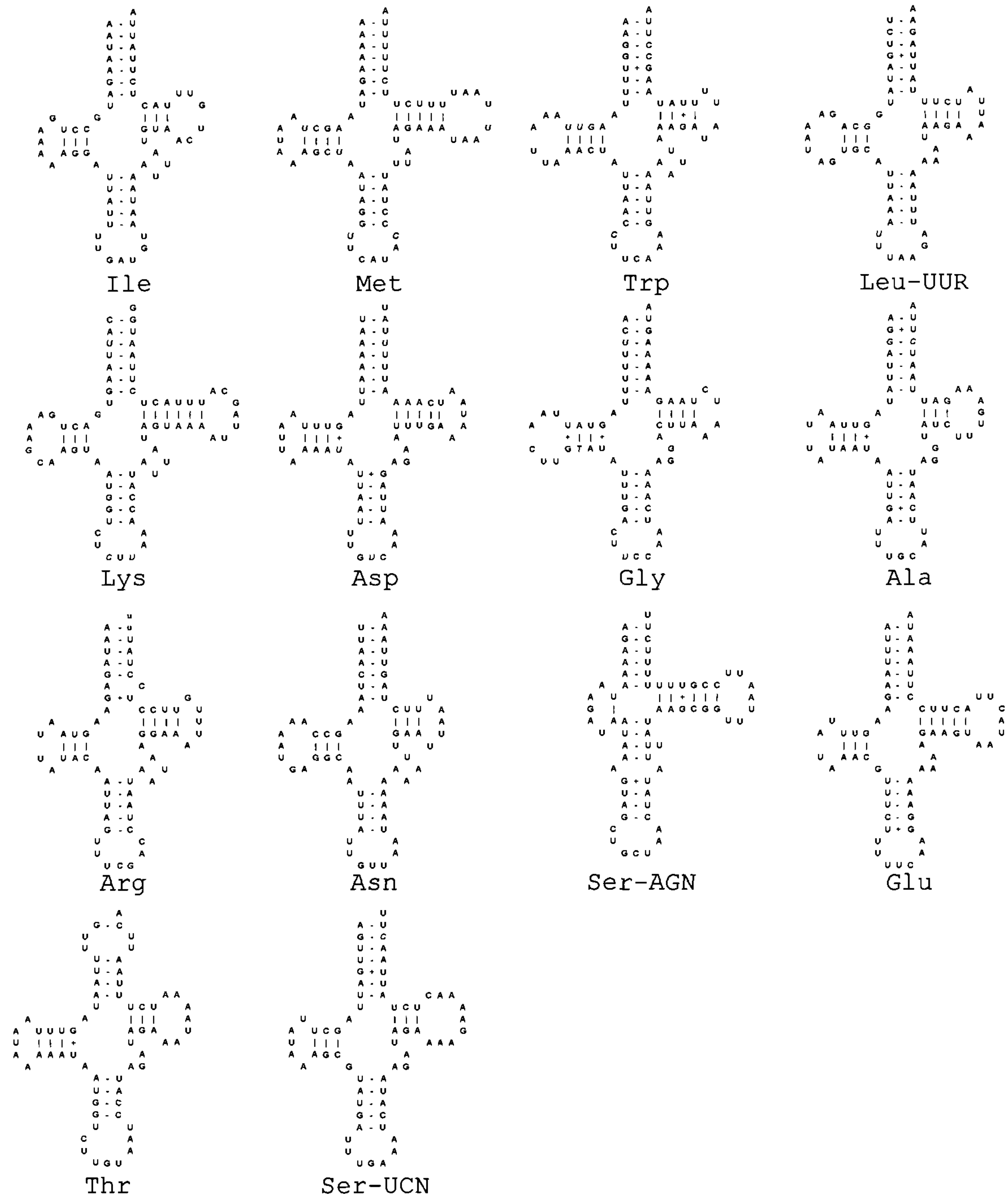


Figure 2.3. Putative secondary structure folds for the tRNA genes of the *Philaenus spumarius* mt-genome. Watson-Crick basepairs is designated by "–", and G-T basepairs, by "+". A. Minority coding strand tRNAs. B. Majority coding strand tRNAs.

(*continued on next page*)

(Figure 2.3, *concluded*)



repeated 19 times. Repeat unit six has a single A to C transversion at position 30, repeat 18 has a deletion of an A at position 4 and an insertion of a C at position 12. The terminal three repeats also have an additional A at the 3' end.

The nucleotide composition of this repeat region is 75.9% A+T, compared to 78.9% in the entire A+Trich region, and 81.9% for the A+T region excluding this repeat. Nucleotide compositions for various genomic features are summarized in Table 2.3.

The spittlebug A+T rich region was compared to the A+T rich region of the kissing bug using DOTTER (Sonnhammer and Durbin, 1995). No significant sequence similarity was identified for this region for the two hemipteran species.

Potential secondary structure folds were observed for this repeat region. Stem-loop structural units may consist of folds within a single repeat unit or between the eight 3' nucleotides of one repeat unit and the adjacent repeat unit (see Figure 2.4). Interestingly, all nucleotide substitutions and insertions noted in the repeat units occur in non-basepairing regions of these predicted secondary structures.

The A+T rich region also contains a second repeat sequence of 146 bp in length. This A+T rich repeat unit (79.5% A+T) occurs four times, with a fifth truncated repeat unit near the *tRNA*^{Ile} gene. A map of the region between the small ribosomal subunit (*srRNA*) and *nad2* genes is given in Figure 2.5. The largest open reading frame found within the A+T rich region was only 35 amino acids in length, and is assumed to have no functional significance.

Our long PCR reactions utilizing the primer pair TI-N18 and SR-J14766, spanning the control region also amplified two other fragments when lower annealing

temperatures were utilized in the annealing/elongation step of the Long-PCR cycling reaction. We do not believe these larger fragments represent mt-genome encoded copies, as two independently designed primers, specific for sequence contained only within the larger fragments were unable to produce amplification products when used in PCR with primers known to amplify within the *srRNA* sequence. Also, primers designed for the shorter amplification product (DL-J and DL-N, see Table 2.1) amplified fragments containing the same number of repeats.

2.3.2. Protein Coding Genes

Four different initiation codons appear to function in the spittlebug genome. This annotation includes six genes encoding ATG start codons (*cox1, atp6, cox3, nad4, nad4L* and *cob*), three with TTG (*nad2, cox2,* and *nad3*), two with ATA (*nad6* and *nad1*), and two with ATT (*atp8* and *nad5*). The TAG stop codon was not observed within this mt-genome. Only three protein coding genes (*atp6, nad5* and *nad1*) encoded a stop codon that did not overlap with the abutting gene. As is common to many insect mt-genomes sampled to date, the *atp8/atp6* and *nad4l/nad4* gene pairs appear to overlap seven nucleotides (ATGNTAA) in different reading frames, and a number of the protein coding genes have a single nucleotide overlap with the abutting genes. Four genes (*cox1, cox2, cox3* and *nad4*) appear to have their termination signals constructed through the addition of the poly-A tail to the mature mRNA (Ojala, *et al.*, 1980a; Ojala, *et al.*, 1981).

Table 2.3. Nucleotide Composition of Meadow Spittlebug (*P. spumarius*) Genome Features.

Feature	Proportion nucleotides					Number of
	A	C	G	T	A+T	Nucleotides
Whole Genome (J-strand)	0.409	0.124	0.106	0.361	0.770	16324
Protein Coding Genes – (coding nts) •	0.319	0.113	0.134	0.434	0.753	11148
- J-strand Encoded •	0.349	0.126	0.121	0.404	0.753	6802
- N-strand Encoded •	0.281	0.096	0.131	0.492	0.773	4219
- 1st Codon Position *	0.353	0.105	0.178	0.364	0.717	3710
- 2nd Codon Position •	0.197	0.176	0.153	0.474	0.671	3710
- 3rd Codon Position •	0.408	0.059	0.069	0.464	0.872	3710
tRNA genes – total	0.407	0.083	0.124	0.385	0.792	1454
- J-strand encoded	0.413	0.090	0.120	0.377	0.790	918
- N-strand encoded	0.397	0.071	0.132	0.399	0.797	536
rRNA Genes – (coding, N-strand)	0.315	0.082	0.133	0.470	0.785	1999
Intergenic Nucleotides (J-strand) &	0.429	0.036	0.036	0.500	0.929	28
Repeat Region 1	0.413	0.132	0.108	0.347	0.759	877
Repeat Region 2	0.330	0.084	0.121	0.466	0.795	704
A+T Rich Region	0.378	0.103	0.108	0.411	0.789	1835

Note: overlapping nucleotides are counted in each feature in which they appear.

* Excluding stop codons.

& Does not include A+T region or repeat regions.

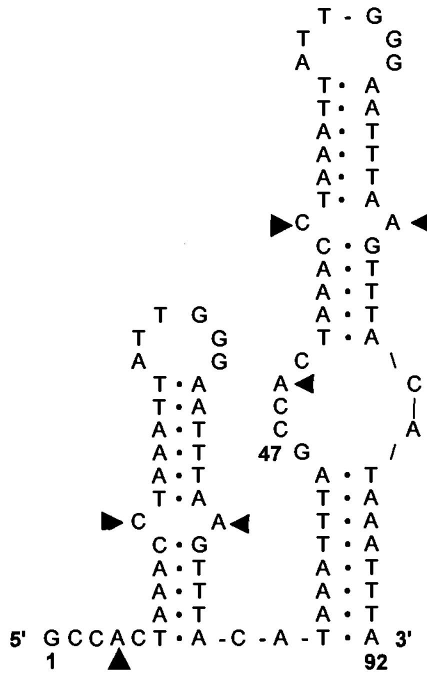


Figure 2.4. Putative stem-loop structures for two repeat units of Repeat Region 1. Arrows (▶) denote nucleotide positions were variations in repeat unit sequence were observed. Dots represent predicted basepair interactions. Secondary structure was predicted using DNA *mfold* Version 3.1 (Zuker, 2003) with folding temperature of 22°C.

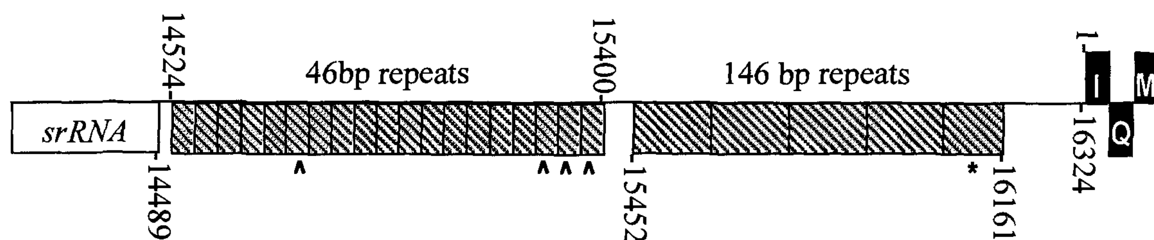


Figure 2.5. Map of the Region Between the *srRNA* and the *nad2* Genes. Rectangles above the central line represent J-strand encoded features, below are N-strand encoded features, and centered boxes have no directionality. The tRNA genes are labeled with their one-letter code amino acid code. Vertically printed numbers represent the genomic position of the vertical dash. The asterisk (*) represent the truncated repeat, while repeats containing nucleotide substitutions are designated by (^). Figure not drawn to scale.

The relative synonymous codon usage (RSCU) was calculated for all protein coding genes, and then for the major coding strand genes and minor coding strand genes separately (Table 2.4). Comparison of four-fold degenerate amino acid codon usage shows a clear bias towards an A in the third codon position for J-strand encoded genes, and a bias for the use of T in the third codon position for N-strand encoded genes. Only two of the four-fold degenerate amino acids do not clearly demonstrate this bias, ACN-serine (AGA preferred on both strands) and CTN-leucine (CTA and CTT equally utilized on N-strand).

2.3.3. Genomic Comparisons

One of the most distinctive features of the paraneopteran mitochondrial genomes, relative to other insect groups, is the extent of rearrangement of protein coding genes and the ribosomal RNA genes. The rearrangement of tRNA genes within insect mt-genomes is not uncommon, and has been discussed for the migratory locust (Flook, *et al.*, 1995), the honeybee (Crozier and Crozier, 1993), various hymenoptera (Dowton and Austin, 1999; Dowton, *et al.*, 2003), and the silkworm moths (Yukuhiro, *et al.*, 2002). Within the paraneoptera, representatives of three of the four orders display rearrangements of tRNAs and protein coding genes (Shao, *et al.*, 2001; Shao and Barker, 2003; Shao, *et al.*, 2003). The hemipterans (the kissing bug and the spittlebug), however, maintain the ancestral genome arrangement.

I compared nucleotide strand bias between the spittlebug and the kissing bug, with the locust sequence used as a non-hemipteran comparison (Table 2.4). The four-fold degenerate amino acids in all three insects demonstrate a clear bias towards third codon position A nucleotides in J-strand encoded protein coding genes, while the bias is towards

T in the third codon position on the N-strand. For two-fold degenerate amino acids, the locust and spittlebug show a bias towards the codons with A or T in the third codon position for both strands. The kissing bug, however shows a J-strand third codon position bias towards A or T nucleotides, but a bias towards G and T in the third codon position on the N-strand. Dotson and Beard (2001), in their description of the kissing bug genome noted a reduction in A+T ratio in that species. This unusual bias towards third codon position G nucleotides on the N-strand may help to explain this reduced A+T richness.

Examination of codon usage reveals that the cognate codon for each four-fold degenerate tRNA is more commonly utilized only on the J-strand of the mt-genomes examined. The bias towards the use of third codon position T on the N-strand outweighs the presumed advantage of improved efficiency during translation by the use of cognate codons by the mRNAs. For the three organisms analyzed, the two-fold degenerate tRNAs utilize the cognate codon only approximately a third of the time, on either strand, while six of the thirteen tRNAs never preferentially utilize the cognate codon.

2.4. Conclusions

The mt-genome of the spittlebug is relatively conservative in terms of gene organization and nucleotide composition. The genome organization is the same as observed in *Drosophila yakuba*, with the addition of a relatively A+T deficient repeat region located within the A+T rich region. This conservation of genome structure is consistent with the other mt-genome example observed within the Hemiptera, but varied from the other species sampled from the other paraneopteran orders. The nucleotide composition and patterns of nucleotide strand biases are more similar to those observed in other insects than the other hemipteran, the kissing bug.

Table 2.4. Relative Synonymous Codon Usage for Meadow Spittlebug, the Kissing Bug and the Migratory Locust Mitochondrial Genomes. RSCU values were calculated for all 13 genes, then major strand encoded only and minor strand encoded only genes. Values in bold type represent the most commonly utilized codon for the given amino acid. Underlined codons represent the cognate codon for each amino acid.

Amino Acid	Codon	Spittlebug			Kissing bug			Locust		
		All	J	N	All	J	N	All	J	N
Lys	AAA	**1.766**	**1.896**	**1.471**	**1.766**	**1.452**	0.435	**1.414**	**1.829**	0.414
	AAG	0.234	0.104	0.529	0.234	0.548	**1.565**	0.586	0.171	**1.586**
Asn	AAC	0.301	0.342	0.237	0.301	0.840	0.082	0.484	0.632	0.080
	AAT	**1.700**	**1.658**	**1.763**	**1.699**	**1.160**	**1.918**	**1.516**	**1.368**	**1.920**
Gln	CAA	**1.667**	**1.730**	**1.529**	**1.667**	**2.000**	0.714	**1.905**	**2.000**	**1.571**
	CAG	0.333	0.270	0.471	0.333	0	**1.286**	0.095	0	0.429
His	CAC	0.328	0.415	0	0.328	0.952	0	0.704	0.906	0.111
	CAT	**1.672**	**1.585**	**2.000**	**1.672**	**1.048**	**2.000**	**1.296**	**1.094**	**1.889**
Glu	GAA	**1.174**	**1.726**	**1.692**	**1.714**	**1.724**	0.759	**1.714**	**1.960**	**1.259**
	GAG	0.286	0.275	0.308	0.286	0.276	**1.241**	0.286	0.040	0.741
Asp	GAC	0.100	0.143	0	0.100	0.936	0.231	0.325	0.520	0
	GAT	**1.900**	**1.857**	**2.000**	**1.900**	**1.064**	**1.769**	**1.675**	**1.480**	**2.000**
Tyr	TAC	0.331	0.487	0.198	0.331	0.961	0	0.378	0.692	0.137
	TAT	**0.167**	**1.513**	**1.802**	**1.669**	**1.039**	**2.000**	**1.622**	**1.308**	**1.863**
Trp	TGA	**1.726**	**1.821**	**1.500**	**1.726**	**1.855**	0.839	**1.920**	**1.972**	**1.793**
	TGG	0.274	0.179	0.500	0.274	0.145	**1.161**	0.080	0.028	0.207
Cys	TGC	0.140	0	0.200	0.140	**1.273**	0	0.364	**1.000**	0.067
	TGT	**1.861**	**2.000**	**1.800**	**1.860**	0.727	**2.000**	**1.636**	**1.000**	**1.933**
Mey	ATA	**1.710**	**1.791**	**1.540**	**1.710**	**1.850**	0.989	**1.687**	**1.836**	**1.270**
	ATG	0.290	0.210	0.460	0.290	0.150	**1.011**	0.313	0.164	0.730
Ile	ATC	0.159	0.194	0.088	0.159	0.605	0.058	0.225	0.311	0.049
	ATT	**1.841**	**1.807**	**1.912**	**1.841**	**1.395**	**1.942**	**1.775**	**1.689**	**1.951**
Phe	TTC	0.164	0.200	0.122	0.164	0.946	0.071	0.515	0.938	0.125
	TTT	**1.836**	**1.800**	**1.878**	**1.836**	**1.054**	**1.929**	**1.485**	**1.062**	**1.875**
Leu	TTA	**1.728**	**1.777**	**1.672**	**1.728**	**1.890**	**1.059**	**1.757**	**1.966**	**1.584**
	TTG	0.272	0.223	0.328	0.287	0.110	0.941	0.243	0.034	0.419
Leu	CTA	**1.913**	**2.000**	**1.765**	**1.913**	**1.906**	0	**2.149**	**2.706**	0.375
	CTG	0.261	0.207	0.353	0.261	0.235	0	0.119	0.157	0
	CTC	0.174	0.207	0.118	0.174	0.729	0	0.149	0.157	0.125
	CTT	1.652	1.586	**1.765**	1.652	0.129	**4.000**	1.582	0.980	**3.500**

(Continued on next page)

(Table 2.4 *concluded*)

Amino Acid	Codon	Spittlebug All	J	N	Kissing bug All	J	N	Locust All	J	N
T	ACA	**1.961**	**2.300**	0.800	**1.961**	**1.784**	0.300	**2.556**	**3.066**	0.316
	ACG	0.077	0.100	0	0.077	0.081	0.200	0.059	0.072	0
	ACC	0.258	0.233	0.343	0.258	0.973	0.100	0.234	0.287	0
	ACT	1.703	1.367	**2.857**	1.703	1.162	**3.400**	1.151	0.575	**3.684**
P	CCA	1.536	**1.822**	0.800	1.536	**1.560**	0.242	**2.294**	**2.969**	0.615
	CCG	0.096	0.089	0.114	0.096	0.080	0.242	0.088	0.082	0.103
	CCC	0.320	0.311	0.343	0.320	0.800	0.121	0.118	0.165	0
	CCT	**2.048**	1.778	**2.743**	**2.048**	**1.560**	**3.394**	1.500	0.784	**3.282**
A	GCA	1.500	**2.000**	0.556	1.500	**1.857**	0.750	**2.107**	**2.781**	0.533
	GCG	0.039	0.059	0	0.038	0	0.333	0.053	0.038	0.089
	GCC	0.385	0.294	0.556	0.385	1.214	0.083	0.160	0.190	0.089
	GCT	**2.077**	1.647	**2.889**	**2.077**	0.929	**2.833**	1.680	0.990	**3.289**
S	TCA	1.770	**2.228**	1.117	1.770	**2.094**	0.336	1.897	**3.194**	0.548
	TCG	0.030	0.051	0	0.030	0.094	0	0.063	0.062	0.065
	TCC	0.312	0.329	0.288	0.312	0.719	0.112	0.111	0.124	0.097
	TCT	**1.889**	1.392	**2.595**	**1.888**	1.094	**3.551**	**1.929**	0.620	**3.290**
S	AGA	**2.688**	**3.015**	**2.333**	**2.688**	**3.139**	**2.143**	**2.945**	**3.491**	**2.400**
	AGG	0.128	0.000	0.267	0.128	0	0.643	0.073	0	0.145
	AGC	0.256	0.308	0.200	0.256	0.354	0	0.109	0.073	0.145
	ACT	0.928	0.677	1.200	0.928	0.506	1.214	0.873	0.436	1.309
R	CGA	**1.726**	**2.125**	1.053	**1.725**	**2.824**	1.636	**2.473**	**3.657**	0.400
	CGG	0.471	0.625	0.211	0.471	0.235	0.364	0.145	0.114	0.200
	CGC	0.157	0.125	0.211	0.157	0.118	0	0.073	0.114	0
	CGT	1.647	1.125	**2.526**	1.647	0.824	**2.000**	1.309	0.114	**3.400**
G	GGA	**1.757**	**2.045**	1.284	**1.757**	**3.030**	0.375	**2.019**	**3.079**	0.390
	GGG	0.785	0.812	0.741	0.785	0.273	0.750	0.192	0.159	0.244
	GGC	0.131	0.060	0.247	0.131	0.333	0.375	0.019	0.000	0.049
	GGT	1.327	1.083	**1.728**	1.327	0.364	**2.500**	1.769	0.762	**3.317**
V	GTA	1.653	**1.918**	1.275	1.653	**2.087**	0.566	1.749	**2.731**	0.456
	GTG	0.287	0.367	0.174	0.287	0.174	0.496	0.066	0.038	0.101
	GTC	0.240	0.327	0.116	0.240	0.661	0.071	0.109	0.154	0.051
	GTT	**1.820**	1.388	**2.435**	**1.820**	1.078	**2.867**	**2.077**	1.077	**3.392**
Number of Codons		3670	2260	1410	3668	2271	1417	3725	2296	1429

Chapter 3. The Complete Mitochondrial Genome Sequence of a Giant Stonefly, *Pteronarcys princeps* (Plecoptera; Pteronarcyidae)

3.1. Introduction

Mitochondrial DNA sequence has a long history of use in molecular studies of populations (Wrischnik, *et al.*, 1987) and molecular phylogenetic analyses between taxonomic groups (Kocher, *et al.*, 1989). With improvements in amplification, isolation and sequencing technology, groups studying phylogenetic relationships of more deeply diverging taxa have begun to use complete mt-DNA data sets for their analyses (for examples and discussion see Curole and Kocher, 1999; Inoue, *et al.*, 2001; Delsuc, *et al.*, 2003; Friedrich and Muqim, 2003; Nardi, *et al.*, 2003b; Stewart and Beckenbach, 2003; Cameron, *et al.*, 2004; Helfenbein, *et al.*, 2004; Papillon, *et al.*, 2004; Scouras, *et al.*, 2004; Lavrov, *et al.*, 2005 and others).

Complete mitochondrial (mt-) genome sequencing of insects and other hexapods has recently received renewed attention from a number of research groups. The first seven insect sequences were released between 1985 (Clary and Wolstenholme, 1985a) and 1999 (Lee, *et al.* unpublished, Accession AF149768). Since 2000, 38 new insect species have had complete or near-complete mt-genome sequences released. Of the thirty recognized insect orders (Kristensen, 1991; Klass, *et al.*, 2002), complete mt-genome sequences are available in GenBank for eleven orders. Near-complete genome sequences (those missing some tRNAs, small sections of *srRNA* and / or *nad2*, and all or

part of the A+T rich region) increase the number of orders represented by mt-DNA sequence by two (Odonata and Blattodea). This sampling still leaves seventeen insect orders not represented in the mitochondrial sequence databases.

The subsequent use of the resulting "mitogenomic" data sets is only one motivation to obtain complete mt-genome sequence. Interesting molecular evolutionary phenomenon are also being reported in insect mt-genomes. Contrary to the generally stable gene order and arrangement observed in most insect mt-genomes, it is becoming apparent that some insect lineages appear to have evolved a predisposition to major genome rearrangements. One such group is the Paraneoptera, where many different patterns of gene re-arrangement have been observed, (Shao, *et al.*, 2001; Shao and Barker, 2003; Shao, *et al.*, 2003; Thao, *et al.*, 2004), including the apparent evolution of up to four independent rearrangements within the whitefly family (Hemiptera, Sternorrhyncha, Aleyrodidae). A similar phenomenon has been discussed in detail regarding tRNA clusters within the Hymenoptera (Dowton and Austin, 1999; Dowton, *et al.*, 2003).

Obtaining sequence in the larger context of the whole or majority of the mt-genome has also allowed for the identification of unusual coding of functional genes that may otherwise have been interpreted as the amplification of a nuclear pseudogene copy if only partial gene sequences been amplified and sequenced. One such example is the report of programmed translational framshifting in the mitochondrial *cytb* gene of an ant genus (Beckenbach, *et al.*, 2005).

An important grouping of orders referred to as the Neoptera *incertae sedis* by Kristensen (1991), or as the "basic" or "lower" Neoptera have very poor representation by

complete mt-DNA sequence (See Chapter 1, Figure 1.2). At present, the complete mt-genome sequence of the migratory locust, *Locusta migratoria* (Flook, *et al.*, 1995) and the oriental mole cricket, *Gryllotalpa orientalis* (Kim, *et al.*, 2005) represent the single order Orthoptera. The near-complete mt-DNA sequence of a cockroach, *Periplaneta fuliginosa* (Order Blattodea) increases the representation to only two of the eleven Neoptera *incertae sedis* orders. Given the ancient origin and unknown inter-order relationships of most of these groups, it is unusual that these order have been neglected this long.

The stoneflies (order Plecoptera) are an ancient group of insects. Their fossil record extends into the Lower Permian (285 to 240 million years before present), where five to six distinct families can be described, implying an earlier origin for the group (reviewed by Wootton, 1981). Some of the competing evolutionary hypotheses place the stoneflies as a single basal group, diverging early after the split between the Paleoptera and the Neopteran ancestors, and before other Neopteran diversification (see Figure 3.1).

In this work, I present the complete mitochondrial genome sequence of a giant stonefly, *Pteronarcys princeps* (Plecoptera; Pteronarcyidae). This sequence provides the first complete mt-genome from a plecopteran.

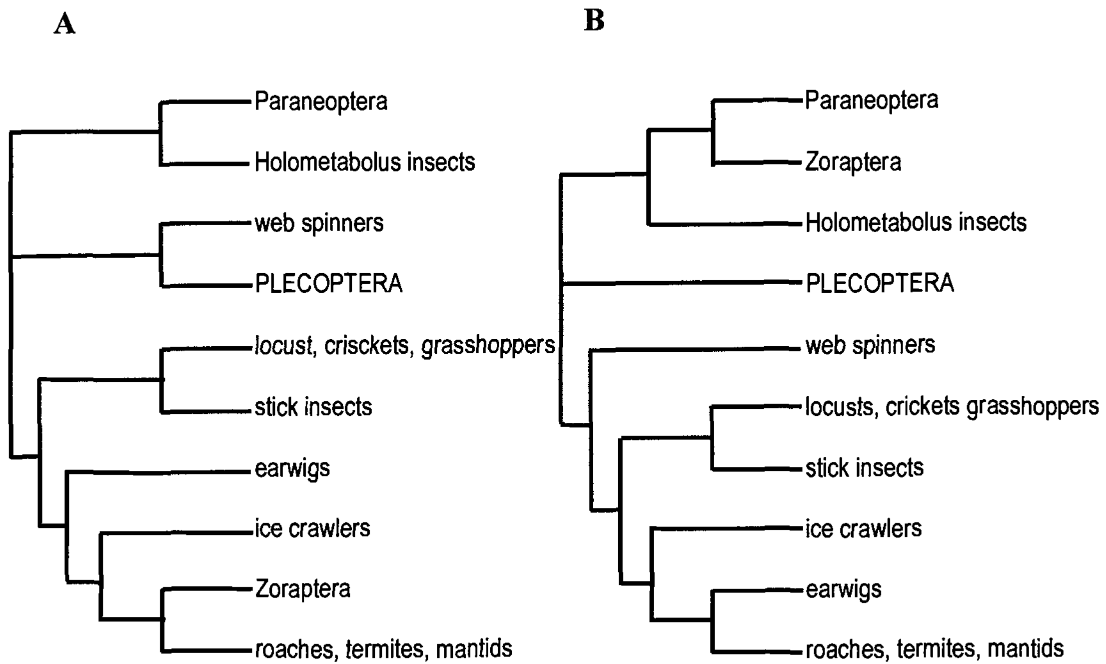


Figure 3.1. Differing Phylogenetic Views of the Placement of Plecoptera within the Insects. A - Summary of relationships presented in (Boudreaux, 1979). B – Summary of relationships presented in (Hennig, 1981).

3.2. Methods

3.2.1. Specimen Collection and Identification

A giant stonefly was collected from the foliage surrounding Stoney Creek, in Burnaby, British Columbia, Canada on April 22, 2002. The specimen was identified as an adult male of the species *Pteronarcys princeps* (Plecoptera; Pteronarcyidae) using the key by Baumann *et al.* (1977). After the specimen was killed by exposure of ethyl acetate vapours, it was pinned and allowed to dry prior to DNA extraction.

3.2.2. DNA isolation

Two legs were removed from the specimen and each was ground and digested separately in 100 μl protease buffer (0.01M Tris pH 7.8, 5mM EDTA, 0.5%SDS, 50ng/μl proteinase K) for 10 minutes at 60°C. The solution was extracted with 1 volume of Tris-buffer saturated phenol (pH 7.6), then with 1 volume chloroform: isoamyl alcohol (24:1). The aqueous layer was again removed to a new Beckman ultrafuge 1.5ml Eppendorf tube, and the volume measured. Tubes were balanced to volume by the addition of 10mM Tris. A volume of cold 95% ethanol 2.1 times the adjusted aqueous layer volume was added. The solution was mixed by hand agitation, and left to precipitate overnight at -20°C. The DNA was pelleted by centrifugation in a Beckman desktop ultra-centrifuge with a TLA-45 rotor for 30 minutes at 40,000 rpm. The pellet was washed with cold 70% ethanol and centrifuged in a desktop Eppendorf centrifuge at maximum speed for 10 minutes. The pellet was allowed to air dry on the bench. The pellet was finally dissolved in 100μl of ddH_2O. 1/10 were dilutions used in PCR reactions.

3.2.3. PCR and Long-PCR Amplification

The genome, except for the A+T rich region, was amplified in overlapping PCR fragments, using QIAGEN *Taq* DNA polymerase (QIAGEN Inc.). Heterologous primers used were designed using aligned insect and hexapod mt-genome sequences. At later stages in the sequencing effort, primers designed from the obtained stonefly sequence were used in PCR and sequencing reactions. Primer pairs are listed in Table 3.1. PCR amplifications were carried out in an Eppendorf Mastercycler 5333 thermocycler. QIAGEN® *Taq* DNA polymerase was utilized in 50µl reactions with $MgCl_2$ concentrations adjusted to 2.0 mM, 0.2 mM of each dNTP, 400 nM of each primer and 1.25 U of *Taq* polymerase in a 1/10 dilution of the supplied reaction buffer.

The PCR cycling consisted of a 2 minute denaturation step at 94°C, followed by five 30 second cycles of denaturation at 94°C, annealing for 30 seconds at 45°C, then 120 second elongation at 72°C. Thirty additional cycles were carried out as above, with the annealing temperature increased to 50°C.

Some primer pairs required additional optimization, normally an adjustment of annealing temperatures upward to 50 - 55°C. In these cases, cycling was conducted as described above, but with 35 cycles, with the same annealing temperature for all cycles.

PCR reactions were visualized by agarose gel electrophoresis on a 0.8% agarose gel with 0.5X TAE (20mM Tris-HCl, 5mM EDTA, 10mM acetic acid) as buffer. Four µl of each of the reactions was run against 200ng of 100 bp Ladder (Invitrogen™ Life Technologies) as a size and concentration standard.

Table 3.1. Primers Used in the Amplification of the Mitochondrial Genome of the Giant Stonefly *Pteronarcys princeps*.

	Forward Primer&	Sequence (5'-3')	Reverse Primer&	Sequence (5'-3')
1	TI-J34	GCCTGATAAAAAGGRTTAYYTTGATA	N2-N993	CCATTTTTAGGATTTTTACCNCC
2	TM-J210	AATTAAGCTACTAGGTTCATACCC	C1-N1859	GGAGGCTAAACWGTTCANCCTGTNCC
3	N2-J574	ATAGGAGCCGCTCCNTTYCA	C1-N1828	CAAGTAGGACAWGGWCGWGG
4	C1-J1709	AATTGGWGGWTTYGGAAAYTG	C1-N2353	CGAGCACATAGTTGCAGATAWGG
5	C1-J1709	(see above))	C1-N2776	ATTAGACTTATTGCAGCNCC
6	C1-J2222	TTAATTTTACCAGGATTYGG	C2-N3389b	ATGAGTATYCTAGTYATAGTRAC
7	TL-J3034	TAATATGGCAGATTAGTGCA	A8-N4067	GACTCTTATTCAAACAATAGT
8	TK-J3790	CATTAGATGACTGAAAGCAAGTA	C3-N4911	AAGAGCACTTTATAGAGCAGTRGT
9	A8-J3927	TTTTATTCCTCAAATAGCWCC	C3-N4911	(see above)
10	C3-J4792	GTTGATTATAGACCWTGRCC	N3-N5731	AATCCCAGTTTAGGYGTRAG
11	TG-J5584	AGTATATTTGACTTCCAATC	N5-N6551$_{SF}$	CCTACTTCACTCATAACACCTC
12	TN-J6172	AGAGGTATATCACTGTTAATGA	N5-N7211	GTTCCGAAAWAAYAAATAYACRCG
13	N5-J7077	AAATCCTTTGARTAAAANCC	N5-N7793	AATCCAACYCTACCNRATCC
14	N5-J7077	(see above)	N5-N8727	CGAAATTAACGTATRAGNAG
15	N5-J7714$_{SF}$	GTAAAGCAACATCTCCAATACGA	N4-N9154	ACTCCAATAGTTGGNCTYGC
16	N4-J8944	GGAGCTTCAACATGAGCTTT	N6-N10035$_{SF}$	TCGATTGCCTACTAAATAAACTC
17	N4-J8944	(see above)	CB-N10198	TATAAAGTAAACACCGTACAG
18	TT-J9885	AAAAAAAACATTGGTCTTGTA	CB-N10917	GTTTTACTATTAACWGGAGTNCC
19	TT-J9885	(see above)	N1-N12067	TTAGCWTGAGGWAAACTAAAACG
20	CB-J10933	GTTCTACCTTGAGGNCAAATRTC	N1-N12067	(see above)
21	CB-J11335	CATATTCAACCWGAATGRTA	N1-N12560	TAAGTTTAAGCATTYCCNGG
22	CB-J11335	(see above)	LR-N12886	TGTACTAGACTCAAGTYTGGCC
23	N1-J12260	AATYTCATAAGAAATAGTTTGWGC	LR-N13003	AATGGAATCCCTATTGTCGC
24	LR-J12967	TTAATCCAACATCGAGGTCGCAA	SR-N14220	AACATGTGTAGCGGGCAG
25	LR-J13342	CCTTCGCACRGTCAAAATACYGC	SR-N14744b	CACGGTCGTNYRCGCCAATNTG
26	LR-J13900	TTTGATAAACYCTGATACAMAAG	SR-N14588	TTTGATCCTAATCTATGGGATAATA
27	SR-J14247*	CGATGTGTGCATATTTYAGAGC	DL-N15188$_{SF}$*	CAATAAATAAATTATTTACCCCG
28	SR-J14618$_{SF}$*	CTTTAAATTTCACAGCTTCAACAACC	TM-N193$_{SF}$*	CGATTACCCAAGTATGGGGTAAAT
29	DL-J15121$_{SF}$*	GACAATAAATTTACCTAACCAAC	N2-N313$_{SF}$*	CGTGGTTTAGTACTTAACCGAC

*Long-PCR reaction utilized with this primer pair.

&Primer names followed by $_{SF}$ were designed specifically from the obtained stonefly sequence.

The A+T rich region of the genomes was amplified using the GeneAmp® XL PCR Kit (Applied Biosystems). Specific, high annealing primers were designed for use with the XL PCR reaction (see Table 2.1). DNA template concentrations were 10 times as concentrated as that utilized in standard PCR reactions. Reaction mixture concentrations were as suggested by the supplier, but scaled to 50μl. Final concentrations for the reaction mixture were 0.9mM dNTP mix, 1.25mM magnesium acetate, 0.25 μM of each primer in 1-x supplied reaction buffer.

Long-PCR Cycling was conducted in an Eppendorf Mastercycler 5333 thermocycler, and consisted of a one-minute 94°C incubation of the template and the primer master mixes, before they were combined, then a 94°C denaturation for 90s after the combination of the two master mixes. The first 16 cycles consisted of a denaturation step of 90°C for 20s, followed by an anneal/elongation step at 62°C for 6 minutes. The final 23 cycles were similar, but with an addition of 15 seconds per cycle to the anneal/elongation time.

3.2.4. Purification of PCR Products

Successful amplification products (single bands of approximately expected size) were isolated using the QIAquick® PCR Purification kit (QIAGEN®) as directed by the supplier, with minor alterations to the protocol. This purification kit was designed to remove excess nucleotides and primers from the PCR reaction, so that the products could be sequenced.

Briefly, the PCR reaction was mixed with five volumes of the proprietary chaotropic salt mixture supplied with the kit. The PCR products were bound to a membrane within the column by centrifuging the solution through the column. Next, 0.75ml of an ethanol

containing "wash buffer" was passed over the membrane by centrifugation. The column was then dried by centrifugation for 1 minute with no solutions passing through the membrane. The membranes was then allowed to thoroughly dry by leaving it on the bench top for at least 1 hour. The DNA was eluted by the addition of 40μl of either ddH_2O or a 1/10 dilution of the supplied elution buffer, then centrifuging the eluted DNA into a fresh 1.5mL Eppendorf tube.

If optimization of the PCR protocol was not able to eliminate secondary bands from the PCR reaction, the band of the expected size was separated by agarose-gel electrophoresis. The target band was then cut from the gel and transferred to a 1.5ml Eppendorf tube. The gel slice was allowed to freeze at -20°C overnight. The next morning, the gel slice was subjected to centrifugation at top speed in a desktop centrifuge for 20 minutes. The resulting liquid was pipetted into a fresh tube and saved.

The remaining agarose in the original tube was then extracted using the QIAquick® Gel Extraction kit (QIAGEN®). The agarose was dissolved in three volumes of the supplied PB buffer, and dissolved at 50°C for at least 10 minutes. This solution was then passed over a column through centrifugation, then washed with 0.75ml of the supplied ethanol containing wash buffer. The column was then dried by centrifugation for 1 minute with no solutions passing through the membranes, to speed drying. The column was then allowed to sit on the bench top for at least one hour to dry. The liquid recovered after centrifugation was treated as a PCR reaction and purified using the same column as was used above, in the agarose extraction.

3.2.5. DNA Sequencing

Sequencing was conducted by the University of Calgary DNA Sequencing Center. The sequencing center uses an ABI Prism model 377 sequencing machine, with the BigDye Terminators Version 3.0 or Version 3.1 sequencing kits (Applied Biosystems). Pre-mixed primer (3.2pmol) and template (100ng / 1 kb template length) were diluted into 12μl total volume, then mailed to the sequencing center. The sequencing center supplied the ABI trace files generated by the sequencer, text files of each individual sequence read, and a .pdf file of all sequencing runs.

3.2.6. Sequence Assembly, Annotation and Analysis

Sequence alignments, assembly of PCR fragments, and nucleotide sequence analysis was conducted using BioEdit (version 7.0.1, August 17, 2004) (Hall, 1999). Relative Synonymous Codon Usage calculations (RCSU) were conducted using MEGA2 (version 2.1, 2001) (Kumar, *et al.*, 2001). Alignment of amino acid sequences, as well as rRNA sequences and A+T region sequences was carried out using CLUSTAL W (version 1.81) (Thompson, *et al.*, 1994). Clustal aligned sequences often required manipulation by eye after the completion of the alignment.

Dot plots were conducted to search for repeats and inverted repeats in sequence analyses, especially in the major non-coding region (also called the A+T rich region) of mitochondrial genomes. Dot plots were conducted using DOTTER for windows (October 1999) (Sonnhammer and Durbin, 1995).

Searches for tRNA genes were conducted using tRNAScan-SE (version 1.23, April 2002) (Lowe and Eddy, 1997), installed on a Sun Microsystems UNIX machine running SunOS 5.8. Searches were constrained to organellar tRNA's only, using the

invertebrate mitochondrial genetic code, with a COVE cut-off score of 5. For searches on small sections of DNA sequence the tRNAScan-SE Search Server Version 1.21 was used online at (http://www.genetics.wustl.edu/eddy/tRNAscan-SE/).

Sequence statistics were calculated using features of BioEdit, while relative synonymous codon usage (RSCU) values were calculated using MEGA 2.1 (Kumar, *et al.*, 2001). Serine TCN and AGN codons were re-calculated, by hand, as two 4-fold degenerate codon families (instead of the 8-fold family computed by MEGA), since two simultaneous non-synonymous nucleotides substitutions in first and second codon positions would be required for the synonymous movement between the two serine codon families. The leucine CUN and UUR codons were also separated into a four-fold and a two-fold degenerate family, as the UUA codon was so strongly over-represented that the six-fold codon RSCU obscured the contribution of the CUN codons.

Codon usage for A-T and G-C skew calculation was calculated using CodonW (version 1.4.2 May 2005, available at http://codonw.sourceforge.net/) (Peden 1999).

3.2.7. Sequence Comparisons

Primer design, gene annotation and comparative analyses relied on the alignment of other complete mitochondrial genomes from other hexapods that were deposited in GenBank. Table 3.2 lists the genomes, released in GenBank prior to April 22, 2005, that were utilized in this study.

A number of independent species groups within the insects appear to be undergoing strongly accelerated rates of evolution, accompanied with gene-order rearrangements. These groups include the social bees and wasps (Crozier and Crozier, 1993; Foster, *et al.*, 1997; Foster and Hickey, 1999; Castro and Dowton, 2005), and some of the

paraneopteran lineages (Shao, *et al.*, 2001; Shao and Barker, 2003; Shao, *et al.*, 2003; Thao, *et al.*, 2004). Use of various phylogenetic reconstruction techniques to estimate insect phylogenies identify these species as "long-branch attracting" taxa (Hendy and Penny, 1989). Given the increase rate of sequence evolution, primer design often did not consider sequence variation from the remaining insect alignment observed for the sequences of *Apis*, *Melipona, Neomaskellia, Pachypsylla, Schizaphis, Trialeurodes, Heterodoxus,* Lepidopsocid and *Thrips.*

The A+T rich region of two stoneflies from genus *Peltoperla* (family Peltoperlidae) were obtained from Genbank (Accessions AY142073 and AY142074). The giant stoneflies (family Pteronarcyidea) are considered a sister group to the Peltoperlidae+family Styloperlidae (Thomas, *et al.*, 2000; Zwick, 2000).

The complete mt-genome sequence was submitted to GenBank, under accession number AY687866 (with associated RefSeq NC_006133).

Table 3.2. Complete and Near-Complete Hexapod Mitochondrial Genome Sequences Utilized in this Study, with their GenBank Accession Numbers.

Accession Number	RefSeq Number	Species	Common Name	Reference
Collembola: Arthropleona				
AY191995.1	NC_005438	*Gomphiocephalus hodgsoni*	springtail	(Nardi, *et al.*, 2003b)
AF272824.1	NC_002735	*Tetrodontophora bielanensis*	giant springtail	(Nardi, *et al.*, 2001)
Insecta: Archaeognatha				
AY793551.1	NC_006895	*Nesomachilis australica*	jumping bristletail	(Cameron, *et al.*, 2004)
Insecta: Blattaria				
AB126004.1	NC_006076	*Periplaneta fuliginosa*	smokybrown cockroach	(Yamauchi, *et al.*, 2004)
Insecta: Coleoptera				
AF467886.1	NC_003372	*Crioceris duodecimpunctata*	asparagus beetle	(Stewart and Beckenbach, 2003)
AF452048.1	NC_003970	*Pyrocoelia rufa*	firefly	(Bae, *et al.*, 2004)
AJ312413.2	NC_003081	*Tribolium castaneum*	flour beetle	(Friedrich and Muqim, 2003)
Insecta: Diptera				
AY072044.1	NC_006817	*Aedes albopictus*	Asian tiger mosquito	Ho *et al.*, 2004, unpublished
L20934.1	NC_002084	*Anopheles gambiae*	mosquito	(Beard, *et al.*, 1993)
L04272.1	NC_000875	*Anopheles quadrimaculatus*	mosquito	(Mitchell, *et al.*, 1993)
AY210702.1	NC_005333	*Bactrocera oleae*	olivefly	(Nardi, *et al.*, 2003a)
AJ242872.1	NC_000857	*Ceratitis capitata*	Mediterranean fruitfly	(Spanos, *et al.*, 2000)
AF352790.1	NC_002697	*Chrysomya putoria*	latrinefly	(Junqueira, *et al.*, 2004)
AF260826.1	NC_002660	*Cochliomyia hominivorax*	screwworm fly	(Lessinger, *et al.*, 2000)
AY463155.1	NC_006378	*Dermatobia hominis*	human botfly	unpublished
AF200830.1	NC_005779	*Drosophila mauritiana*	-	(Ballard, 2000b)
U37541.1	NC_001709	*Drosophila melanogaster*	-	(Lewis, *et al.*, 1995)
AF200828.1	-	*Drosophila melanogaser**	Oregon R strain	(Ballard, 2000b)
AF200832.1	NC_05780	*Drosophila sechellia**	-	(Ballard, 2000b)
AF200833.1	NC_05781	*Drosophila simulans**	-	(Ballard, 2000a)
X03240.1	NC_001322	*Drosophila yakuba*	-	(Clary and Wolstenholme, 1985a)
DQ029097.1	NC_007102	*Haematobia irritans irritans*	horn fly	Lessinger *et al.*, 2005, unpublished
Insecta: Hemiptera				
AY572538.1	NC_006160	*Aleurochiton aceris*	whitefly	(Thao, *et al.*, 2004)
AY521251.1	NC_005939	*Aleurodicus dugesii*	Doogie Howzer whitefly	(Thao, *et al.*, 2004)
AY521259.2	NC_006279	*Bemisia tabaci*	sweet potato whitefly	(Thao, *et al.*, 2004)
AY875213.1	NC_006899	*Homalodisca coagulata*	glassy-winged sharpshooter	Baumann and Baumann, 2005, unpublished
AY572539.1	NC_006159	*Neomaskellia andropogonis*	sugarcane whitefly	(Thao, *et al.*, 2004)
AY278317.1	NC_006157	*Pachypsylla venusta*	psylid	(Thao, *et al.*, 2004)
AY630340.1	NC_005944	*Philaenus spumarius*	meadow spittlebug	(Stewart and Beckenbach, 2005)

(Continued next page)

(**Table 3.2** ***concluded***)

Accession Number	RefSeq Number	Species	Common Name	Reference
AY531391.1	NC_006158	*Schizaphis graminum*	greenbug (aphid)	(Thao, *et al.*, 2004)
AY521262.2	NC_06292	*Tetraleurodes acaciae*	acacia whitefly	(Thao, *et al.*, 2004)
AY521265.2	NC_006280	*Trialeurodes vaporariorum*	greenhouse whitefly	(Thao, *et al.*, 2004)
AF301594.1	NC_002609	*Triatoma dimidiata*	kissing bug	(Dotson and Beard, 2001)
Insecta: Hymenoptera				
L06178.1	NC_001566	*Apis mellifera ligustica*	honeybee	(Crozier and Crozier, 1993)
AF466146.1	NC_004529	*Melipona bicolor**	stingless bee	Silvestre,D and Arias, 2003, unpublished
AY787816		*Perga condei**	spitfire grub (sawfly)	(Castro and Dowton, 2005)
Insecta: Lepidoptera				
AY242996.1	NC_004622	*Antheraea pernyi*	Chinese oak silkmoth	Lui *et al.*, 2003, unpublished
AB070263.1	NC_003395	*Bombyx mandarina*	wild silkmoth	(Yukuhiro, *et al.*, 2002)
AF149768.1	NC_002355	*Bombyx mori*	domestic silkmoth	Lee *et al.*, 2000, unpublished
AF467260.1	NC_003368	*Ostrinia furnacalis**	Asian cornborer	(Coates, *et al.*, 2005)
AF442957.1	NC_003367	*Ostrinia nubilalis**	European cornborer	(Coates, *et al.*, 2005)
Insecta: Odonata				
AB126005.1	-	*Orthetrum triangulare melania**	dragonfly	(Yamauchi, *et al.*, 2004)
Insecta: Orthoptera				
AY660929.1	NC_006678	*Gryllotalpa orientalis**	Oriental mole cricket	(Kim, *et al.*, 2005)
X80245.1	NC_001712	*Locusta migratoria*	migratory locust	(Flook, *et al.*, 1995)
Insecta: Phthiraptera				
AF270939.1	NC_002651	*Heterodoxus macropus*	wallaby louse	(Shao, *et al.*, 2001)
Insecta: Plecoptera				
AY687866.1	NC_006133	*Pteronarcys princeps*	ebony salmonfly	This study
Insecta: Psocoptera				
AF335994.2	NC_004816	Lepidopsocid RS-2001	scaly-winged barklouse	(Shao, *et al.*, 2003)
Insecta: Thysanoptera				
AF335993.2	NC_004371	*Thrips imaginis*	plague thrips	(Shao and Barker, 2003)
Insecta: Thysanura				
AY639935.1	NC_006080	*Thermobia domestica*	firebrat	Cook and Akam, 2004, unpublished
AY191994.1	NC_005437	*Tricholepidion gertschi*	silverfish	(Nardi, *et al.*, 2003b)

Species names followed by * represent near-complete mt-genome sequences.

3.3. Results and Discussion

3.3.1. Genome Annotation

The genome sequencing methodology used, based on amplification with insect-derived heterologous primers gave an immediate crude map of the genome based on the identity of the genes from which the primers were designed. Refinement of this annotation began with the identification of the tRNA genes, which are known to punctuate the larger genes within insect mt-genome (Ojala, *et al.*, 1980a; Clary, *et al.*, 1982)

3.3.1.1. Annotation of tRNA Genes

As sequence accumulated, the assembled genome fragments were examined for tRNA genes. The tRNAScan software (Lowe and Eddy, 1997) (Version 1.23, complied for a Sun Microsystems UNIX machine running SunOS 5.8) was able to identify 20 of the 22 tRNAs for the stonefly. The remaining two, $tRNA^{Ser\text{-}AGN}$ and $tRNA^{Arg}$, were identified by inspection of the two approximately 70bp gaps in the tRNA cluster located between the *nad3* and *nad5* genes, where the orthologues of these tRNAs are found in most other insect mt-genomes. The identity of the $tRNA^{Arg}$ gene was confirmed by the folding of the putative tRNA by hand to ensure a cloverleaf structure was observed. The sequence was then aligned to orthologues from other insect mt-genomes for further confirmation. Similar to most animal mitochondria sequenced to date, the $tRNA^{Ser\text{-}AGN}$ for the stonefly mitochondria lacked a dihydrouridine loop and stem (Wolstenholme, 1992). Due to the unusual secondary structure for the $tRNA^{Ser\text{-}AGN}$ gene, the sequence was first identified by alignment to orthologous tRNA genes, and then by finding a putative

secondary structure fold that was constrained by tertiary structure considerations (Steinberg and Cedergren, 1994; Steinberg, *et al.*, 1994). The proposed secondary structures of the tRNA genes are shown in Figure 3.2.

3.3.1.2. Protein Coding Genes

With the tRNA genes identified, the large sequence gaps between them was examined for open reading frames, yielding amino acid sequences that would align to the known insect mitochondrial protein coding genes. All thirteen expected protein-coding genes were identified in their conserved relative position.

Five putative start codons are utilized by the stonefly mitochondrial genome. The most commonly used start, ATG, was found for six of the protein coding genes (*cox2, atp6, cox3, nad4, nad4L,* and *cytb*). These six protein genes include the four that are abutted to their 5' end by another protein coding gene that overlaps with the start codon by at least one nucleotide (*cox3* and *cytb*), or are overlapped by seven nucleotides (*atp6* and *nad4*). The ATT codon is utilized to initiate three protein genes (*atp8, nad3* and *nad6*), GTG by two (*nad2* and *nad5*), and TTG by one (*nad1*).

The *cox1* gene has been tentatively annotated to start with a CGA codon at position 1445-1447. It can be argued that the start should instead be annotated to start at the ATT at position 1436-1438, requiring the overlapping of the *cox1* and $tRNA^{Tyr}$ genes by seven nucleotides, and extending the 5' end of *cox1* by nine nucleotides. Since these two genes are encoded on opposite strands, there is no functional argument to dismiss this possible annotation. Common practice has, however, been not to overlap these genes, and to base the annotation of the *cox1* the open reading, or on a four (or greater) base initiation signal,

A

```
        Acceptor                                    Anti-                                   Acceptor
        Stem      DHU- Stem and Loop               codon          TψC Stem and Loop         Stem
        1111111..2222 .........  2222.33333 ..***.. 33333      44444........ 44444.1111111.
I       AATGAAGTGCCT  GATAAA      AGGATTATC GTGATAG GATAA ATTATGTAAATTATA    TTTAC CTTCATTA
Q       TATGTTCTAGTGT ATGGT      GCACGAAAGT TTTTGAT ACTTT CGGA AATAGTTCAAGT  CTATT GGACATAA
M       AGAAAGATAAGCT AATAA      AGCTAATGGG TTCATAC CCCAT TTAT AGAGGTTTTACT  CCTCT TCTTTCTA
W       AAGGCTTTAAGTT AATTA      AACTAATAGC CTTCAAA GCTGT AAAT AAAG TATCTAT   CTTT AAGCCTTA
C       AGCTTTATAGTCA ATGA       TGATATTAGA CTGCAAT TCTAA AGGT GTAA GTGATC    TTAC TAAGGCTT
Y       GGTAGAATGGCT  GAGATATAA   GGTGATAGA TTGTAAA TCTAT TTAG GAGG TATACC    CCTC TTCTACCA
L-UUR   TCTAATATGGCA  GATAAG      TGCAATGGA TTTAAGC TCCAT ATAT AAAGGCTATT    CCTTT TGTTAGAA
K       CATCAGATGACT  GAAAGCA     AGTATTGGT CTCTTAA ACCAT TTTATAGTAATCTCGCAA TTACT TCTGATGA
D       AAAGAATTAGTTA AAT        TAACATTAGT ATGTCAT GCTAA AATT ACTAGTACTAGC  CTAGT ATTCTTTA
G       ATTTGTCTAGTAT AATAA      GTATATTTGA CTTCCAA TCAAA AGGT CTGA ATATCT    TCAG GACAAATA
A       AGGACTGTAGTTA AGTA       TAACATTTGG TTTGCAT CCAAA AAGT ATTG AGATT     CAAT CTGTCTTA
R       GAATAAGAAGCGA TTTT       TTGCACTCAG TTTCGAC CTGAA TCTA GATAACTCAG    TTATC CTTATTCT
N       TTAATTGAAGCC  AAATAGA     GGCATATCA CTGTTAA TGATA TAATTGAAT AACTT     ATTC CAATTAAG
E       ATTTATATAGTTT AATTCA     AAACATTACA TTTTCAC TGTAA AAAT AAAG GCTTTA    CTTT TATAAATA
F       ATTCAAGTAGCTT ATAAATA    GAGCATGACA CTGAAGA TGTTA GGGA GATT AATGT     AATC TTTGAATA
T       GTTTTAATAGTTT AAAGA      AAACATTGGT CTTGTAA ATCAA AAATTAAGG CCTGGA    CCTT TTAGAACT
S-UCA   AGTTAATAAGCTT ATTGTA     AAGCATATGT TTTGAAA ACATA AGAC AGAAGTTTAAAT  CTTCT ATTAACTT
P       CAGAGAATAGTTT ATTTA      AAATATTAGT TTTGGGG ATTAA TGATAGGGGGTTT      CTCTT TTCTCTGA
L-CUN   TCTATTTTGGCA  GATAAG      TGCAATGGG TTTAGAA TCCAT TTAT GTAA AGGGATT    TTAC AGATAGAA
V       CAAGGCAGAGCTT GATTAGTT   AAGCATTTCA TTTACAC TGAGA AGTC ACTG TGCAATT    CAGTTTGTTTTGA

        11111111    2  ...   2 . 3.3.33333 ..***.. 33333.3.3 .. 44444          44444 11111111.
S-AGN   GAAATATG    A  TGT   T C AAGAAAAAG CTGCTAA CTTTTCTCT TT AGCGGTTAAATTCCGTT TATATTTCT
```

B

```
                 .
                1-1
                1-1
                1-1   Acceptor
                1-1   Stem
                1-1
                1-1
                1-1     ...
              .    44444   .
    ...    .       |||||   .  TψC Stem
  .  2222          44444   .  and Loop
  .  ||||          .      ...
  .  2222            .
    ...       .   ..
DHU Stem        3-3
and Loop        3-3
                3-3
                3-3
                3-3
              .   .
              .   .
               ***
           Anti-codon
```

C

```
                                     U
                                    G-C
                                    A-U
                         Acceptor   A-U
                         Stem       A-U
                                    U-A
                                    A-U
                              ┌──── U-A
                              │      ▲
                        U─────┼────►A-U
                        │     │     G-U
                        │     │     C-G
                        │     │     G-C
                        │     │     G-C    TψC Stem
                        │     │    U   U
                        │     │    U   U
                 AA     │     ▼     AAA
                U  CUUUUUCUU  GA U
                C  |||||||||  ||   G
   Anti-codon   G  GAAAAAGAA  CU U
                 CU
                                DHU-Loop
```

Figure 3.2. Proposed Secondary Structure for the tRNA Genes of the Giant Stonefly. The tRNA genes are listed by their one letter codes, with the codon listed for leucine and serine. A - Alignment of the primary sequences of the tRNA genes. B - Guide to fold all tRNAs except *tRNA$^{Ser\text{-}AGN}$* into potential clover-leaf structures. C - Tertiary structure compensation fold of the *tRNA$^{Ser\text{-}AGN}$* gene, from (Steinberg and Cedergren, 1994; Steinberg, *et al.*, 1994). Watson-Crick and G-T basepairs are denoted by a dash for B and C.

beginning with an A (Clary and Wolstenholme, 1983a; de Bruijn, 1983; Beard, *et al.*, 1993; Mitchell, *et al.*, 1993; Flook, *et al.*, 1995). Data to be discussed in a later chapter will clarify my choice of the shorter, and non-standard start codon (see chapter 4).

Stop codons that do not overlap with the abutting gene were only observed for two of the protein coding genes, for *nad4* (a TAA stop) and for *nad1* (a TAG stop). Three protein genes (cox1, *cox2,* and *nad5*) do not encode their stop codons in their DNA sequence, but evidently make use of single in-frame T nucleotides that are abutted by a tRNA genes and are presumed to have their stop codons completed by the polyadenylation of the mature mRNA (Ojala, *et al.*, 1980a; Ojala, *et al.*, 1981). Six protein coding genes encode a putative stop codon that overlaps by one or two nucleotides with the 5' end of their downstream tRNA or protein coding genes (a potential TAA for *nad2, atp6, cox3* and *nad6* and a TAG for *nad3* and *cytb*). It is equally likely that these six overlapping stop codons are not functional, but coincidentally appear to be stop codons. These genes could be interpreted to encode only the in-frame T or TA nucleotides that are then polyadenylated as mRNAs to construct their functional stop codons.

As is observed in most arthropod mt-genomes, the stop codons for the short *atp8* and *nad4L* genes are encoded as part of a seven nucleotide overlap with the coding region of their 3' neighbouring genes, *atp6* and *nad4*, respectively. It has been demonstrated in *Drosophila* that these gene pairs are expressed via bicistronic mRNA molecules (Berthier, *et al.*, 1986). It appears, by the conservation of this overlap that the stonefly maintains a bicistronic expression system for these gene pairs.

3.3.1.3. Ribosomal RNA Genes

Finally the *lrRNA* gene and the *srRNA* gene were identified based on conserved relative genome position and sequence alignment to the orthologues in other insect mitochondria. My annotation of the ribosomal genes presented here is based on the assumption that there are no non-coding nucleotides between the rRNA genes and their abutting tRNA genes. The boundary between the 5' end of *srRNA* and the major non-coding region (or the A+T rich region) could be detected by the alignment of a conserved "TTNAAGTTNTAARANCG" 5' motif for hexapod *srRNA* sequences.

3.3.2. Genome Architecture

The complete genome of the giant stonefly *Pteronarcys princeps* was observed as a circularly amplifiable molecule through the PCR methods employed. The genome size was found to be 16004 bps in total length. The same gene content and relative order observed in the *Drosophila yakuba* mitochondrial genome was maintained in the stonefly sequence (Clary and Wolstenholme, 1985a). This sequence organization has been termed the ancestral genome arrangement, due to the conservation of this exact gene order in non-insect hexapods (Nardi, *et al.*, 2003b) and crustaceans (Crease, 1999).

3.3.3. Nucleotide Composition

Analyses of nucleotide composition for various genome features are summarized in Table 3.3. The complete genome is observed to consist of approximately 71.5% A and T nucleotides. High A+T richness is observed for the 1158 bp A+T rich region (81.3%), the 3739 third codon position nucleotides (84.6%), and the 60 non-coding, intergenic nucleotides (85.0%). A+T content values less than those observed for the complete

Table 3.3. Nucleotide Composition of Stonefly Mitochondrial Genome Features. Bolded values represent the most frequently used nucleotide for the described feature, while italicized numbers represent the least frequently used nucleotide.

Annotated Genome Feature	number of nucleotides	A	T	G	C	A+T
Whole Genome (J strand)	16004	**0.3710**	0.3435	*0.1108*	0.1746	0.7146
J strand protein genes	6896	0.3008	**0.3872**	*0.1299*	0.1821	0.6879
1st codon position	2298	0.2881	**0.3042**	0.2293	*0.1784*	0.5923
2nd codon position		0.1915	**0.4339**	*0.1375*	0.2372	0.6254
3rd codon position		**0.4230**	**0.4230**	*0.0231*	0.1310	0.8460
N strand protein genes	4325	0.2717	**0.4603**	0.1734	*0.0946*	0.7320
1st codon position	1441	0.2713	**0.4282**	0.2124	*0.0881*	0.6995
2nd codon position		0.1811	**0.4684**	*0.1749*	0.1756	0.6459
3rd codon position		0.3622	**0.4844**	0.1332	*0.0201*	0.8466
tRNA genes – coding	1413	0.3369	**0.3631**	0.1670	*0.1316*	0.7146
tRNA genes – J	870	0.3563	**0.3609**	0.1425	*0.1402*	0.7172
tRNA genes – N	543	0.3094	**0.3665**	0.2063	*0.1179*	0.6759
rRNA genes - N	2121	0.3310	**0.3885**	0.1862	*0.0943*	0.7195
A+T rich region (J strand)	1158	**0.4387**	0.3739	*0.0639*	0.1235	0.8126
Non-coding nucleotides	60	**0.5000**	0.3500	*0.0500*	0.1000	0.8500

genome are found for the protein coding gene first and second codon positions, and the rRNA and tRNA genes (59.2% to 72.0%).

The Relative Synonymous Codon Usage (RSCU) values were calculated for the protein coding genes of the giant stonefly (Table 3.4). The results for the two-codon tRNA families show a distinct bias for the use of the A or T nucleotide in the third codon position, regardless of the identity of the anti-codon encoded by the tRNA. If one inspects the majority-coding strand alone (the J strand), the third codon position sites for four-codon tRNA families show a preference for A nucleotides (T nucleotides for the minority-coding or N-strand proteins). This preference for A on the J-strand is common for most insect mt-genomes investigated (see silverfish and locust comparisons in Table 3.4).

Positional asymmetric nucleotide bias has been carefully explored for mammalian mt-DNA (Reyes, *et al.*, 1998; Saccone, *et al.*, 1999). In mammals, it is hypothesized that the slow initiation of lagging strand replication leaves the nucleotides exposed (see figure 1.3A), increasing the rate of deamination of C to U, resulting in a bias towards the pyrimidine T on the lagging replication strand (Francino and Ochman, 1997). Also, A is a target for deamination to hypoxanthine, which is complemented on the leading strand by by C. The majority of replicating mt-DNA molecules in *Drosophila* have been observed, using electron microscopy coupled with restriction mapping, to undergo an extreme asymmetric replication (Figure 1.3B). Up to 99% of the leading strand (the J-strand of the genome) is replicated before replication is the initiated for the lagging strand (the N-Strand) (Goddard and Wolstenholme, 1978; 1980). Such a replication mechanism is expected to leave obvious patterns of nucleotide substitution, in which genes near the

Table 3.4. Comparison of Relative Synonymous Codon Usage for the Giant Stonefly, a Silverfish and a Locust

Amino Acid	Codons	Stonefly All	Stonefly J	Stonefly N	Silverfish All	Silverfish J	Silverfish N	Locust All	Locust J	Locust N
K	AAA	**1.135**	**1.556**	0.737	**1.506**	**1.704**	**1.043**	**1.414**	**1.829**	0.414
	AAG	0.865	0.444	**1.263**	0.494	0.296	0.957	0.586	0.171	**1.586**
N	AAC	0.297	0.384	0.122	0.464	0.653	0.080	0.484	0.632	0.080
	AAT	**1.703**	**1.616**	**1.878**	**1.536**	**1.347**	**1.920**	**1.516**	**1.368**	**1.920**
Q	CAA	**1.684**	**1.922**	**1.200**	**1.726**	**1.852**	**1.368**	**1.905**	**2.000**	**1.571**
	CAG	0.316	0.078	0.800	0.274	0.148	0.632	0.095	0	0.429
H	CAC	0.420	0.515	0	0.775	0.848	0.429	0.704	0.906	0.111
	CAT	**1.580**	**1.485**	**2.000**	**1.225**	**1.152**	**1.571**	**1.296**	**1.094**	**1.889**
E	GAA	**1.619**	**1.960**	**1.118**	**1.526**	**1.773**	**1.188**	**1.714**	**1.960**	**1.259**
	GAG	0.381	0.040	0.882	0.474	0.227	0.813	0.286	0.040	0.741
D	GAC	0.338	0.490	0	0.685	0.958	0.160	0.325	0.520	0
	GAT	**1.662**	**1.510**	**2.000**	**1.315**	**1.042**	**1.840**	**1.675**	**1.480**	**2.000**
Y	TAC	0.408	0.707	0.080	0.517	0.709	0.294	0.378	0.692	0.137
	TAT	**1.592**	**1.293**	**1.920**	**1.483**	**1.291**	**1.706**	**1.622**	**1.308**	**1.863**
W	TGA	**1.886**	**2.000**	**1.676**	**1.714**	**1.853**	**1.459**	**1.920**	**1.972**	**1.793**
	TGG	0.114	0	0.324	0.286	0.147	0.541	0.080	0.028	0.207
C	TGC	0.191	0.500	0.067	0.500	**1.077**	0.258	0.364	**1.000**	0.067
	TGT	**1.809**	**1.500**	**1.933**	**1.500**	0.923	**1.742**	**1.636**	**1.000**	**1.933**
M	ATA	**1.656**	**1.796**	**1.462**	**1.553**	**1.727**	**1.184**	**1.687**	**1.836**	**1.270**
	ATG	0.344	0.204	0.538	0.447	0.273	0.816	0.313	0.164	0.730
I	ATC	0.256	0.340	0.081	0.412	0.532	0.155	0.225	0.311	0.049
	ATT	**1.744**	**1.660**	**1.919**	**1.588**	**1.468**	**1.845**	**1.775**	**1.689**	**1.951**
L	TTA	**1.772**	**1.974**	**1.592**	**1.453**	**1.893**	**1.118**	**1.757**	**1.966**	**1.584**
	TTG	0.228	0.026	0.408	0.547	0.107	0.882	0.243	0.034	0.419
F	TTC	0.348	0.580	0.045	0.510	0.834	0.150	0.515	0.938	0.125
	TTT	**1.652**	**1.420**	**1.955**	**1.490**	**1.166**	**1.850**	**1.485**	**1.062**	**1.875**
L	CTA	1.643	1.694	1.333	**2.170**	**2.792**	0.625	**2.149**	**2.706**	0.375
	CTG	0.119	0.083	0.333	0.395	0.226	0.813	0.119	0.157	0
	CTC	0.310	0.333	0.167	0.269	0.327	0.125	0.149	0.157	0.125
	CTT	**1.928**	**1.889**	**2.167**	1.166	0.654	**2.438**	1.582	0.980	**3.500**
V	GTA	1.636	**1.926**	1.177	**1.558**	**2.433**	0.566	1.749	**2.731**	0.456
	GTG	0.363	0.207	0.612	0.478	0.233	0.755	0.066	0.038	0.101
	GTC	0.346	0.415	0.235	0.407	0.500	0.302	0.109	0.154	0.051
	GTT	**1.655**	1.452	**1.977**	**1.558**	0.833	**2.377**	**2.077**	1.077	**3.392**

(Continued on next page)

(Table 3.4 *concluded*)

Amino Acid	Codons	Stonefly All	J	N	Silverfish All	J	N	Locust All	J	N
T	ACA	1.425	1.646	0.853	**1.929**	**2.154**	0.952	**2.556**	**3.066**	0.316
	ACG	0.146	0.025	0.459	0.089	0.044	0.286	0.059	0.072	0
	ACC	0.384	0.506	0.066	0.804	0.923	0.286	0.234	0.287	0
	ACT	**2.046**	**1.823**	**2.623**	1.179	0.879	**2.476**	1.151	0.575	**3.684**
P	CCA	1.227	1.382	0.800	**2.306**	**2.815**	0.778	**2.294**	**2.969**	0.615
	CCG	0.133	0.146	0.100	0.139	0.074	0.333	0.088	0.082	0.103
	CCC	0.640	0.836	0.100	0.611	0.593	0.667	0.118	0.165	0
	CCT	**2.000**	**1.636**	**3.000**	0.944	0.519	**2.222**	1.500	0.784	**3.282**
A	GCA	0.873	0821	0.986	**1.696**	**2.281**	0.571	**2.107**	**2.781**	0.533
	GCG	0.127	0	0.406	0.196	0.066	0.444	0.053	0.038	0.089
	GCC	0.491	0.689	0.058	0.826	1.091	0.317	0.160	0.190	0.089
	GCT	**2.509**	**2.490**	**2.551**	1.283	0.562	**2.667**	1.680	0.990	**3.289**
S	TCA	1.550	**1.714**	1.301	**1.934**	**2.759**	0.701	1.897	**3.194**	0.548
	TCG	0.115	0.032	0.241	0.083	0.110	0.041	0.063	0.062	0.065
	TCC	0.402	0.635	0.048	0.529	0.800	0.124	0.111	0.124	0.097
	TCT	**1.933**	1.619	**2.410**	1.455	0.331	**3.134**	**1.929**	0.620	**3.290**
S	AGA	**2.050**	**2.109**	**2.000**	**2.029**	**2.571**	**1.650**	**2.945**	**3.491**	**2.400**
	AGG	0	0	0	0.412	0.071	0.650	0.073	0	0.145
	AGC	0.298	0.363	0.242	0.235	0.357	0.150	0.109	0.073	0.145
	ACT	1.653	1.527	1.758	1.324	1.000	1.550	0.873	0.436	1.309
R	CGA	**2.323**	**2.769**	1.565	**1.931**	**2.526**	0.800	**2.473**	**3.657**	0.400
	CGG	0.258	0.103	0.522	0.828	0.632	1.200	0.145	0.114	0.200
	CGC	0.129	0.205	0	0.207	0.316	0	0.073	0.114	0
	CGT	1.290	0.923	**1.913**	1.034	0.526	**2.000**	1.309	0.114	**3.400**
G	GGA	**2.050**	**2.282**	**1.708**	**1.600**	**1.971**	1.121	**2.019**	**3.079**	0.390
	GGG	0.571	0.254	1.042	0.980	0.870	1.121	0.192	0.159	0.244
	GGC	0.235	0.366	0.042	0.196	0.261	0.112	0.019	0	0.049
	GGT	1.143	1.099	1.208	1.224	0.899	**1.645**	1.769	0.762	**3.317**
Number of Codons		**3729**	**2291**	**1438**	**3724**	**2289**	**1435**	**3725**	**2296**	**1429**

origin of replication that are exposed longest accumulate more nucleotide substitutions and exhibit a more extreme nucleotide bias (Reyes, *et al.*, 1998). This nucleotide substitution pattern thus supports the asymmetric unidirectional replication model (Goddard and Wolstenholme, 1978; 1980; Clayton, 2000), and implies a limited role in the bi-directional, symmetric and semidiscontinious replication model (Goddard and Wolstenholme, 1978; 1980; Yang, *et al.,* 2002; Bowmaker, *et al.,* 2003). For insects, a pattern of more A and C nucleotides are predicted to accumulate on the J-strand (leading) of the insect mitochondria as you move from the *nad2* gene towards the ribosomal subunit genes.

To test if this nucleotide bias signature was present in the giant stonefly, third codon position nucleotides were compared for the four-fold degenerate codon families. The *nad2* and *cox1* genes are coded approximately 780 to 3040 bps from the center of the A+T rich region, and would be single stranded only a short time during replication. In comparison, the *nad6* and *cytb* genes, which are coded approximately 3880 to 5550 bp from the A+T region center on the opposite side, would be expected to be in a single stranded form for a much longer period of time. As these genes are encoded on the J-Strand, an expected relative increase in A and C nucleotides at neutral sites is expected in response to the reduction in A and C on the complement N-strand. Comparisons show a strong bias towards the use of A in the third codon position for the J-strand encoded *nad6* and c*ytb*, with a sharp decrease in the use of G nucleotides. A more centrally encoded gene cluster of *atp6, cox3* and *nad3* show an intermediate usage of A and C nucleotides (see Figure 3.3).

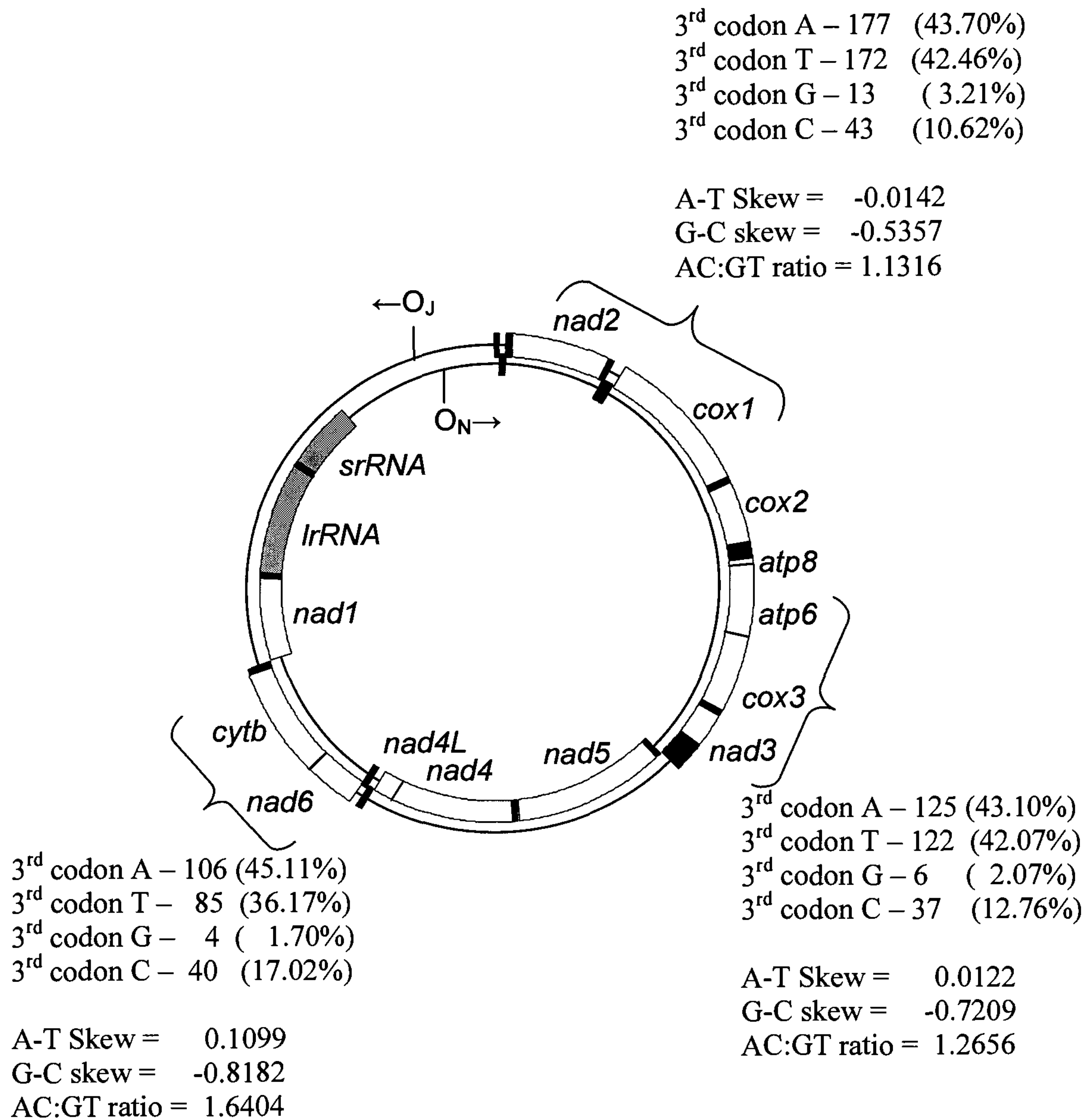


Figure 3.3. Illustration of the Directional Mutational Pressure Observed in the Mitochondrial genome of the Giant Stonefly. Hypothesized locations of the origins of replication (O_J and O_N) are labeled. Due to the extended time spend single stranded during replication, the N-strand is expected to experience increased levels of C to T substitutions and selection against A nucleotides in neutral sites. Genes closest to (clockwise from) O_N will remain single stranded for less time, resulting in lower mutational bias towards T and C nucleotides at neutral sites. The nucleotide composition of four-fold degenerate third codon position nucleotides and the calculated A-T and G-C skews, as well as AC:GT ratios for the analyzed gene clusters are shown.

The ratio of A and C nucleotides to G and T third codon position nucleotides was calculated for each of the three gene clusters, and showed an increase as you moved towards the *nad6* and *cytb* gene cluster, as expected in this model (Figure 3.3). The A-T skew (the difference between number of A and T nucleotides, divided by the sum) and G-C skew were calculated for these three gene clusters of the giant stonefly mt-genome (Perna and Kocher, 1995). Similar to the situation observed in mammals, the A-T skew increases as the estimated time that the particular region of DNA is expected to persist in the single stranded state increases (see Figure 3.3) (Reyes, *et al.*, 1998).

The analysis was repeated for eighteen additional representative insect mt-genomes, from ten insect orders. Mt-genome sequences which lacked complete *nad2* sequence were excluded, as were mt-genomes with protein gene rearrangements relative to the *D. yakuba* mt-genome. Generally, the trend of increasing AC:GT ratio and A-T skew from the *nad2* – *cox1* genes versus the *nad6* – *cytb* genes was observed for the insects examined (Table 3.5). Six of the nineteen insect mt-genomes observed did not show the expected directional mutation bias. These six organisms represent four of the insect orders (Lepidoptera, Hymenoptera, Coleoptera and Hemiptera). The mt-genomes currently sequenced for the Lepidoptera and Hymenoptera are known to have extreme A+T biases (79.78% to 86.62%, excluding the A+T rich regions) (Crozier and Crozier, 1993; Yukuhiro *et al.* 2002) and may be undergoing additional mutational pressures that could obscure the directional mutation effects. There has been the suggestion of a second origin of replication in the honeybee, between the *cox1* and *cox2* genes, which would have a strong effect on the observed substitution pattern (Cornuet *et al.* 1991). Further investigation into paraneopteran and coleopteran mt-DNA replication and nucleotide

Table 3.5. Positional Nucleotide Bias in the Mt-Genome of the Giant Stonefly and other Representative Insects. Values were calculated using third codon postiton nucleotides of fourfold degenerate codons for the three gene clusters (see Figure 3.3). The ratio of A and C to G and T nucleotides, as well as A-T and G-C skews are also presented. Bolded numbers represent trends in AC: GC ratio and skew values based on the directional mutation pressure driven by an asymmetric unidirectional replication model.

Species	AC:GT Ratio			A+T Skew			G+C Skew		
	nad2 & cox1	*atp6 – cox3*	*nad6 & cytb*	*nad2 & cox1*	*atp6 – cox3*	*nad6 & cytb*	*nad2 & cox1*	*atp6 – cox3*	*nad6 & cytb*
Diptera									
Drosophila yakuba	**1.0679**	1.0165	**1.2159**	**0.0349**	0.0085	**0.0880**	**0.0000**	0.0000	**-1.0000**
Chrysomya putoria	**1.3691**	**1.5784**	**2.1343**	**0.1078**	**0.2097**	**0.3403**	-1.0000	-0.4667	-0.5790
Ceratitis capitata	**1.0343**	1.0000	**1.7887**	**-0.0119**	-0.0202	**0.2553**	**-0.5000**	-0.3846	**-0.8000**
Anopheles gambiae	**1.3611**	1.2072	**1.8116**	**0.1615**	0.0644	**0.2865**	**0.0000**	-0.6667	**-0.3333**
Lepidoptera									
Antheraea pernyi	1.1793	0.9908	1.0689	-0.0145	-0.0778	-0.0307	-0.7500	-0.5833	-0.6471
Bombyx mori	**2.7073**	**3.6428**	2.8372	**0.4456**	0.5714	**0.4737**	-0.6842	-0.5385	-0.5385
Hymenoptera									
Apis mellifera	**2.3452**	**2.8409**	2.2391	**0.3806**	**0.4731**	0.3566	**-0.8462**	-1.0000	**-1.0000**
Coleoptera									
Tribolium castaneum	2.4786	2.1765	2.1918	0.3851	0.3484	0.3107	-0.5765	-0.4694	-0.5714
Crioceris duodecimpunctata	**1.3642**	1.2800	**1.4024**	0.1761	0.0872	0.1228	**-0.0462**	**-0.3333**	**-0.4615**
Pyrocoelia rufa	**1.9160**	**2.7742**	**3.5952**	**0.3141**	**0.4579**	**0.5778**	**-0.3143**	-0.6000	**-0.3846**
Hemiptera									
Triatoma dimidiata	**2.4286**	**3.1343**	2.5797	0.3531	0.4570	0.3552	**-0.6543**	-0.7500	**-0.6875**
Philaenus spumarius	**0.9940**	**1.8493**	**2.8627**	**0.0239**	**0.3222**	**0.5439**	0.2105	-0.1429	-0.0769
Pachypsylla venusta	**2.1217**	**2.2609**	1.4878	0.3311	0.3503	0.1345	-0.5185	-0.6429	-0.5151
Blatteria									
Periplaneta fuliginosa	**2.3000**	**2.4189**	**3.3750**	**0.4030**	**0.4250**	**0.5612**	-0.2857	-0.2308	-0.2857
Orthoptera									
Locusta migratoria	**2.4255**	**3.6481**	**5.0625**	**0.4890**	**0.5949**	**0.6757**	**-0.2421**	-0.1429	**-0.5556**
Gryllotalpa orientalis	**1.2022**	**1.6600**	**2.8360**	**-0.0533**	**0.1142**	**0.3708**	**-0.5146**	-0.8723	**-0.8214**
Plecoptera									
Pteronarcys princeps	**1.1316**	**1.2656**	**1.6404**	**-0.0143**	**0.0122**	**0.1099**	**-0.5357**	**-0.7209**	**-0.8182**
Thysanura									
Tricholepidion gertechi	**2.9720**	**3.4032**	**4.9070**	**0.5164**	**0.5333**	**0.7059**	**-0.4222**	-0.5873	**-0.4800**
Thermobia domestica	**1.5412**	1.9074	**1.7684**	**0.1057**	0.1342	**0.1237**	**-0.5644**	-0.8072	**-0.7101**
Archeognatha									
Nesomachilis australica	**1.4824**	**1.9694**	**2.4211**	**0.0859**	**0.2105**	**0.3300**	**-0.5625**	-0.7460	**-0.7000**

substitution patterns are required to understand their deviation from the expected directional mutation bias.

G-C skew values generally did not reveal the trend as clearly at the A-T values. In most cases, the number of third codon G and C nucleotides was very low (typically <10% and never >30%), providing very few G or C containing sites, limiting the power of the analysis.

3.3.4. Stonefly A+T Rich Regions

The 1158bp non-coding region was observed in the conserved location, between the *srRNA* and the $tRNA^{Ile}$ genes. The sequence was found to be highly A+T rich at 81.26%, approximately 10% higher than the entire genome content. Only the third codon positions of protein coding genes and the 60 intergenic non-coding nucleotides had higher A+ T compositions.

The region was searched for open reading frames (ORFs) of 50 amino acids or larger, with the seven potential start codons suggested for insect mitochondrial genomes (ATN and NTG codons). The fifty amino acid limit was chosen, as it would produce a product approximately the size of the stonefly's *atp8* gene (51 amino acids), the smallest known functional protein coding gene in insect mt-genomes. Four open reading frames were found meeting these criteria, a 64aa ORF (positions 15928 to 15737), a 56aa ORF (positions 14898 to 15065), a 54aa ORF (positions 15157 to 14996), and 50aa ORF (positions 15362 to 15511). Blastp searches of these ORFs gave a small number of hits (1, 2 or 3) with high E values (2.3 to 8.6), to fragments of annotated genes within the database, and were not consider significant matches. The small size and lack of similarity to any known proteins imply that these are non-functional ORFs.

The region was searched against itself, using DOTTER (Sonnhammer and Durbin, 1995) to identify repeat and inverted repeat regions within the A+T rich region (see Figure 3.4A). Two regions were identified, stem-loop 1 (with positions 15126 – 15178) and stem-loop 2 (positions 15252 – 15290) showed sequence similarity to each other, and the potential ability to fold into stem-loop structures. The identified regions were folded using DNA *mfold* (Zuker, 2003) to look for putative secondary structures (see Figure 3.5).

A comparison of two stonefly A+T rich regions from genus *Peltoperla* has been published previously, and revealed conserved repeat structures and secondary structures within that genus (Schultheis, *et al.*, 2002). Interestingly, the sequence encompassing one of the putative stem-loops was aligned by Clustal X to the regions coding "inverted repeat 1" for both *Peltoperla* species (Figure 3.6). The other putative *P. princeps* stem-loop aligned between the two remaining inverted repeats for *Pe. tarteri*.

I compared the A+T region of the giant stonefly to those from *Pe. arculata* and *Pe. tarteri*. Dot plot analysis revealed five conserved sequence blocks shared by the A+T rich regions of the two different genera (see Figure 3.4B). Additionally, the dot plots reveal a shared low sequence complexity region spanning 168nt (in *P. princeps*) to 188nt (in *Pe. tarteri*), where the A+T compositions are elevated to 94.15% to 96.43%. Generally, the A+T regions of these two *Peltoperla* species are not immediately alignable to that of the giant stonefly. Clustal X alignment (Thompson, *et al.*, 1994) show very low levels of sequence identity when comparing the giant stonefly to *Pe. arcuata* (61.44%) or *Pe. tarteri* (60.44%), despite the very high A+T richness for all three species (81.10% to 82.56% A and T nucleotides).

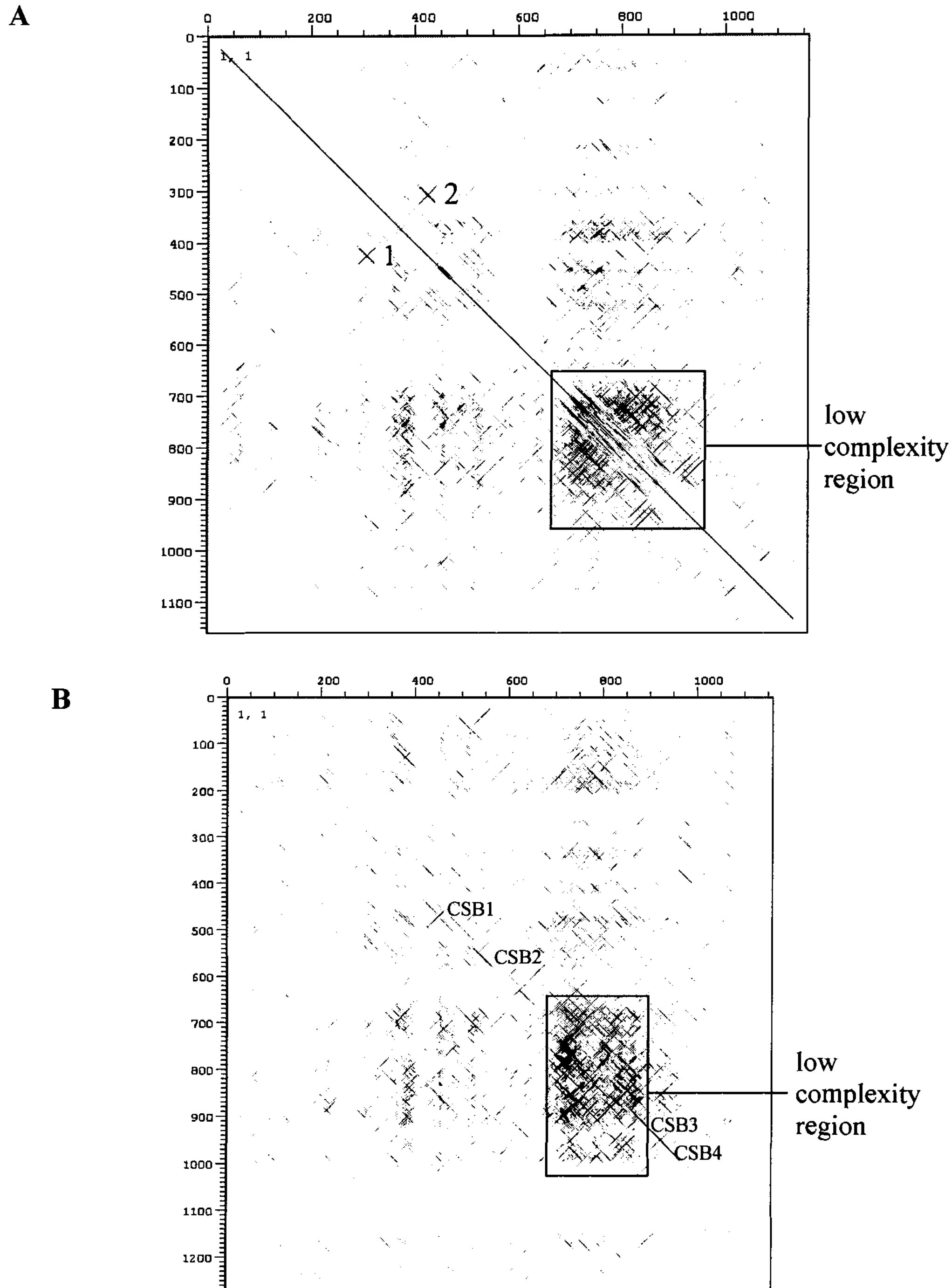
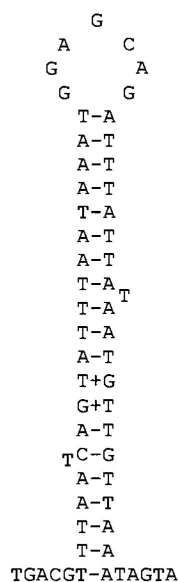


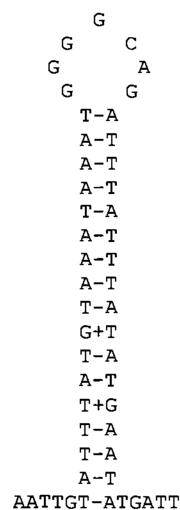
Figure 3.4. Dot Plot Analyses of the A+T Rich Region of *Pteronarcys princeps.* Dot-plots were generated using DOTTER. A – comparison of *P. princeps* against itself identified inverted repeat units (labeled 1 and 2) and a low-complexity sequence region. B – comparison of *P. princeps* (horizontal) versus *Pe. tarteri* (vertical). Conserved Sequence Blocks (CSBs) are identified and the low-complexity sequence region are identified.

A



ΔG = -8.7

B



ΔG = -7.9

Figure 3.5. Predicted Secondary Structure Folds for Potential Stem-Loop Regions Identified within the A+T Rich Region of the Giant Stonefly. Secondary structure folds and ΔG values (kcal/mol) were calculated by DNA *mfold* (Zuker, 2003). Numbers represent the genome positions of the noted nucleotides. Watson – Crick basepairs represented by "-" and G-T pairing represented by "+". A – Stem-loop from positions 15124 – 15175, shown on the N-strand. B – Repeat from position 15252 – 15291.

```
   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Pp GGACTTAGACAAAAAAAAGACTTTG--CCTTAAATCATTCAACTCAATAGTTAATTAAGTGTATATTAGAAAGTTGAAAA 14925
Pt -----..T.TT......CC.T.AA.AC...CT...T.AC..CAC...A.T...CA.GTAAAACA..A.T...A.AA.TT.  1196
Pa -----..T.TT......CC.T.AA.AC...CT...T.AC..CAC...A.T...CA.GT.AAACA..A.T...A.AA.TT.   998

   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Pp ATAACAAATAGGGGGGCCCAAAAAATTTTTTTAAACAAAATCAAAAGCCAAATT---TTAAATAGACTTTCTTACTTACC 15002
Pt .C..GC..G.ATAAAA.TTTG....A.AAAAC...-..T..GGG..TAT...A.AAA...G...C.AA.AAAC.TA...T  1117
Pa .C..GC..G.ATAAAA.TTTG...GA.AAGAC...-..T..AGG..C.....A.AAA.....C.T.AA.AA.C.T.A..T   919

   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Pp ---TCTCCACCTGATTAAACAGTCA-----AATTCCCTCCCTAAAAAAAATCTAAATCTTCATATAACATGA--TAATTA 15072
Pt GCACT.AA.AAA..A....T.A.A.ATTAA...A.AA.TTAA...C....AAC..GCAAA..A....A.CA.ACCGG.C.  1037
Pa GAACT.AA.AAA..A...GT.A.A.ATAAA..AC.AA.TTAAG..T....AAC....AAA..A..T.A.C..ACCGG.C.   839

   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Pp ATCAA-TAATCAACAAACA-CATACCATTTGCCTAAAGTTTGCGGTATAAATACTATTTAAC----AACTTATTATAAT 15142
Pt CCT.GG.G.C.CT......AT.C.AT.C..AAT...TAA..T.ATC..T.TG..A.A..TG.CCCCA..T......C...   957
Pa .CT.GG.G.C.CT.G....ATGA.AT.-----------------------------------------------------   812

   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Pp AAATCTGC---------TCCATTTATTAAATAAACTGATTAAACGACAATAAATTTACCTAACCAACTTTAAATATTAAA 15216
Pt ....AAAACCATTAACAAAT.C...A.T.TA.TCTCA....--TT.T..C......G..C--....A....TTACAAT..   881
Pa -----------------AAT..C..A.T.....TT.C....--TT.T.........C.G..C--....A....TTATAAT..   753

   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Pp ATAAAAATTATATTATATTCAAATTAACTTATTAACTATTCATCTAATAAATCTGCCCC---ATTTATTA-AATAAATAA 15292
Pt ....-..CC..-.A.C.AAT.CT.A.GTA..A.CT.A...AT.A...C....T......CAA..C.....T..........   803
Pa ....T..CC..-.A.C.AAT.CT.A.TTA..A.CT.A...ATCA...C....T.....CAA..........T..........   674
                                     (CSB1)TAAyAAATyTGCCCC---ATyTATTA-AATAAATAA

   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Pp CTAAATTTGTTTATTTTAATAAAAATAACTAAATAACCTACTCAATAAAACCATTAAATATATAATTTAAATAACTGAAC 15372
Pt AATT.C..T.A.....-....T...CCTT.T..A..A.ACT.........T------..CAT.G.AC..T.A.TTAATGT   730
Pa AATT.C..T.A.....-....T...ACTTAT..A..ATACT.........T------...AT...AC..T...TTCAT.T   601

   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Pp TATTAATAGTAAAATTAGTTAAAGTTTATTGGAAAGACTTAATATTTACTTTTTCAAATGGATTTTTTATCTTAATGGAT 15451
Pt A.A...C--...........................AAG..AG--A.T.-----..GATA..AAA...GGA.....-..   660
Pa AGG...C--..............A...........GAAG..A.A.A.A.-----G..ATA..AAA...GGA.-..T...   529
     (CSB2)TAAAATTAGTTAAArTTTATTGGAAAGr

   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Pp GATCCTCATTTTTTTCTTATTTATAAAATGAGGATCATCCCATTAATTTATTATTTAAATTAATAACCTCTTGTTAATAA 15532
Pt AC.TTC.........TA..GGAC.......GAA.GT...AATA.....AC....A.T..A..-...TTATA.A.G.T.TT   581
Pa AC.TTC.........TA..GAAC.......GAA.GT...AATA...A.ACC...G.T..A..-...T..TA.A.A...TT   450
```

(Continued on next page)

Figure 3.6. The A+T Rich Regions of *Pteronarcys princeps* and Two *Peltoperla* Species. Dots represent a nucleotide identical to the *P. princeps* (*Pp*) sequence at top, and dashes represent introduced gaps for the alignment. Numbers represent the genomic position of the right-hand most nucleotide. *Pe. tarteri* (*Pt*) and *Pe. arcuata* (*Pa*) are show as the reverse-complement of the Genbank Flatfiles (AY142074 and AY142073). Boxed sequences represent areas containing putative stem-loop structures. Conserved sequence blocks (CSB) are highlighted by the consensus sequence beneath the aligned regions.

(Figure 3.6. *concluded*)

```
   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Pp GAACTTAATTTAA-TATATATATTTAATTAAATA-ATAATTATATTTATATATTTATTTATATAATATAAAATTTAATTT 15610
Pt A..---.......A...T....G.AT.A......C...TA.....AC..--.AA...A.TA......AGC.TACAT..A.   506
Pa A..---....C..-...T......AT.A......C...TA.....A......AA...A.TA......AGC-TC.AT..A.   375
         (-----------------------------------------low complexity region-----------

   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Pp TTA-ATATAATTA-ATAATTA-ATCATATATAAATTAAATTCTTATT-------------AATATATA----AATAATTA 15670
Pt A..T......A..T.....C.T..A.......T..AT...AAGA...TAATTAATTAATA...G....ATAT.TA...AG   426
Pa A..T.....TA..T...--..T.-A.......T..AT....A.....----------ATA........-TAT.TA...AG   309
   -----------------------low complexity region-----------------------------------

   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Pp TATAATAACAATTAAGTTTATA-TATATA---ATCTATATATA-AATAAGAATTTAATATTAAATATATCTATTTTTTGC 15746
Pt .....AT.T.TA.GTA.....TA...-..TAA..A........T........................C.T.TA.....T.   347
Pa .....AT.T.TA.GTAA.....T......AAT..A.G......C..........................T.-A.....TT   230
   -------------------------------------)TATATA-AATAAGAATTTAATATTAAATAT(CSB3)

   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Pp CTTTAACTATAAACATAATAGGTATATAAATATGATGTATATAATATAAGCTCTTATTATACACCCAAATTTTTAACAAA 15825
Pt T...TTA.....G................GG..A..........AT...AT......-.....A.AGT.A....T.A...   268
Pa ....TTA.....G................GA..A..........AT...AT......-.....A.A.T.A....TTAT..   140
    (CSB4)TATAArCATAATAGGTATATAArnATrATGTATATAAnnTAAryTCTTAT-ATACA

   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Pp AAATTACACCTTCTGAACAATTTACTTTTTGTTCATAAAATTAAGGGTTTTGGGTTTTTAAAAAAAGGGTAAAAAAAATC 15905
Pt C..AA.A.AAAA....TAGG...TAG...C..GG.C..CT.....AA...G...GA.GGGTT...CCCCCC....CCCC.   188
Pa T..AA.A.AAAA....TAGG...TAG......GG.C..CT.....AA...G...GAAGGGTT...CCTCCC....CCCC.    71

   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Pp CA--AAATCCTCAATTTTTTGAACAAAAATCATAAA----AGAAAACATTCCCTCTTCCCCTTTTAAAGGGCCATATTCA 15979
Pt .CCC......CT..AA..G.CCC.....CC.TC..TTTTTG....TTT.CAG...C.TT.AA.....G.TAAA.A.A.A.   108
Pa .CCC......CT..AA..G.CCC.....CC.TC..TTTTTG.....TT.CAG..AC.TT.AA......AT----------     1

   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Pp -ATTACATTTAGATAAAAAAATAAAA------------------------------------------------------ 16004
Pt T.A..A..C..T....C..TCC.T.CTCTATCAAGTTCCCGTAAATAATAAAAAAAAAAATGAAACCCTGATAAAAAGGAT    28
Pa --------------------------------------------------------------------------------     1

   ....|....|....|....|....|..
Pp ---------------------------(tRNA^Ile) 16004
Pt TATCGTGATAGGATAAATCATGTAAAT(tRNA^Ile)     1
Pa ---------------------------(tRNA^Ile)     1
```

The four conserved sequence blocks (CSBs) identified in the dot-plot analyses were compared in the sequence alignment. CSB1 is a 32 to 36 nt sequence (positions 15381 – 15408) that co-localizes with putative stem-loop structures predicted to occur at the same aligned position in all three species. CSB2 is 28nt in length and is identical in sequence to that in *Pe. tarteri*, and codes only two A – G nucleotide transitions relative to the *Pe. arcuata* sequence. CSB4 occurs immediately after the low complexity regions, and encodes a single nucleotide insertion in the two *Peltoperla* species over the region's 29nt. CSB4 occurs closest to the $tRNA^{Ile}$ gene, and encodes only 8 transitions and one insertion over the 55nt length.

Though the conservation of these sequences and structures are intriguing, it is important to remember that the mechanism of replication initiation in insect mt-DNA is still not characterized. As such, it is not possible to assign and functionality to these conserved segments, nor infer that the observed secondary structures may have homologous function to the D-loop structures observed the replication of other animal mt-genomes (Jacobs, *et al.*, 1989).

3.4. Conclusions

The complete mitochondrial genome of the first plecopteran representative is presented. The size, nucleotide composition and genome arrangement are very typical of a "standard" insect mitochondrial genome, such as the *D. yakuba* mt-genome (Clary and Wolstenholme, 1985a). The coding of the rRNA, tRNA, and protein coding genes is also canonical, including the apparent absence of a start codon for the *cox1* gene, as first observed in *Drosophila* species (Clary and Wolstenholme, 1983a; de Bruijn, 1983).

Nucleotide biases are observed to be strand specific and positional in nature. These observations argue that the replication of the mt-genome in the giant stonefly follows the same highly asymmetrical and asynchronous pattern as replication in *Drosophila* species (Goddard and Wolstenholme, 1978; 1980). Positional bias also argues that the N-strand is replicated first, from the A+T region toward the ribosomal gene cluster, as in *Drosophila.*

Comparisons of the A+T rich region from two other stoneflies of the genus *Peltoperla* reveal conservation of some putative secondary structure elements, as well as regions of very high sequence conservation. These results are surprising, given the highly variable nature of the insect A+T rich region. Sampling of a broader selection of stonefly families may reveal deeper conservation of these elements, which may lead to investigations of their involvement in insect replication and potentially transcription initiation.

With the addition of this plecopteran sequence to the mt-genome databases, further sampling of the Neoptera *incertae sedis* orders must be undertaken. With the mt-genome sequences available, further insights into the diversification of these early diverging groups can help improve our conflicted understanding of early insect evolution. With the immensely broad diversity of insect evolution available to sample, who knows what other interesting molecular evolutionary phenomenon may be waiting to be described in insect mitochondria.

Chapter 4. Characterization of Mitochondrial Transcripts in *Drosophila melanogaster.*

4.1. Introduction

4.1.1. The *cox1* Initiation Mystery in Insect Mitochondria

One of the most striking features of insect mt-genomes has been the apparent lack of a coded start codon for the essential mitochondrial gene *cytochrome c oxidase subunit 1* (*cox1*) in most insects sampled to date. The first descriptions of this phenomenon were published by investigators involved in the sequencing of the *Drosophila melanogaster* (de Bruijn, 1983) and *D. yakuba* (Clary and Wolstenholme, 1983a) mt-genomes. These studies identified the presence of an in-frame TAA stop codon at the 5' end of the *cox1* open reading frame, followed by a TCG serine and the remaining open reading frame of the *cox1* gene. In order to satisfy the conditions of a start signal for the gene, the researchers proposed a four-base ATAA start codon (the in-frame TAA plus a non-coding A to the 5'). This four-base start was assumed to undergo an unusual basepairing interaction during the initiation of transcription (Clary, *et al.*, 1983; de Bruijn, 1983; Clary and Wolstenholme, 1985a). Further analyses within *Drosophila* species showed a lack of conservation of the ATAA signal (Satta, *et al.*, 1987), and the first mosquito mt-genome sequences showed that a four base start signal was not conserved within the Diptera (Beard, *et al.*, 1993; Mitchell, *et al.*, 1993). Beard *et al.* (1993) suggested that the lack of conservation of the multi-base starts were not indicative of a conserved, unusual start mechanism and proposed that the TCG instead may serve as an unusual start codon

for insect mt-genomes. This proposed novel start codon has been questioned due to the frequency by which nucleotide transitions converting the TCG to a CCG codon has been observed in other *Drosophila* species (Ballard, 2000a; 2000b).

Characterization of *cox1* from other insect orders has added to the confusion regarding the start of the *cox1* gene. Some groups of Lepidoptera, such as the *Bombyx* and *Ostrinia* moths also share an aligning in-frame stop with the Diptera (Yukuhiro, *et al.*, 2002; Coates, *et al.*, 2005). In contrast, the butterflies of the genus *Papilio* appear to encode a highly variable sequence region that would align to the *Bombyx* and *Ostrinia* moths' in-frame stop codons and dipterian first nucleotide of the TCG serine (Caterino and Sperling, 1999).

Within most other species studies to date, neither a start nor stop codon can be identified adjacent to the first serine (Flook, *et al.*, 1995; Lunt, *et al.*, 1996; Friedrich and Muqim, 2003; Stewart and Beckenbach, 2003; Bae, *et al.*, 2004). Continuation of the open reading frame into the small non-coding region, or the adjacent $tRNA^{Tyr}$ gene eventually finds either a wobble-start codon or a stop codon. This is most likely due to random occurrence because of the high proportion of A and T nucleotides utilized in insect mt-genomes, not due to the coding of highly variable *cox1* reading frame sizes. It is becoming increasingly common in these circumstances to annotate the first in frame codon as the start of the *cox1* gene and simply annotate the gene with "start codon unknown". The four orders comprising the Paraneoptera in contrast appear to have evolved and maintain a methionine start codon (Shao, *et al.*, 2001; Shao and Barker, 2003; Shao, *et al.*, 2003; Thao, *et al.*, 2004; Stewart and Beckenbach, 2005). An alignment of insect *cox1* starts is presented in Figure 4.1.

This unusual *cox1* initiation is also very ancient. Examples are known in non-insect hexapods (Nardi, *et al.*, 2001; Nardi, *et al.*, 2003b). So far, the phenomenon is confined to the hexapods, as crustaceans appear to utilize ATN or NTG codons for initiation (Valverde, *et al.*, 1994; Crease and Little, 1997; Crease, 1999).

4.1.2. Mitochondrial Transcription and RNA Processing in *Drosophila*

Prior to the determination of a complete *Drosophila* mt-genome sequence, Spradling *et al.* (1977) reported twelve poly-T binding RNAs and one RNA that did not bind poly-T, that were synthesized in the presence of the antibiotic Actinomycin D and were found to hybridize to isolated mitochondrial DNA. Merten and Pardue (1981) also mapped the two ribosomal RNAs and "9 or 10" isolated mitochondrial RNAs to mt-DNA through the use of electron microscopy and hybridizing restriction digested mt-DNA fragments to the RNA in Northern Blot analyses.

After the release of the complete mitochondrial DNA sequence of *D. yakuba* (Clary and Wolstenholme, 1985a), Berthier *et al.* (1986) were finally able discern the gene identities of these mitochondrial RNAs through Northern blotting, and determine the relative abundance of the mature molecules. This study also described 28 minor RNA species that appeared to correspond to RNAs being processed into the mature mRNAs, tRNAs and rRNAs. Based on the pools of molecules observed, the authors proposed the existence of five transcriptional start sites in *Drosophila* mitochondria, two on the majority coding strand (one upstream of $tRNA^{Ile}$, and one upstream of $tRNA^{Thr}$) and three

```
                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|...
amino acid                                                                ?  ?  ?  W  l  f  S  T  N  H  K  D  I  G  T  l  y
D. melanogaster     acttaatc------------------------------------aaTAAtcgcgacaatgattatttctacaaatcataaagatattggaactttatat
D. yakuba           acttaatc------------------------------------caTAAtcgcgacaatggttatttctacaaatcataaagatattggaactttatat
screwworm fly       attTAAtc-------------------------------------------gcgacaatggttatttctactaatcataaagatattggtactttatat
Meditterian fly     attTAAtc-------------------------------------------gcaacaatggctatttcaactaatcataaagatattggaactttatat
olive fly           attTAAtc-------------------------------------------gcgacaatggctatttcaacaaatcacaaagatattggaactttatac
mosquito            attTAAtc-------------------------------------------gcgacaatgattatttcaacaaatcataaggatattggaactttatat
oakmoth             attttatt--------------------------------------aatTAGcgaaaatgacttattctacaaatcataaagatattggaactttatat
silkwormmoth        attttatt------------------------------tttttcttttTAGcgaaaatgaatttattctacaaatcataaagatattggaacattatat
cornborermoth       attttatt-----------------------------------ttaattTAGcgaaaatgactatattcaaccaatcataaagatattggaactttatat
sawfly*             ATTctacc-------------------------------------------gattaaatgattatattcaactaatcacaaagatattggaactttatat
honeybee*           aaatctta---------------------------------------ATAataaagtgattcatatcaaccaatcataaaaatattgggatcttgtat
flourbeetle         cacttcaa------------------------------------ctaaaCTGaataaatgattatttcaactaaccacaaagatattggaacacttat
asparagusbeetle     ATTtcacc-------------------------------------------gaataaatggctgtttctactaaccataaagatattggaacattatac
fireflybeetle       ATTttact-------------------------------------------aaataaatgattttattctactaatcataaaaatattggaactttatat
kissingbug          actttatc----------------------------------------cATGaataaatgactttctcaacaaaccacaaggatatcggaactctgtat
spittlebug          actttagt-----------------------------------------ATGaaaaaatgaatattctctactaatcacaaagatatcggaactttatat
leaf hopper         accttaat------------------------------------agcaATGaataaatgattatttctacaaatcataaggacattggaactctatac
whitefly            ctttattc------------------------------------- aATGaaaaaatgatttttctactaatcataaggatattggtatttatat
aphid               attttatc----------------------------------------tATA---aaatgaatctttcaactaatcataaagatattggaactttatat
psylid              cattttat------------------------------------ctgaATG---atttggctttttagaacaaaccacaaactattggtactttatac
barklouse*          ATTctaac------------------------------------------taataagtggatatttctacaaatcataaagatattggtactttatat
louse*              ctaattca----tgtaaaattttcatattacttagtacctctgcATGttttctcggtgattttattcttctaatcacaaaatattggtatttatat
thrips*             cattggtt---------------------------ATAacaaaaaatttagaaaaatgactttttcaacaaatcataaagacattggaatttatac
locust              ATCttatc------------------------------atcctcattaacgcaaaaatgattattctcaacaaaccacaaggacattggtacactgtat
mole cricket        attttatc----------------------------------------tATGcaacgatgattattctccaccaatcataaagatatcggaaccctttat
cockroach           attctact------------------------------------taccCTGcaacgatgaatatttcgacaaatcataaggatattggaactttatac
dragonfly           atcttacctaaattatattaaggggtcataatataatatttaagataTTGcgacgatggttatttttctactaatcataaggatattggaaccttatac
stonefly            ATTctacc-------------------------------------------gcgacaatgattattctcaacaaatcacaaggatattggaactttatac
silverfish          accctaCT-------------------------------------------Gaaacgatgattattctcgactaaccacaaagacattggtaccctatat
firebrat            atcctatt---------------------------------taatttcatTTGcaacgatgatatttttctacaaaccataaagatattggtacaatatac
archeognatha        ATTttatc-------------------------------------------gcaacgatggctattctcaacaaaccataaagacattggtaccttgtat
giantspringtail     actttatt----------------------------------------------taaatgaatctttcaacaaatcacaaagatattggaactatatat
springtail          actctATT--------------------------------------------ttacgctgaatatttcaacaaatcataaagatattggaactatatac
Daphnia             atttgact-------------------------------agcgtattaTTGcgacgttgactcttctcaactaatcataaggacattggcactctctat
```

Figure 4.1. Alignment of the 5' End of the *cox1* Gene for Various Insect, Springtail and Crustacean Representatives. Asterisks after the organism's common name represent a non-standard 5' abutting tRNA gene ($tRNA^{Tyr}$ for species without asterisks). Shaded nucleotides represent 80% or higher identity. Upper case nucleotides represent the first in frame start or stop codon for *cox1*. The first line represents a consensus amino acid translation where capital letters represent an amino acid always coded at that position in the alignment, lower case represents amino acids coded for over 80% of the time in the alignment, and "?" no amino acid is utilized over 80% of the time. The thick bar at the bottom highlights nucleotides encoding the last eight nucleotides of the abutting tRNA gene.

in the minority coding strand (one upstream of *srRNA*, one within *NAD6* gene upstream of $tRNA^{Pro}$, and one within the *cox1* gene, upstream of $tRNA^{Cys}$and $tRNA^{Tyr}$). Figure 4.2 summarizes the results of this study.

Transcription is thought to initiate at any of the five sites described above, then proceed around to the transcriptional termination site. At this point, the pre-mRNAs are recognized by an undescribed processing complex that identifies and removes the tRNA genes from within the transcript. The RNA fragments that are left behind correspond to the protein coding genes. These are believed to be identified, polyadenylated, and utilized as the mature mRNAs. The mature mRNAs encode no, or only a few, noncoding nucleotides, are not 5' capped as in nuclear transcript maturation, and have an exposed 5' monophosphate (Ojala, *et al.*, 1980b; Montoya, *et al.*, 1981). This processing model has been termed the tRNA punctuation model, and was first described for the human mitochondrial genome (Ojala, *et al.*, 1980a; Ojala, *et al.*, 1981).

4.1.3. 5' Amplification and Sequencing

The amplification of the 5' ends of linear DNA or of RNA molecules has long been problematic, due to the primer-dependence of the polymerases utilized in amplification. The most commonly used method to sequence 5' end of RNA molecules in to use a protocol called RACE (Rapid Amplification of cDNA Ends) (Frohman, *et al.*, 1988). Briefly, cDNA is produced via a reverse transcriptase enzyme. The resulting cDNA is then modified by the addition of a polynucleotide "tail" to its 3' end using the enzyme Terminal deoxynucleotidyl Transferase (TdT). This reaction extends a mononucleotide tail beyond the original RNA's 5' end. The modified cDNA is then

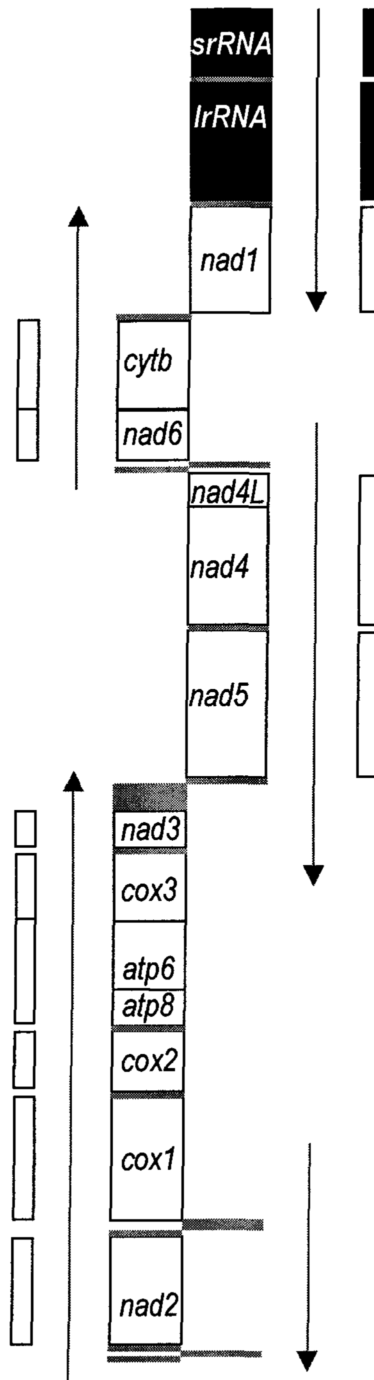


Figure 4.2. Summary of the Analysis of Mitochondrial Transcripts from Drosophila, as Described by Berthier *et al.* (1986). The genes are labeled in the center of the figure, with genes on one strand offset from the other. Arrows depict transcription units, with the start of the line at the predicted transcript initiation site and the arrow head at the predicted transcript termination site. White rectangles to the outside represent mRNAs, black the rRNA genes.

amplified through PCR, using a gene specific primer and a primer complementary to the nucleotide tail that was added to the cDNA.

The enzyme T4 RNA ligase has the ability to ligate together single stranded RNAs, or RNAs to DNAs (Tessier, *et al.*, 1986). T4 RNA ligase accomplishes this by covalently linking a 5' monophosphate to a free 3' hydroxyl group from the two oligonucleotides. T4 RNA ligase has been used to generate circularized tRNAs for amplification and sequencing. The protocol for the circularization of RNA molecules using T4 RNA ligase was originally developed for the sequencing of tRNA molecules to study RNA editing of these molecules in some mitochondrial systems (Janke, *et al.*, 1994; Yokobori and Paabo, 1995). This protocol has been adapted to protein coding genes for nuclear genes (Couttet, *et al.*, 1997) and mitochondrial genes in plants (Kuhn and Binder, 2002). Kuhn and Binder have dubbed this technique CR-RT-PCR. (CiRculariztion and RT-PCR).

In this chapter, I describe and characterize the mitochondrial mRNA and rRNA molecules from *Drosophila melanogaster*. The RNAs are amplified by a combination of 5' and 3' RACE protocols (Frohman, *et al.*, 1988) and CR-RT-PCR (based on the protocol of Couttet (1997). Characterization of the mature RNAs provide insights into mitochondrial transcript processing, maturation and expression in insect mitochondria. A potential model for the expression of the enigmatic *Drosophila cox1* gene is discussed.

4.2. Materials and Methods

4.2.1 Blast Searches for Nuclear Copies of the *cox1* Gene in *Drosophila*

The 511aa long open reading frame of two *cox1* mt-genes (see below) were tBlastn searched against the NCBI sequence database, restricted to *Drosophila melanogaster* sequence hits, and mitochondrial chromosome hits excluded. The standard genetic code was used to translate the *D. melanogaster* nuclear genome sequences.

The first *cox1* amino acid query sequence was derived from the *D. melanogaster* Oregon R mt-genome sequence (Accession AF200828.1, positions 1474 – 3006). The second from the ATG-containing *cox1* sequence (discussed in Chapter 2) of the spittlebug *Philaenus spumarius* mt-genome sequence (Accession AY630340.1, positions 1382 – 2914). The most recent search was conducted on May 18, 2005.

4.2.2. Maintenance of *Drosophila* Fly Stocks

Oregon R strain *Drosophila melanogaster* flies were maintained in a population cage at lab temperature (19-24°C range). Vials of fly food (See below) were placed in the population cage, as needed. Each vial was allowed to stay in the cage for approximately 3 weeks before they were removed. This typically allowed the majority of the pupae attached to the vial to emerge, but minimized the levels of mold and bacteria inside the population cage.

To make the fly food, 1L of tap water was brought to a boil. 3.25g of agar was slowly sprinkled into the water, with constant stirring with a wooden spoon. Next, 40g of sucrose, 20g of D-glucose, 0.36g of calcium chloride, and 2.8g of sodium-potassium-tartrate was dissolved in the mixture. Twenty grams of inactivated brewers yeast was then

slowly stirred into the mixture. The mixture was brought to a vigorous boil again, and then 47.6g of cornmeal was added and left to boil for 5 minutes. The mixture was removed from the heat and allowed to cool at room temperature for 7-10 minutes, until the mixture had thickened. 5ml of acid mixture (51.5% H_2O, 44.1% propionic acid, 4.4% phosphoric acid) was stirred in to the mixture, as a mold and bacterial inhibitor. The mixture was distributed into glass vials (approximately 20-25 ml per vial), and allowed to cool and solidify with cheesecloth covering the vials. The hardened mixtures were stored in a bag or with parafilm covering the vial mouth at 4°C until used.

Vials were left on the bench top to warm and excess water that had condensed and pooled on top of the food was poured off before use.

4.2.3. Maintenance of *Drosophila* Schneider-2 Cell Lines

Drosophila melanogaster Schneider-2 cell (Schneider, 1972) were generously provided by Darrell Bessette, from the lab of Dr. Esther M. Verheyen.

4.2.3.1. Complete S2/M3 Media

The entire vial of Shields and Sang M3 Insect Media Powder (39.4g, enough for 1L of culture media) was dissolved in 600ml of water, then the solution was topped up to 675ml. Next, 200ml of 5x BPYE (12.5g Difco Bacto-peptone and 5g Difco TC-yeastolate in 1L of water, autoclaved and stored at 4°C) and 0.5g potassium bicarbonate was added, and the mixture was filter sterilized through a Filtropur BT25 0.22μ bottle-top filter (Sarstedt), into bottle that had been autoclaved. Prior to use, FBS (Fetal Bovine Serum, Gibco™) and a Penicillin/Streptomycin antibiotic mixture (Invitrogen™ Life Technologies®) were thawed in a 28°C water bath. The sterile mixture of M3 media and BPYE were also warmed to 28°C. In the tissue culture hood, 125ml of FBS was added to

the M3+BPYE mixture, followed by the addition of 10ml of the Penicillin/Streptomycin antibiotic mixture. The bottle was capped again, and shaken to mix thoroughly. After the bubbles had disappeared, the media was used in cell culture.

The media was kept for up to 4 months at 2-4°C.

4.2.3.2. Cell Line Growth Conditions

All cell culture manipulations were carried out in a Laminar Flow Hood that had been wiped down with 70% ethanol prior to use. Cells were grown in T75 size culture flasks at 25°C in approximately 50ml of Complete S2/M3 Media.

Cells were allowed to grow for 2 weeks, then were split by the transfer of 5ml of old cell culture (containing cells) into an new culture flask containing 45ml of fresh media. As the cells adhere weakly to the culture flask, simply swirling of the flask was sufficient to free the cells and allow the splitting of the culture.

Culture flasks were maintained at 25°C in an incubator with atmospheric air conditions, with humidity maintained by placing open beakers of sterilized water into the incubator and allowing the water to evaporate.

4.2.4. Mitochondrial Isolation Protocols

4.2.4.1. Crude Mitochondria Isolation from Cell Culture

The Schneider-2 cells were transferred from the T75 culture dish to a 50ml Falcon Tube, which was spun at 1000rpm for 2 minutes to pellet the cells. The cells were resuspended in 250µl of cold MSB buffer per 0.1g of recovered wet cell mass. The MSB / cell suspension was transferred to an ice chilled Wheaton homogenizer. The cells were disrupted by ten strokes of the homogenizer. All stages of the mitochondrial

isolation was carried out at cold temperatures, either in a 4°C cold room or with tubes were kept on ice.

The aqueous homogenate was transferred to 1.5ml Eppendorf tubes in 1.0ml fractions, leaving the large tissue debris in the glass homogenizer. The homogenate was centrifuged in a desktop centrifuge for 5 minutes at 4000rpm in a 4°C cold room to pellet the cellular debris. The supernatant was drawn off to a fresh, pre-chilled tube that was centrifuged again for 5 minutes at 4500rpm in a 4°C cold room. The supernatant was then transferred to Beckmann 1.5ml polyallomer, Eppendorf-style ultracentrifuge tubes.

In some experiments, the homogenized solution was treated with 1μl of 35mg/ml chloramphenicol. Other experimental conditions did not include the chloramphenicol treatment. The supernatant was centrifuged for 20 minutes and 20,000 rpm in a Beckmann TLA-45 centrifuge rotor to pellet the crudely purified mitochondria. Discoloured (red, dark brown or milky white) pellets were resuspended again in 1.0ml MSB (containing chloramphenicol if the original solution was subjected to that treatment) and re-centrifuged for 20 minutes at 20,000 rpm.

Clean pellets containing the crude mitochondrial fraction were resuspended in 100μl of 1x TE (10mM Tris-Cl, 1mM EDTA, pH 7.4) and used immediately for RNA isolation. Alternately, the mitochondrial pellet was exposed to 100μl of 1x TE and 50μl of Denaturation Solution (from the Ambion Inc. TōTALLY RNA™ RNA isolation kit), vortexed to resuspend the mt-pellet, then placed at -80°C until the RNA extraction could be undertaken (see section 4.2.5).

4.2.4.2. Crude Mitochondrial Isolation for Adult flies

Adult flies were transferred to a glass Wheaton homogenizer that had been chilled on ice. When the insects had been slowed from the cold, approximately 10ml of cold MSB buffer (210.0mM mannitol, 70.0mM sucrose, 50.0mM Tris-Cl pH 7.5, 10.0mM EDTA) per gram of insects was added to the homogenizer. The tissues were disrupted by 10 strokes of the homogenizer, or more until the fly bodies were well homogenized.

The aqueous homogenate was recovered by centrifugation of the disrupted tissue, and was alloquated into 1.5ml Eppendorf tubes, leaving behind the organismal remnants. Mitochondria were then isolated by centrifugation, as described above (section 4.2.4.1).

4.2.5. RNA Extraction from Isolated Mitochondria

4.2.5.1. Special Precautions for the Handling of RNA

All protocols that involved the use of RNA were carried out is such a way to reduce the potential of contamination of the RNA samples by RNase. Isolation of RNA and the subsequent handling of RNA solution were carried out on a dedicated bench top with dedicated pipetters. The bench top and pipetters were cleaned with RNase Erase (ICN Biomedicals Inc.) prior to use. All pipette tips and tubes used were purchased from suppliers, with assurances of the RNase-free nature of the products. Primer stocks to be used in RT-PCR were dissolved in, and diluted to stock concentrations with DEPC-treated H_2O.

4.2.5.2. RNA Extraction Protocol

The Ambion Inc. TōTALLY RNA™ RNA isolation kit was utilized for the isolation of total mitochondrial RNA. The protocol followed was that specified by the instruction manual supplied with the kit, with minor modifications.

The 100μl crude mitochondrial pellet in TE was lysed by the addition of 50μl of the kit's Denaturation Solution. This mixture was vortexed for at least one minute to mix the solution and disrupt the mitochondria.

The disrupted mitochondria were extracted with 150μl of Phenol: Chloroform: Isoamyl Alcohol solution (supplied by kit). The mixture was vortexed for 1 minute, then allowed to sit on ice for 5 minutes. The extraction mixture was centrifuged for 5 minutes at full speed in an Eppendorf 5415C desktop centrifuge, at room temperature. The aqueous phase was removed by careful pipetting, and was transferred to a new, clean tube. Care was taken not to transfer any of the organic phase with the aqueous phase.

Sodium acetate solution (3.0M, pH 4.5, supplied with kit) was added to the aqueous phase, at one-tenth the volume of the recovered aqueous phase. The solution was then extracted with 150μl of Acid-Phenol: Chloroform solution (supplied by kit). The mixture was vortexed for 1 minute, left on ice for 5 minutes, then centrifuged, as described above. The aqueous phase was then transferred to an RNase free 0.6ml Eppendorf tube, and mixed with an equal volume of isopropanol. The solution was allowed to sit at -20°C overnight to precipitate.

RNA was pelleted by centrifugation at full speed in an Eppendorf 5415C desktop centrifuge for 20 minutes. The isopropanol was removed with a pipetter, and the pellet was washed with 200μl of 70% ethanol. The pellet was centrifuged again for 15 minutes, and the ethanol was removed with a pipetter. The pellet was allowed to dry within 30cm of a Bunsen burner for at least 20 minutes, until no trace of liquid could be seen in the tube. For long-term storage of RNA, the pellets were stored dry at -80°C. Samples to be used immediately were dissolved in 100μl of kit-supplied DEPC treated water with

0.1mM EDTA, and allowed to dissolve on the bench top until the RNA pellet was no longer visible.

4.2.5.3. Removal of Contaminant DNA by Lithium Chloride Precipitation.

Lithium chloride precipitation is employed to remove gross DNA contamination from the samples, but also removes small RNA molecules, such as tRNAs or 5s rRNA. One half volume of kit-supplied LiCl was added to the dissolved RNA sample and left at -20°C for at least 1 hour. The large RNA molecules were then pelleted by centrifugation in a desktop centrifuge at full speed for 20 minutes. The supernatant containing small RNAs and DNA was removed and stored in a separate RNase-free Eppendorf tube. The RNA pellet was then washed with cold 70% ethanol, and allowed to dry as described above.

Removal of DNA by DNase became the preferred method to obtain clean RNA samples. As a result, the less effective Lithium Chloride precipitation method was discontinued in the early stages of the project, and was only used on the Schneider-2 cell line RNA isolations.

4.2.5.4. Removal of Contaminant DNA by DNase Digestion.

A second method for the removal of contaminant DNA was to treat the dissolved RNA with Ambion® TURBO™ DNase (RNase-free). RNA samples were adjusted to a 1x concentration, using the 10x DNase buffer supplied. Two units (1μl) of TURBO™ DNase was added per 49μl of RNA-DNase buffer solution. The solution was incubated 30 minutes at 37°C. The solution was extracted using the Ambion® TōTALLY RNA™ RNA isolation kit as described previously, but using one-tenth volume of denaturation solution, for the extraction of the DNase-treated samples.

4.2.6. Reverse-Transcription PCR Protocols

4.2.6.1. Primer Design

A set of primers were designed to specifically amplify the RNA molecules, based on comparison of the *D. melanogaster* complete mitochondrion sequence (GenBank Accession U7541.1; (Lewis, *et al.*, 1995)) to the *D. melanogaster* Oregon R strain near-complete mitochondrial sequence (GenBank Accession AF200828; (Ballard, 2000b)). Some heterologous primers (described in Chapters 2 and 3) which matched exactly the two *D. melanogaster* mt-DNA sequences were utilized as well. Both sequences were utilized to re-confirm the identity of nucleotides for primer design, and to check that the chimeric *D. melanogaster* mt-DNA sequence completed in 1995 did not contain strain-specific sequence variations that may have affected primer performance.

Primers are listed in Table 4.1 and Table 4.2.

4.2.6.2. Standard RT-PCR Protocol

The Enhanced Avian HS RT-PCR Kit (Sigma®) was used for Reverse-Transcription PCR (RT-PCR) reactions. Reaction volumes used varied depending on the amount of amplified product required. For direct sequencing of the generated RT-PCR amplification products, 50µl reactions were used, while 25µl were sufficient for products that were to be cloned, or visualized but not sequenced.

Final concentrations for the reactions were 3.0mM $MgCl_2$, 200µM of each dNTP, 0.4 units/µl of the supplied RNase Inhibitor Enzyme, 0.4µM of each primer, 0.4 units/µl of eAVV-RT reverse transcriptase, 0.05 units/µl of the JumpStart AccuTaq LA DNA polymerase, in a 1x reaction buffer. Approximately 0.3ng (in 1µl) of total RNA extract was utilized per 25µl of RT-PCR reaction buffer volume.

Table 4.1. Primers Used in 5' and 3' RACE Amplification.

Gene	3' Race Primer	Sequence (5'-3')	5' Race Primer	Sequence (5'-3')
nad2	N2-J953	AAATTTTTTATCATTAGGAGGATT	N2-N562	AAATCAAAAATGGAAAGGAGCG
cox1	C1-J2203	GGTCACCCTGAAGTTTATATT	C1-N1701	AAAGGGCACTAATCAATTTCC
	C1-J2636	ATAGGRGCWGTATTTGCYATTAT	C1-N1828	GTTCATCCTGTWCCWGCWCC
	C1-J2791	ATACCTCGACGTTATTCAGA	C1-N1843	AAGGGAGGGTACACAGTYCA
cox2	C2-J3198	TAATAATTACAGTATTGGTGGG	C2-N3152	CTAAAATTAATAATGCATGATCATG
	-		C2-N3389	TCATAACTTCAGTATCATTG
	-		C2-N3665	CCACAAATTTCTGAACATTG
atp8	A8-J3927	TTTTATTCCTCAAATAGCWCC	-	
atp6	A6-J4526	ATTAGAAATATWATYCGRCC	-	
cox3	C2-J4777	AATCACCCMTTTCAYTTAGTWGA	C3-N4980	ATAAATARAATTATTCCTCATCG
	C3-J5188	TACTGTAACTTGAGCCCACCA	-	
nad3	N3-J5723	AATCGACCGAGAAAAAAGATC	N3-N5677	GCTTTTTTTGATAAAATTGAAGC
nad5	N5-N7076	GGAATACCATTTTTAGCTGG	-	
nad4	N4-N8871	TACTAAAATTAGGAGGTTATGG	-	
nad4L	N4L-N9771	TTGTTTCTAATCGGAAAC	-	
nad6	N6-J10176	AATTTTTTTAGGAGGAACATTTG	-	
cytb	CB-J11335	CATATTCAACCWGAATGRTA	-	
nad1	N1-N12059	CCTTTTGATTTTGCTGAAGGAG	-	
lrRNA	LR-N13396	CGCCTGTTTANCAAAAACATG	-	
	N3-J5723	AATCGACCGAGAAAAAAGATC	-	
	pT-L01	ATTGCGATTGCGTTGCGA(T)18	pT-L01	ATTGCGATTGCGTTGCGA(T)18
	-		pA-L01	ATTGCGATTGCGTTGCG(A)18

Table 4.2. Primers Utilized in CR-RT-PCR Reactions

Gene	5' Facing Primer	Sequence	3' Facing Primer	Sequence
nad2	N2-N562	AAATCAAAAATGGAAAGGAGCG	N2-J965	AAATTTTTTATCATTAGGAGGATT
cox1	C1-N1540	TGTTCCAACTATTCCAGCTCA	C1-J2791	ATACCTCGACGTTATTCAGA
	C1-N1701	AAAGGGCACTAATCAATTTCC	C1-J2203	GGTCACCCTGAAGTTTATATT
cox2	C2-N3152	CTAAAATTAATAATGCTAGATCATG	C2-J3198	TAATAATTACAGTATTGGTGGG
tRNA^{Lys}	-		TK-J3790	CATTAGATGACTGAAAGCAAGTA
tRNA^{Asp}	-		TD-J3837	GTTAAAATCATAACATTAGTATGTC
	-		TD-J3862	ATAAAAAATTAGTTAAAATCATAAC
atp8	A8-N4014	TTTAATTCATTAGATTTAGGTG	-	
atp6	A6-N4456	AAGGTATAAGAATAGCGGGTGT	A6-J4522	CGACCTGGAACATTAGCTGTTCG
cox3	C3-N5322	GTTGCTATAAAAAATGTTGATCC	C3-J5375	CGGAATTCATGTATTAATCGGA
nad3	N3-N5821	AGGTAGAATTAATGCAATCTC	N3-J5723	AATCGACCGAGAAAAAAGATC
	N3-N5677	GCTTTTTTTGATAAAATTGAAGC	-	
tRNA^{Ala}	-		TA-J6016	ATATAACATTTGATTTGCATTC
tRNA^{Arg}	-		TR-J6089	TTGATTGCAATTAGTTTCGACC
tRNA^{Asn}	-		TN-J6148	GCCAAAAAGAGGCATATCACTG
tRNA^{Ser}	TS-N6191	TTAGCAGCTTTTACTTGATCATC	TS-J6240	TTTCTTTTAATGGTTAAATTCCA
tRNA^{Glu}	-		TE-J6288	AAATAAAACCTTACATTTTCATTG
tRNA^{Phe}	-		TF-J6361	CCAAAGATTTAATAATCTCCAT
nad5	N5-J7806	GAMACAARACCTAACCCATCYCA	N5-N6881	AAAGTTGAATTATACTCCGTGG
nad4	N4-J9157	ATAAACCAGCTTGTAAACGTT	N4-N8871	TACTAAAATTAGGAGGTTATGG
	N4L-J9597	TTGAAAATAATCATTTCCATG	N4-N8283	TTAGAAGAGGTAAAATTCGAGA
nad4L	N4L-J9648	TTGTTTCTAATCGGAAAC	-	
nad6	N6-N10488	AGATATTATTCGAATAGGTCC	N6-J10176	AATTTTTTTAGGAGGAACATTTG
	N6-N10292	AAAATAAAGAAGAAGAAGTTTTATC	N6-J10409	AAATAAATTATATAATTTTCCCACA
cob	CB-N11114	ATTAGATCCTGTTTGATGAAGG	CB-J11278	CAAATTTATTGGGAGACCCTGA
nad1	N1-J12099	AAGGATCCGATTAGTTTCAGCT	N1-N12059	CCTTTGATTTTGCTGAAGGAG
	N1-J12278	GAGCCACAGCTCGCAAACCYCC	-	
	N1-J12337	ACTATAACAGTATAAACCCCCA	-	
lrRNA	LR-J13900	TTTGATAAACYCTGATACAMAAG	LR-N12866	ACATGATCTGAGTTCARACCGG
srRNA	LR-J14766	TAACCGCAACTGCTGGCAC	SR-N14472	TTTAATCGATAATCCACGATGG

The reaction was incubated at 42°C for 60 minutes to allow for the reverse transcription reaction. The temperature was then elevated to 94°C for 2 minutes, followed by 35 cycles of 94°C for 15 seconds, 50-55°C for 30 seconds, and 68°C for 120 seconds. The reaction was finished by a 5-minute soak at 68°C before reducing the reaction temperature to 8°C to end the reaction.

The large tRNA cluster between the *nad3* and the *nad5* genes appeared to be capable of forming secondary structures that inhibited reverse transcription. To amplify this region by RT-PCR the RT-PCR reaction was set up without the enzymes present, then exposed to 92°C for 30 seconds then transferred directly to ice. The enzymes were then added and the RT-PCR reaction carried out as described above.

4.2.6.3. Control Reaction for DNA Contamination of RNA Samples

Each total RNA sample underwent the same control experiment to ensure the presence of RNA in the sample, and to ensure no contamination by DNA. Two RT-PCR reactions were set up (as per section 4.2.7.2), one that had its RNA digested by the addition of 10ng of RNase AI for one half hour on the bench top prior to the addition of the RT-PCR mixture, while the second vial contained intact RNA as template.

The region amplified by the RT-PCR covered a region spanning the genes *nad2*, $tRNA^{Trp}$, $tRNA^{Cys}$, $tRNA^{Tyr}$, and *cox1*, with the amplification product measuring 759bp in total length. The primers used for this control were N2-L965 and C1-H1702 (see Table 4.1).

If both tubes showed an absence of amplification, the RNA extract was assumed to have failed, and the template was disposed of. If amplification of appropriate sized bands

for RNase digested and undigested samples, the sample was interpreted as having contaminating DNA. The RNA sample was again subjected to DNase digestion (section 4.2.5.2). Amplification only in the RNase-free tube showed the absence of contaminant DNA in the sample, and the presence of RNA through the RT-PCR amplification of the target region. These samples were utilized is subsequent experiments.

4.2.7. Protocol for 5'-Race Amplification

The RACE protocol (Rapid Amplification of cDNA Ends) is modified slightly from that originally presented by Frohman *et al.* (1988). Approximately 0.3ng of total mitochondrial enriched RNA template was reverse transcribed using either the enzyme supplied in the Enhanced Avian RT-PCR kit (eAMV-RT by Sigma®) or by the M-MLV RT enzyme (Ambion®). Conditions for the reactions were as directed by the enzyme supplier, with either a gene-specific primer for amplification of the target *Drosophila* mitochondrial gene, or a poly-$T_{(18)}$ primer, designed to anneal to mRNA's poly-A tail. Reverse transcription was conducted for one hour at 42°C (for eAMV-RT) or at 37°C (for M-MLV-RT).

The cDNAs were precipitated using 2 volumes of 95% ethanol over night. In the morning, the pellets were washed with 70% ethanol, then dissolved in 10µl H2O containing 1ng RNase A1 to remove the residual RNA from the solution. A TdT reaction was then used to add a mononucleotide tail of a single nucleotide to the 3' end of the generated cDNA molecule. The reaction was carried out in a total volume of 25µl, containing 1x of the supplied reaction buffer, 1.65mM of the chosen dNTP, and 6U of the rTdT enzyme (Invitrogen™ Life Technologies). The solution was incubated for 60 minutes at 30°C, then passed through a QIAGEN® QIAquick® PCR column (see Chapter

3, section 3.2.4). The QIAquick® purified cDNA was then precipited overnight at -20°C with 2 volumes of 95% ethanol. PCR of the modified cDNA was then conducted with a gene specific primer as well as a mononucleotide-binding primer. The primers used are listed in Table 4.1.

4.2.8. Protocol for 3'-RACE Amplification

Reactions for 3'-RACE were set up as described above (section 4.2.7), except that a single gene-specific primer was used. The second primer consisted of 18 T's (poly-T_{18}) at the 3' end, with a 5' unique sequence motif not found the *D. melanogaster* mitochondrial genome (primer pT-L01). The unique sequence was designed to raise the annealing temperature of the primer, after the reverse transcription step had incorporated the primer into the cDNAs, allowing for annealing of 50°C or more, depending on the nature of the matched gene-specific primer.

4.2.9. RNA Circularization for RT-PCR

RNA samples were circularized with T4 RNA ligase (Fermentas Life Sciences) using the supplied buffers, ATP and BSA. To decrease the probability of ligating different RNA molecules together, and increasing the pool of circularized RNA molecules, I modified the protocol from Couttet *et al.* (1997). Approximately 6.0ng of total RNA was diluted in a 200µl reaction (20µl of 10x reaction buffer, 1µl SUPERase•In RNase Inhibitor (Ambion®), 8µl of 1mg/ml BSA, 2µl of the 10mM ATP and 0.5 µl of 10U/µl T4 RNA ligase). The mixture was incubated at 16°C for at least 16 hours. After circularization, the RNA was extracted as described in section 4.2.5.

The RNA was the precipitated with an equal volume of isopropanol, incubation at -20°C for at least 4 hours, precipitation for 20 minutes at top speed in a bench top centrifuge, and the pellet was washed with 70% ethanol.

The RNA was diluted to approximately 0.3 ng/µl and utilized in a standard RT-PCR reaction (Section 4.2.6.2). The primers utilized were selected such that a primer near the 3' of the RNA would produce the first cDNA strand from the ligated RNA circle, through the 3' region of the molecule, across the ligation site, then to the 5' end of the molecule (see Figure 4.3).

4.2.10. Cloning of Amplification Products

The TOPO TA Cloning® Kit, Version R (Invitrogen™ Life Technologies) was used for cloning. This kit utilizes the Topoisomerase enzyme, linked to the plasmid via a 3' phosphate to each end of the cloning vector's insertion site to carry out the ligation of the insert into the plasmid vector.

Four µl of PCR or RT-PCR products was added to 1µl of the supplied salt solution and 1µl of the kit's vector. The reaction was carried out at room temperature for up to 30 minutes, before being transfected in to the supplied One Shot® chemically competent *E. coli* cells, as directed by the kit's manual.

Cells were plated onto LB agar plates containing 50µg /ml of kanamycin, 10mg / ml of X-gal, and 100mM of IPTG, using 10µl, 50µl and 150µl of the transfected cells. Putative colonies were selected via blue / white screening, and grown up overnight in 3ml LB cultures containing 50µg /ml of kanamycin. After isolation of the plasmid by alkaline lysis miniprep methods (Sambrook, *et al.*, 1989), the plasmids were selected for sequencing based on the size of the insert that was excised by digestion of two the

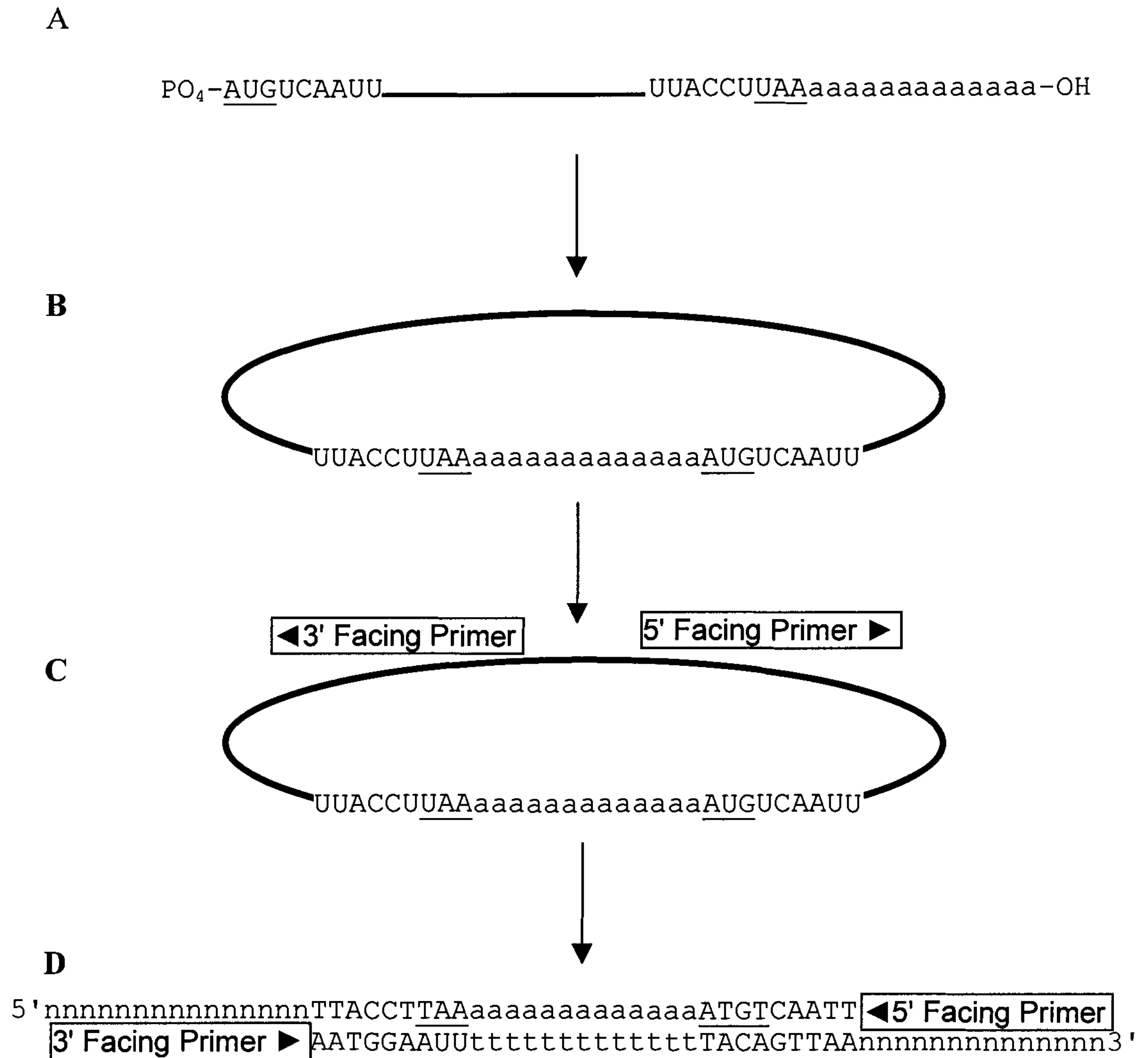


Figure 4.3 A Schematic Representation of the Circularization RT-PCR Reaction. A – Mono-phosphate mitochondrial mRNA molecules. B – After incubation with T4 RNA ligase in dilute conditions, the molecule has been 5' to 3' ligated to produce a circular RNA. C –RT-PCR is conducted using gene specific primers.
D - Simultaneous generation of 5' and 3' cDNA amplification products of the original mRNA molecule.

flanking restriction sites with the restriction enzyme *EcoRI* (Invitrogen™ Life Technologies).

4.2.11. Sequencing of cDNAs and Sequence Alignment

All amplification products were isolated using the QIAquick PCR purification kit (see Chapter 3, Section 3.2.4) eluted in 40μl glass-distilled ddH_20. Samples were prepared for sequencing, and sequenced by the University of Calgary DNA Sequencing Center (Chapter 3, section 3.2.5).

Obtained cDNA sequences were aligned to both the *D. melanogaster* mt-DNA sequence (Genbank Accession AF200828.1) and the complete *D. melanogaster* complete mitochondrion sequence (Benbank Accession U37541.1) in order to characterize the boundaries of the mature RNA molecules.

4.3. Results

4.3.1. Search for a Nuclear Copy of *cox1*

The tblastn analysis of the complete *D. melanogaster* genome sequence from GenBank revealed no putative nuclear encoded copies of the *cox1* gene. Blast searches with the translated open reading frame from both *D. melanogaster* and the spittlebug revealed the same small regions of similarity. Three of the hits appear to represent the same DNA region on chromosome 4, and two hits appear to originate from chromosome 2L (see Table 4.3). Both hits revealed only small portions of the *cox1* open reading frame (35 and 53aa versus the *D. melanogaster* query). Inspection of the Blast-generated alignment lead to the conclusion that these are most likely the remnants of two nuclear

Table 4.3. Results of tBlastn Searches for a Nuclear *coxI* Gene. The Results of a tBlastn search of the translated mitochondrial *coxI* amino acid sequence (511 amino acids) from *D. melanogaster and Philaenus spumarius*. Blast hits were limited to *D. melanogaster* sequence only, and excluded mitochondrial DNA. E Value, Hit Length, Identities and Positives list the values from both *D. melanogaster* (left number) and *Philaenus spumarius* (right number). The result at the bottom (AF200828.1) shows the best hit obtained by tBlastn search against mitochondrial DNA only.

Accession	Description	E Value	Hit Length	Identities	Positives
AC006467.13	Drosophila melanogaster clone BACR03L08, complete sequence	7e-07/ 3e-05	35 / 41	27 / 25	21 / 31
AC150553.1	Drosophila melanogaster clone BACR09H20, complete sequence	7e-07/ 3e-05	35 / 41	27 / 25	21 / 31
AE003781.5	Drosophila melanogaster chromosome 2L, section 80 of 83 of the complete sequence	7e-07/ 3e-05	35 / 41	27 / 25	21 / 31
AE003844.5	Drosophila melanogaster chromosome 4, section 2 of 5 of the complete sequence	4e-04/ 1.2	53 / 58	35 / 28	40 / 28
AC114394.1	Drosophila melanogaster, chromosome 4, region 101B-101C, BAC clone BACR33O09, complete sequence	4e-04/ 1.2	53 / 58	35 / 28	40 / 28
AF200828.1	Drosophila melanogaster isolate Oregon R mitochondrion, complete genome	0.0 / 0.0	511/ 507	420/ 329	439/ 393

pseudogenes of the mitochondrial *cox1* gene, complete with TGA-stop codons aligning to the TGA-Trp codons of the query sequences. Nuclear pseudogene sequences of mitochondrial origin are commonly encountered in insects (Sunnucks and Hales, 1996; Bensasson, *et al.*, 2000) and other groups of organisms (reviewed in Bensasson, *et al.*, 2001; Bensasson, *et al.*, 2003).

These blast results imply that the translatable sequence from the *D. melanogaster* complete genome do not contain an open reading frame of suitable composition to suggest a nuclear version of the *cox1* gene. This adds further evidence to the belief that the mitochondrial version is the functional version of the protein.

4.3.2. RNA Extraction from Mitochondria

Thiamphenicol and chloramphenicol have been shown to inhibit protein synthesis in animal mitochondria (Jager and Bass, 1975). Studies with thiamphenicol have shown that this class of antibiotic's inhibition of protein synthesis results in an increase in steady-state levels of mature mitochondrial RNAs (Chrzanowska-Lightowlers, *et al.*, 1994; Selwood, *et al.*, 2001).

Mitochondrial extractions were carried out both with and without chloramphenicol being added after the tissues were homogenized to compare levels of mitochondrial RNA obtained. Comparisons of mt-RNA yield between chloramphenicol-treated mitochondria and untreated mitochondria showed very little difference in the amount of RNA obtained. Total mt-RNA yields ranged from 100 to 150 ng per 1.0g of adult flies homogenized, for both treatments. Variation between duplicate tubes from the same homogenization equaled the difference between samples treated with chloramphenicol and samples not treated with chloramphenicol. As a result, further

experimentation with chloramphenicol added to the mitochondrial preparation was determined to be irrelevant and was discontinued.

4.3.3. Detection of Unprocessed pre-mRNA Molecules in Flies

RT-PCR was conducted on the adjacent protein-coding and ribosomal RNA genes to determine if large pre-processing transcripts could be observed at RT-PCR detectable levels. I tested whether a gene-specific primer from each gene, from *nad2* (position 965) around the genome to *srRNA* (position 14746 in OreR *D. melanogaster*), could amplify in a RT-PCR reaction with primers from either of their adjacent genes. This observation included RNA molecules that encoded genes that are on opposite strands of the mt-DNA molecule at the *nad3-nad5, nad4L-nad6* and *cytb–nad1* gene boundaries. The bands were sequenced from a single primer to confirm the identity of the amplification products. Amplification from the *nad2* to *srRNA* was not attempted.

Detection of RNA molecules containing three or more protein coding genes were not detected by this methodology. The inability to detect such molecules could be due to the inability of the reverse transcriptase enzyme to produce cDNAs of sufficient size to serve as a PCR template, or the absence of the molecules in the total RNA pool. Multi-gene pre-mRNA molecules have been observed in *Drosophila* by others (see (Berthier, *et al.*, 1986)), therefore it appears that the RT enzymes lack the ability to generate sufficiently large cDNAs for amplification.

4.3.4. Mapping of mRNA and rRNA Boundaries by RACE

The 3' end of all mRNA and the *lrRNA* molecules were mapped by the sequencing of cDNAs generated by 3' RACE. The presence of the post-transcriptionally added poly-A

tail made it impossible to determine whether A-nucleotides at the 3' end of the transcripts were the result of polyadenylation versus those that were transcribed from the mitochondrial DNA.

4.3.4.1. Characterization of RNA 3' Ends Detected by 3' RACE Techniques

The 3' ends of eleven mRNAs were determined using 3' RACE protocols. Two of the molecules were bicistronic messages, with one coding both the *atp8* and *atp6* genes, and the other encoding both *nad4L* and *nad4*. The remaining nine mRNAs were single gene mRNAs for the remaining nine mitochondrial genes. An alignment of the 3' cDNA sequences is presented in Figure 4.4. .

A polyadenylated *lrRNA* was amplified and sequenced as well. The presence of a polyadenylated *lrRNA,* has been discussed in mammal and insect mitochondrial genomes (Ojala, *et al.*, 1980a; Merten and Pardue, 1981; Berthier, *et al.*, 1986; Benkel, *et al.*, 1988). A cDNA from the *srRNA* could not be amplified using 3' RACE techniques.

An interesting population of molecules was detected at the 3' end of the *nad3* gene. Distinct polyadenylated RNA species were detected that contained just the *nad3* gene, as well as the *nad3* gene with one or more of the 3' abutting tRNA genes still included in the RNA. To further investigate these molecules, tRNA specific primers were designed and used in CR-RT-PCR (see below, section 4.3.6.2).

4.3.4.2. Characterization of RNA 5' Ends Detectable by 5' RACE Techniques

The 5' sequence for five RNA species was determined by 5' RACE methods. The *nad2*, *cox1, cox2 cox3* and *nad3* genes were amplified and sequenced using this protocol. The *nad2, cox1, cox2,* and *nad3* genes were detected after the addition of a poly-T stretch upstream of the first nucleotide of the message, and *cox1* and *cox3* detected after the

```
....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|

nad2
TCAATTTTTGGATTATTTTTAATTTCTTTATTTTATTTTATATTTTaaggctttaagttaataaaactaataaccttcaa   1297
UCAAUUUUUGGAUUAUUUUUAAUUUCUUUAUUUUAUUUUAUAUUUUAA(Poly-A)
 S  I  F  G  L  F  L  I  S  L  F  Y  F  M  F  *
cox1
CCGCCAGCTGAACATAGATATTCTGAATTACCACTTTTAACAAATTAATTtctaatatggcagattagtgcaatagattt   3041
CCGCCAGCUGAACAUAGAUAUUCUGAAUUACCACUUUUAACAAAUUAAUU(Poly-A)
 P  P  A  E  H  S  Y  S  E  L  P  L  L  T  N  *
cox2
GTTCCTGTAAATTACTTTATTAAATGAATTTCTAGAAATAACTCTTcattagatgactgaaagcaagtactggtctctta   3801
GUUCCUGUAAAUUACUUUAUUAAAUGAAUUUCUAGAAAUAACUCUU(Poly-A)
 V  P  V  N  Y  F  I  K  W  I  S  S  N  N  S  *
atp8/atp6
TATGTGTTTGCTGTATTAAGAACTTTATATTCTAGAGAAGTAAATTAatgtctacacactcaaatcaccctttccattta   4771
UAUGUGUUUGCAGUAUUAAGAACUUUAUAUUCUAGAGAAGUAAAUUAA(Poly-A)
 Y  V  F  A  V  L  S  T  L  Y  S  S  E  V  N  *
cox3
GTAGTTTGATTATTTTTATATATCACAATTTACTGATGAGGAGGATAATTATATTATTAATTAAATatctatatagtata   5656
GUAGUUUGAUUAUUUUUAUAUAUCACAAUUUACUGAUGAGGAGGAUAAUUAUAUUAUUAAUUAAAUA(Poly-A)
 V  V  W  L  F  L  Y  I  T  I  Y  W  W  G  G  *
nad3
GGATTATACCATGAATGAAATCAAGGAATATTAAATTGATCAAACTAATATATATTTATATATATATAgggttgtagtta   6093
GGAUUAUACCAUGAAUGAAAUCAAGGAAUGUUAAAUUGAUCAAACUAAUAUAUAUUUAUAUAUAUAUA(Poly-A)
 G  L  Y  H  E  W  N  Q  G  M  L  N  W  S  N  *
nad5
TTATTTGTATTTTGAATTTTAATTTTATTAATTTTATTATTTTTATatccaaatagcttatactagagtttgacattgaa   6368'
UUAUUUGUAUUUUGAAUUUUAAUUUUAUUAAUUUUAUUAUUUUUAUA(Poly-A)
 L  F  V  F  W  I  L  I  L  L  I  L  L  F  L  *
nad4L/nad4
TTAAATTTATTAATTTTAAAAAGAGAATCATTTATATTATGATTATatttaaatagtttaaaaaaaatactaatttgtgg   8172'
UUAAAUUUAUUAAUUUUAAAAAGAGAAUCAUUUAUAUUAUGAUUAUA(Poly-A)
 L  N  L  L  I  L  K  S  E  S  F  M  L  W  L  *
nad6
GTAAAAATTACAAAATTATTTAAAGGACCTATTCGAATAATATCTTAATTAatgaataaacctttacgaaattcccatcc  10526
GUAAAAAUUACAAAAUUAUUUAAAGGACCUAUUCGAAUAAUAUCUUAAUUAA(Poly-A)
 V  K  I  T  K  L  F  K  G  P  I  R  M  M  S  *
cytb
TTAGTAAACCCATTAATTACAAAATGATGAGATAATTTATTAAATTAAATagttaatgagcttgaataagcatatgtttt  11666
UUAGUAAACCCAUUAAUUACAAAAUGAUGAGAUAAUUUAUUAAAUUAAAUA(Poly-A)
 L  V  N  P  L  I  T  K  W  W  D  N  L  L  N  *
nad1
TATTTATTTTTTATTGGGTTTAAAATTTTATTATTTTCTTTATTATAGTGAATTTTTTTTAGTAAaagttaatagaaaat  11688'
UAUUUAUUUUUUAUUGGGUUUAAAAUUUUAUUAUUUUCUUUUUUAUAGU(Poly-A)
 Y  L  F  F  I  G  F  K  I  L  L  F  S  L  L  *

lrRNA
AATATCAAAATAATTATATTTTTTATAAGAATATTATTAATATAAAactattttggcagattagtgcaataaatttagaa  12700'
AAUAUCAAAAUAAUUAUAUUUUUUAUAAGAAUAUUAUUAAUAUAAAA(Poly-A)

srRNA
ATAAGTCGTAACATAGTAGATGTACTGGAAAGTGTATCTAGAATGAcaatttaaagcttattaagtaaagtatttcattt  14097'
AUAAGUCGUAACAUAGUAGAUGUACUGGAAAGUGUAUCUAGAAUGAaaa(aaa)
```

Figure 4.4 Alignment of 3' cDNA Sequence to the Mitochondrial DNA Sense Strand for *Drosophila melanogaster*. The top sequence is derived from the *D. melanogaster* OreR sequence. Lower case sequence represents nucleotides that encode the abutting gene. Numbers to the right represent the genome position of the right-hand most nucleotide (with ' representing a minority strand encoded sequence). The second strand represents the sequence obtained from the cDNA (illustrated as an RNA sequence to differentiate from the genomic sequence).

```
      ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|

nad2
223   taatccttttctttttaATTTTTAATAATTCGTCAAAAATTTTATTTATTACAATTATAATTATTGGGACATTAATTACA
#                      AUUUUUAAUAAUUCGUCAAAAAUUUUAUUUAUUACAAUUAUAAUUAUUGGGACAUUAAUUACA
                        I  F  N  N  S  S  K  I  L  F  I  T  I  M  I  I  G  T  L  I  T
cox1
1453  agccacttaatcAATAATCGCGACAATGATTATTTTCTACAAATCATAAAGATATTGGAACTTTATATTTTATTTTTGGA
#                      UCGCGACAAUGAUUAUUUUCUACAAAUCAUAAAGAUAUUGGAACUUUAUAUUUUAUUUUUGGA
                        S  R  Q  W  L  F  S  T  N  H  K  D  I  G  T  L  Y  F  I  F  G
cox2
3066  cttttattagaaAATAAATGTCTACATGAGCTAATTTAGGTTTACAAGATAGAGCTTCTCCTTTAATAGAACAATTAATT
cox-2a        aaaaUAAAUGUCUACAUGAGCUAAUUUAGGUUUACAAGAUAGAGCUUCUCCUUUAAUAGAACAAUUAAUU
# cox2-b               AUGUCUACAUGAGCUAAUUUAGGUUUACAAGAUAGAGCUUCUCCUUUAAUAGAACAAUUAAUU
                      *  M  S  T  W  A  N  L  G  L  Q  D  S  A  S  P  L  M  E  Q  L  I
atp8/atp6
4045  ataattaatattttttaATTCCACAAATAGCACCTATTAGATGATTATTATTATTTATTATTTTTTCTATTACATTTATT
                      aAUUCCACAAAUAGCACCUAUUAGAUGAUUAUUAUUAUUUAUUAUUUUUUCUAUUACAUUUAUU
                        I  P  Q  M  A  P  I  S  W  L  L  L  F  I  I  F  S  I  T  F  I
cox3
4719  tctagagaagtaaattaATGTCTACACACTCAAATCACCCTTTCCATTTAGTGGATTATAGTCCATGACCATTAACAGGA
                      aaUGUCACCACACUCAACUCACCCUUUUCAUUUAGUGGAUUAUAGUCCAUGACCAUUAACAGGA
                        M  S  T  H  S  N  H  P  F  H  L  V  D  Y  S  P  W  P  L  T  G
nad3
5590  aattaatagtatagataATTTTTTCTATTATTTTTATTGCTTTATTAATTTTACTAATTACAACTATTGTTATATTTTTA
#                      AUUUUUUCUAUUAUUUUUAUUGCUUUAUUAAUUUUACUAAUUACAACUAUUGUUAUAUUUUUA
                        I  F  S  I  I  F  I  A  L  L  I  L  L  I  T  T  I  V  M  F  L
nad5
8148' aatattcattttaaatcATGAAATATTTATCTATTTGTAGAATTAGATTTGTTAATTTAATTTCTATAAGTTTATCATGT
                       aUGAAAUAUUUAUCUAUUUGUAGAAUUAGAUUUGUUAAUUUAAUUUCUAUAAGUUUAUCAUGU
                        M  K  Y  L  S  I  C  S  I  S  F  V  N  L  I  S  M  S  L  S  C
nad4L/nad4
9843' tatttcttttaaaactAATGATTATAATTTTATATTGAAGTTTACCTATAATTTTATTTATTTTAGGCTATTTTGTTTT
                      aaUGAUUAUAAUUUUAUAUUGAAGUUUACCUAUAAUUUUAUUUAUUUUAGGCUAUUUUGUUUU
                        M  I  M  I  L  Y  W  S  L  P  M  I  L  F  I  L  G  L  F  C  F
nad6
9945  atttcttttctcttgaAATTATTCAATTAATATTATACTCATTAATTATTACTACTTCCATTATTTTTCTAAATATAATT
                     aaaUUAUUCAAUUAAUAUUAUAUUCAUUAAUUAUUACUACUUCCAUUAUUUUUCUAAAUAUAAUU
                        I  I  Q  L  M  L  Y  S  L  I  I  T  T  S  I  I  F  L  N  M  I
cytb
10471 gaataatatcttaattaATGAATAAACCTTTACGAAATTCCCATCCTCTATTTAAAATTGCCAATAATGCTTTAGTAGAT
                      aaUGAAUAAACCUUUACGAAAUUCCCAUCCUCUAUUUAAAAUUGCCAAUAAUGCUUUAGUAGAU
                        M  N  K  P  L  R  N  S  H  P  L  F  K  I  A  N  N  A  L  V  D
nad1
12675'tttattacaaatagtaCTTGTTTTATATAGAATTTATTTTGTCATTAATTGGTAGTTTATTATTAATTATTTGTGTATTA
                       UUGUUUUAUAUAGAAUUUAUUUUGUCAUUAAUUGGUAGUUUAUUAUUAAUUAUUUGUGUAUUA
                        L  F  Y  M  E  F  I  L  S  L  I  G  S  L  L  L  I  I  C  V  L
lrRNA
14063'aaatcaatataaattgaGTTAGTTTTTATTTATTAATTTTTATTATTTTTTAAAAAATTATTAGAAATAACTATAAAATT
                      aGUUAGUUUUUAUUUAUUAAUUUUUAUUAUUUUUUAAAAAAUUAUUAGAAAUAACUAUAAAAUU
srRNA
14922'aaatcaaaaataaaaatTTAAAGTTTTATTTTGGCTTAAAAATTTGTTATTAATTTGATTTATATGTAAATTTTTGTGTG
                         aaaGUUUUAUUUUGGCUUAAAAAUUUGUUAUUAAUUUGAUUUAUAUGUAAAUUUUUGUGUG
```

Figure 4.5. Alignment of 5' cDNA Sequence to the Mitochondrial DNA Sense Strand for *Drosophila melanogaster*. The top sequence is derived from the *D. melanogaster* OreR sequence, with the sequence to the 5' of *srRNA* inferred from the chimeric *D. melanogaster* sequence. Lower case sequence represents nucleotides that encode the abutting gene or feature. Numbers to the left represent the genome position of the left-hand most nucleotide (with ' representing a minority strand encoded sequence). The second strand represents the sequence obtained from the cDNA (illustrated as an RNA sequence to differentiate from the genomic sequence). The number symbols (#) represent cDNA sequence confirmed by CR-RT-PCR and 5' RACE. The *cox2-a* and *cox2-b* RNA species are labeled.

addition of a poly-A nucleotide stretch. These results are summarized in Figure 4.5. Only the *cox-2b* species was identified by 5' RACE.

4.3.5. Detection of RACE Amplified RNAs from Schneider-2 Cells

Four RNA species were characterized from the mitochondria of Schneider-2 cells. The *cox1* and *cox2* genes had their 5' ends sequenced through 5' RACE amplification products, and the *cox1*, *nad4* and *nad1* had their 3' ends sequenced through 3' RACE amplification products. The RNA products characterized matched precisely to those later determined in the mitochondria isolated from adult flies, with the *cox-2b* cDNA sequence only being observed in the cell lines (see below).

4.3.6. Characterization of RNAs by CR-RT-PCR

4.3.6.1. Characterization of Mature RNA Molecules

The CR-RT-PCR was a far more efficient method for the amplification and sequencing of the 5' ends of the mitochondrial genes than the 5' RACE protocol. The 3' ends and 5' ends of all eleven mRNAs and two rRNAs were examined using the CR-RT-PCR protocols (see Figure 4.5 for 5' sequence alignments, Figure 4.4 for 3' alignments). The same eleven mRNAs characterized by 3' RACE were characterized by RNA circularization.

The *cox2* gene was unusual in that a single population of RNA molecules was not observed. Cloning and sequencing of the amplification products generated revealed the *cox2* mRNA exists in the RNA pool as two mature mRNAs. The first mRNA, which I will refer to as *cox2-a*, starts at the approximate location predicted by the removal of the upstream *tRNA*$^{Leu\text{-}UUR}$ by the tRNA punctuation model (Ojala, *et al.*, 1980a; Ojala, *et al.*,

1981). The second molecule starts five nucleotides downstream of this point, and will be referred to as *cox2-b* (see Figure 4.5).

The poly-A tailed *lrRNA* was also detected by circularization. The CR-RT-PCR protocol, unlike 3' RACE, does not rely on the presence of the 3' poly-A tail in the amplification of an RNA molecule. Consequently, if molecules lacking a poly-A tail are present in the pool, they should be detected as well. Though one cannot conclusively deny the presence of a molecule through lack of its observation, it is interesting to note that no copies of the *lrRNA* gene, or any of the protein-coding genes lacking a post-transcriptionally added poly-A tail were detected in the CR-RT-PCR analyses.

The *srRNA* was amplified and both ends sequenced. The *srRNA* had a very small poly-A tail added to its 3' end. Depending on the annotation of the 5′ end or the srRNA, the molecule has either 3-4 or 6-7 non-transcribed A's on the 3' end of the molecules (see Figure 4.4 and Figure 4.5).

For the poly-adenylated RNAs, each CR-RT-PCR reaction generated a pool of sequences, each containing poly-A tails of varying length. This variation in the A-tail length made it impossible to obtain the complete 3' and 5' sequence in a single sequencing reaction. Careful examination of the chromatogram obtained from the sequencing center allowed for the observation of an estimate of the poly-A tail length, by looking for the accumulation and distribution of the first non-A nucleotide encountered in the sequencing read (Figure 4.6). Cloning and sequencing of representative amplification products allowed for sequencing reaction that could read through the poly-A tail region, plus gave accurate measurements of the size of representative poly-A tails.

A

B

Figure 4.6. Example of Sequencing Results Obtained from CR-RT-PCR Generated cDNAs. Two sequencing reaction reads are presented. A – Sequence from the 5' facing LR-J13900 primer. B – Sequence from the 3' facing LR-N12866 primer.

Due to the ligation of the 3' encoded poly-A tail to the 5' end of the molecule using the CR-RT-PCR method, it is not possible in all cases to unambiguously determine which A nucleotides originate from the transcript and which nucleotides are post-transcriptionally added as part of the poly-A tail. To accurately map the 5' starting position of all mRNAs, it would be necessary to obtain 5' RACE sequences in which a primer other than a poly-A based 5' RACE primer was used to amplify the modified cDNA. Nucleotide positions of the transcribed mRNA and rRNA boundaries are found in Table 4.4.

4.3.6.2 Polyadenylation of Pre-mRNAs

CR-RT-PCR was employed to further characterize the pool of poly-A molecules detected for the *nad3* genes and the 3' encoded tRNAs. Gene-specific primers for each of the tRNA genes were paired with a gene-specific *nad3* primer (see Table 4.2). Amplification was observed for each of the Majority Strand encoded tRNA genes ($tRNA^{Ala}$, $tRNA^{Arg}$, $tRNA^{Asn}$, $tRNA^{Ser\text{-}AGN}$, and $tRNA^{Glu}$) circularizing to the 5' end of the *nad3*gene. No amplification of the Minority Strand encoded $tRNA^{Phe}$ circularized to the *nad3* gene was detected.

These *nad3*-tRNA CR-RT-PCR products were cloned and sequenced. Excluding $tRNA^{Glu}$, the length of the poly-A tail varied widely, from 14 to 53 in the sampled clones. The $tRNA^{Glu}$ gene had a substantially smaller A tail on the 3' end, numbering only 4-5 A's. Figure 4.7A depicts the pool of molecules amplified.

A similar situation was observed with the *cox2* gene, which is abutted at the 3' end by the two tRNA genes $tRNA^{Lys}$ and $tRNA^{Asp}$. CR-RT-PCR amplified three molecules: one RNA encoding the polyadenylated *cox2* gene with five non-coding

Table 4.4. Positional Annotation of Transcribed RNA Nucleotides as Determined Through cDNA Sequencing. GenBank Annotation positions refer to annotations from *D. melanogaster* Oregon R strain mt-genome sequence (Accession AF200828).

Gene	5' Position		3' Position	
	GenBank Annotation	Observed in cDNA*	GenBank Annotation	Observed in cDNA*
nad2	240	240	1263	1263-1265
cox1	1470	1474	3009	3011
cox2	3083	3083	3767	3767
atp8	3907	3906-3907	4068	(as for *atp6*)
atp6	4062	(as for *atp8*)	4735	4734-4736
cox3	4736	4735-4736	5524	5542-5543
nad3	5608	5608	5961	5978-5979
nad5	8116	8131	6400	6400-6399
nad4	9563	(as for *nad4L*)	8198	8198-8197
nad4L	9826	9827-9826	9537	(as for *nad4*)
nad6	9961	9959-9961	10485	10488-10489
cytb	10489	10488-10489	11625	11627-11628
nad1	12649	12658	11711	11710
lrRNA	14046	14048-14047	12725	12728-12724
srRNA	14905	14903-14901	14120	14121-14120

* Ranges of values represent ambiguous identification of transcriptional versus post-transcriptionally derived A nucleotides (see discussion).

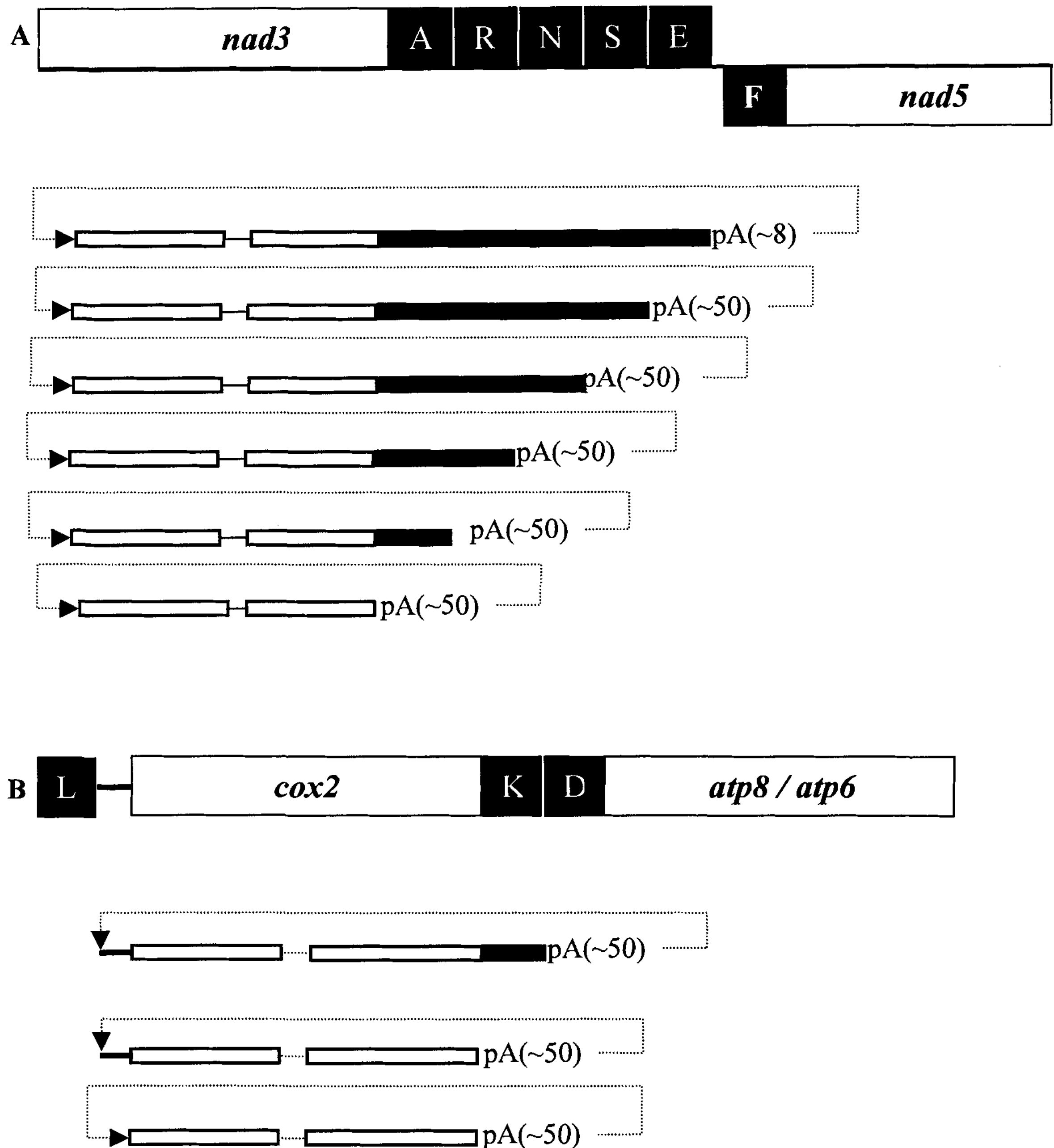


Figure 4.7. Characterization of Polyadenylated mRNA and pre-mRNA Molecules from *Drosophila* mitochondria. CR-RT-CPR products are shown with a wrapping dotted arrow, representing the ligation of the mRNA. Unfilled boxes represent protein gene regions and black boxes represent tRNA gene regions. A – Schematic representation of poly-A RNA molecules originating from the *nad3-nad5* gene junction, detected by CR-RT-CPR and 3' RACE. B – Schematic representation of poly-A RNA molecules originating from the *cox2-atp8* gene junction The thick line at the 5' end of the gene represents a 5nt non-coding region between *cox2* and $tRNA^{Leu}$.

nucleotides, *cox2* and the $tRNA^{Lys}$ with a poly-A tail. Attempts to amplify a molecule that contained $cox2+tRNA^{Lys}+tRNA^{Asp}$ were carried out with two separate $tRNA^{Asp}$ primers paired with the 5' facing *cox2* primer, but failed to produce an amplification product. Either this product is not produced, or is rapidly removed from the RNA pool (see Figure 4.7B)

4.4. Discussion

4.4.1. Schneider-2 Cells Versus Adult Fly Data

The DSMZ German Collection of Microorganisms and Cell Cultures reports that the Schneider-2 cell line currently contains 60-80% tetraploid cells (DSMZ, 2005). Though polyploidy of the nuclear genome does not necessarily mean that there would be an effect on mitochondrial transcription or transcript processing, such a substantial alteration of the genetic system within the cells may have any number of unknown effects on the system.

Supporting my distrust of cell lines is a study on the effects of overexpression of the *Drosophila* mitochondrial polymerase γ-α subunit (pol γ-α) in whole organism versus Schneider cell lines (Lefai, *et al.*, 2000). The authors observed that overexpression of pol γ-α in their cell cultures by more than 50-fold over normal expression levels resulted in no obvious affects on cell physiology, growth rate, relative mt-DNA content, or mt-DNA integrity. In contrast, overexpression of pol γ-α to similar levels turned out to be lethal to late stage pupae, and resulted in decreased levels of mt-DNA and mitochondrial transcripts observed. With this knowledge, I felt it important to switch to studies on the

whole organism whenever possible, and to reconfirm all results obtained from cell-line mitochondria.

Despite my concerns, the small number of cDNA sequences obtained from experimentation with Schneider-2 showed no variation from the sequences obtained from adults. The only variation was the detection of a single 5' processed *cox2* mRNA instead of the two differing molecules observed in the mt-RNAs of adult flies. As studies had been abandoned on cell cultures, I was unable to determine if this variation was due to a difference in the transcript processing. It is more likely that this variation is the result of increased examination of the transcripts from the adult flies, allowing me by chance to detect molecules that existed at lower frequencies.

4.4.2. Detection of Pre-mRNA Molecules

RT-PCR reactions using total RNA from mitochondrially enriched extracts were able to amplify between any two adjacent protein coding genes, protein gene and rRNA gene, and rRNA genes. The notable exception to this was the *srRNA* and *nad2* genes that are separated by the 4.6 kb A+T rich region. RT-PCR amplification was not attempted due to the large size of this gene junction.

The methodology I used to amplify the RNAs was not able to determine which strand the RNA molecule amplified originated from. Due to this, comparisons to the pre-RNA pools observed by Berthier *et al.* (1986) are not possible across the protein coding regions of the mt-DNA. I do, however, characterize the existence of a pre-rRNA molecule that encodes the $srRNA{+}tRNA^{Val}{+}lrRNA$. This observation, in the absence of a similar molecule being observed by Berthier *et al.* (1986) opens the possibility that the radiolabeling of RNAs may not have detected all mitochondrial transcripts. The RT-PCR-

only detectable level of these molecules implies that they are very quickly turned over by the RNA processing machinery, and are therefore not detected by standard analyses, such as RNA labeling or Northern blot analysis.

As it is assumed that the *Drosophila* mt-genome is transcribed from five transcription initiation sites, based on the pre-mRNA pool observed by Berthier *et al.,* it is important that confirmation of their transcription units be undertaken with a more sensitive assay. It may be possible to determine strand specific origin of the RNAs through RT-PCR, by splitting the reaction into two steps. The reverse transcription step would be carried out with only the majority-strand primer, and a parallel reaction conducted with only the minority strand primer. After the RT step, the RNA is removed by RNase, the second primer added, and the PCR amplification step initiated.

4.4.3. Description of Mitochondrial mRNAs

I detected the same eleven mitochondrial mRNA species for *D. melanogaster* as reported previously (Spradling, *et al.*, 1977; Merten and Pardue, 1981; Berthier, *et al.*, 1986). The cDNAs were sequenced on their 5' and 3' ends, resulting in a more precise map of the mRNA transcripts than was previously available.

4.4.3.1. Bicistronic Mitochondrial mRNAs

The *atp8-atp6* and *nad4L-nad4* genes are reported here to be expressed as two bicistronic messages, confirming the earlier prediction by (Berthier, *et al.*, 1986). Expression of bicistronic genes in known from three well studied nuclear genes in *Drosophila*, as well as up to 44 more cases predicted from the *Drosophila* genome project annotations (reviewed in Blumenthal, 1998; 2004). In the nuclear instances, the genes are separated by a non-coding region, and are thought to be either translated by the

recruitment of the ribosome to an internal ribosome entry site (IRES), or by the ribosome re-initiating translation of the second gene after completion of the first transcript.

In contrast to these nuclear examples, the mt-bicistronic messages overlap their reading frames by one nucleotide (*nad4L – nad4*) or overlap by seven nucleotides (*atp8 – atp6*). Besides very small regions of T-rich sequence, there is very little observable sequence similarity between the two mRNAs that could be interpreted as a conserved IRES. An alternate mechanism for the expression of these genes may involve the re-recruitment of a ribosome that has finished translation of the first protein (see Sarabhai and Brenner, 1967) coupled with a –1 or –7 slippage of the ribosome to initiate at the ATG start codon of the second gene. The establishment of an *in vitro* translation system for mitochondria is still in its early stages (Hanada, *et al.*, 2000; Hanada, *et al.*, 2001), so further investigations of this translational system are currently not feasible.

4.4.3.2. Mapping of mRNA 3' Ends

For the 3' ends of the mRNA genes, sequence was originally obtained using 3' RACE amplification, and was confirmed through sequencing CR-RT-PCR products with the 3' facing primer. Generally, processing the 3' end of the mRNA appears to be mediated by a tRNA punctuation, as proposed by Attardi and colleagues in the early 1980's (Ojala, *et al.*, 1980a; Ojala, *et al.*, 1981). The *cox3* and *nad3* genes encode relatively large non-coding regions between their stop codons and their distally encoded tRNAs. Examinations of their transcripts show that this large non-coding sequence is present on the mature mRNA molecules. Examination of the *cox2, nad5* and *nad4* genes reveal that these genes lacking stop codons are processed as predicted by the tRNA punctuation model (Ojala, *et al.*, 1981) (see Figure 4.4). The genes are processed at the

in-frame 3' T nucleotide that directly abuts the tRNA genes encoded downstream, and polyadenylation completes the predicted stop codons. For the *nad5* and *nad4* genes, it is not possible to discern whether the first A nucleotide annotated at the 5' end of the downstream tRNA is encoded as part of the transcribed mRNA genes or is added through polyadenylation.

One exception to the 3' tRNA processing model was observed. The *nad1* gene encodes a 17nt non-coding region to the 3' of its encoded TAG stop codon (see Figure 4.4). The observed mRNAs for the *nad1* gene lacked 16 of these bases, encoding only the single in-frame T that immediately follows the encoded stop codon. Two factors regarding this gene boundary make it distinct from the majority of the protein gene – tRNA boundaries observed in the mitochondrial genome. First, it is one of four locations in the mt-genome where a protein-coding gene is abutted by a tRNA gene encoded on the opposite strand. The other three locations are all instances where the tRNA is to the 5' side of the protein gene ($tRNA^{Tyr}$ – *cox1*, $tRNA^{Thr}$ – *nad4L*, and $tRNA^{Pro}$ – *nad6*). The *nad1*-$tRNA^{Ser\text{-}TCN}$ is the only 3' tRNA to 3' protein gene boundary within the mt-genome. This complement of the $tRNA^{Ser\text{-}TCN}$ may not be able to adopt the appropriate structures to be recognized by mitochondrial RNase P, and therefore may not be able to serve as a punctuating signal. The human mitochondrial RNase P has been well characterized, and appears to excise tRNAs via a precise 5'-then-3' excision mechanism that would serve as accurate punctuation for the mammalian mRNAs (Rossmanith, *et al.*, 1995; Rossmanith and Karwan, 1998; Puranam, *et al.*, 2001). It has been suggested that the human $tRNA^{Ser\text{-}AGY}$, which lacks a dehydrouridine-stem and loop, may not itself be a substrate of the RNase P, but is liberated by the processing of the two flanking tRNA genes

(Rossmanith, 1997). This claim implies that the RNase P may be very discriminating in its substrates, making it a reasonable assumption that a "reverse-complement" tRNA may not be recognized as a substrate.

Secondly, it has been shown that the mitochondrial transcription termination factor DmTERF1 (formerly DmTTF) binds to two DNA sequences within *Drosophila* mt-DNA (Roberti, *et al.*, 2003). One of these binding sites is the non-coding junction between the *nad1* and $tRNA^{Ser\text{-}TCN}$, and the second is the junction between the 3' to 3' abutting tRNA genes, $tRNA^{Phe}$ and $tRNA^{Glu}$. DNase I protection studies revealed that DmTTF1 protects a stretch of DNA that includes 6nt of *nad1*, 5 nt of $tRNA^{Ser\text{-}TCN}$, and the entire intervening non-coding sequence (Roberti, *et al.*, 2003). No further information regarding how DmTERF1 interacts with the mitochondrial transcription machinery is currently available.

It is tempting to speculate that the transcription is stopped by DmTERF1 after the *nad1* gene. The terminated transcript could then be polyadenylated, and become a functional mRNA by the removal of the $tRNA^{Leu\text{-}UUR}$ at the 5' end. The transcript termination would, however, need to be very precise to generate only the one observed *nad1* mRNA species. A processing based mechanism where excision of the 3' encoded sequence produces the RNA to be polyadenylated would seem to be a more precise mechanism. A processing based model would also seem more plausible, based on the tight linkage observed between transcript processing and polyadenylation (Taanman, 1999; Fernandez-Silva, *et al.*, 2003). It will be very interesting to see if further advances in our understanding of mitochondrial transcript initiation and termination can begin to address this one exception of the 3' tRNA processing model for *Drosophila* mt-mRNAs.

CR-RT-PCR detected a distribution of the observed poly-A tail lengths for the various mRNAs. By identifying the first non-A nucleotide to appear in the 5' of the cDNA sequence, it is possible to observe a distribution of the poly-A tail lengths in the chromatogram obtained from the sequencing of the cDNAs (see Figure 4.6). Coupled with the sampling of cDNAs that were cloned and sequenced, thereby giving an exact poly-A tail length, it was possible to generally characterize the poly-A tail length for the mRNAs. The poly-A tails typically ranged from 35 A's to 60, with the peak of most observed distributions occurring at approximately 50 nucleotides. This poly-A size correlates well to the approximately 55nt A-tail observed in mammals (Ojala, *et al.*, 1980a).

No RNA species containing a protein-coding gene sequence were observed in the absence of a poly-A tail. The CR-RT-PCR methods do not have a preference between poly-A or non-poly-A RNAs. Therefore, if an intermediate RNA existed in the total RNA pool, the methods presented here should have detected such a pool of molecules. Agarose gel electrophoresis of the CR-RT-PCR products did resolve slightly thicker bands than observed in a standard PCR reaction. These bands were not thick enough to infer a pool of amplified products that differed by the approximate length of the poly-A tail. Nor did the sequencing reaction chromatograms or sequenced clones infer that there was potentially an A-tailless portion of cDNAs in the sequencing pool. These results are consistent with the current belief that the polyadenylation of mitochondrial RNAs is very tightly linked to the processing cleavage reactions that generate the free 3'-hydoxyl of the mRNA (discussed in Taanman (1999) and Fernandez-Silva *et al.* (2003)).

4.4.3.3. Mapping of mRNAs 5' Ends

The precise 5' nucleotide for four of the mRNA transcripts was mapped using a combination of CR-RT-PCR and 5' RACE (*nad2, cox1, cox2-b,* and *nad3*). Additionally, the *nad1* gene could be accurately mapped by the lack of a 5' A nucleotide in its sequence. The *nad5* 5' can be inferred by the absence of a C nucleotide to the 5' end of a 5' ATG codon. Conclusive evidence that the 5' A for *nad5* is transcribed instead of being the terminal poly-A nucleotide in the circularized mRNA was not obtained, but transcription of this A nucleotide is consistent with the 5' processing pattern observed for the other mitochondrial mRNAs (discussed below).

The remaining five mRNAs plus *cox2-a* are mapped to within two 5' A's encoded on the mt-DNA (*atp8, cox3, nad4L* and *cytb*), three A's (*nad6*), or four A's (*cox2-a*). Repeated attempts to obtain 5' RACE sequence with non-A nucleotides were unsuccessful for these genes. Further attempts to obtain 5' RACE sequence, plus a proposed alteration to the CR-RT-PCR protocol may be attempted to complete the characterization of their 5' ends (see Chapter 5, section 5.3.1).

For the six precisely mapped mRNAs sequences, processing occurred at the first codon position of the first in-frame recognized invertebrate mitochondrial start codon. The one notable exception is the *cox1* gene (discussed in section 4.3.6.1) which starts at the first codon position of an in-frame TCG-Serine codon. This close proximity of the first codon to the very 5' end of the mRNA is also consistent with the observations in mammals (first reported by Montoya, *et al* (1981), reviewed by Taanman (1999) and Fernandez-Silva *et al.* (2003)).

Three of the gene boundaries within the *Drosophila* mt-genome are structured so that a tRNA gene on the opposite strand abuts the protein-coding gene in a 5' to 5' fashion ($tRNA^{Tyr}$ – *cox1*, $tRNA^{Thr}$ – *nad4L* and $tRNA^{Pro}$ – *nad6*). Earlier, I discussed the potential difficulties with a complementarily encoded tRNA gene adopting an appropriate structure for RNase P to identify and process (see section 4.4.3.2). These three gene junctions would be expected to encounter similar problems during transcript processing. Preliminary inspections of the *nad4L* and *nad6* 5' cDNA sequences show that punctuation at the complementary tRNA could produce the observed 5' for the mRNA (considering the 2-nucleotide ambiguous annotation of these 5' mRNAs). The 5' end of the *cox1* mRNA, however cannot be produced by a processing of the complementarily encoded tRNA sequence. Five additional nucleotides (AATAA) have to be removed from the transcript to produce the observed mRNA. Interestingly, these extra nucleotides encode the in-frame stop codon that was so troubling to earlier researchers (Clary and Wolstenholme, 1983a; de Bruijn, 1983).

Two of the gene boundaries within the *Drosophila* mt-genome are protein-protein gene boundaries (*atp6 – cox3* and *nad6 – cytb*). Both of the *cox3* and *cytb* genes appear to be precisely processed to leave their start codons within 1 to 2 nucleotides of the 5' end of the mRNA. The processing of protein-protein boundaries has not been discussed in great detail for mitochondrial processing. It is assumed that a mechanism must be employed to facilitate this processing, but mechanistic details are not available.

The *cox2* mRNA is uncharacteristic in that two distinct species were identified. 5' RACE techniques detected a sequence that was processed precisely at the first codon position of the in-frame ATG proposed as the start codon for the gene (*cox2-b*). CR-RT-

PCR identified *cox2-b*, but also a second species (*cox2-a*). The *cox2-a* species encodes another 3 to 7 nucleotides of transcribed sequence at the 5' end.

More accurate mapping of the *nad4L* and *nad6* will be required to discuss their processing, but in the remaining nine mitochondrial mRNAs, it appears the mature mRNAs do not encode non-coding nucleotides from the gene junctions. For mRNAs of *cox1, cox2-b, nad4L, nad6* and *nad1*, their processing would involve the removal of nucleotides other than those predicted to be removed by tRNA processing model. The open reading frames of the mitochondrial mRNAs begin with the first nucleotide of the mRNA.

A mutation derived from a human patient suffering a mitochondrial disorder may provide and interesting model that implicates the 5' sequence of the downstream protein gene in the accurate processing of the transcript boundaries. Multiple mutations have been observed in mitochondria from that patient, so it is currently unclear which mutations are responsible for the mitochondrial misfunction and disease (Chrzanowska-Lightowlers, *et al.*, 2004). One homoplasic mutation observed results in the deletion of two nucleotides (TA) at the 3' end of *atp6* that directly abut the ATG-start of the *cox3* gene (polyadenylation is assumed to complete the *atp6* stop codon, similar to what I observe for *Drosophila*) (Temperley, *et al.*, 2003). Analyses of a cell line derived from the patient revealed that the mRNA for *cox3* was processed normally, and maintained the appropriate start codon (Temperley, *et al.*, 2003). These results imply that the mechanism by which protein-protein gene junctions are processed operates either by recognition of the 5' end of the downstream-encoded gene (*cox3*), or by an as of yet unknown

mechanism that is tolerant of the deletion of two important nucleotides from the 3' end of the upstream-encoded *atp6* gene.

4.4.4. Initiation of the *cox1* Gene

As discussed above (Section 4.4.3), the six precisely mapped mRNA molecules occur at the first codon position of the first in-frame codon. This codon is not recognized as a start codon in any currently proposed genetic code. Nor is it easy to model a first and second codon position wobble interaction between U-U and A-C nucleotides that would allow for the methionine to bind to this codon and initiate translation. The TCG codon is not a remarkable codon. It is underrepresented in use by the *Drosophila* mt-genome, as are the other codons with third position G for four-fold degenerate amino acids. The $tRNA^{Ser\text{-}TCN}$ is not noted for any unusual structures or features.

One alternative explanation is that this TCG is not the start codon, and that the open reading frame of *cox1* begins 36 nt downstream, at an in-frame ATT codon (see Figure 4.5). Though ignoring the first thirteen codons would find a suitable in-frame codon, the evolution of the 36nt intervening sequence implies that it is a functional protein sequence. Alignment to other insect *cox1* show high levels of conservation of the translated amino acids for this region, and that nucleotide substitutions tend to be synonymous substitutions at third codon positions. How then does the *cox1* initiate?

First, can translation initiate in the absence of a start codon? It has been known since the earliest experiments to determine the genetic code that start codons are not necessary for translation initiation. In the famous work by Nirenberg and Matthaei (1961) it was shown that poly-U, poly-C and poly-A RNAs do direct the translation of their encoded mono-amino acid peptides. This transcript initiation is much less efficient than

when an ATG codon is used, but these experiments were conducted using isolated *E. coli* translation machinery.

Mitochondrial ribosomes are very unusual in comparison to the other ribosomes observed throughout living organisms. Despite a drastic reduction in the amount of RNA sequence for the ribosome, the mitochondrial ribosome is larger than the prokaryotic counterpart (Sharma, *et al.*, 2003). The increased size is due to the apparent replacement of ribosomal RNA components with those of protein components, imported from the nuclear genome. Given this unusual evolution of the mitochondrial ribosome, it may be possible that novel function has also evolved for the insect ribosome.

As discussed above, the *Drosophila* mitochondrial mRNAs observed in this study appear to initiate with the first codon of the first in-frame codon (except for *nad4L* and *nad6,* where prediction of the precise 5' nucleotide is not possible, and the *cox2-a* RNA species, discussed in section 4.4.3.3). By processing the mRNAs to begin their open reading frames at the first nucleotide the mitochondria has eliminated the need to code for a special initiation signal to find the correct open reading frame. The first nucleotide of the mRNA is sufficient for this function. With the ribosome only exposed to these simplified translational templates, it becomes easy to imagine the ribosome losing the specific requirement of the methionine tRNA for translational initiation.

The ability of the ribosome to translate all open reading frames in the mitochondria is speculative at this point, however another alternative translational mechanism to start the *cox1* gene cannot be envisioned. Cell free mitochondrial translation systems are being developed, but are currently not capable of full-length protein synthesis (Hanada, *et al.*, 2000; Hanada, *et al.*, 2001). Until the development of a robust cell free mitochondrial

translation system, experimentation to clarify the mechanism of mitochondrial translation initiation will need to wait. Also, further investigation of the RNAs of other insect groups would tell us how conserved the 5' processing at the initiation of the open reading frame is within insects. Of great interest would be the study of organisms, such as the Paraneoptera, who appear to have re-evolved the consistent use of a start signal for their mitochondrial genes.

4.4.5. Characterization of the Ribosomal RNAs

The inability to obtain 5' RACE sequences of the rRNA genes limits discussion of the ribosomal RNA molecules. If the ambiguous 5' A nucleotide for the *lrRNA* gene is actually part of the poly-A tail, the sequence of the *lrRNA* described here corresponds to that predicted by S1 nuclease protection mapping of the *D. yakuba lrRNA* (Clary and Wolstenholme, 1985b). The absence of the potentially transcribed A would also be the predicted molecule based on tRNA punctuation.

The sequence reported here for the *srRNA* differs from that predicted by *D. yakuba* S1 nuclease protection assays by 2 to 5 nucleotides, depending on the origin of the three 5' ambiguously annotated A nucleotides (Figure 4.5). The 5' abutting sequence to the *srRNA* is the A+T rich region, and our lack of understanding regarding the function of the region give us no insights into how transcript processing occurs at the 5' end of the *srRNA.* The fact that the *srRNA* could be sequenced by circularization reveals that the molecule has an exposed 5' monophosphate. RNA transcripts observed in prokaryotic and eukaryotic systems (prior to 5' capping) have a 5' triphosphate. The presence of this monophosphate implies that the *srRNA* is also the result of a transcript processing, where by a primary transcript containing A+T region sequence was processed to remove the

non-*srRNA* nucleotides (including the two T's that were originally annotated as the 5' most nucleotides of the RNA). Though it would have been of interest to pursue the identity of this pre-rRNA molecule, the nucleotide composition and structure of the A+T rich region complicated any attempts to isolate the molecule.

4.4.6. Polyadenylated Pre-mRNAs

The 3' boundaries of the *nad3* and *cox3* genes are unique in the *Drosophila* mt-genome in that they are the only cases where multiple tRNAs encoded on the same strand abut the 3' ends of protein coding genes. The observation of a pool polyadenylated RNAs corresponding to these two regions may provide insight to the processing mechanisms and tRNA maturation of insect mitochondria. A diagram of the transcripts observed for the *nad3* and *cox2* regions is presented in Figure 4.7.

4.4.6.1. The nad3 gene

Five polyadenylated RNA species were identified containing large (approximately 50 nt) poly-A tails. These included the fully processed transcript for the *nad3* gene alone, and four species where the *nad3* gene is observed with the next one through four tRNAs still encoded in the RNA (the largest being *nad3* with $tRNA^{Ala}$, $tRNA^{Arg}$, $tRNA^{Asn}$ and $tRNA^{Ser\text{-}AGN}$). A sixth species was identified encoding *nad3* through to $tRNA^{Glu}$, but this species encoded a much smaller poly-A tail (up to 8 nt).

Attempts to obtain a molecule that encoded *nad3*, all the J-stand encoded tRNAs plus the N-strand encoded $tRNA^{Phe}$ were unsuccessful. Also, attempts to circularize a small molecule containing just the tRNA cluster in the absence of *nad3* were unsuccessful.

Given the pool of molecules observed, a simple pattern of processing of this region emerges (see Figure 4.8). The original transcript is produced, and the $tRNA^{Gly}$ to the 5' of

nad3 is removed. This leaves a molecule containing *nad3* and five tRNAs (to the $tRNA^{Glu}$). How the processing at the 3' end of this molecule is undertaken is unknown. The DmTERF1 termination factor has been shown to DNase I protect the region of DNA encoded between the $tRNA^{Glu}$ and $tRNA^{Phe}$, and may serve to terminate transcription before the transcription of the complement of $tRNA^{Phe}$. The mechanism by which the short poly-A tail is added or what processing of this 3' end occurs could not be deduced by these investigations.

The next step in the process in the recognition of the $tRNA^{Glu}$ by the processing machinery, and its subsequent liberation by mitochondrial RNase P. A poly-A tail is added to the 3' end of the $tRNA^{Ser\text{-}AGN}$. The processing of tRNAs then continues, with each tRNA being removed serially until the $tRNA^{Ala}$ is finally removed. The serial processing model is preferred over a model that targets specific tRNAs or randomly processes any of the remaining tRNAs as the cleavage product of two or more 3' tRNAs that would be lost by such a mechanism were never observed despite attempts at their CR-RT-PCR amplification.

It is also interesting to note that the *nad3* gene encodes its own in-frame stop codon, then non-coding 3' nucleotides. As such any of these molecules at any stage of the tRNA processing could serve as mRNAs for the *nad3* gene, that would terminate at the coded stop codon, and when liberated again from the ribosome, would be further processed to remove more tRNAs for maturation.

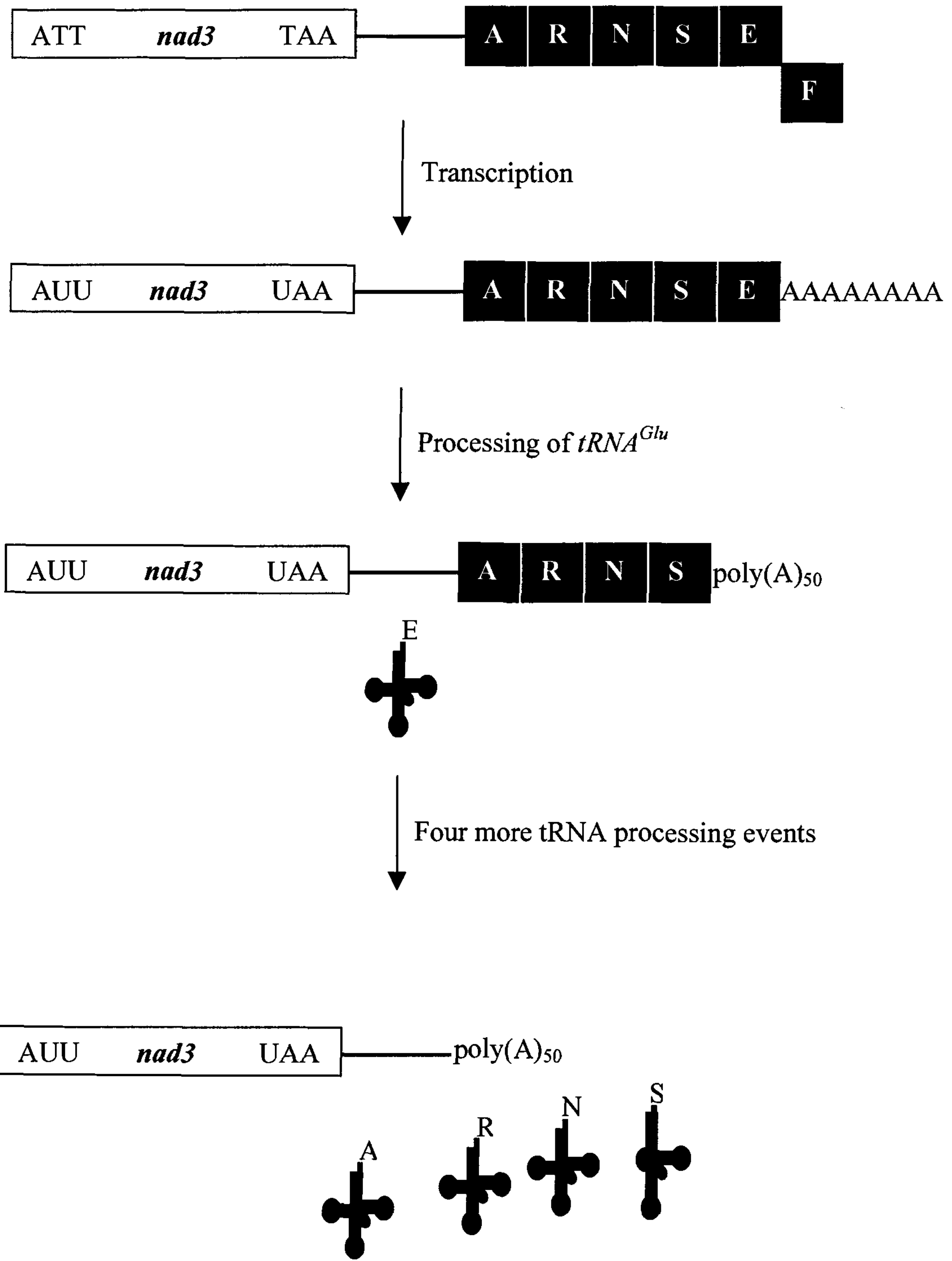


Figure 4.8. A Model for Processing of the *nad3 - tRNA*^{Glu} RNA Transcript.

4.4.6.2. The cox2 gene

The pool of observed *cox2* molecules is distinct from the *nad3* pool in two ways. First, the *cox2* gene requires the processing of the 3' encoded $tRNA^{Lys}$ and polyadenylation to complete its stop codon. Second, there are two distinct species that were observed, with differing 5' ends.

One advantage of the CR-RT-PCR protocol over RACE methods is that by linking together the 5' and 3' ends of the RNA in the circularization, it is possible to simultaneously obtain information about each end of the molecule. This is not possible with 5' RACE, as the use of a gene specific primer in amplification causes the loss of all information downstream of the primer after amplification.

The *cox2-a* 5' RNA species was observed with two distinct 3' sequences. One sequence included the $tRNA^{Lys}$ followed by a large poly-A tail. The second *cox2-a* species was observed with the $tRNA^{Lys}$ removed and the stop codon of *cox2* completed by the poly-A tail. *cox2-b* in contrast, the species that lacks an in-frame stop codon encoded to the 5' of the ATG start codon for the gene, was only observed with its 3' end lacking the tRNA, and being processed to include the post-transcriptional stop.

Figure 4.9 diagrams a model to describe the processing and maturation of the *cox2* mRNA. The polycistronic pre-mRNA for the region spanning *cox2* through *atp8/atp6* is recognized by the processing machinery and the $tRNA^{Asp}$ is removed from the transcript, freeing the *cox2* portion from the *atp8/atp6* portion. The lack of observation by CR-RT-PCR of a *cox2* species containing the $tRNA^{Asp}$ supports the removal of this tRNA as important to generating a free 3' end lacking the *atp8/atp6*. The

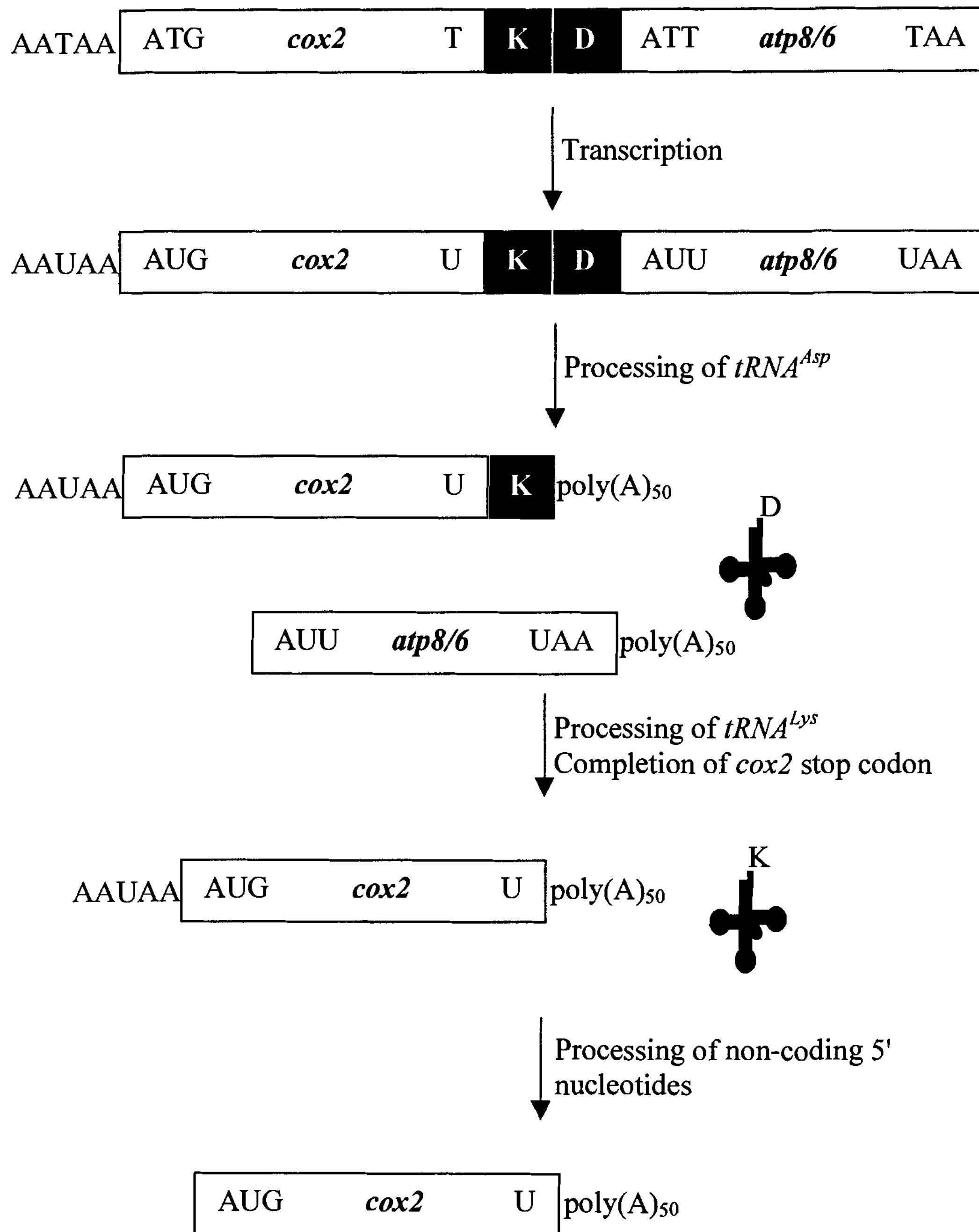


Figure 4.9. A Model for the Processing of the *cox2-atp8/atp6* Transcript.

polyadenylated *cox2-a+tRNA*^{Lys} is now present in the mitochondrial. If the ribosome were able to recognize this molecule, it would produce a *COX2* with eleven extra amino acids from translation of the *tRNA*^{Lys} sequence. Give the very conserved size of the *cox2* gene through insect evolution, it would be assumed that these extra nucleotides would be deleterious. Fortunately, the *cox2-a* 5' sequence encodes an in frame stop codon prior to the reading frame of *cox2*. If translation were initiated on *cox2-a*, a stop codon would be encountered at the ninth codon, and the release factors would rapidly be employed and the pre-mRNA liberated.

The molecule is now further processed by the removal of the *tRNA*^{Lys}, and polyadenylation completes the stop codon for the 228aa *cox2* open reading frame. This pre-mRNA still encodes the *cox2-a* 5' end, so is untranslatable if my assumptions above are correct. Finally, the *cox2-a* 5' end is processed, revealing the *cox2-b* 5' end. This mRNA now encodes, as is observed with the other mRNAs for *Drosophila* mitochondria, the start of the open reading frame being defined by the first nucleotide. The *cox2-b* RNA is the mature mRNA.

The observation of non-coding RNAs that are polyadenylated has been reported in the cytoplasm of *Drosophila* (Tupy, *et al.*, 2005), but the use of large poly-A tails on pre-mRNAs in mitochondrial systems is not known. The observation that not all poly-A RNAs are matured mRNAs, with the notable exception of the *lrRNA* (Spradling, *et al.*, 1977; Merten and Pardue, 1981; Berthier, *et al.*, 1986; Benkel, *et al.*, 1988), has not been recognized in mitochondrial research. This new recognition of mature and incompletely processed mt-mRNAs with poly-A tails alters our previous view that all poly-A mt-RNAs represent mature and functional message.

Chapter 5. Conclusions and Discussion

5.1. Transcript Characterization in *Drosophila* Mitochondria

The results I have presented here increase our understanding of the structure of mature mRNA and rRNAs in the mitochondria of *D. melanogaster* beyond that presented in the study by (Berthier, *et al.*, 1986). Figure 5.1 depicts the additional information obtained (compare to Figure 1.4 or 4.1).

5.1.1. Characterization of mt-rRNAs

Sequence level mapping of *lrRNA* is in agreement with the current *lrRNA* annotation (Clary and Wolstenholme, 1985a). There is a slight annotation conflict with the *lrRNA* annotation given for the Oregon R strain mt-genome (Accession AF200828) which does not include the 5' G nucleotide in the *lrRNA*, instead annotating it as a non-coding nucleotide (Ballard, 2000b), while the chimeric *D. melanogaster* (Accession U37541) does include the G as part of the rRNA. The polyadenylated state of the mature *lrRNA* is confirmed for *D. melanogaster* mitochondria, as no non-poly-A *lrRNA* molecules were observed, even at RT-PCR detectable levels (Spradling, *et al.*, 1977; Merten and Pardue, 1981; Berthier, *et al.*, 1986; Benkel, *et al.*, 1988).

For the *srRNA*, a minor variation was observed relative to the S1 nuclease protection map obtained by Clary and Wolstenholme (1985b). The sequence obtained through CR-RT-PCR shows that two unpaired T nucleotides from the very 5' end of the *srRNA,* which they annotated as part of the *srRNA,* are not present. Identification of the precise 5' end of *srRNA* was not possible since the A residues may have originated from the poly-A

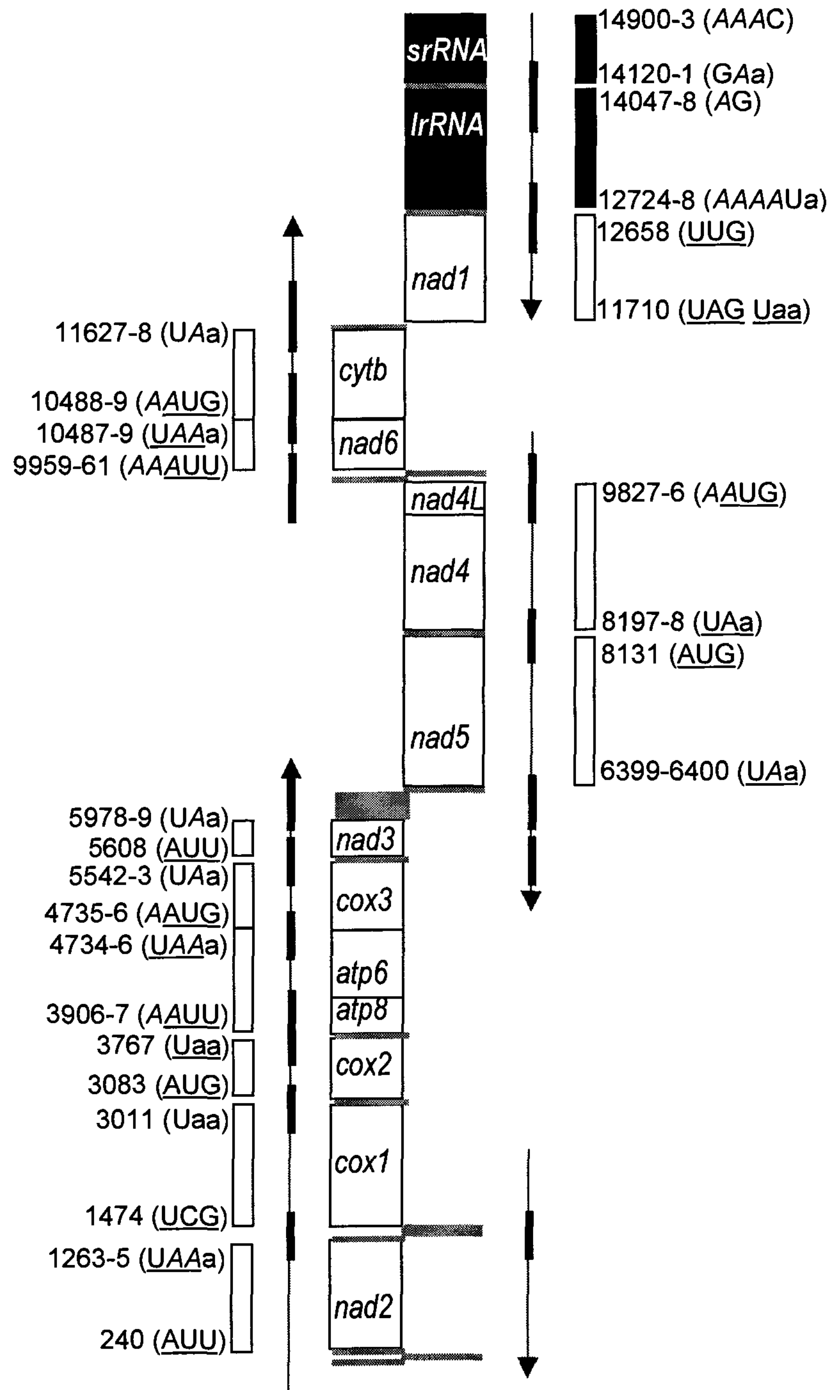


Figure 5.1. Transcription Map for the *D. melanogaster* Mitochondrial Chromosome. Blocks to the outside represent mRNAs (unfilled) and rRNAs (filled black) sequenced in this study. Numbers represent mapped OreR strain mt-genome positions of transcribed nucleotides, with ranges representing ambiguous annotations due to methods used. The letters following the numbers represent the encoded nucleotides, with italicized letters represented the ambiguously annotated nucleotides, and underlines represent codons, and lower case "a" representing known post-transcriptionally added poly-A tail nucleotides. Arrows represent predicted transcription units based on Berthier *et al.* (1996), with thickening of the arrows representing transcripts detected in this study by RT-PCR.

tail, artificially linked during circularization. The complementary 5' RACE experiments were not successful with the rRNAs. The S1 nuclease protection-based secondary structure by Clary and Wolstenholme encodes three unpaired nucleotides at the 5' end of the secondary structure. This implies that if the transcribed sequence begins at either nucleotide positions 14092 or 14093 as I suggest, there would be no effect on the 5' encoded stem-loop structure predicted for positions 14092-14886.

It is interesting to note that a small poly-A tail was also found on the *srRNA* molecule. The small size of this poly-A tail (5-9 nt) would have excluded it from being detected by its ability to bind poly-T, and the S1 nuclease protection assays would not have detected it, based on its inability to anneal to the $tRNA^{Val}$ sequence, and therefore would not be observed previous experiments (Spradling, *et al.*, 1977; Merten and Pardue, 1981; Clary and Wolstenholme, 1985b; Berthier, *et al.*, 1986). All RNA molecules detected in this study had at least three post-transcriptionally added A nucleotides on their 3' ends, implying that polyadenylation is a feature of all mRNAs, rRNAs and pre-mRNAs in insect mitochondria.

5.1.2. Characterization of mt-mRNAs

The sequence of the mRNAs reveals a number of features of mitochondrial transcript processing. First, the 3' end of *Drosophila* mRNAs are processed where predicted by the tRNA punctuation model (Ojala, *et al.*, 1980a; Ojala, *et al.*, 1981). The only 3' exception to this is the apparent processing of the *nad1* mRNA at the DmTERF1 binding sequence (see Chapter 4, section 4.4.3.2). This processing leaves six mRNA molecules that encode 3' UTR sequences (*cox1, cox3, nad3, nad6, cytb* and *nad1*). These UTRs show a great variation in size (1nt to 19nt) and show very little sequence similarity.

Given that the remaining five mRNAs completely lack a 3' UTR, it is likely that they have no functional significance to the *Drosophila* mRNAs. Alignment of the mt-genomes of the various insects reveals that these UTR regions are very poorly conserved, even within the *Drosophila* species. These observations appear to suggest that these UTRs have arisen by random insertional mutations, independently in different insects. As the processing at the 3' end of the mRNAs appears to be directed by the abutting tRNA sequences, it appears that the conservation of the intact stop codon neutralizes any potential effects these UTRs could have on the amino acid sequence.

The 5' cDNA sequences of the nad5 and nad1 genes imply that an update to the *Drosophila* annotation of these genes is warranted. The *nad5* mRNA appears to encode five additional codons not included in the current gene annotation, and initiates with an AUG codon. If the boundaries at the 5' end I have observed for *D. melanogaster* can be applied to *D. yakuba*, these extra amino acids align well, and the *D. yakuba* sequence would initiate with an in-frame GUG codon that aligns with the *D. melanogaster* AUG codon (Clary, *et al.*, 1984). It appears that this potential initiator, which would not have been recognized as an initiator at the time these sequences were obtained, has been ignored as subsequent sequences were annotated with the objective of aligning to the *D. yakuba nad5* amino acid sequence. Similarly, the *nad1* mRNA appears to encode three additional amino acids, and initiates with a UUG codon that is well conserved among the Diptera represented by complete mt-genome sequence.

The 5' ends of the molecules do not conform to what is predicted by the tRNA punctuation model. Nucleotides other than those encoding the abutting tRNA sequence were observed to be removed for the *cox1*, *cox2* and *nad1* genes. In these three cases the

extra nucleotides removed leave the mRNA with the first nucleotide defining the ORF for the gene. Precise annotation was not available for the *cox3, nad4L/4, nad6* or *cytb* mRNAs in this study because of ambiguity arising from circularizing polyadenylated transcripts. These four genes also may encode extra nucleotides other than the ORF for the protein. Precise mapping for *nad2* and *nad3* was also accomplished, but these mRNAs, as well as *nad5* and *atp8/6*, are predicted to encode only the ORF predicted by tRNA punctuation.

Additionally, characterization of the processing intermediates of the $cox2+tRNA^{Lys}+tRNA^{Asp}+atp8+atp6$ transcript may imply that the 5' "trimming" to the ORF follows the tRNA punctuation processing. RNA molecules with *cox2* sequence were observed as matured poly-A mRNAs, and as a pre-mRNA with the $tRNA^{Lys}$ sequence still encoded. No molecules with both the $tRNA^{Lys}$ and $tRNA^{Asp}$ sequences still associated were observed, implying that the removal of the $tRNA^{Asp}$ gene is the first processing step, followed by the removal of the $tRNA^{Lys}$, then finally the five 5' nucleotides that precede the start (see Chapter 4, Figure 4.9).

Determination of the precise 5' nucleotide for the ambiguously annotated mRNAs is a high priority (see section 4.4.3.3). I discussed in chapter 4 (section 4.4.4) the possibility of the mitochondria of insects adopting a 5' processing model that leaves just ORFs, which eventually may have evolved to no longer require start codons to define the ORF for proper translation. Completion of the sequencing of the mRNAs is required to test whether this model of 5' mRNA processing is universally applied in the *Drosophila* mitochondria. Such a mechanism can be interpreted as a simplification of the ancestral translational machinery as opposed to the evolution of a novel function.

5.1.3. Processing of tRNAs Versus mRNAs in Insect Mitochondria

The standard methodology for the annotation of animal mitochondrial genes has been to annotate just the ORF of the predicted amino acid as the gene. Some researchers take into account the tRNA punctuation model, and therefore do not allow the overlap of stop codons from predicted ORFs with abutting tRNAs, while others annotate such scenarios as gene overlaps. Though for gene identification these differing methods do not result in important differences in the genome annotation, they imply an important distinction between two models of RNA processing.

The first method implies a system in which one pre-mRNA, which could be matured to a poly-A mRNA, is produced for each of the transcribed copies of the tRNA genes. Such a model would infer that mRNAs would be considerably less stable than the tRNA genes, as translation would require that an equilibrium be maintained where tRNAs are far in excess of the mRNAs, despite their 1 to 1 transcription ratio. The second model implies that a mechanism exists that determines whether a particular transcript encoding a protein gene / tRNA gene overlap is to be matured into the tRNA or the mRNA. Inclusion of the overlapping nucleotides in the mRNA would remove one to two 5' nucleotides from the amino acid acceptor stem of the potential tRNA, rendering that tRNA molecule non-functional. This would necessitate multiple rounds of transcription in order to maintain tRNA levels for translation.

The methods presented in this thesis are not able to infer which of the above models is at work in insect mitochondria, but the molecules inferred to be processing intermediates of the *nad3* and *cox2* genes may imply that the first model described is the one utilized. If an *nad3* mRNA was directly produced from the pre-mRNA molecule that

contained the *nad3* gene plus the five 3' encoded tRNA genes, CR-RT-PCR should have been able detect a circularized product using primers for the tRNA gene sequences. I was unable to detect any molecules that encoded only these tRNA genes in the absence of the *nad3* sequence. Though this lack of a result cannot prove the absence of a molecule, it is intriguing that the CR-RT-PCR methods were able to detect a circularization of the entire $srRNA+tRNA^{Val}+lrRNA+tRNA^{Leu}+nad1$ transcript, which was not detectable by direct RNA labeling or Northern Blot analysis by Berthier *et al.* (1986). This implies that CR-RT-PCR is much more sensitive at detecting RNAs. If the 5-tRNA molecule is produced by mRNA specific transcript processing, it must be actively recognized and degraded within the mitochondria.

Also, characterization of the $cox2+tRNA^{Lys}+tRNA^{Asp}$ region supports a model of 1 to 1 processing. A circularized product containing $cox2+tRNA^{Lys}+tRNA^{Asp}$ was never identified. That product would only be expected if the 3' encoded *atp8/6* bicistronic message was actively identified and removed by another undescribed processing mechanism. Instead, I was only able to observe the *cox2* only and the $cox2+tRNA^{Lys}$ products, as would be predicted by the active processing of the tRNA genes.

To determine whether tRNAs and mRNAs are matured via a 1 to 1 model, or through specific targeting will require direct experimentation on the rates of degradation of mRNAs and tRNA molecules, or through the complete characterization of the mitochondrial RNA processing machinery.

5.2. Insect Mitochondrial Genomics

5.2.1. Two Insect Mitochondrial Genomes

The mt-genomes of *Philaenus spumarius* and *Pteronarcys princeps* are presented in Chapters 2 and 3, respectively. Generally the two genomes were quite typical in terms of genome content, genome arrangement, nucleotide composition, codon usage and A+T bias.

The stonefly sequence is significant in that it now represents the second order of early branching Neoptera (Neoptera excluding the Endopterygota and Paraneoptera) for which complete mt-genome sequence is available (Flook, *et al.*, 1995; Kim, *et al.*, 2005), or the third if you include the near-complete cockroach sequence (Yamauchi, *et al.*, 2004). The positional bias analysis conducted based on asymmetric replication caused mutational bias should become another standard analysis done on new mt-genomes, so that unusual evolutionary alterations in the mode of replication of insect mitochondria can be detected.

The spittlebug sequence is also significant as it represented the first auchenorrhynchan Hemiptera sequence in the database. With the impressive diversity of other Hemiptera now available, it is now possible to begin to investigate some of the unusual phenomenon within the Paraneoptera, such as gene order instability and the reversion to the strict coding of recognizable initiation codons for the *cox1* gene.

5.2.2. Annotation of Mitochondrial Genes

The annotation of the start for *cox1* has long been problematic. The Diptera and Lepidoptera encode a conserved, in-frame TAA or TAG stop codons in place of a

recognizable start codon (Clary and Wolstenholme, 1983b; de Bruijn, 1983; Beard, *et al.*, 1993; Mitchell, *et al.*, 1993; Caterino and Sperling, 1999; Ballard, 2000a; 2000b; Lessinger, *et al.*, 2000; Spanos, *et al.*, 2000; Yukuhiro, *et al.*, 2002; Nardi, *et al.*, 2003a; Junqueira, *et al.*, 2004). Comparative analyses began to question the use of the proposed four-base codons based on the lack of conservation of such an unusual coding signal within other organisms (Beard, *et al.*, 1993). In contrast, the Paraneoptera (excluding the psocopteran barklouse) faithfully code an ATN start codon downstream of the $tRNA^{Tyr}$, and most commonly utilize the ATG start codon (Shao, *et al.*, 2001; Shao and Barker, 2003; Thao, *et al.*, 2004), including the spittlebug sequence presented in Chapter 2. This confusing array of comparative observations has further contributed to a very diverse set of annotations of the *cox1* gene.

Pfam searches of mitochondrial *cox1* from insects show a strong conservation of a *cox1* domain. This domain initiates with the tryptophan at amino acid 4 of the translated *cox1* open reading frame from *Drosophila,* predicted from the obtained cDNA sequence. This domain, starting with the tryptophan, is very well conserved through mitochondrial *cox1* and the known bacterial homologues. The crystal structure of the bull (*Bos taurus*) mitochondrial complex IV has been solved (Tsukihara, *et al.* 1996). This crystal structure reveals that the 5' end of *cox1* forms a small alpha helix, initiating with amino acid 2 and ending with the conserved tryptophan mentioned above. By alignment, this helical region is not present in the *Drosophila* mRNAs, implying that its function has been lost or compensated for by other structural changes in the insect *cox1*.

My proposed annotation and alignment of insect the *cox1* genes is presented in Figure 5.2. Based on the cDNA sequence obtained from *D. melanogaster*, I propose that

annotation of the gene should initiate with the first in-frame nucleotide of the *cox1* open reading frame whether a putative initiation signal can be detected or not. This annotation should not overlap with the complementary sequence to the *tRNA*^{Tyr} gene at the 5' end, or any other 5' abutting genes in the case of genome rearrangements relative to the conserved insect arrangement.

RNA sequencing should be expanded to better sample the diversity of *cox1* gene structure represented by the insects to help us come to a clearer understanding of the molecular structure of this gene. Obvious candidates for sequencing would include start-codon containing Hemiptera (such as *Triatoma* or *Philaenus*) (Dotson and Beard, 2001). Also, taxa that appear to encode neither a start nor stop should be investigated, such as a beetle or the giant stonefly. A dragonfly would also an interesting taxon, as they encode an extensive sequence between the *tRNA*^{Tyr} and *cox1* that is presumably not part of the mature mRNA. I would also like to sequence the *cox1* mRNA from another *Drosophila* species for which the TCG serine found to initiate the *cox1* of *D. melanogaster* is known to overlapped the *tRNA*^{Tyr} gene to examine the effects of this overlap of the composition of the mature mRNA (Ballard, 2000a; 2000b).

5.2.3. Genome Structure and Rearrangements

After the release of the GenBank file representing the mt-genome of *Philaenus spumarius,* a large number of mt-genomes representing the remaining hemipteran group, the Sternorrhyncha, were made public (Thao*, et al.*, 2004). These species included six whiteflies, an aphid and a psylid. A second mt-genome from an auchenorrhyncan, the leafhopper *Homalodisca coagulatta* was also released in GenBank (Baumann and Baumann, unpublished).

```
                   ....|....|....|....|....|....|....|....|....|....|....|..
                                      ______________________________________
amino acid                            W  l  f  S  T  N  H  K  D  I  G  T  l  y
D. melanogaster    ------tcgcgacaatgattatttctacaaatcataaagatattggaacttatat
D. yakuba          ------...........g.....................................
screwworm fly      ---------........g..........t................t.........
Mediterian fly     ---------.a......gc......a..t..........................
olive fly          ---------........gc......a........c...................c
mosquito           ---------................a...........g.................
oakmoth            ---------...a....c.t.a.................................
silkwormmoth       ---------...a....a.t.a..........................a......
cornborermoth      ---------...a....c...a...a..c..........................
sawfly             ---------atta........a...a..t.....c....................
honeybee           ------ataat.a.g....ca.a..a..c........a.......g.tc..g...
flourbeetle        ------ct.aata............a..t..c..c.............ac.t...
asparagusbeetle    ---------aata....gc.g.......t..c................a.....c
fireflybeetle      ---------aata......t.a......t........a.................
kissingbug         ------at.aata....c.t..c..a.....c..c..g.....c.....c.g...
spittlebug         ------at.aa.a....a....c.....t.....c........c...........
leaf hopper        ------at.aata........................g..c........c....c
aphid              ------ata---a....a.c.....a..t..........................
psylid             ------at.---att..gc.t...aga....c..c..ac......t........c
barklouse          ---------aata.g..ga..........................t.........
louse              ---atg.tttct.gg....t.a....t.t.....c..a.......t.t.......
thrips             ------.taga.a....c.t.....a..............c......t......c
locust             ------a...a.a.........c..a.....c..c..g..c....t..ac.g...
mole cricket       ------at..a..g........c..c..c..............c....cc.t...
cockroach          ------ct..a..g...a.......g...........g................c
dragonfly          ------.t....g..g...........t........g..........c....c
stonefly           ---------.............c..a........c..g................c
silverfish         ---------aa..g........c..g..t..c..c.....c....t..cc.....
firebrat           ------.t..a..g.................c.............t..aa....c
archeognatha       ---------.a..g...gc...c..a.....c........c....t..c..g...
giantspringtail    ------------a....a.c.....a........c..............a.....
springtail         ---------tt..gc..a.......a.......................a....c
Daphnia            ------.t....gt..c....c..a..t........g..c....c...c.c...

Bos taurus         atgttcattaac.gc..c....c..a..c..c.............t..cc.t...
                    M  F  I  N  R  W  L  F  S  T  N  H  K  D  I  G  T  L  Y
```

Figure 5.2. Proposed Annotation of Insect *cox1* 5' ends for Representative Insect and Hexapod Taxa. Nucleotide identities to the *D. melanogaster* sequence are represented by a dot. Underlined nucleotides represent putative wobble-interacting start codons or ATG start codons. The amino acid consensus of the hexapod alignment is above the *D. melanogaster* sequence, with capital letters representing amino acids always present, and lowercase representing amino acids present in >80% of the hexapod sequences. The black line above the alignment represents sequence within the conserved *cox1* domain predicted by Pfam. The bull (*Bos taurus*) *cox1* nucleotide sequence inferred from the crystal structure is compared below, with its amino acid translation.

The genome arrangement of the kissing bug, and the two auchenorrhynchan mt-genomes have conserved the ancestral mt-gene arrangements. Of the sternorrhynchan sequences, only the psylid and aphid sequences appear to conserve the ancestral insect mt-genome arrangement. The whitefly mt-genomes all appear to have undergone at least minimal tRNA re-arrangements (such as in *Trialeurodes, Aleurochiton* and *Aleurodicus*), while the others have undergone major tRNA and protein coding gene rearrangements (*Bemisia* and *Tetraleurodes*) or rearrangements of all gene types (in *Neomaskelia*) (Thao, *et al.*, 2004). For the non-hemipteran orders within the Paraneoptera, protein gene rearrangements are observed in all three representative mt-genomes currently available (Shao, *et al.*, 2001; Shao and Barker, 2003; Shao, *et al.*, 2003). What is unusual about the mitochondria from these four insect orders that could lead to this predisposition to re-arrange, when other insect groups appear to conserve the gene arrangement?

One feature is the increase in reported tandem repeat units observed in these non-rearranged taxa within the Paraneoptera. The kissing bug sequence is very unusual in that it encodes a large non-coding region between the $tRNA^{Ser\text{-}UCN}$ and *nad1* genes (Dotson and Beard, 2001). This region is 309bp in length and consists of eight direct repeat units between the $tRNA^{Ser\text{-}UCN}$ and *nad1* genes. The *Philaenus* and *Homalodisca* sequences are also noted for their repeat structures in their A+T regions.

Tandem repeats have been demonstrated to be involved in bacterial genome rearrangements (Rocha, 2003) and are implicated in human disease-causing mitochondrial deletions (Samuels, *et al.*, 2004), and can, if they form stem-loop structures, lead to duplications that can result in gene order rearrangement (Stanton, *et al.*, 1994).

Diptera also have extended repeat units within their A+T rich regions, and tRNA gene rearrangements and translocations are also observed in the mosquitoes (Beard, *et al.*, 1993; Mitchell, *et al.*, 1993), plus the duplication of the $tRNA^{Ile}$ has been observed in *Chrysomya* blowflies (Lessinger, *et al.*, 2004).

A predictive example of this theory could also be examined in the *Pyrocoelia* firefly beetles. The complete mt-genome of *Pyrocoelia rufa* revealed a large expansion of a repeat motif between the *nad2* gene and the $tRNA^{Trp}$ genes (Bae, *et al.*, 2004). Amplification and sequencing of this mt-genome region in other *Pyrocoelia* species may reveal the genome rearrangements and further support the association between large repeats and predisposition of genome rearrangement.

5.3. Future Directions

5.3.1. Modification of the CR-RT-PCR Protocol

The CR-RT-PCR methodology was more efficient than the 5' RACE methods at obtaining 5' sequence of the mitochondrial RNAs. Unfortunately, the ligation of the poly-A tail to the 5' end of the RNA molecules obscured precise annotation of the 5' transcribed nucleotide in nine of the eleven mRNAs and both of the rRNAs due to the linkage of a variable length 3' polyadenylated sequence to one or more 5' A residues. For this reason, it was important to continue to attempt to obtain cDNA sequence through 5' RACE amplification. By comparison of CR-RT-PCR sequence obtained from a 5' RACE sequence in which a nucleotide stretch other than A was added to the 5' end of the cDNA, it became possible to unambiguously identify the 5' transcribed nucleotides.

As CR-RT-PCR was a more efficient means of amplification, it would be advantageous to modify this protocol to introduce non-A nucleotides to the terminal end of existing poly-A tails, thereby speeding the precise identification of the processed ends of the RNAs. The poly(A) polymerase enzyme has long been utilized to add poly-A tails to bacterially derived transgene RNAs *in vitro*. It has been noted that in the absence of rATP molecules, the poly A polymerase has the capacity to add nucleotides other than A to the 3' end of an RNA (Martin and Keller, 1998). By combining a poly-A polymerase step to add non-A nucleotides to the mitochondrial RNAs, followed by RNA circularization, it may be possible to develop a rapid protocol to generate circularized RNAs, with a non-A nucleotide linked to the 5' end of the RNA molecule. Comparisons to the non-modified circularized mRNA would lead to the easy discrimination of the exact 5' boundaries of all of the RNAs.

Also, the model presented for the transcription of the *cox1* gene should be investigated further. The obvious next investigation would be the isolation and peptide sequencing of the *cox1* protein to determine which amino acids are translated by the *Drosophila* mitochondrial translation machinery.

5.3.2. Transcription of A+T Rich Region Sequence

The *D. melanogaster* mitochondrial A+T rich region is a very large region of DNA (4601bp in size) that contains a large number of tandemly repeated sequence units (Lewis, *et al.*, 1995). As the name implies, the A+T region is heavily over-represented by A and T nucleotides, making PCR amplification and sequencing technically challenging. Also the large and repetitive structure makes primer design within the A+T region very difficult. As a result, RT-PCR reactions linking *srRNA* and *nad2* were not attempted. I

was also unable to construct primers that would test for the presence of RNA species that encoded sequence from the A+T rich region and the flanking genes. Given the continued assumption that the A+T region contains regulatory elements that are important to replication and transcript initiation, it is unfortunate that *D. melanogaster* mt-DNA encodes such an unusual A+T rich region. Unlike the D-loop region in vertebrate mitochondria, there have been no demonstrations of the role of the major non-coding region in *Drosophila* mitochondrial transcription, nor reports of A+T region transcripts.

The CR-RT-PCR protocol utilized in this study should have been able to overcome any limitations of primer design. It is expected that the amplification of the *nad2* and *srRNA* genes after circularization would have detected pools of the matured forms of the RNAs, as well as the unprocessed primary transcripts that would encode A+T region sequence that was transcribed immediately downstream of the transcriptional promoter regions.

Not all insects encode such problematic mitochondrial A+T rich regions. Many of the other *Drosophila* species have much smaller A+T regions, and are amenable to growth in laboratory conditions. Comparative analyses between the A+T region of *Drosophila* species has been undertaken, and conserved sequence regions have been identified that may be candidates for replication or transcriptional control (Lewis, *et al.*, 1994). Further experimentation should be undertaken with another *Drosophila* species to examine what role A+T rich sequences may have on mitochondrial transcription in flies.

The exact composition of the 5' most nucleotide of mitochondrial transcripts is currently not known for animal mitochondria. In a study of mitochondrial RNA from HeLa cells it was determined that 5' caps were not present in human mitochondrial

RNAs, and that only mono-phosphate (rNMPs) and a di-phosphate (rADP), presumed to be from the 5' of the RNAs digested in the study, were present in observable levels (Grohmann, *et al.*, 1978). If these finding could be extended to insect mitochondria, it could be suggested that primary transcripts either start with a di-phosphate or mono-phosphate nucleotide. If they do initiate with mono-phosphate nucleotides, primary transcripts should have been detectable by CR-RT-PCR, and therefore imply that the large *srRNA-nad1* transcript is the primary transcript. This would imply that the sequence directly upstream of the observed *srRNA* cDNA sequence is the transcriptional promoter for this gene cassette.

If the di-phosphate is the natural state for the primary transcript, the primary transcript would not have been circularized (due to the T4 RNA ligase specificity for 5' mono-phosphate). The large *srRNA-nad1* molecule observed would then be the product of a 5' removal of a transcribed leader sequence from this multi-gene transcript.

It should be feasible to explore these two described scenarios by a simple modification of the protocols used in this study. If the mitochondrial total RNA pool was first de-phosphorylated by an RNA phosphatase, then the molecules mono-phosphorylated by an RNA kinase, it should be feasible to detect the primary transcript by CR-RT-PCR, providing the correct 3' facing primer could be chosen. By adopting this protocol, it may be possible to determine the transcript initiation sites in the A+T rich region, without switching to another *Drosophila* species as the model organism.

5.3.3. Further Characterization of Nuclear Transcription and Processing Factors

In vitro transcription initiation is now possible, using mammalian mitochondrial transcription factor alpha TFAM, (Garstka, *et al.*, 1994), either of the two mt-

transcription factor betas (TFB1 or TFB2), and the mitochondrial RNA polymersae POLMT, (Tiranti, *et al.*, 1997) (Falkenberg, *et al.*, 2002). This simple transcription system cannot yet function on isolated mitochondrial DNA, but can transcribe from mitochondrial sequences cloned into plasmids (Gaspari, *et al.*, 2004a). A mitochondrial transcription termination factor DmTERF1 has been characterized (Fernandez-Silva, *et al.*, 1997), and three additional mTERFs are being investigated (Gustafsson, C.M., personal communication).

Additionally, human mitochondrial poly(A) polymerase has been characterized (Tomecki, *et al.*, 2004; Nagaike, *et al.*, 2005). Given the tight association between the polyadenylation of mt-mRNAs and the transcript processing machinery, it may now be possible to identify the processing factors through protein-protein interactions with mt-poly(A) polymerase. I was able to quickly identify a putative *Drosophila* homologue of the mammalian poly(A) polymerase using a tBlastn search of the *Drosophila* complete genome sequence. The 612 amino acid candidate, CG11418, was examined to determine if the predicted protein sequence had a mitochondrial import signal using MitoProtII (Claros and Vincens, 1996). The first 33 amino acids were determined to be a mitochondrial import signal (with p=0.833). The sequence was then analyzed using Pfam (version 17.0, March 2005) and found to contain a 406 amino acid poly(A)polymerase associated domain (Sonnhammer, *et al.*, 1997). A putative homologue for *Anopheles* (XM_319519) was also identified using tBlastn of the human sequence against the *Anopheles* complete genome. XM_319519 was 0.46 identical and 0.68 similar to *Drosophila* CG11418, and had the same poly(A) polymerase domain. The ORF for the *Anopheles* sequence was incomplete at the 5' end, so the presence of a mitochondrial

import signal could not be investigated. This simple analysis has lead to potential candidate for biochemical and molecular characterization to determine if it is the *Drosophila* mt-poly(A) polymerase. If the molecule does turn out to be the *Drosophila* mt-poly(A) polymerase gene, it will be of great interest to determine what other proteins interact with this enzyme. Given the close association between transcript processing and polyadenylation, it would be very interesting to determine if other enzymes involved in mitochondrial transcript processing could be isolated by co-immunoprecipitation or other such molecular methods. These and other studies should be undertaken in attempt to isolate, characterize and eventually reconstitute the components of the mitochondrial processing machinery for *in vitro* studies.

In *Drosophila,* homologues for TFAM (Goto, *et al.*, 2001; Takata, *et al.*, 2001), TFB2 (Matsushima, *et al.*, 2004), RNA polymerase (Goldenthal and Nishiura, 1987) appear to make up the minimal transcription unit. The first *Drosphila* mitochondrial termination factor, DmTERF1 has been characterized (Roberti, *et al.*, 2003; Roberti, *et al.*, 2005), with three other paralogues identified and awaiting characterization (Gustafsson, C.M., personal communication).

This body of knowledge strongly points towards the possibility of complete *in vitro* mitochondrial transcription and transcript processing system being developed. The *in vitro* system would have profound impacts upon our understanding of mitochondrial transcription and transcript maturation, as it is still not possible to introduce recombinant DNA or exogenous mRNAs into animal mitochondria. Also, due to the fundamental biochemical processes being carried out by the protein subunits expressed from mt-DNA, the isolation and characterization of non-fatal mitochondrial mutations in animal model

organisms has been a very slow and random process. As the field progresses towards a cell-free mitochondrial transcription system, the possibility of experimental observation of these processing mechanisms becomes an exciting possibility.

Appendix A. GenBank Flatfile for *Philaenus spumarius.*

```
LOCUS       AY630340               16324 bp    DNA     circular INV 11-MAY-2005
DEFINITION  Philaenus spumarius mitochondrion, complete genome.
ACCESSION   AY630340
VERSION     AY630340.1  GI:47933676
KEYWORDS    .
SOURCE      mitochondrion Philaenus spumarius (meadow spittlebug)
  ORGANISM  Philaenus spumarius
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Paraneoptera; Hemiptera; Euhemiptera; Cercopoidea;
            Aphrophoridae; Philaenus.
REFERENCE   1  (bases 1 to 16324)
  AUTHORS   Stewart,J.B. and Beckenbach,A.T.
  TITLE     Insect mitochondrial genomics: the complete mitochondrial genome
            sequence of the meadow spittlebug Philaenus spumarius (Hemiptera:
            Auchenorrhyncha: Cercopoidae)
  JOURNAL   Genome 48 (1), 46-54 (2005)
   PUBMED   15729396
REFERENCE   2  (bases 1 to 16324)
  AUTHORS   Stewart,J.B. and Beckenbach,A.T.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAY-2004) Molecular Biology and Biochemistry, Simon
            Fraser University, 8888 University Drive, Burnaby, British Columbia
            V5A 1S6, Canada
FEATURES             Location/Qualifiers
     source          1..16324
                     /organism="Philaenus spumarius"
                     /organelle="mitochondrion"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:36667"
                     /dev_stage="adult"
     tRNA            1..63
                     /product="tRNA-Ile"
                     /note="codons recognized: AUY"
                     /anticodon=(pos:29..31,aa:Ile)
     tRNA            complement(61..129)
                     /product="tRNA-Gln"
                     /note="codons recognized: CAR"
                     /anticodon=(pos:complement(97..99),aa:Gln)
     tRNA            129..198
                     /product="tRNA-Met"
                     /note="codons recognized: AUR"
                     /anticodon=(pos:160..162,aa:Met)
     gene            199..1191
                     /gene="nad2"
     CDS             199..1191
                     /gene="nad2"
                     /codon_start=1
                     /transl_table=5
                     /product="NADH dehydrogenase subunit 2"
                     /protein_id="AAT39431.1"
                     /db_xref="GI:47933677"
                     /translation="MTMMNSTKMIFLLFMFVSPLISLSSNNWFGSWMGLEINLISFTP
                     LMPSKNNYYSSESSMKYFIIQSVSSMILLLGIISLSLMMSYSYLIMICALMAKLGVAP
                     FHMWAPSVMEGISWFNCFIFLTWQKLAGLMLMSYIIINSMVVFPVIASLIIGSFGGIN
                     QSSMKKLMAYSSINNMGWIIMGMTISFSLWMSYFLIYTFMVYNLIYLFMIMEINYLNQ
                     FLINMHNSSFKIILSLLLFSFGGVPPLLGFLPKLIIIQSLMLNSNFLILLIMIMTSLI
                     TLFYYIRVTTMMLMVNSSKIKFNNYLLLNSKLFYLMTFMNLFGFGLIFIFKTIN"
     tRNA            1191..1255
```

```
                     /product="tRNA-Trp"
                     /note="codons recognized: UGR"
                     /anticodon=(pos:1222..1224,aa:Trp)
     tRNA            complement(1248..1314)
                     /product="tRNA-Cys"
                     /note="codons recognized: UGY"
                     /anticodon=(pos:complement(1281..1283),aa:Cys)
     tRNA            complement(1317..1381)
                     /product="tRNA-Tyr"
                     /note="codons recognized: UAY"
                     /anticodon=(pos:complement(1351..1353),aa:Tyr)
     gene            1382..2915
                     /gene="cox1"
     CDS             1382..2915
                     /gene="cox1"
                     /note="TAA stop codon completed by the addition of 3' A
                     residues to the mRNA"
                     /codon_start=1
                     /transl_except=(pos:2915,aa:TERM)
                     /transl_table=5
                     /product="cytochrome c oxidase subunit I"
                     /protein_id="AAT39432.1"
                     /db_xref="GI:47933678"
                     /translation="MKKWMFSTNHKDIGTLYFLFGIWSGMIGTTLSLLIRVELGQPGS
                     FIGDDQIYNVIVTSHAFIMIFFMVMPIMIGGFGNWLVPLMIGAPDMAFPRMNNMSFWM
                     LPPSLTLLLSSSMIDNGVGTGWTVYPPLSSGVAHAGACVDLAIFSLHLAGISSILGAV
                     NFITTIFNMRSSGMKMDRTPLFVWAVLITTILLLLSLPVLAGAITMLLTDRNINTSFF
                     DPSGGGDPILYQHLFWFFGHPEVYILILPGFGLISHIISQESGKNESFGSLGMIYAMM
                     AIGLLGFVVWAHHMFTVGMDVDTRAYFTSATMIIAVPTGIKIFSWLATMHGMPFKLSS
                     PILWSIGFVFLFTIGGLTGIVLSNSSIDIILHDTYYVVAHFHYVLSMGAVFAILGSFI
                     QWYPLFTGLTMNSKWLKMQFMIMFVGVNLTFFPQHFLGLSGMPRRYSDYPDAYMSWNI
                     LSSIGSMISFIGILLLIFIVWESLISKRKSIFSKNMISSIEWLQMMPPSEHSYNELPM
                     LTN"
     tRNA            2916..2980
                     /product="tRNA-Leu"
                     /note="codons recognized: UUR"
                     /anticodon=(pos:2946..2948,aa:Leu)
     gene            2981..3650
                     /gene="cox2"
     CDS             2981..3650
                     /gene="cox2"
                     /note="TAA stop codon completed by the addition of 3' A
                     residues to the mRNA"
                     /codon_start=1
                     /transl_except=(pos:3650,aa:TERM)
                     /transl_table=5
                     /product="cytochrome c oxidase subunit II"
                     /protein_id="AAT39433.1"
                     /db_xref="GI:47933679"
                     /translation="MATWSNFGLQDGMSPLMEQMIFFHDHTLMILIMITMMVGYTMVS
                     LTYNNLINRYLLEGQMIELLWTIIPAITLVFIALPSLKLLYLLDEMSNPSITIKSIGH
                     QWYWSYEYSDFKNLEFESYMSLYEKTKFRLLDVDNRIILPFLIQIRMLVCSMDVIHSW
                     TIPSLGVKLDALPGRLNQMSFLMSRPGLMFGQCSEICGSNHSFMPIVIESISMEKFVK
                     WLKSY"
     tRNA            3651..3721
                     /product="tRNA-Lys"
                     /note="codons recognized: AAR"
                     /anticodon=(pos:3681..3683,aa:Lys)
     tRNA            3723..3789
                     /product="tRNA-Asp"
                     /note="codons recognized: GAY"
                     /anticodon=(pos:3753..3755,aa:Asp)
     gene            3791..3946
```

```
                     /gene="atp8"
     CDS             3791..3946
                     /gene="atp8"
                     /codon_start=1
                     /transl_table=5
                     /product="ATP synthase FO subunit 8"
                     /protein_id="AAT39434.1"
                     /db_xref="GI:47933680"
                     /translation="MPQMAPMWWLSLFMMFILSFFLFNSIIYFFFKSNKINTFNKFII
                     NQMIWKW"
     gene            3940..4599
                     /gene="atp6"
     CDS             3940..4599
                     /gene="atp6"
                     /codon_start=1
                     /transl_table=5
                     /product="ATP synthase FO subunit 6"
                     /protein_id="AAT39435.1"
                     /db_xref="GI:47933681"
                     /translation="MMTNLFSVFDPSTGILSLNWVSTLIGLIIIPMPLWILPSRLFIM
                     NLIIFQKINNELKILMTMNSKGNSLIFVSMFFFILYNNIMGLLPYIFTSSSHLVFTLS
                     LALPMWLMLMIFGWLMNTNKMFYHLIPVGTPSILMPFMVCIETISNMIRPGSLAVRLT
                     ANMIAGHLLMTLLGNMSINSSFIMLIIIIQLMLMIFETAVSIIQAYVFTILSTLYSSE
                     I"
     gene            4602..5382
                     /gene="cox3"
     CDS             4602..5382
                     /gene="cox3"
                     /note="TAA stop codon completed by the addition of 3' A
                     residues to the mRNA"
                     /codon_start=1
                     /transl_except=(pos:5382,aa:TERM)
                     /transl_table=5
                     /product="cytochrome c oxidase subunit III"
                     /protein_id="AAT39436.1"
                     /db_xref="GI:47933682"
                     /translation="MMKNHPFHLVDASPWPITGSIGAMTITSGMVMWFHQLNLFLFML
                     GLIIIIILTMFQWWRDVTRESTFQGNHTIIVTKGMKWGMLLFIVSEVFFFLSFFWGFFH
                     SSLSPSMEIGMMWPPKGIVTFNPMQIPLLNTMILLSSGITITWSHHSMMNSNHSQSVQ
                     GLFLTIILGIYFSMLQGYEYYESPFTISDSIYGSTFFMSTGFHGLHVLIGTTFIIVTM
                     ARQMNFHFSNKHHFGFEAAAWYWHFVDVVWLFLYISIYWWGS"
     tRNA            5383..5447
                     /product="tRNA-Gly"
                     /note="codons recognized: GGN"
                     /anticodon=(pos:5414..5416,aa:Gly)
     gene            5448..5801
                     /gene="nad3"
     CDS             5448..5801
                     /gene="nad3"
                     /codon_start=1
                     /transl_table=5
                     /product="NADH dehydrogenase subunit 3"
                     /protein_id="AAT39437.1"
                     /db_xref="GI:47933683"
                     /translation="MMNLIMSSFMIITILILLIILLTLISKKSFMDREKSSPFECGFD
                     PINSSRIPFSLHFFLMAVIFLIFDVEIVLILPITIIFKYSVIWEWMLSSIMFIMILIL
                     GLYHEWYHGMIEWAN"
     tRNA            5801..5864
                     /product="tRNA-Ala"
                     /note="codons recognized: GCN"
                     /anticodon=(pos:5830..5832,aa:Ala)
     tRNA            5865..5927
                     /product="tRNA-Arg"
```

```
                     /note="codons recognized: CGN"
                     /anticodon=(pos:5892..5894,aa:Arg)
     tRNA            5926..5990
                     /product="tRNA-Asn"
                     /note="codons recognized: AAY"
                     /anticodon=(pos:5957..5959,aa:Asn)
     tRNA            5991..6055
                     /product="tRNA-Ser"
                     /note="codons recognized: AGN"
                     /anticodon=(pos:6015..6017,aa:Ser)
     tRNA            6057..6122
                     /product="tRNA-Glu"
                     /note="codons recognized: GAR"
                     /anticodon=(pos:6084..6086,aa:Glu)
     tRNA            complement(6121..6185)
                     /product="tRNA-Phe"
                     /note="codons recognized: UUY"
                     /anticodon=(pos:complement(6149..6151),aa:Phe)
     gene            complement(6185..7894)
                     /gene="nad5"
     CDS             complement(6185..7894)
                     /gene="nad5"
                     /codon_start=1
                     /transl_table=5
                     /product="NADH dehydrogenase subunit 5"
                     /protein_id="AAT39438.1"
                     /db_xref="GI:47933684"
                     /translation="MNYKINLGFITSLMLILFSLIFFLMFLYFVIHEKLILLEWSIIF
                     LNSSSVYMSIILDWMSLIFMSFVLGISSFVIFYSNEYMIVDKNINRFIYLIILFVASM
                     MFLIISPNLISILLGWDGLGLISYCLVVYYQNNKSYNAGMLTALMNRIGDVALLMSIG
                     WMMNFGGWNYIFYLDKMFNFEFISFMIILAAFTKSAQLPFSSWLPAAMAAPTPVSALV
                     HSSTLVTAGVYLLIRFNLLIMNFFMIDFFLILSLLTMIMSSFGANFEFDLKSIIALST
                     LSQLGFMMSVLMMGFPSMAFFHLLTHALFKALLFLCAGVVIHSMWGNQDIRFMGGLIN
                     QLPIITSCMNISNLALCGFPFLTGFYSKDLIMESMMMSNLNFLSYMLFLLSVGLTVSY
                     SFRLSYYTLSGNLNSFSFLCIYDCSSNMSWSIFIMVFFSIIGGSMLNWLMFSSPVVIM
                     LPLDLKLTTLTMIVMGSLIGYELSMYKNNFFYYFPLLNYFSGMMWFTPLFSTFFFTNH
                     LFKFSLIFSLVINQGWGEYWGPKGLSNLVVNFSKFNQIIQMNSFKFYLMSLIIWILFM
                     FMI"
     tRNA            complement(7895..7959)
                     /product="tRNA-His"
                     /note="codons recognized: CAY"
                     /anticodon=(pos:complement(7927..7929),aa:His)
     gene            complement(7960..9280)
                     /gene="nad4"
     CDS             complement(7960..9280)
                     /gene="nad4"
                     /note="TAA stop codon completed by the addition of 3' A
                     residues to the mRNA"
                     /codon_start=1
                     /transl_except=(pos:complement(7960),aa:TERM)
                     /transl_table=5
                     /product="NADH dehydrogenase subunit 4"
                     /protein_id="AAT39439.1"
                     /db_xref="GI:47933685"
                     /translation="MLKFIVFIIFLIPLYFNFWLIQSMMFLLCILLMFSNYGFNYTMM
                     SYFFSIDLISLGFIILSIWISCLMVMASFMIKKNNNYPGSFILLVNLLMFFLIVSFST
                     NNYLLFYFFFESSLVPTLILIFGWGYQPKRLSSGYYLLFYTLLASLPLLLSIIYLYNL
                     NGSFMFMINSYNSNCLYIYIGLIMAFLVKLPMFMFHFWLPKAHVEAPISGSMVLAGIL
                     LKLGGYGLMRFMYFVPNYFITYNYIWIVVSLYGGVLISLVCLVQSDVKSLIAYSSVCH
                     MSLVLMGIMTLFCWGESGSFMLMLGHGLCSSGLFCLSNIMYERLNSRSFFISKGMMSF
                     MPSMSLFWFMFCSSNMASPPSMNLLGEIMIINSLISWSSFSVLFLGLMSFLSACYSLY
                     LFSYSQHGNFYSGYNSFSCGYVREYLLLILHWLPLNFFVLKLNMFLLL"
     gene            complement(9274..9561)
```

```
                     /gene="nad4l"
     CDS             complement(9274..9561)
                     /gene="nad4l"
                     /codon_start=1
                     /transl_table=5
                     /product="NADH dehydrogenase subunit 4L"
                     /protein_id="AAT39440.1"
                     /db_xref="GI:47933686"
                     /translation="MNFFIIYMYMYIIGLWTLCSIRKHILLCLLSLEFIVLSLFLMLY
                     FYLVNFNYELYFSMIYLSFSVCEGVLGLSVLVAMVRSHGNDYLNSFNSLLC"
     tRNA            9564..9628
                     /product="tRNA-Thr"
                     /note="codons recognized: ACN"
                     /anticodon=(pos:9594..9596,aa:Thr)
     tRNA            complement(9631..9696)
                     /product="tRNA-Pro"
                     /note="codons recognized: CCN"
                     /anticodon=(pos:complement(9663..9665),aa:Pro)
     gene            9698..10216
                     /gene="nad6"
     CDS             9698..10216
                     /gene="nad6"
                     /codon_start=1
                     /transl_table=5
                     /product="NADH dehydrogenase subunit 6"
                     /protein_id="AAT39441.1"
                     /db_xref="GI:47933687"
                     /translation="MKMMIMMSSIMMSTIFSMMKHPLSMGMILLIQTLITCLMNGLNS
                     YSMWFSYILFITFIGGMLILFIYIASIASNEKFLFSMKMMMIIFMMFMLLIMSMMDM
                     TIVVNNHMIIETMSYNKSVNEMNKEINSIMKMFNMPTTMITLMTIIYLLFTMITVVKV
                     TNMKDGPLRMKN"
     gene            10216..11349
                     /gene="cytb"
                     /note="synonym: cob"
     CDS             10216..11349
                     /gene="cytb"
                     /codon_start=1
                     /transl_table=5
                     /product="cytochrome b"
                     /protein_id="AAT39442.1"
                     /db_xref="GI:47933688"
                     /translation="MFKPMRKKELLSIINSTIVDLPTPCNLSSWWNFGSLLGLCLMIQ
                     VVTGIFLAMHYSSNVEMAFNSISHICRDVNFGWLIRIMHANGASLFFICMFMHIGRGM
                     YYGSYKSIKTWVMGVMIMLTVMATAFLGYVLPWGQMSFWGATVITNLLSAIPYLGESM
                     VKWIWGGFAVDNATLTRFFTLHFLMPFIVLVMSSLHIFFLHKKGSNNPLGISSNIDKI
                     PFHPYFSIKDLMGFTIAMMMFLILNLMEPYILGDPDNFIPANPLVTPAHIKPEWYFLF
                     AYAILRSIPNKLGGVIALIMSILILLTLPFSNMSKFKGMKFYPISQMLFWMLVSTTML
                     FTWIGARPVEEPYIWTGMNLTYIYFSYFIIEPLSKKLWDNFIN"
     tRNA            11349..11413
                     /product="tRNA-Ser"
                     /note="codons recognized: UCN"
                     /anticodon=(pos:11377..11379,aa:Ser)
     gene            complement(11430..12347)
                     /gene="nad1"
     CDS             complement(11430..12347)
                     /gene="nad1"
                     /codon_start=1
                     /transl_table=5
                     /product="NADH dehydrogenase subunit 1"
                     /protein_id="AAT39443.1"
                     /db_xref="GI:47933689"
                     /translation="MYFINVLILVIFIMLGVGFLTLVERKILGYIQVRKGPNKLGFLG
                     IFQPFSDAIKLFTKEQMVPLKSNFLIYYFSPILSLFISLFLWLVYPFLINCFSFIYGI
```

```
                     LFLLCCMSMGVYTIMISGWASCSTYSLLGSLRSVAQTISYEVSLSMIFLSFIFLINSY
                     SLIDFYYYQLNLWFLFICFPLSMCFISSCMAETNRSPFDFAEGESELVSGFNVEYSGG
                     GFSYIFLSEYSSIIFMSVLYSIIFLGCNLSSYTFFLSVTFVCFLFVWVRGTLPRYRYD
                     KLMYMAWSSYLPISLNYFVFFLGFKSLII"
     tRNA            complement(12349..12414)
                     /product="tRNA-Leu"
                     /note="codons recognized: CUN"
                     /anticodon=(pos:complement(12382..12384),aa:Leu)
     rRNA            complement(12417..13661)
                     /product="large subunit ribosomal RNA"
     tRNA            complement(13662..13735)
                     /product="tRNA-Val"
                     /note="codons recognized: GUN"
                     /anticodon=(pos:complement(13698..13700),aa:Val)
     rRNA            complement(13736..14489)
                     /product="small subunit ribosomal RNA"
     misc_feature    14490..16324
                     /note="A+T rich region; putative control region"
     repeat_region   14524..15400
                     /rpt_unit="14524..14569"
     repeat_region   15452..16161
                     /note="second part of repeat is incomplete"
                     /rpt_unit="15452..15597"
ORIGIN
        1 aataagatgc ctgaaaaagg attattttga tgtaataaat tatgtaactg tttactctta
       61 ttataatatt tagaattaaa ctaaatctat taatttcaaa aattaaggtt catcttaaac
      121 tacattataa aaagataagc taataaagct aataggttca taccctattt atagaaataa
      181 ttaattttct tctttttatt gactataata aattcaacta aaataatttt cttattattt
      241 atatttgtta gacctttgat ttccttgtct tctaataatt gatttggttc ttgaatagga
      301 ttagaaatta atttaatttc ttttactcct ctaataccct caaaaaataa ttattattct
      361 tctgaatcaa gaataaaata ttttattatt cagagagtaa gatcaatgat tctattacta
      421 ggaattatta gcttatcatt aataataaga tactcatatt tgattataat ttgtgcactt
      481 atagcaaaat tgggagtcgc accattccat atatgagctc ctagagtaat ggaaggtatt
      541 agatgattta attgttttat tttttaaca tggcaaaaat tagctggttt aatattaata
      601 agttatatta ttattaattc catagttgtt tttcctgtaa ttgcttcttt aattattggt
      661 tcatttgggg gaatcaatca aagttcaata aaaaaattaa tagcttattc ttctattaat
      721 aatataggat gaatcattat aggtataact attagttttt cattatgaat aagatatttt
      781 ttaatctata cattcatagt ttataatcta atttatttat ttataattat agaaattaat
      841 tatttaaatc aatttttaat taatatacac aattcttctt ttaaaattat tttatcctta
      901 ttattatttt cttttggagg agtacctcct ctattaggat ttttacctaa attaattatt
      961 attcagtcat taatattaaa ttcaaatttt ttaattttgt taattataat tataacttca
     1021 ttgattactc tattttatta tattcgggtt acaacaatga tactgatagt taattcatcc
     1081 aaaatcaaat ttaataacta tctattacta aatagaaaat tattttattt aataactttt
     1141 ataaatttat ttggatttgg tttaatcttt atttttaaaa caattaatta aaggttttaa
     1201 gttaaattaa actattaacc ttcaaagtta aatatataat attttataag ccttagtatt
     1261 aatttttact cctttagaat tgcagtctaa aatcatttta ttgactataa ggcttatact
     1321 aaaggaaatc tattatttca tttataaatt tacaatttat tactctttca gccactttag
     1381 tatgaaaaaa tgaatattct ctactaatca caaagatatc ggaactttat attttctatt
     1441 tgggatttga tctggaataa ttgggactac tctaagatta ttaattcggg ttgaattggg
     1501 tcaacctggg tcatttattg gggatgatca aatttataat gtaattgtaa cttcccatgc
     1561 ttttatcatg attttttta tagttatacc aattataatt ggtggttttg gaaattgatt
     1621 agttccatta ataattggtg ctcctgacat agcatttcca cgaataaata atataagatt
     1681 ttgaatgctt cctccttcat taacgcttct tctttcaaga agaataattg ataatggagt
     1741 agggacggga tgaacagttt atcctccttt atcaagaggt gtagcacatg ctggtgcatg
     1801 tgttgattta gctattttct ctttacattt agcaggaatt tcttctattc taggtgctgt
     1861 aaattttatt actactattt ttaatatacg ttcttcaggt atgaaaatag atcgaactcc
     1921 tttatttgta tgagcagttt tgattactac aattttactt ttactttcct taccagtctt
     1981 ggctggagct attacaatgt tattaacaga tcgaaatatt aatacttcat tttttgatcc
     2041 gtcgggggga ggagatccaa ttttatacca acatttattt tgattttttg ggcacccaga
     2101 ggtttatatt ttgattttac cagggtttgg tttaatctct catattatta gacaagaaag
     2161 aggaaaaaat gaatcttttg gatctttagg tataatttat gctataatag caattggttt
     2221 gctaggtttc gtggtttgag ctcatcatat atttactgta ggtatagatg ttgatacacg
     2281 tgcatatttt acttcagcca caataattat tgctgtaccc acaggtatca aaatttttag
```

2341 ttgattggct acaatacatg gaataccatt caaattgtct tctcctattt tatgatcaat
2401 tgggtttgta tttttattta caattggggg tttgacagga attgttttat ctaattcttc
2461 tattgatatt attcttcatg atacttacta tgtagtagct cattttcatt atgtattatc
2521 tataggggca gtatttgcaa ttttagggag atttattcaa tgataccctt tatttactgg
2581 attaacaatg aattcaaaat ggttaaaaat acaatttata attatatttg ttggggtgaa
2641 cctaacattt tttcctcaac attttttagg tttaagagga atacctcgtc gatactctga
2701 ttacccagat gcttacatat cttgaaatat tttatcttca attggaagaa taatttcatt
2761 tattggaatt ttattattaa tttttattgt ttgagaaaga ttaatttcaa aacgaaaaag
2821 aatttttca aaaaatataa tttcttcaat tgagtgactc caaatgatac cccatccga
2881 gcattcttat aacgaattac caatattaac taatttctga tatggcagaa atagtgcatt
2941 aaatttaaga tttaaaaata agaaattatc tttattagaa ttggcaactt gatcgaattt
3001 tggtttacaa gatggaatat cccctcttat agaacaaata atttttttc atgatcacac
3061 attaataatt ctaattataa ttacaataat agtaggttat actatagtta gattgacgta
3121 taataatttg attaatcgtt accttcttga aggtcaaata attgaattat tatgaacaat
3181 cattccagcc attacattag tttttattgc cctgccatca ttaaaattat tgtacctatt
3241 agatgaaata agtaatccat ctattacaat taaatctatt ggtcatcaat ggtactggtc
3301 atatgaatat tcagatttta aaaatttaga atttgaatca tatatatcat tatatgaaaa
3361 aacaaaattc cgattattag atgttgataa tcgtatcatt cttccttttc taattcaaat
3421 tcgaatatta gtctgttcta tggatgttat tcattcatga acaattccaa gattaggagt
3481 aaaacttgat gctttaccgg ggcgattaaa tcaaatgaga tttttaatat cccgtcccgg
3541 tttaatattt ggtcagtgtt cagaaatttg tggatcaaat catagattta tacctattgt
3601 aattgaaagt attagaatag aaaaatttgt taaatgatta aaatcttact cattaagtga
3661 ctgaaagcaa gtaatggtct cttaaaccat attatagtaa attagcattt actcttaatg
3721 gataaaaatt agtttaatta aaatattaat ttgtcaaatt agagaatttg aaataattca
3781 aaatttttat attcctcaaa tagctccaat atgatgatta tccttattta taatatttat
3841 tttatctttc tttttattta attcaattat ttattttttc tttaaatcaa ataaaattaa
3901 cacatttaat aaatttatta ttaatcaaat aatttgaaaa tgataacaaa tttattttca
3961 gtattcgatc cttcaacagg aattttatca ttaaattgag taagaacatt aattggatta
4021 attattattc ctataccttt atgaatttta ccctcacggt tatttattat aaatttaatt
4081 atttttcaaa aaattaataa cgaattaaaa atcttaataa caataaattc aaaaggaaat
4141 tctttaattt ttgtatcaat attttttttt attttatata ataacattat aggcttatta
4201 ccatatatct ttacaagttc aagacattta gtatttacct taagattagc tttacctata
4261 tgattaatat taataatctt tggatgatta atgaatacaa acaaaatgtt ttatcatctt
4321 attcctgtgg ggactccttc aattttaata ccatttatag tatgtatcga aacaattaga
4381 aacatgattc gcccaggatc tttagctgtt cgattaacag caaatataat tgctggtcat
4441 ttactaataa cattactagg aaatatatca attaatagat catttattat attaatcatt
4501 attattcaat taatattaat gatttttgag acagcagtct caattattca agcatatgtg
4561 tttacaattc ttagaactct ttactcaaga gaaatttaaa tatgataaaa aatcatcctt
4621 ttcatttggt agatgcaagc ccttgaccaa ttactggctc tatcggagcc ataacaatca
4681 cctcaggtat agtaatatga tttcaccaac taaatttatt tttatttata ttagggttaa
4741 ttattattat tttaaccata tttcaatgat gacgtgatgt aacacgagag tcaacttttc
4801 aagggaatca tactattatt gtaactaagg gaataaaatg aggaatatta ttattcattg
4861 tatcagaagt atttttttc ttatcttttt tttgaggatt ttttcacaga agattatctc
4921 cttctataga aattggtata atatgacctc caaaaggaat tgtaactttt aatcctatac
4981 aaattccact tttaaataca ataattttgt tatcttcagg gattacaatt acatgatcac
5041 atcatagaat aataaatagt aatcatagcc agtctgttca gggacttttt ttaaccatta
5101 ttttaggtat ttattttct atattacaag gatatgaata ttatgagtcc ccatttacaa
5161 tttcagattc aatttatggt tcaacatttt ttatatcaac cggatttcat gggcttcatg
5221 tattgattgg aactacattt attattgtaa caatggcacg gcaaataaat tttcattttt
5281 caaataagca tcattttgga tttgaagcag ctgcttgata ctgacatttt gttgatgttg
5341 tttgattatt tttatatatt tctatctatt gatgaggtag atactttttt agtattaact
5401 tgtatatttg acttccaatc aaaaggactt aaatctaaga aaaagtattg ataaatttga
5461 ttatatcaag attcataatt attactattt taattttatt aattatttta ttaacattaa
5521 tttctaaaaa aagatttata gatcgggaaa agtcttcccc atttgaatgt ggatttgatc
5581 caattaactc atcacgtatc ccattttcat tacatttttt tttaatagca gtgatttttt
5641 taatttttga tgtagaaatt gtcttaattt taccaattac aattattttt aaatattctg
5701 ttatttgaga atgaatatta tctagaatta tatttattat gattttaatt ttagggttat
5761 accatgaatg ataccatgga ataattgagt gagcaaacta aggattatag ttaattataa
5821 taattgattt gcattcaata ggtatctttg aaagattaat cttaaataga gaagtaatta
5881 tacaattagt ttcgacctaa taataaggaa atttgttcct cctatttaac taaagccaaa
5941 atagaggcaa tttattgtta ataaaaaaat tgaattaatt tctagttaaa agaaaataag
6001 ataataaagt agctgctaac tatatttata agcggtttaa ttccgttttt ttcttaattt
6061 aagaagttta taaacgtttc tttttcaagg aaaaaaagaa gtaatactta cttccttaaa

```
6121 tattttagag taacaactat attaatatct tcaatattat gcttttaatt ttaagctact
6181 aaaattaaat tataaacata aataaaattc aaataattaa tcttattaaa taaaatttaa
6241 atctatttat ttgaataatt tgattaaatt tggagaaatt aacaactaaa ttagacaaac
6301 ctttaggtcc tcaatattcc cctcaacctt gattaattac taaagaaaaa attaatgaaa
6361 acttaaataa atgattagta aaaaaaaaag ttgagaatag gggggtaaac catattattc
6421 cagaaaaata atttaataaa ggaaaataat aaaaaaaatt attcttatac attcttaatt
6481 cataaccaat taaagagcct ataacaatta ttgttagagt agttaatttt aaatctaaag
6541 gtaatataat aactacagga ctactaaata ttaatcaatt tagtatagaa cctccaataa
6601 tagaaaaaaa cactataata aaaattgatc aagatatatt ccttgaacaa tcatagatac
6661 ataaaaaaga aaatgagttt aaattaccag ataaggtata atatcttaga cgaaaagaat
6721 aacttacagt taaaccaact cttaataaaa ataatatata ggatagaaaa tttaaattac
6781 ttattattat agattccata attaaatcct ttgagtaaaa tccagttaaa aaaggaaatc
6841 cacataatgc caaatttgaa atattcatac aagaggtaat aataggtaat tgattaatta
6901 aacctcctat aaaacgaata tcctgattcc ctcatattga atgaataaca acacctgcac
6961 ataaaaataa taaagcttta aataaagcat gtgttaataa atgaaaaaaa gccatagaag
7021 gaaatcctat tatcaaaaca gatattataa atcctaattg gcttaaagta gataaagcaa
7081 taatcctctt taaatcaaat tcaaaattgg ctccaaatct tgatataatt atagttaata
7141 aagacaaaat taagaaaaaa tcaattatga aaaaattcat aattaataaa ttaaatcgaa
7201 ttaatagata aaccccagca gtaactaaag ttgaagaatg aactaaagct gatactggag
7261 ttggggcagc catggctgcc ggcagccatg aagaaaaggg taattgagct ctcttagtaa
7321 aagctgctaa aataattata aaagaaataa attcaaaatt aaatatttta tcaagataaa
7381 aaatatagtt tcaacctcca aaatttatta ttcaaccaat agatattaat aaggctacat
7441 ctccaattcg gtttattaaa gctgttaata tcccagcgtt ataagattta ttattttgat
7501 aataaaccac taaacaataa gaaatcaaac ctaaaccatc tcaacctaac agaattctaa
7561 ttaagttagg actaataatt aaaaatatta tagaagcaac aaatagaata attaaataaa
7621 taaaccgatt aatattttta tcaacaatca tgtattcatt actataaaag attacaaaag
7681 aggaaatacc taatacaaat gatataaaaa ttaaagacat tcaatctaaa ataattgata
7741 tatatacaga actagaattt aaaaaaataa tagatcattc taataaaatg agtttttcat
7801 gaataacaaa atacaaaaat attaaaaaaa aaatcaatga gaataaaatt aacattaatg
7861 aagtaataaa gcccaaatta attttataat taataactta aaatcaatag cagatatcta
7921 tgaatccaca aatcataatt tttattaaac tatttaagta taataataaa aatatattca
7981 actttaatac aaaaaaattt aaaggaagtc aatgaagaat taataataaa tactctcgaa
8041 catacccaca agaaaatcta ttataaccag aataaaaatt gccatgctga ctataagaaa
8101 ataaatacaa tctataacaa gctcttaaga aagacattaa acctaaaaat aaaacactaa
8161 atgaagatca agaaattaat ctattaataa ttataatttc acccaataaa tttattcttg
8221 gaggtctagc tatatttgaa gaacaaaata taaatcagaa tagggatatt gaaggtataa
8281 aacttattat ccccttagaa ataaaaaaac tgcggctatt taaacgttca tatataatgt
8341 tagaaagaca aaaaaggcca gaagaacata gcccatgacc taatattaat ataaaagatc
8401 ctgactctcc ccaacaaaat aaagttataa tacccattaa aactaatctt atatgacaaa
8461 ctgaagaata ggcaattaaa gattttacat ctctttgaac taagcatact aatctaatta
8521 atacccccc atataatctg actacaatcc aaatataatt atatgtaata aaataattag
8581 gtacaaaata cataaaacgt attaaaccgt aacctcctaa tttaagtaaa attccagcta
8641 aaactattga tccagaaata ggagcttcta catgagcttt aggaagtcaa aaatgaaata
8701 taaatatagg taatttaacc aaaaaagcta taattaaacc aatataaata tataaacaat
8761 tagaattata agaattaatc ataaatatga atctaccatt aaggttatat aagtaaatga
8821 tagataataa caaaggtaaa gaagctaata aagtataaaa caataaataa taaccagaac
8881 ttaatcgttt tggttggtac cctcaaccaa aaattagaat taatgttggt actaatcttg
8941 attcaaaaaa aaaataaaat aacaaataat tatttgtact aaaagaaaca atcaaaaaaa
9001 atataagaag gtttactagt aaaataaaac tccctggata attattattt tttttaatta
9061 taaatcttgc cataaccatt aagcatctaa tccaaattct tagaataata aaacctaaag
9121 aaattaaatc aatactaaaa aaatagctta ttatagtgta attaaatcca taattagaaa
9181 atattaacaa aatacataat aaaaatatta ttgactggat taatcaaaaa ttaaaatata
9241 gaggaatcaa aaaaataata aaaacaataa attttaacat aataaggaat taaaagaatt
9301 taagtaatca ttgccatgac tacgaactat agcaactaaa acagataatc ccaacacacc
9361 ttcacaaact ctaaaagata aataaattat tctaaaatat aattcatagt taaaattaac
9421 caaataaaaa tataatatta aaaataaaga taaaacaata aattctaatc ttaatagaca
9481 tagtaaaata tgttttcgaa ttgaacataa agttcataaa ccaataatat atatatatat
9541 ataaatgata aaaaaattca ttagttttaa tagtttaata aaaataatgg tcttgtaatc
9601 catagataga aataaaatct ttaattcatt tcagcaaaaa aaaaaactct tatcttaaat
9661 ctccaaaatt tatattttat aataaactat ttgctgtata aaaataataa ttataatatc
9721 atcaattata atatcaacta ttttttcaat aataaaacac cctttatcaa taggaataat
9781 tttattaatc caaacactaa ttacatgttt aataaatgga ttaaattctt attcaatatg
9841 attttcatac atcttattta ttacatttat tggaggaata ttaattctat ttatttatat
```

```
 9901 tgcaagaatt gcaagaaatg aaaaattttt attttccata aaaataataa taattattat
 9961 ttttataata tttatactat taattatgag aataatagat ataacaattg tggtaaataa
10021 ccacataatt attgaaacaa tatcatataa caaaagcgta aacgaaataa ataaggaaat
10081 taactcaatt ataaaaatat ttaatatacc cacaacaata atcacattaa taacaattat
10141 ttatttacta tttacaataa ttacagtggt aaaagtaact aacataaaag atggtcctct
10201 acgtataaaa aactaatgtt taaaccaata cgaaaaaaag aattacttag tattattaat
10261 agtacaattg tagatttacc aactccatgt aatttatcaa gatgatgaaa ttttggatca
10321 ctattaggat tatgtttaat aattcaagta gttacaggga tttttcttgc tatacactat
10381 tcatctaatg tagaaatagc atttaataga attagacata tttgtcgaga tgtaaatttt
10441 ggatggttaa ttcgaattat acatgcaaat ggagcatcat tatttttcat ttgtatattt
10501 atacatattg gacgaggaat atattatggg tcatataaat ctattaaaac atgagtcata
10561 ggagtaataa ttatactgac agttatagca actgcatttc tagggtatgt cctaccatga
10621 ggtcaaatat cattttgagg ggcaacagta attaccaatt tattatcagc aattccttac
10681 ctcggagaat caatagtaaa atgaatttga ggaggatttg ctgttgataa tgcaacatta
10741 acacgatttt tcactttaca tttttttaata ccattcattg tgttagttat aagaagctta
10801 catatttttt ttcttcacaa aaaaggatca aataatccac ttggaatttc atcaaatatt
10861 gataaaattc catttcatcc ctatttttca attaaagatt taataggatt tactattgct
10921 ataataatat ttttaatcct aaatttaata gaaccttata ttttaggaga ccctgataat
10981 tttatcccag caaatccatt agtcacccct gcccacatta aaccagaatg atacttctta
11041 tttgcgtatg ctattctacg atcaattccc aacaaattgg ggggggtaat tgcattaatc
11101 atatcaatct taattttatt aactttacct ttttcaaata taagaaaatt taaaggaata
11161 aaattttatc caattagtca aatattattt tgaatattag tgtcaacaac aatattattt
11221 acttgaattg gagcacgacc agttgaagaa ccttacattt gaacaggaat aaatctcaca
11281 tatatttatt tctcttattt tattattgaa ccattaagaa aaaaattatg ggacaacttt
11341 attaattaag ttgattagct tataaagggt atgatttgaa atcataagat agaaaagaaa
11401 actttattaa cttaactata attaatgaat taaataatta aagacttaaa ccctaaaaaa
11461 aaaacaaaat aatttaaaga aatgggtaga tatctcctcc aagctatata tataagctta
11521 tcataacgat aacgaggtaa agtaccacgt actcatacaa ataaaaaaca aacaaaagta
11581 acccttaaaa aaaatgtgta agaactcaaa ttgcatccca aaaaaataat tgaatacaga
11641 actctcataa aaataattct tgaatactca gacaaaaaaa tataagaaaa tccccctcca
11701 ctatattcaa cattaaatcc agagactaat tcagattcac cttcagcaaa atcaaatgga
11761 gaacgattag tttcagctat acaagaggaa ataaaacata tagataaagg aaaacaaata
11821 aataaaaatc acaaatttaa ctgataataa taaaaatcaa ttaatctata tctattaatc
11881 aaaaaaataa atgataaaaa aattatagat aatcttacct cataagaaat agtttgagct
11941 actgaacgaa gtctccccaa taatgaataa gtagaacagg aagcccaacc agaaatcata
12001 atagtataaa cacccattct tatacaacaa agcaaaaaca aaataccata aataaatgaa
12061 aaacaattaa ttaaaaaagg ataaaccaat cataaaaata aagaaataaa caatcttaaa
12121 ataggggaaa aataataaat taaaaaatta gattttaaag gaaccatttg ttctttagta
12181 aataatttaa tagcatcaga gaaaggttga aaaattccta aaaatccaag tttattaggc
12241 cctttacgaa cttggatata accaagaatt ttacgttcaa ccaaagttaa gaatccaact
12301 cccaacataa taaaaataac taaaattaat acattaataa aatatatgta ctgaatatag
12361 taaaacctat ataatcaaat tctaaattta atgcactatt tctgccaatt cagttaatta
12421 taattaaagt aaaattaata aattattgaa ttggtccttt cgtactgata tcaaatataa
12481 aatctaagga tagaaaccaa cctggctcac gccggtctga actcagatca tgtaagaatt
12541 taaaggtcga acagacctag aaataaaaaa actgcccctt attctaatct taatccaaca
12601 tcgaggtcgc aaactatttt atcgatttga actctccaaa ataattacgc tgttatccct
12661 aaggtaactt attcttttaa tcaaaaaaaa aataaaggat caaataaaca tttatttatg
12721 aataaaaaaa aataagaaaa aataattaat ccatcacccc aattaaaaaa tcactacata
12781 tcaaaaaata tataaaacaa atttgaaata aataaaaata aagctctata gggtcttatc
12841 gtcccttaaa aaaatttaag ctttttcact taaaaattaa gttttaataa aaaaatgaag
12901 aaagttgttt tctcatccaa tcattcattc cagactccaa ttaaaagact aattattatg
12961 ctacctttgc acagtcaaat tactgcggcc attaaaaaat cattgggcag actaaatctt
13021 taaaataaaa tctaaaaacc atgtttttga taaacaggtg gaaaaaagaa ttgccgaatt
13081 cctacaaaaa tattgataat tatactaatt taatcattat taattaaaaa aataattaac
13141 tttaaaatct taataaaaaa aaaataaaac caaataataa attataaaaa taaacaatta
13201 cgtaatcaaa tgaatgctgt ttctaaccct ccaaaaatta ttataataaa aaattataaa
13261 tattaaagca agattatccc tcctttttta atatcaaaaa aattattaaa ataaaatttt
13321 taaaaaactt aataatccaa aattaaattt aaatcttaag aaaccagata ccattaaggg
13381 cgaataatat ttcactacta aaaaaaaatt attaataatt atgaaacatt aattaaaata
13441 aaaatttaaa taatttaaat tcgggaaatt aaaatatatt ataattaatt aattaaccct
13501 gatacaaaag gtacaataaa aaaatactta taatcaaaat aaaaatatcc tatacagtac
13561 atataaacta tagaaataag aattttttca aaagttatac ttaaataaaa aaaatacttt
13621 atttaaaaaa taaataataa aaaaaaatca aaatattata ttcaaattaa attgaattgc
```

```
13681 acaacttatt tattagtgta aataataagc ttaaactata taaagcttta atttgaactt
13741 tcagcttcac ctccggtaaa actactttgt tacgacttat cccaagttat tgggagtgac
13801 gggcgatgtg tgcataatta agagctaaaa tcaaaataac tttactagtt aaaattactg
13861 ttaaatccaa aatctaatag ttaattacaa actataacaa attcaaaaat tattgtaacc
13921 catttctact ttaatacaaa ctgcaccttg acctaacatt tttctcaaac agaaaaaaga
13981 aaataatcta atgatacatt caatgacaga gatatacaag taacattaaa gtaaaatcaa
14041 tcgtggatta tcaattatag aacaggttcc tctaaaaaga attaattacc gccaaattct
14101 ttgaatttca agatcatagc taataataat caacctaaaa tatttatatt tcaaaataat
14161 agggtatcta atcctagttt aattaagaat ttctcaaaat aaaaaagaaa taaaaataaa
14221 taaaaatttc acttaattaa tttatttttt aatttctata aaaacaattt ttgaaagata
14281 atatatttga atattaataa atgtataacc gcgattgctg gcacaaaatt aatcataaaa
14341 attatcaatt aacaaaatca tactaaaata aattaataga attaattact gcgaattatg
14401 ttatattctt ttagaacaat ataaacccca aaaaatttta catgtaaaat aatgataaat
14461 tcaaatttaa agcaagaatc aaactttaat ttaaaaaaag attagttatg gaacataaat
14521 ttagccaact aaacctaaat tattgggaat ttaagtttac ataaatttag ccaactaaac
14581 ctaaattatt gggaatttaa gtttacataa atttagccaa ctaaacctaa attattggga
14641 atttaagttt acataaattt agccaactaa acctaaatta ttgggaattt aagtttacat
14701 aaatttagcc aactaaacct aaattattgg gaatttaagt ttacataaat ttagccaact
14761 aaacctaaat tattgggaat ttcagtttac ataaatttag ccaactaaac ctaaattatt
14821 gggaatttaa gtttacataa atttagccaa ctaaacctaa attattggga atttaagttt
14881 acataaattt agccaactaa acctaaatta ttgggaattt aagtttacat aaatttagcc
14941 aactaaacct aaattattgg gaatttaagt ttacataaat ttagccaact aaacctaaat
15001 tattgggaat ttaagtttac ataaatttag ccaactaaac ctaaattatt gggaatttaa
15061 gtttacataa atttagccaa ctaaacctaa attattggga atttaagttt acataaattt
15121 agccaactaa acctaaatta ttgggaattt aagtttacat aaatttagcc aactaaacct
15181 aaattattgg gaatttaagt ttacataaat ttagccaact aaacctaaat tattgggaat
15241 ttaagtttac ataaatttag ccaactaaac ctaaattatt gggaatttaa gtttacataa
15301 atttaagcca ctaaacccta aattattggg aatttaagtt tacataaatt taagccaact
15361 aaacctaaat tattgggaat ttaagtttac ataaatttaa aaaatttaat tttttttttc
15421 atttataaga taaaaatata attatggttt ccaaattcac tttaatggtt atttccgggt
15481 tttcttcttt ttcttaaata aaagaagaaa acaattaatt attgattatt tttgggaatt
15541 tctcacttaa agttaggaat taattatttg attattagat aatatttatt atggtttcaa
15601 attcacttta atggttattt ccgtgttttc ttctttttct taaataaaag aagaaaacaa
15661 ttaattattg attatttttg ggaatttctc acttaaagtt aggaattaat tatttgatta
15721 ttagataata tttattatgg tttcaaattc actttaatgg ttatttccgg gttttcttct
15781 ttttcttaaa taaaagaaga aaacaattaa ttattgatta tttttgggaa tttctcactt
15841 aaagttagga attaattatt tgattattag ataatattta ttatggtttc aaattcactt
15901 taatggttat ttccgggttt tcttcttttt cttaaataaa agaagaaaac aattaattat
15961 tgattatttt tgggaatttc tcacttaaag ttaggaatta attatttgat tattagataa
16021 tatttattat ggtttcaaat tcactttaat ggttatttct gtgttttctt ctttttctta
16081 aataaaagaa gaaaacaatt aattattgat tatttttggg aatttctcac ttaaagttag
16141 gaattaatta tttgattatt atcaattgat attactaata gatattaatt atacttttgt
16201 tataaaataa tttaatatat ttttgatctg tgggtagttt attaaattat tttttattaa
16261 gaataaattt taattcaatc atttaattaa gttccatcta tttttttttc taaaaataag
16321 atat
//
```

Appendix B. GenBank Flatfile for *Pteronarcys princeps.*

```
LOCUS       AY687866               16004 bp    DNA     circular INV 11-MAY-2005
DEFINITION  Pteronarcys princeps mitochondrion, complete genome.
ACCESSION   AY687866
VERSION     AY687866.1  GI:50402839
KEYWORDS    .
SOURCE      mitochondrion Pteronarcys princeps (ebony salmonfly)
  ORGANISM  Pteronarcys princeps
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Plecoptera; Perloidea; Pteronarcyidae; Pteronarcys.
REFERENCE   1  (bases 1 to 16004)
  AUTHORS   Stewart,J.B. and Beckenbach,A.T.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2004) Biological Sciences, Simon Fraser
            University, 8888 University Drive, Burnaby, British Columbia V5A
            1S6, Canada
FEATURES             Location/Qualifiers
     source          1..16004
                     /organism="Pteronarcys princeps"
                     /organelle="mitochondrion"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:285953"
                     /tissue_type="leg"
                     /dev_stage="adult"
     tRNA            1..67
                     /product="tRNA-Ile"
                     /note="codons recognized: AUY"
                     /anticodon=(pos:30..32,aa:Ile)
     tRNA            complement(65..133)
                     /product="tRNA-Gln"
                     /note="codons recognized: CAR"
                     /anticodon=(pos:complement(101..103),aa:Gln)
     tRNA            142..210
                     /product="tRNA-Met"
                     /note="codons recognized: AUR"
                     /anticodon=(pos:172..174,aa:Met)
     gene            211..1245
                     /gene="nad2"
     CDS             211..1245
                     /gene="nad2"
                     /codon_start=1
                     /transl_table=5
                     /product="NADH dehydrogenase subunit 2"
                     /protein_id="AAT76648.1"
                     /db_xref="GI:50402840"
                     /translation="MLNNPTKFLFLMTLMSGTLISVSANSWFGAWVGLEINLLSFIPL
                     MTNSNNLLSTEASLKYFLTQALASATLLFTIILAALIFSSPSSLLISNPFLITLVNSS
                     LLLKMGAAPFHFWFPGVMEGLNWMNGLMLMTWQKLAPLMLLSYNLKIDIFIALAIILS
                     VIIGSLGGLNQTSLRKILAYSSINHLGWLVAALTLGENLWGLYFMIYSFLTATIIFML
                     HSFKLFYFNQIFSLNFTPPVIKFSLFMTLLSLGGLPPFIGFLPKWIIIQSMVEANLTL
                     LITIMVVMTLITLFYYLRMGFGAFMLTYSESLWNYSYLQNNFTHMTALTLTMCSTLGL
                     VICATLYHIF"
     tRNA            1244..1310
                     /product="tRNA-Trp"
                     /note="codons recognized: UGR"
                     /anticodon=(pos:1274..1276,aa:Trp)
     tRNA            complement(1303..1367)
                     /product="tRNA-Cys"
                     /note="codons recognized: UGY"
                     /anticodon=(pos:complement(1336..1338),aa:Cys)
     tRNA            complement(1376..1443)
                     /product="tRNA-Tyr"
                     /note="codons recognized: UAY"
                     /anticodon=(pos:complement(1409..1411),aa:Tyr)
     gene            <1444..2975
                     /gene="cox1"
```

```
CDS             <1444..2975
                /gene="cox1"
                /note="start codon not determined; TAA stop codon
                completed by the addition of 3' A residues to the mRNA"
                /codon_start=2
                /transl_except=(pos:2975,aa:TERM)
                /transl_table=5
                /product="cytochrome c oxidase subunit I"
                /protein_id="AAT76649.1"
                /db_xref="GI:50402841"
                /translation="RQWLFSTNHKDIGTLYFIFGAWSGMVGTSLSLLIRAELGQPGSL
                IGDDQIYNVIVTAHAFVMIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFWLL
                PPSLTLLLASSLVENGAGTGWTVYPPLSAGIAHAGASVDLAIFSLHLAGVSSILGAVN
                FITTVINMRSNGMTLDRMPLFVWAVAITALLLLLSLPVLAGAITMLLTDRNLNTSFFD
                PAGGGDPILYQHLFWFFGHPEVYILILPGFGLISHIVSQESGKKEAFGALGMIYAMLS
                IGLLGFVVWAHHMFTVGMDVDTRAYFSSATMIIAVPPGIKIFSWLATLHGTQLNYSPA
                LLWAMGFIFLFTIGGLTGVVLANSSLDIILHDTYYVVAHFHYVLSMGAVFAIMAGFIQ
                WYPLFTGLTMNPHWLKIQFLTMFIGVNMTFFPQHFLGLAGMPRRYSDYPDAYTTWNVV
                SSIGSMISLIGVIMLVFIVWESMAANRIALFPLQMSTSIEWYQNLPPAEHSYAELPML
                TG"
tRNA            2976..3041
                /product="tRNA-Leu"
                /note="codons recognized: UUR"
                /anticodon=(pos:3005..3007,aa:Leu)
gene            3046..3733
                /gene="cox2"
CDS             3046..3733
                /gene="cox2"
                /note="TAA stop codon completed by the addition of 3' A
                residues to the mRNA"
                /codon_start=1
                /transl_except=(pos:3733,aa:TERM)
                /transl_table=5
                /product="cytochrome c oxidase subunit II"
                /protein_id="AAT76650.1"
                /db_xref="GI:50402842"
                /translation="MATWANLGLQDSASPLMEQLTYFHDHALLILIIITVLVSYVMAT
                LFYNSNLDRFLLDGQTIETVWTILPAITLLFIALPSLRLLYLLDEVAYPTLTLKTIGH
                QWYWSYEYSDFLNIEFDSYMTPYNEMGLEGFRLLDVDNRAVLPMNTEIRILVTAADVL
                HSWTVPALGVKVDATPGRLNQTHFTLLRPGLFFGQCSEICGANHSFMPIVIESVPTNI
                FINWVSSLSEA"
tRNA            3734..3804
                /product="tRNA-Lys"
                /anticodon=(pos:3764..3766,aa:Lys)
tRNA            3804..3870
                /product="tRNA-Asp"
                /note="codons recognized: GAY"
                /anticodon=(pos:3832..3834,aa:Asp)
gene            3871..4029
                /gene="atp8"
CDS             3871..4029
                /gene="atp8"
                /codon_start=1
                /transl_table=5
                /product="ATP synthase FO subunit 8"
                /protein_id="AAT76651.1"
                /db_xref="GI:50402843"
                /translation="MPQMAPISWLTLFIAFSLILLIFNCMNYYSFLPNTPETRAKSIS
                QTPMNWKW"
gene            4023..4700
                /gene="atp6"
CDS             4023..4700
                /gene="atp6"
                /codon_start=1
                /transl_table=5
                /product="ATP synthase FO subunit 6"
                /protein_id="AAT76652.1"
                /db_xref="GI:50402844"
                /translation="MMTNLFSVFDPSTSVLNLSLNWISTILGLTLIPSIYWFMPSRYT
                LIWNKILLTLHNEFKTLLGPTGHAGSTFMFISLFSLILFNNFLGLFPYVFTSSSHLTL
                TLALALPLWLSFMIYGWINHTQHMFAHLVPQGTPAVLMPFMVCIETISNVIRPGTLAV
```

```
                     RLAANMIAGHLLLTLLGNTGPSLTYSILSLLIVAQIALLVLESAVAIIQSYVFAVLST
                     LYSSEVN"
     gene            4700..5488
                     /gene="cox3"
     CDS             4700..5488
                     /gene="cox3"
                     /codon_start=1
                     /transl_table=5
                     /product="cytochrome c oxidase subunit III"
                     /protein_id="AAT76653.1"
                     /db_xref="GI:50402845"
                     /translation="MSTHANHPYHLVDQSPWPLTGALGALVTVSGLLKWFHHYNSSLL
                     ALGLLITTLTMIQWWRDITREGTFQGLHTYPVTLGLRWGMILFIVSEVFFFLSFFWAF
                     FHSSLAPTTELGVMWPPVGIIPFNPLQIPLLNTAILLASGVTVTWAHHGLMEGNHSQS
                     LQGLFFTVILGIYFTILQAYEYIEAPFAISDAVYGSTFFVATGFHGLHVLIGTTFLLI
                     CLIRHASFHFSVNHHFGFEAAAWYWHFVDVVWLFLYISIYWWGS"
     tRNA            5488..5553
                     /product="tRNA-Gly"
                     /note="codons recognized: GGN"
                     /anticodon=(pos:5518..5520,aa:Gly)
     gene            5554..5907
                     /gene="nad3"
     CDS             5554..5907
                     /gene="nad3"
                     /codon_start=1
                     /transl_table=5
                     /product="NADH dehydrogenase subunit 3"
                     /protein_id="AAT76654.1"
                     /db_xref="GI:50402846"
                     /translation="MLSISAIATIIITLSIIVMSLASILSKKSIIDREKSSPFECGFD
                     PKSSSRLPFSLRFFLIAVIFLIFDVEIALLLPMIIILPSSSLLIWLTVSFFFLAILLI
                     GLYHEWNQGALEWAE"
     tRNA            5906..5969
                     /product="tRNA-Ala"
                     /note="codons recognized: GCN"
                     /anticodon=(pos:5935..5937,aa:Ala)
     tRNA            5970..6035
                     /product="tRNA-Arg"
                     /note="codons recognized: CGN"
                     /anticodon=(pos:5999..6001,aa:Arg)
     tRNA            6037..6101
                     /product="tRNA-Asn"
                     /note="codons recognized: AAY"
                     /anticodon=(pos:6067..6069,aa:Asn)
     tRNA            6104..6170
                     /product="tRNA-Ser"
                     /note="codons recognized: AGN"
                     /anticodon=(pos:6129..6131,aa:Ser)
     tRNA            6171..6237
                     /product="tRNA-Glu"
                     /note="codons recognized: GAR"
                     /anticodon=(pos:6202..6204,aa:Glu)
     tRNA            complement(6243..6309)
                     /product="tRNA-Phe"
                     /note="codons recognized: UUY"
                     /anticodon=(pos:complement(6275..6277),aa:Phe)
     gene            complement(6309..8044)
                     /gene="nad5"
     CDS             complement(6309..8044)
                     /gene="nad5"
                     /note="TAA stop codon completed by the addition of 3' A
                     residues to the mRNA"
                     /codon_start=1
                     /transl_except=(pos:complement(6309..6310),aa:TERM)
                     /transl_table=5
                     /product="NADH dehydrogenase subunit 5"
                     /protein_id="AAT76655.1"
                     /db_xref="GI:50402847"
                     /translation="MLWSLSLCLISFIFLFIISLFLVITGFYFVMQDLVYFIEWEVLS
                     MNSGSIVMTFLFDWMSLIFMGFVLFISSLVIYYSQEYMDGDLNINRFIILVLMFVLSM
                     MFLIISPNLISILLGWDGLGLVSYLLVIYYQNVKSYNAGMLTALSNRIGDVALLLAIA
```

```
                     WMLNFGSWNYIYYLECSKSSTEMLIIGSLVMLAAMTKSAQIPFSSWLPAAMAAPTPVS
                     ALVHSSTLVTAGVYLLIRFNVLLANSSLGSFLLLFAGLTMFMAGLGANFEFDLKKIIA
                     LSTLSQLGLMMSILAMGFPKLAFFHLLTHALFKALLFMCAGAIIHNMKDSQDIRFMGG
                     LVVQMPLTSSCFNLSNLALCGMPFLAGFYSKDLILEIVSLSYLNVFSFILYFFSTGLT
                     VCYSLRLVYYSMSGDFNVNSLHPLNDEGWTMLKGMMTLSLMAVLGGSLLSWTLFPTPS
                     MICLPLSLKTLALSVSVIGGWLGYELAKFSLSYELQTLRFHFMTSFMGSMWFLPFLST
                     YGVSYTPLQLGSYVFKSFDQGWSEYCGAQGLYKTFTSWSTINQGWQNNDLKTYLISFM
                     LWIIILLVLVAL"
     tRNA            complement(8045..8112)
                     /product="tRNA-His"
                     /note="codons recognized: CAY"
                     /anticodon=(pos:complement(8080..8082),aa:His)
     gene            complement(8119..9459)
                     /gene="nad4"
     CDS             complement(8119..9459)
                     /gene="nad4"
                     /codon_start=1
                     /transl_table=5
                     /product="NADH dehydrogenase subunit 4"
                     /protein_id="AAT76656.1"
                     /db_xref="GI:50402848"
                     /translation="MLKLLVSLVFLIPLCFMKNMWWTVQSILFLLTFMFMLSNVFSTF
                     FGNLSYFMGCDILSYGLILLSFWICVLMMTASESVYRTRNYEGFFLFMIIMLLILLYC
                     TFSTMNLFMFYLFFESSLIPTLFLILGWGYQPERLQAGVYLLFYTLLASLPLLVGIFY
                     LYSLNSTLYFPLMNKISISHIICYLGLILAFLVKMPMFLVHLWLPKAHVEAPVSGSMI
                     LAGVLLKLGGYGLLRIFKILLIAGLSFNFVWVGISLVGGVLVSLICLRQTDLKALIAY
                     SSVAHMGIVLGGLMTLSYWGFCGSFTLMIAHGLCSSELFCLANISYERLGSRSLLINK
                     GLMNFMPSMTLWWFLLSSCNMAAPPSLNLLSEISLLNSLVGWTWSTMLMLSFLSFFSA
                     AYTLYLYSYSQHGKIYSGVYSCCMGYSREYLLLFLHWVPLNLLILKSEPCMLWL"
     gene            complement(9453..9749)
                     /gene="nad4L"
     CDS             complement(9453..9749)
                     /gene="nad4L"
                     /codon_start=1
                     /transl_table=5
                     /product="NADH dehydrogenase subunit 4L"
                     /protein_id="AAT76657.1"
                     /db_xref="GI:50402849"
                     /translation="MLTSFYWAFPLFLFMCGGWVFTSKRKHLLLTLLSLEFIVLSLFL
                     ILFIYLNLLNYELFFSMVFLTFSVCEGALGLSILVSMIRTHGNDYFQSFSVLQC"
     tRNA            9752..9818
                     /product="tRNA-Thr"
                     /note="codons recognized: ACN"
                     /anticodon=(pos:complement(9782..9784),aa:Thr)
     tRNA            complement(9820..9885)
                     /product="tRNA-Pro"
                     /note="codons recognized: CCN"
                     /anticodon=(pos:complement(9853..9855),aa:Pro)
     gene            9887..10411
                     /gene="nad6"
     CDS             9887..10411
                     /gene="nad6"
                     /codon_start=1
                     /transl_table=5
                     /product="NADH dehydrogenase subunit 6"
                     /protein_id="AAT76658.1"
                     /db_xref="GI:50402850"
                     /translation="MSHLIIMTASSILSIIFTQMNHPLAMGLMLLLQTLVICLMTGFM
                     TQSFWFSYILFLVFLGGLLVLFIYVTSLASNEMFSLSMSTLILTSGPFMFMLMLMMFL
                     DPLPWLTSIVSNDMMTISNLMIYQEESITSLMKLYNQPTSLITLMLVVYLFLTLIAVV
                     KITSIFYGPLRQKN"
     gene            10411..11547
                     /gene="cytb"
                     /note="synonym: cob"
     CDS             10411..11547
                     /gene="cytb"
                     /codon_start=1
                     /transl_table=5
                     /product="cytochrome b"
                     /protein_id="AAT76659.1"
                     /db_xref="GI:50402851"
```

```
                     /translation="MNKPLRTSHPLFKIANNALVDLPAPVNISAWWNFGSLLGLCLVV
                     QIATGLFLAMHYTAHIDLALSSVAHICRDVNYGWLLRTLHANGASFFFICLYLHTGRG
                     MYYGSYMFMHTWMIGVIILFLVMGTAFMGYVLPWGQMSFWGATVITNLLSAVPYLGID
                     MVQWVWGGFAVDNATLTRFFTFHFVLPFIVAAAVLIHLLFLHQTGSNNPLGLNSDVDK
                     IPFHPYFTFKDVVGFLVLIMILTLLTLLEPYMLGDPDNFIPANPLVTPVHIQPEWYFL
                     FAYAILRSIPNKLGGVIALILSIAILLILPFSNKSNFRGMQFYPINQIMFWFLLVIVI
                     LLTWIGARPVEDPYILVGQILTVLYFLYYILNPLMFKLWDNLLK"
     tRNA            11546..11615
                     /product="tRNA-Ser"
                     /note="codons recognized: UCN"
                     /anticodon=(pos:11577..11579,aa:Ser)
     gene            complement(11637..12587)
                     /gene="nad1"
     CDS             complement(11637..12587)
                     /gene="nad1"
                     /codon_start=1
                     /transl_table=5
                     /product="NADH dehydrogenase subunit 1"
                     /protein_id="AAT76660.1"
                     /db_xref="GI:50402852"
                     /translation="MFLYDMILPLVGALILIICVLVGVAFLTLLERKVLGYIQIRKGP
                     NKVGFAGIPQPFCDAIKLFTKEQTYPVTSNYLPYYFSPIFSLFLALLVWMIIPYLSGL
                     FTFNLGLLFFLSCTSLGVYTVMIAGWSSNSNYALLGGLRGVAQTISYEVSLALILLSF
                     VFLIGGYCLMDFLMYQKYLWFVVMTFPLALSWFASCLAETNRTPFDFAEGESELVSGF
                     NVEYSSGGFALIFLAEYASILFMSMLFVVIFLGCNIYSWEFFVKLAFISFMFIWVRGT
                     LPRFRYDKLMYLAWKSFLPLSLNYLFMFVGLKLLIMSILF"
     tRNA            complement(12589..12654)
                     /product="tRNA-Leu"
                     /note="codons recognized: CUN"
                     /anticodon=(pos:complement(12623..12625),aa:Leu)
     rRNA            complement(12655..13982)
                     /product="large subunit ribosomal RNA"
     tRNA            complement(13983..14053)
                     /product="tRNA-Val"
                     /note="codons recognized: GUN"
                     /anticodon=(pos:complement(14018..14020),aa:Val)
     rRNA            complement(14054..14846)
                     /product="small subunit ribosomal RNA"
     misc_feature    14847..16004
                     /note="A+T rich region; putative control region"
ORIGIN
        1 aatgaagtgc ctgataaaag gattatcgtg ataggataaa ttatgtaaat tatatttacc
       61 ttcattatgt ccaatagact tgaactattt ccgaaagtat caaaaacttt cgtgcaccat
      121 acactagaac atataatata aagaaagata agctaataaa gctaatgggt tcatacccca
      181 tttatagagg ttttactcct cttctttcta gtgctaaata atccaaccaa attcttattt
      241 ttgataacat taataagtgg tactttaatc tctgtttcag ccaattcatg atttggagca
      301 tgagtagggt tagaaattaa tcttttatct tttatcccct taataactaa ttctaataat
      361 ttattatcta ctgaagcttc tttaaaatac tttttaactc aagccctagc ttctgctaca
      421 ctattattta caattattct tgctgctcta atttttagat cccctagatc attattaatt
      481 agaaatcctt tcttaattac cttagttaat tcatctttat tattaaaaat gggagctgct
      541 cccttccatt tttgattccc aggtgtaata gagggattaa actgaataaa tggtcttata
      601 ctaataacct gacaaaaact tgctccatta atattacttt cttataatct aaaaattgac
      661 atctttattg ctttagcaat tatcttatca gtaattattg gatctttagg aggcctcaat
      721 caaacttcat tacgtaaaat tcttgcttat tcctctatta atcatttagg ttgattagta
      781 gcagcattaa ccttagggga aaatttatga ggactttatt ttataattta ttcttttta
      841 actgctacaa ttattttcat acttcattca tttaaattat tctattttaa tcaaatcttt
      901 tcattaaatt ttactccccc tgttattaaa ttttcattat ttataacttt actatccctt
      961 ggtggactcc ctccatttat tggatttctt ccaaagtgaa tcattattca atctatagtt
     1021 gaagctaatc taaccctact aattactatt atagtagtaa taacactaat tactctattt
     1081 tactatctac gaataggatt tggagccttt atattaactt attctgaatc tttatgaaat
     1141 tattcctatt tacaaaataa tttcactcac ataactgccc taaccttaac aatatgttca
     1201 accttagggt tagttatttg tgcaactcta taccacattt tctaaggctt taagttaatt
     1261 aaactaatag ccttcaaagc tgtaaataaa gtatctatct ttaagcctta gtaagatcac
     1321 ttacaccttt agaattgcag tctaatatca tcattgacta taaagctatt tactatggta
     1381 gaagaggggt atacctccta aatagattta caatctatca ccttatatct cagccattct
     1441 accgcgacaa tgattattct caacaaatca caaggatatt ggaactttat acttcatttt
     1501 tggagcatga tcaggaatag ttggaacatc cctaagtctt ttaattcgag ctgaattagg
     1561 acaaccgggt tccttaattg gagatgatca aatttataat gtaattgtta ctgctcacgc
     1621 ttttgtaatg attttttta tagtaatacc tattataatt ggaggattcg gaaattgatt
     1681 agtaccttta atactaggtg ctcccgatat agctttcccc cgaataaata atataagttt
     1741 ttgattactt cctccttctc taactttact gttagcaagt agactagtgg aaaatggggc
```

```
1801 tggtacaggt tgaacagttt accctccact atctgcagga attgctcatg ctggagcatc
1861 tgtcgactta gctattttct cattacattt agctggagtc tcatccatct taggtgctgt
1921 aaattttatt accactgtaa tcaacatacg atctaacgga ataacattag atcgtatgcc
1981 attatttgta tgagctgtag caattacagc tctactttta ctttatctt taccagtttt
2041 agcaggggct atcacaatac ttctaacaga tcgaaatctt aatacatcat tttttgatcc
2101 agccgggggga ggtgatccta ttctttatca acatcttttt tgattttttg gtcaccctga
2161 agtatacatt ttaattttac caggatttgg gcttatttca catattgtta gacaagaaag
2221 aggtaaaaag gaagcctttg gagcattagg aataatttat gctatacttt ctattggttt
2281 attaggattt gttgtatgag ctcatcacat atttactgta ggcatagatg ttgatacacg
2341 agcttacttt tcttctgcca ctataattat tgctgtccct cctggaatta aaatttttag
2401 ttgactagct acattacatg gaactcaatt aaactatagc ccagccctct tatgagctat
2461 aggatttatt ttcctattta ctattggagg attaacaggt gtagttctag ctaattcttc
2521 tttagatatc attttacatg atacttatta tgtagtagct cattttcatt atgtgttatc
2581 cataggagca gtatttgcta ttatggctgg atttattcaa tgatacccat tatttacagg
2641 ccttactata aaccctcact gattaaagat ccaattcctt acaatattta ttggagttaa
2701 tataacattc ttccctcaac atttcttagg attagcaggt atgcctcgac gttactctga
2761 ttacccagac gcctatacta catgaaatgt agtttcatca attggttcca taatctcttt
2821 aattggtgta attatattag ttttattgt ttgagaaaga atagccgcta atcgaattgc
2881 tttatttcct ttacaaataa gtacgtcaat tgaatgatac caaaatctgc cccctgcaga
2941 acatagttat gctgaattac ctatactaac tggattctaa tatggcagat aagtgcaatg
3001 gatttaagct ccatatataa aggctattcc ttttgttaga aattaatggc aacatgagca
3061 aatttaggtc ttcaagatag tgcttcccct ttaatagaac aattaacata ttttcatgat
3121 catgcacttt taattttaat tattattaca gttctagtaa gttatgtaat agcaacttta
3181 ttttataatt ctaatcttga tcggttttta ctagatggtc aaactattga aacagtatga
3241 actattctac cagctattac tcttcttttt attgctttac cttctttacg tcttctttat
3301 ttactagatg aagttgctta cccaaccctt actttaaaaa caattggtca tcaatgatac
3361 tgaagttatg aatattccga tttcctaaat attgaatttg actcatatat aacaccttat
3421 aacgaaatag gattagaagg tttccgttta ttagatgttg ataatcgagc cgtccttcct
3481 ataaatactg aaattcgtat tcttgttaca gctgcagacg tattacattc ttgaacagta
3541 ccagccctcg gagtaaaggt ggatgctaca ccaggacgac taaatcaaac tcattttact
3601 ctccttcgtc ctggcttatt ctttggtcaa tgttcagaaa tctgtggagc taatcataga
3661 tttataccta ttgtaattga aagtgttcca actaatattt tcattaattg agtctcttcc
3721 ttaagtgaag cttcatcaga tgactgaaag caagtattgg tctcttaaac cattttatag
3781 taatctcgca attacttctg atgaaagaat tagttaaatt aacattagta tgtcatgcta
3841 aaattactag tactagccta gtattcttta attccacaaa tagccccaat tagttgatta
3901 actttattta ttgctttttc attaatttta ttaattttta attgtataaa ttattattca
3961 tttttaccca atacacctga aactcgagcc aagtctatta gacaaactcc cataaactga
4021 aaatgataac aaaccttttt tccgttttg atccttcgac aagcgtatta aatctttcct
4081 taaactgaat tagtacaatt ttaggactaa ctctgattcc ctcaatttac tgatttatgc
4141 catcacgata tactttaatt tgaaataaaa ttttattaac acttcataat gaatttaaaa
4201 cattattagg cccaactggt catgcaggaa gaacatttat attcatctct ttattttcat
4261 taattttatt caataatttc ctaggccttt tcccttatgt ttttactagt tcaagtcatt
4321 taactttaac cttagcctta gctttacctc tttgattaag atttataatt tacggttgaa
4381 ttaatcacac tcaacatata tttgctcatt tagttcctca aggaactcca gcagttttaa
4441 taccatttat agtatgtatt gaaactatta gaaatgtaat tcgcccagga acacttgctg
4501 tccgattggc tgcaaatata attgctggac acttattact aactttatta ggtaatactg
4561 gcccaagctt aacttactca attctatcat tactcattgt agctcagatt gctttattag
4621 tcctagaatc tgctgtggct atcattcaat cttacgtttt tgctgtttta agaactttat
4681 attctagcga agtaaattaa tgtcaacaca cgctaaccat ccttatcatc ttgtagatca
4741 aagaccttga cccttaactg gagctctcgg agctttagtt actgtttcag gacttcttaa
4801 atgattccac cattacaatt catcattatt agccctaggt ttattaatta ctactcttac
4861 tataatccaa tgatgacgtg atatcacacg agaaggaaca ttccaaggac ttcacaccta
4921 tccagtcact ttaggattac gatgaggaat aattttattt attgtatctg aagtattctt
4981 tttcttgtct ttttttgag cattcttcca cagaagtctt gcccctacaa ctgaattagg
5041 tgtaatatga cctccggtag gaattattcc atttaatcct ttacaaattc cacttctaaa
5101 tacagctatc cttttagcct ctggtgttac agtaacatga gctcatcatg gcctaataga
5161 aggtaatcat agtcaaagtc ttcaaggcct attttttaca gtaattttag gtatttactt
5221 tactatttta caagcctatg aatatattga agccccttt gctatttcag atgctgttta
5281 cggatcaaca ttttttgtag ctacaggctt ccatggatta catgtcctta ttggaacaac
5341 tttccttcta atttgcctaa ttcgacatgc ttccttccat ttttctgtaa atcatcattt
5401 tggatttgaa gcagctgctt gatattgaca ctttgttgat gtagtttgat tatttttata
5461 tatctcaatt tactgatgag gtagctaatt tgtctagtat aataagtata tttgacttcc
5521 aatcaaaagg tctgaatatc ttcaggacaa ataattttat ctatttcagc aattgcaaca
5581 attattatta cactttctat tattgttata tccttagcat caattctttc taaaaaatct
5641 attattgatc gagaaaaaag atccccttt gaatgtggat ttgaccctaa aagatcctca
5701 cgacttccgt tttctttacg attcttctta atcgctgtga ttttcttaat ttttgatgtt
5761 gaaattgctc ttctacttcc aataatcatt atcttaccat catcaagatt attaatttga
5821 ttaacagtaa gatttttctt tttagccatt ctattaattg gcctctatca tgaatgaaac
5881 caaggagctc tagaatgagc cgaataggac tgtagttaag tataacattt ggtttgcatc
5941 caaaaagtat tgagattcaa tctgtcttag aataagaagc gattttttgc actcagtttc
6001 gacctgaatc tagataactc agttatcctt attcttatta attgaagcca aatagaggca
```

```
 6061 tatcactgtt aatgatataa ttgaataact tattccaatt aaggaaatat gatgttcaag
 6121 aaaaagctgc taacttttct ctttagcggt taaattccgt ttatatttct atttatatag
 6181 tttaattcaa aacattacat tttcactgta aaaataaagg ctttactttt ataaatacaa
 6241 tctattcaaa gattacatta atctccctaa catcttcagt gtcatgctct atttataagc
 6301 tacttgaata taaagcaacc aagactagca aaataataat tcataatata aaactaatta
 6361 aatatgtctt taaatcatta ttttgtcacc cctgattaat agttgatcaa gaagtaaacg
 6421 ttttatataa tccttgagct ccacaatact cacttcatcc ttgatcaaaa gatttaaata
 6481 cataacttcc taattgtaaa ggagtataac tcaccccata agtcgataaa aaagggagaa
 6541 atcatattga acctataaaa cttgttataa aatgaaaccg taaagtttgt aattcataac
 6601 tcaaagaaaa ttttgccaac tcataaccta atcatccccc aataactcta acacttaaag
 6661 ctaaagtttt taaagataaa ggtaaacaaa tcattctagg tgtaggaaat aaagttcatc
 6721 tcaataaact accccctaaa acagctatta aagataaagt tattattccc ttcaatatag
 6781 ttcaaccctc atcatttaat ggatgtaaag aatttacatt aaaatcacct cttatagaat
 6841 aatacaccaa tcgtaatgaa taacataccg ttaacccagt agaaaaaaaa tataaaataa
 6901 agctaaaaac atttaaatat cttaatgaca caatttccaa aattaaatcc tttgagtaaa
 6961 atcccgccaa aaaaggtata ccacataaag ccaaattaga caaattaaaa caagaagatg
 7021 tcaaaggtat ttgaacgact aaccctccta taaatcgaat atcctgagaa tccttcatat
 7081 tatgaataat cgccccagca catataaata ataaagcctt aaataatgca tgagttaata
 7141 aatgaaaaaa tgctagcttt gggaaaccta tagctaaaat tctcattatt aatcctaatt
 7201 gtcttaatgt agataaagca ataatctttt ttaaatcaaa ttcaaaatta gcccctaagc
 7261 cggctataaa tatagtcaat cctgcaaata ataataaaaa agatcctaat gaactattag
 7321 ctaataaaac attaaatcga attaataaat agactccagc cgttactaaa gttgaagaat
 7381 gaactaaagc agacactggt gttggagctg ctatagcagc tggcaatcat gaagaaaaag
 7441 gaatttgagc tctcttagtt attgcagcca atataactaa tctaccaata attaatattt
 7501 cagttcttct ttttctacat tccaaataat agatataatt tcaactacca aaatttaata
 7561 ttcaagcaat agctaatagt aaagcaacat ctccaatacg attagaaaga gcagttaata
 7621 tcccagcatt atatgatttc acattctgat aataaattac taataaatat gatactaatc
 7681 ctaaaccatc tcaccctaat aaaattctaa tcaggttagg tctaataatc agaaacatta
 7741 ttgataaaac aaacattaat accaaaataa taaaacggtt aatatttaaa tctccatcca
 7801 tatattcctg actataataa atcactaatg aagaaataaa taatacaaat cctataaaaa
 7861 ttaaagatat tcaatcaaat aaaaaagtca taacaatact ccctgaattt atagataata
 7921 cctctcactc aataaaatat accaaatctt gtattacaaa ataaaatcct gtaataacta
 7981 aaaataaaga aataataaac aaaaaaataa atctaattaa acaaagagac aaggaccata
 8041 acacgaccta agatgagact tatcccatat ccttgacacc acaaatcaaa gtttttatta
 8101 aactacctaa gtgattattt ataaccataa tatacatggc tctctcttta aaatcaataa
 8161 atttaaagga actcaatgta aaaataataa taaatactct cgagaataac ctatacaaca
 8221 cgaataaact ccagaataaa tctttccatg ttgtctataa gagtacaaat ataaagtata
 8281 cgcagctcta aaaaatgata aaaatcttaa catcaatata gttgatcaag tccaaccaac
 8341 taaactattt aataaactaa tttctcttaa taaatttaaa gaaggaggtg ccgctatatt
 8401 acatgaactt aataaaaatc atcataatgt catacttggc ataaaattca ttaatccctt
 8461 attgattaac aaacttcgtc ttcctaaacg ttcataagaa atatttgcta aacaaaataa
 8521 ttccgaagaa caaagaccat gagcaatcat taacgtaaaa gatccacaaa atcctcaata
 8581 acttaaagtt attagacctc ctagaacaat ccccatatga gctacagaag aataagcaat
 8641 taaagcctta agatcagtct gacgtaaaca aattaaacta actaatactc cacctactaa
 8701 tctaataccc actcatacaa aattaaatga taaaccagca attaataaaa ttttaaaaat
 8761 acgcaataat ccatatcccc caagctttaa taacacacca gccaaaatta tcgacccaga
 8821 gacaggagct tcaacatgag ccttaggaag tcacaaatga actaaaaata ttggtatttt
 8881 tactaaaaat gctaaaatta atcctaaata acaaataata tgagaaattc taatcttatt
 8941 tattaaagga aaatataaag ttctattcaa tgaatataaa taaaaaatac ctactaataa
 9001 aggtaaagaa gctaataagg tataaaaaag taaatacacc cctgcttgca accgttcagg
 9061 ctgataccct caccccaaaa ttaaaaataa cgtcggaatc aagctgcttt caaagaacaa
 9121 ataaaatata aataaattta ttgtactaaa tgtacaataa agtaaaatta ataatataat
 9181 aattataaat aaaaaaaatc cctcataatt acgagtacga taaactgatt ctctagcagt
 9241 tattattaaa acacaaattc aaaaacttaa taaaattaga ccataagaaa gaatatcaca
 9301 tcctataaaa taacttaaat taccaaaaaa agttctaaat acattactta atataaatat
 9361 aaaagttaat agaaacaaaa ttgactgaac cgttcatcac atatttttca taaaacataa
 9421 ggggatcaaa aaaactaaag aaactaataa ttttaacatt gcaaaactct aaaagactga
 9481 aaataatcat ttccatgagt tcgaattatc gaaaccaaaa ttgataatcc caaagctcct
 9541 tcacaaactg aaaaagttaa aaatactatt ctaaaaaata attcataatt taataaattt
 9601 aaataaataa ataaaattaa aaataatctt aaaacaataa attctagact taataaagtt
 9661 aacaataaat gcttccgctt tgaagtaaat actcaacccc cgcacataaa taaaaataaa
 9721 ggaaatgctc aataaaaaga tgttaacatt agttttaata gtttaaagaa aacattggtc
 9781 ttgtaaatca aaaattaagg cctggacctt ttagaacttt cagagaaaag agaaaccccc
 9841 tatcattaat ccccaaaact aatattttaa ataaactatt ctctgcattt cccatttaat
 9901 cattataact gctagatcaa ttttaagaat tatcttcact caaataaatc atccgttagc
 9961 tataggtctt atattattac tccaaacact agtaatttgt ctaataacag gattcataac
10021 tcaaagattt tgattttctt acatcttatt tttagtcttc ctcggagggc tattagtttt
10081 atttatttat gtaacaagac tagcctcaaa tgaaatattt tctttatcaa tatcaaccct
10141 tattttaact tcaggcccct ttatatttat attaatacta atatatttt tagacccctt
10201 accttgatta acaagtatcg taagaaatga tataataaca atttcaaatt taataattta
10261 tcaagaagaa tcaattacct ctttaataaa actctataat caacccacta gacttattac
```

10321 tctaatatta gtggtttact tattcttaac tctaattgct gtagtaaaaa tcacaagaat
10381 tttttatgga cctttacgac aaaaaaacta atgaataaac ccttacgaac cagacatccc
10441 ttatttaaaa ttgctaacaa tgcattagta gacttacctg ctcctgtcaa catttctgca
10501 tgatgaaact ttggttcctt actaggatta tgcttagtag ttcaaatcgc tactggatta
10561 ttcttagcaa tacattatac agctcatatt gatttagctc tctcaagagt agctcatatt
10621 tgccgagatg taaattatgg ttgactttta cgaactcttc atgctaatgg agcttcattc
10681 tttttttatct gtctatatct acatactggc cgaggaatat actatggatc atatatattt
10741 atacacactt gaataattgg agtaatcatt ttatttcttg ttataggtac agcttttata
10801 ggatacgtat taccctgagg acaaatatca ttttgaggag ccacagttat tacaaattta
10861 ttatcagctg tcccctatct aggtattgac atagttcagt gagtatgagg gggatttgct
10921 gtagataatg ctaccttaac ccgattcttt acatttcatt ttgtcttacc ctttattgta
10981 gctgctgccg tattaattca cttactattc ttacaccaaa caggatcaaa taatccctta
11041 ggattaaata gtgacgtaga taaaatcccc tttcatccat attttacctt caaggatgtt
11101 gtaggatttt tagtattaat tataatttta acccttttaa ctcttctaga accttatata
11161 ttaggtgacc ctgataactt tatcccagct aatccccttg ttaccccagt tcatatccaa
11221 cctgaatgat acttcttatt tgcttacgct attttacgct ctattcccaa caaattagga
11281 ggagttattg cccttatttt atctattgct attttattaa ttctaccatt ttctaataaa
11341 agaaattttc gaggtataca attttatcca attaaccaaa ttatattttg atttttatta
11401 gttattgtta ttttattaac ctgaattgga gctcgtccag tcgaagaccc atatattctt
11461 gttggacaaa tcttaactgt attatacttc ctttactata tcctaaatcc cttaatgttc
11521 aaattatgag ataacttact taagtagtta ataagcttat tgtaaagcat atgttttgaa
11581 aacataagac agaagtttaa atcttctatt aacttaaatt tactaaaaaa aatacactaa
11641 aataagatag atataattaa aagttttaat ccaacaaata taaataagta gtttaaagat
11701 aaaggtaaaa aactctttca agctaaatat atcaatttat catatcgaaa acgaggaagt
11761 gtacctcgaa cccaaataaa tataaaagaa ataaaagcta acttgacaaa aaactcccaa
11821 gaataaatat tacatcccaa aaaaataaca acaaaaagca ttcttataaa taaaatgctt
11881 gcatattctg ctaaaaaaat taaagcaaat cccccctcttc tatactccac attaaaccca
11941 gatacaagct cagactcacc ttccgcaaaa tcaaaaggag tacgattagt ctcagccaaa
12001 catgaagcaa atcaacttaa agccaaagga aaagttataa ccacaaatca taaatacttt
12061 tgatacatca aaaaatctat taaacaatac cccccaatta aaaaaacaaa tgataacaaa
12121 atcaacgcca aactaacttc ataagaaata gtttgagcaa cccctcgtaa acctcctaat
12181 aatgcataat tagaattaga agaccatcct gcaattatta cagtatacac ccccaaactt
12241 gtacaagata aaaaaaataa taaacctaaa ttaaaagtaa ataaacctct taaataagga
12301 ataattattc aaaccaataa cgccaagaac aaactaaaaa taggtgaaaa ataataaggt
12361 aaataatttg aagtaacagg atacgtctgc tccttagtaa ataatttaat tgcatcacaa
12421 aaaggttgag gaataccagc aaatccaacc ttattaggac ccttacgaat ctgaatatat
12481 cccaaaacct tacgctctaa taaagttaaa aaagctactc caactaatac acaaataatt
12541 aaaattaatg cccctactaa aggtaaaatc atatcatata aaaacaagtt ctatctgtaa
12601 aatcccttta cataaatgga ttctaaaccc attgcactta tctgccaaaa tagatcactt
12661 aatattaatc aactaataag aaaatttatt cccaatgatt ggtcctttcg tactaaaaca
12721 tcaattttta ttaaggatag aaaccgacct ggcttacacc ggtctgaact cagatcatgt
12781 aaggatttaa aggtcgaaca gacctaaact ttgaacatct acacccaaaa ttacccttaa
12841 tccaacatcg aggtcgcaac ccttttttgtc gataagaact ctcgaaaaag attacgctgt
12901 tatccctaag gtaacttagt cttataatca ctaataatgg atcaacaatt catatatcaa
12961 tgtaataaaa taaaaagagt ttattatatc ttcctgtcac cccaacaaaa tacttattaa
13021 actattataa aaatacaacc aaataaataa taacttttaa atatcaagct ctatagggtc
13081 ttctcgtcct ttaaatatat ttaagccttt tgacttaaat gttaaattct aatttcaatt
13141 atactgagac agtcaatacc tcgtccaacc attcattcta gccttcaatt aaaagactaa
13201 tgattatgct acctttgcac ggtcaaaata ccgcggccct tcaaatctta tattcagtgg
13261 gcaggtcaga cctcaaatta tcatctagag gacatgtttt tgttaaacag gcgagcatta
13321 atttgcctaa ttccttaata taacctttca attacataat aaaaaacaat tatttatact
13381 aattttatca ttattaccaa ccttaaaaaa ttaaaagtta catttcaata tactatttaa
13441 attacatcat aaaattatat ttaataaaaa ataaattata acatccttaa attgatggct
13501 aattctaagc ctactaaatt tttacataaa attccatttt ataaattatg tattgaagct
13561 tatcccccaa tacattttta tcaaattata ttcacctaaa aaattagata aataatattt
13621 aataactgaa ttaaattcat ttcttgaaaa accagatatc ttaaagaacg aataacattt
13681 catgactaat catcttttat taatgattat tccacatcaa ctaacaaaaa taattagctc
13741 tctttaaatc gggatctata aatatttata aactttttga taaaccctga tacacaaggt
13801 acaataaatt aaatctactt tcacttaatt acattttcaa cttatttcaa taacctctca
13861 caatacaatt taactatcac attaaaaatt ttatctataa cctatacaaa aaccctacat
13921 aaaaaaaaat gattttatta ataaacatct ttaaacaaaa taattctgta aattcacaat
13981 tatcaaaaca aactgaattg cacagtgact tctcagtgta aatgaaatgc ttaactaatc
14041 aagctctgcc ttgaaacttt ccaggggcac cttccggtac acctactatg ttacgactta
14101 tctcgcctta attaacgaga gcgacgggcg atgtgtgcat gttttagagc caaaatcaat
14161 taaataaact taatataatt accatcaaat ccaccttcaa acaactatta caaatcatta
14221 tccataataa ataaatttat tgtaacccat caccacttaa ttattaactg caccttgacc
14281 tgacatcatt tctaattaaa attcaagaaa attttctccc tctaaggaca ttctgacgac
14341 ggcggtatac aagctgatta ctaaattaag taaggtctaa cgtgggctat cgattacggg
14401 acaggttcct ctgaaaggac taaaacaccg ccaaattctt tgagtttcaa gaatataact
14461 attactactc aagtatttat atttaacgtt taaaataata gggtataata atcctagttt
14521 aatctttaaa tttcacagct tcaacaacct ttaaagaagt tatttaaatt taaaatttca

```
14581 cctaataaat tcacttttac cttccttata tatcacttaa ctatacacta aaaatattca
14641 cttgcaccct tcgtataacc gcggcggctg gcacgacttt tgccggtact aacaaaaatt
14701 actaattcca agttaccttg attaataaaa ttaaatactg cgaatcaata acctttaaaa
14761 tttctttaag aaaaactata acgcacgcaa aaactcacat gtaaaataaa tttgaagtaa
14821 attattaagc aagaataaaa ctttaaggac ttagacaaaa aaaagacttt gccttaaatc
14881 attcaactca atagttaatt aagtgtatat tagaaagttg aaaaataaca aatagggggg
14941 cccaaaaaat ttttttaaac aaaatcaaaa gccaaatttt aaatagactt tcttacttac
15001 ctctccacct gattaaacag tcaaattccc tccctaaaaa aaatctaaat cttcatataa
15061 catgataatt aatcaataat caacaaacac ataccatttg cctaaagttt gcggtataaa
15121 tactatttaa caacttatta taataaatct gctccattta ttaaataaac tgattaaacg
15181 acaataaatt tacctaacca actttaaata ttaaaataaa aattatatta tattcaaatt
15241 aacttattaa ctattcatct aataaatctg ccccatttat taaataaata actaaatttg
15301 tttattttaa taaaaataac taaataacct actcaataaa accattaaat atataattta
15361 aataactgaa ctattaatag taaaattagt taaagtttat tggaaagact taatatttac
15421 tttttcaaat ggattttttta tcttaatgga tgatcctcat tttttctta tttataaaat
15481 gaggatcatc ccattaattt attatttaaa ttaataacct cttgttaata agaacttaat
15541 ttaatatata tatttaatta aataataatt atatttatat atttatttat ataatataaa
15601 atttaatttt taatataatt aataattaat catatataaa ttaaattctt attaatatat
15661 aaataattat ataataacaa ttaagtttat atatataatc tatatataaa taagaattta
15721 atattaaata tatctatttt ttgcctttaa ctataaacat aataggtata taaatatgat
15781 gtatataata taagctctta ttatacaccc aaatttttaa caaaaaatta caccttctga
15841 acaatttact ttttgttcat aaaattaagg gttttgggtt tttaaaaaaa gggtaaaaaa
15901 aatccaaaat cctcaatttt ttgaacaaaa atcataaaag aaaacattcc ctcttcccct
15961 tttaaagggc catattcaat tacatttaga taaaaaaata aaaa
//
```

References

Adams, K.L. and Palmer, J.D. (2003). Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Mol Phylogenet Evol* 29, 3: 380-95.

Adrain, C. and Martin, S.J. (2001). The mitochondrial apoptosome: a killer unleashed by the cytochrome seas. *Trends Biochem Sci* 26, 6: 390-7.

Allen, J.F. (1993). Control of gene expression by redox potential and the requirement for chloroplast and mitochondrial genomes. *J Theor Biol* 165, 4: 609-31.

Allen, J.F. (2003). The function of genomes in bioenergetic organelles. *Philos Trans R Soc Lond B Biol Sci* 358, 1429: 19-37; discussion 37-8.

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J., Staden, R. and Young, I.G. (1981). Sequence and organization of the human mitochondrial genome. *Nature* 290, 5806: 457-65.

Andersson, G.E. and Kurland, C.G. (1991). An extreme codon preference strategy: codon reassignment. *Mol Biol Evol* 8, 4: 530-44.

Andersson, S.G., Zomorodipour, A., Andersson, J.O., Sicheritz-Ponten, T., Alsmark, U.C., Podowski, R.M., Naslund, A.K., Eriksson, A.S., Winkler, H.H. and Kurland, C.G. (1998). The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* 396, 6707: 133-40.

Arndt, A. and Smith, M.J. (1998). Mitochondrial gene rearrangement in the sea cucumber genus *Cucumaria*. *Mol Biol Evol* 15, 8: 1009-16.

Bae, J., Kim, I., Sohn, H. and Jin, B. (2004). The mitochondrial genome of the firefly, *Pyrocoelia rufa*: complete DNA sequence, genome organization, and phylogenetic analysis with other insects. *Mol Phylogenet Evol* 32, 3: 978-985.

Ballard, J.W. (2000a). Comparative genomics of mitochondrial DNA in *Drosophila simulans*. *J Mol Evol* 51, 1: 64-75.

Ballard, J.W. (2000b). Comparative genomics of mitochondrial DNA in members of the *Drosophila melanogaster* subgroup. *J Mol Evol* 51, 1: 48-63.

Barrientos, A., Zambrano, A. and Tzagoloff, A. (2004). Mss51p and Cox14p jointly regulate mitochondrial Cox1p expression in Saccharomyces cerevisiae. *Embo J* 23: 3472-82.

Baumann, R.W., Arden, R.G. and Surdick, R.F. (1977). The Stoneflies (Plecoptera) of the Rocky Mountains. *Memoirs of the American Entomological Society* 31, 1-228.

Beagley, C.T., Okimoto, R. and Wolstenholme, D.R. (1998). The mitochondrial genome of the sea anemone *Metridium senile* (Cnidaria): introns, a paucity of tRNA genes, and a near-standard genetic code. *Genetics* 148, 3: 1091-108.

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

Beckenbach, A.T., Robson, S.K. and Crozier, R.H. (2005). Single nucleotide +1 frameshifts in an apparently functional mitochondrial cytochrome b gene in ants of the genus *Polyrhachis*. *J Mol Evol* 60, 2: 141-52.

Benkel, B.F., Duschesnay, P., Boer, P.H., Genest, Y. and Hickey, D.A. (1988). Mitochondrial large ribosomal RNA: an abundant polyadenylated sequence in *Drosophila*. *Nucleic Acids Res* 16, 20: 9880.

Bensasson, D., Feldman, M.W. and Petrov, D.A. (2003). Rates of DNA duplication and mitochondrial DNA insertion in the human genome. *J Mol Evol* 57, 3: 343-54.

Bensasson, D., Zhang, D., Hartl, D.L. and Hewitt, G.M. (2001). Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends Ecol Evol* 16, 6: 314-321.

Bensasson, D., Zhang, D.X. and Hewitt, G.M. (2000). Frequent assimilation of mitochondrial DNA by grasshopper nuclear genomes. *Mol Biol Evol* 17, 3: 406-15.

Berthier, F., Renaud, M., Alziari, S. and Durand, R. (1986). RNA mapping on *Drosophila* mitochondrial DNA: precursors and template strands. *Nucleic Acids Res* 14, 11: 4519-33.

Binder, S. and Brennicke, A. (2003). Gene expression in plant mitochondria: transcriptional and post-transcriptional control. *Philos Trans R Soc Lond B Biol Sci* 358, 1429: 181-8; discussion 188-9.

Blumenthal, T. (1998). Gene clusters and polycistronic transcription in eukaryotes. *Bioessays* 20, 6: 480-7.

Blumenthal, T. (2004). Operons in eukaryotes. *Brief Funct Genomic Proteomic* 3, 3: 199-211.

Boer, P.H. and Gray, M.W. (1988). Scrambled ribosomal RNA gene pieces in *Chlamydomonas reinhardtii* mitochondrial DNA. *Cell* 55, 3: 399-411.

Bogenhagen, D.F. and Clayton, D.A. (2003). The mitochondrial DNA replication bubble has not burst. *Trends Biochem Sci* 28, 7: 357-60.

Boore, J.L. (1999). Animal mitochondrial genomes. *Nucleic Acids Res* 27, 8: 1767-80.

Borner, G.V., Morl, M., Janke, A. and Paabo, S. (1996). RNA editing changes the identity of a mitochondrial tRNA in marsupials. *Embo J* 15, 21: 5949-57.

Boudreaux, H.B. (1979). "Arthropod phylogeny with special reference to insects" Wiley-Interscience, New York.

Bowmaker, M., Yang, M.Y., Yasukawa, T., Reyes, A., Jacobs, H.T., Huberman, J.A. and Holt, I.J. (2003). Mammalian mitochondrial DNA replicates bidirectionally from an initiation zone. *J Biol Chem* 278, 51: 50961-9.

Boxma, B., de Graaf, R.M., van der Staay, G.W., van Alen, T.A., Ricard, G., Gabaldon, T., van Hoek, A.H., Moon-van der Staay, S.Y., Koopman, W.J., van Hellemond, J.J., Tielens, A.G., Friedrich, T., Veenhuis, M., Huynen, M.A. and Hackstein, J.H. (2005). An anaerobic mitochondrion that produces hydrogen. *Nature* 434, 7029: 74-9.

Boyce, T.M., Zwick, M.E. and Aquadro, C.F. (1989). Mitochondrial DNA in the bark weevils: size, structure and heteroplasmy. *Genetics* 123, 4: 825-36.

Bridge, D., Cunningham, C.W., Schierwater, B., DeSalle, R. and Buss, L.W. (1992). Class-level relationships in the phylum Cnidaria: evidence from mitochondrial genome structure. *Proc Natl Acad Sci U S A* 89, 18: 8750-3.

Bullerwell, C.E. and Gray, M.W. (2004). Evolution of the mitochondrial genome: protist connections to animals, fungi and plants. *Curr Opin Microbiol* 7, 5: 528-34.

Burger, G., Forget, L., Zhu, Y., Gray, M.W. and Lang, B.F. (2003a). Unique mitochondrial genome architecture in unicellular relatives of animals. *Proc Natl Acad Sci U S A* 100, 3: 892-7.

Burger, G., Gray, M.W. and Franz Lang, B. (2003b). Mitochondrial genomes: anything goes. *Trends in Genetics* 19, 12 SU: 709-716.

Cameron, S.L., Miller, K.B., D'Haese, C.A., Whiting, M.F. and Barker, S.C. (2004). Mitochondrial genome data alone are not enough to unambiguously resolve the relationships of Entognatha, Insecta and Crustacea *sensu lato* (Arthropoda). *Cladistics* 20, 6: 534-557.

Carmean, D. and Crespi, B.J. (1995). Do long branches attract flies? *Nature* 373, 6516: 666.

Carver, M., Gross, G.F. and Woodward, T.E. (1991). "Hemiptera (Bugs, leafhoppers, cicadas, aphids, scale insects etc.)". Pp. 429-509 *in* Organisation, C. S. I. R., ed The Insects of Australia, a textbook for Students and Research Workers. Melbourne University press, Victoria.

Castro, L.R. and Dowton, M. (2005). The position of the Hymenoptera within the Holometabola as inferred from the mitochondrial genome of *Perga condei* (Hymenoptera: Symphyta: Pergidae). *Molecular Phylogenetics and Evolution* 34, 3: 469-479.

Caterino, M.S. and Sperling, F.A.H. (1999). Papilio Phylogeny Based on Mitochondrial Cytochrome Oxidase I and II Genes. *Molecular Phylogenetics and Evolution* 11, 1: 122-137.

Chrzanowska-Lightowlers, Z.M., Preiss, T. and Lightowlers, R.N. (1994). Inhibition of mitochondrial protein synthesis promotes increased stability of nuclear-encoded respiratory gene transcripts. *J Biol Chem* 269, 44: 27322-8.

Chrzanowska-Lightowlers, Z.M., Temperley, R.J., Smith, P.M., Seneca, S.H. and Lightowlers, R.N. (2004). Functional polypeptides can be synthesized from human mitochondrial transcripts lacking termination codons. *Biochem J* 377, Pt 3: 725-31.

Claros, M.G., Perea, J., Shu, Y., Samatey, F.A., Popot, J.L. and Jacq, C. (1995). Limitations to in vivo import of hydrophobic proteins into yeast mitochondria. The case of a cytoplasmically synthesized apocytochrome b. *Eur J Biochem* 228, 3: 762-71.

Claros, M.G. and Vincens, P. (1996). Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur J Biochem* 241, 3: 779-86.

Clary, D.O., Goddard, J.M., Martin, S.C., Fauron, C.M. and Wolstenholme, D.R. (1982). *Drosophila* mitochondrial DNA: a novel gene order. *Nucleic Acids Res* 10, 21: 6619-37.

Clary, D.O., Wahleithner, J.A. and Wolstenholme, D.R. (1983). Transfer RNA genes in *Drosophila* mitochondrial DNA: related 5' flanking sequences and comparisons to mammalian mitochondrial tRNA genes. *Nucleic Acids Res* 11, 8: 2411-25.

Clary, D.O., Wahleithner, J.A. and Wolstenholme, D.R. (1984). Sequence and arrangement of the genes for cytochrome b, URF1, URF4L, URF4, URF5, URF6 and five tRNAs in *Drosophila* mitochondrial DNA. *Nucleic Acids Res* 12, 9: 3747-62.

Clary, D.O. and Wolstenholme, D.R. (1983a). Genes for cytochrome c oxidase subunit I, URF2, and three tRNAs in *Drosophila* mitochondrial DNA. *Nucleic Acids Res* 11, 19: 6859-6872.

Clary, D.O. and Wolstenholme, D.R. (1983b). Nucleotide sequence of a segment of *Drosophila* mitochondrial DNA that contains the genes for cytochrome c oxidase subunits II and III and ATPase subunit 6. *Nucleic Acids Res* 11, 12: 4211-27.

Clary, D.O. and Wolstenholme, D.R. (1985a). The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. *J Mol Evol* 22, 3: 252-71.

Clary, D.O. and Wolstenholme, D.R. (1985b). The ribosomal RNA genes of *Drosophila* mitochondrial DNA. *Nucleic Acids Res* 13, 11: 4029-45.

Clayton, D.A. (2000). Transcription and replication of mitochondrial DNA. *Hum Reprod* 15 Suppl 2, 11-7.

Clayton, D.A. (2003). Mitochondrial DNA replication: what we know. *IUBMB Life* 55, 4-5: 213-7.

Coates, B.S., Sumerford, D.V., Hellmich, R.L. and Lewis, L.C. (2005). Partial mitochondrial genome sequences of *Ostrinia nubilalis* and *Ostrinia furnicalis*. *Int. J. Biol. Sci.* 1, 1: 13-18.

Conway, D.J., Fanello, C., Lloyd, J.M., Al-Joubori, B.M., Baloch, A.H., Somanath, S.D., Roper, C., Oduola, A.M., Mulder, B., Povoa, M.M., Singh, B. and Thomas, A.W. (2000). Origin of *Plasmodium falciparum* malaria is traced by mitochondrial DNA. *Mol Biochem Parasitol* 111, 1: 163-71.

Cornuet, J.M., Garnery, L. and Solignac, M. (1991). Putative origin and function of the intergenic region between COI and COII of *Apis mellifera* L. mitochondrial DNA. *Genetics* 128, 2: 393-403.

Couttet, P., Fromont-Racine, M., Steel, D., Pictet, R. and Grange, T. (1997). Messenger RNA deadenylylation precedes decapping in mammalian cells. *Proc Natl Acad Sci U S A* 94, 11: 5628-33.

Crease, T.J. (1999). The complete sequence of the mitochondrial genome of *Daphnia pulex* (Cladocera: Crustacea). *Gene* 233, 1-2: 89-99.

Crease, T.J. and Little, T.J. (1997). Partial sequence of the mitochondrial genome of the crustacean *Daphnia pulex*. *Curr Genet* 31, 1: 48-54.

Crews, S., Ojala, D., Posakony, J., Nishiguchi, J. and Attardi, G. (1979). Nucleotide sequence of a region of human mitochondrial DNA containing the precisely identified origin of replication. *Nature* 277, 5693: 192-8.

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.

Curole, J.P. and Kocher, T.D. (1999). Mitogenomics: digging deeper with complete mitochondrial genomes. *Trends in Ecology and Evolution* 14, 10: 394-398.

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

de Grey, A.D. (2005). Forces maintaining organellar genomes: is any as strong as genetic code disparity or hydrophobicity? *Bioessays* 27, 4: 436-46.

Delsuc, F., Phillips, M.J. and Penny, D. (2003). Comment on "Hexapod origins: monophyletic or paraphyletic?" *Science* 301, 5639: 1482; author reply 1482.

Desjardins, P. and Morais, R. (1990). Sequence and gene organization of the chicken mitochondrial genome. A novel gene order in higher vertebrates. *J Mol Biol* 212, 4: 599-634.

Doersen, C.J., Guerrier-Takada, C., Altman, S. and Attardi, G. (1985). Characterization of an RNase P activity from HeLa cell mitochondria. Comparison with the cytosol RNase P activity. *J Biol Chem* 260, 10: 5942-9.

Dotson, E.M. and Beard, C.B. (2001). Sequence and organization of the mitochondrial genome of the Chagas disease vector, *Triatoma dimidiata. Insect Mol Biol* 10, 3: 205-15.

Dowton, M. and Austin, A.D. (1999). Evolutionary dynamics of a mitochondrial rearrangement "hot spot" in the Hymenoptera. *Mol Biol Evol* 16, 2: 298-309.

Dowton, M., Castro, L.R., Campbell, S.L., Bargon, S.D. and Austin, A.D. (2003). Frequent mitochondrial gene rearrangements at the hymenopteran nad3-nad5 junction. *J Mol Evol* 56, 5: 517-26.

DSMZ. (Retrieved March 05, 2005).Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH - Human and Animal Cell Lines - SCHNEIDER-2. <http://www.dsmz.de/mutz/mutz130.htm>.

Dyall, S.D., Brown, M.T. and Johnson, P.J. (2004a). Ancient invasions: from endosymbionts to organelles. *Science* 304, 5668: 253-7.

Dyall, S.D., Yan, W., Delgadillo-Correa, M.G., Lunceford, A., Loo, J.A., Clarke, C.F. and Johnson, P.J. (2004b). Non-mitochondrial complex I proteins in a hydrogenosomal oxidoreductase complex. *Nature* 431, 7012: 1103-7.

Edqvist, J., Burger, G. and Gray, M.W. (2000). Expression of mitochondrial protein-coding genes in *Tetrahymena pyriformis. J Mol Biol* 297, 2: 381-93.

Ernster, L. and Schatz, G. (1981). Mitochondria: a historical review. *J Cell Biol* 91, 3 Pt 2: 227s-255s.

Falkenberg, M., Gaspari, M., Rantanen, A., Trifunovic, A., Larsson, N.G. and Gustafsson, C.M. (2002). Mitochondrial transcription factors B1 and B2 activate transcription of human mtDNA. *Nat Genet* 31, 3: 289-94.

Farr, C.L., Wang, Y. and Kaguni, L.S. (1999). Functional interactions of mitochondrial DNA polymerase and single-stranded DNA-binding protein. Template-primer DNA binding and initiation and elongation of DNA strand synthesis. *J Biol Chem* 274, 21: 14779-85.

Farre, J.C. and Araya, A. (2001). Gene expression in isolated plant mitochondria: high fidelity of transcription, splicing and editing of a transgene product in electroporated organelles. *Nucleic Acids Res* 29, 12: 2484-91.

Fernandez-Silva, P., Enriquez, J.A. and Montoya, J. (2003). Replication and transcription of mammalian mitochondrial DNA. *Exp Physiol* 88, 1: 41-56.

Fernandez-Silva, P., Martinez-Azorin, F., Micol, V. and Attardi, G. (1997). The human mitochondrial transcription termination factor (mTERF) is a multizipper protein but binds to DNA as a monomer, with evidence pointing to intramolecular leucine zipper interactions. *Embo J* 16, 5: 1066-79.

Fish, J., Raule, N. and Attardi, G. (2004). Discovery of a Major D-Loop Replication Origin Reveals Two Modes of Human mtDNA Synthesis. *Science* 306, 5704: 2098-2101.

Flook, P.K., Rowell, C.H. and Gellissen, G. (1995). The sequence, organization, and evolution of the *Locusta migratoria* mitochondrial genome. *J Mol Evol* 41, 6: 928-41.

Foster, P.G. and Hickey, D.A. (1999). Compositional bias may affect both DNA and protein-based phylogenetic reconstruction. *J Mol Evol* 48, 284-290.

Foster, P.G., Jermiin, L.S. and Hickey, D.S. (1997). Nucleotide composition bias affects amino acid content in proteins coded by animal mitochondria. *J Mol Evol* 44, 282-288.

Foury, F. and Kucej, M. (2002). Yeast mitochondrial biogenesis: a model system for humans? *Curr Opin Chem Biol* 6, 1: 106-11.

Francino, M.P. and Ochman, H. (1997). Strand asymmetries in DNA evolution. *Trends Genet* 13, 6: 240-5.

Friedrich, M. and Muqim, N. (2003). Sequence and phylogenetic analysis of the complete mitochondrial genome of the flour beetle *Tribolium castanaeum*. *Mol Phylogenet Evol* 26, 3: 502-12.

Frohman, M.A., Dush, M.K. and Martin, G.R. (1988). Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc Natl Acad Sci U S A* 85, 23: 8998-9002.

Fuller, K.M. and Zouros, E. (1993). Dispersed discrete length polymorphism of mitochondrial DNA in the scallop *Placopecten magellanicus* (Gmelin). *Curr Genet* 23, 4: 365-9.

Gagliardi, D., Stepien, P.P., Temperley, R.J., Lightowlers, R.N. and Chrzanowska-Lightowlers, Z.M. (2004). Messenger RNA stability in mitochondria: different means to an end. *Trends Genet* 20, 6: 260-7.

Garesse, R. (1988). Drosophila melanogaster mitochondrial DNA: gene organization and evolutionary considerations. *Genetics* 118, 4: 649-63.

Garstka, H.L., Facke, M., Escribano, J.R. and Wiesner, R.J. (1994). Stoichiometry of mitochondrial transcripts and regulation of gene expression by mitochondrial transcription factor A. *Biochem Biophys Res Commun* 200, 1: 619-26.

Gaspari, M., Falkenberg, M., Larsson, N.G. and Gustafsson, C.M. (2004a). The mitochondrial RNA polymerase contributes critically to promoter specificity in mammalian cells. *Embo J* 23, 23: 4606-14.

Gaspari, M., Larsson, N.G. and Gustafsson, C.M. (2004b). The transcription machinery in mammalian mitochondria. *Biochim Biophys Acta* 1659, 2-3: 148-52.

Gillespie, D.E., Salazar, N.A., Rehkopf, D.H. and Feagin, J.E. (1999). The fragmented mitochondrial ribosomal RNAs of *Plasmodium falciparum* have short A tails. *Nucleic Acids Res* 27, 11: 2416-22.

Goddard, J.M. and Wolstenholme, D.R. (1978). Origin and direction of replication in mitochondrial DNA molecules from *Drosophila melanogaster. Proc Natl Acad Sci U S A* 75, 8: 3886-90.

Goddard, J.M. and Wolstenholme, D.R. (1980). Origin and direction of replication in mitochondrial DNA molecules from the genus *Drosophila. Nucleic Acids Res* 8, 4: 741-57.

Goldenthal, M. and Nishiura, J.T. (1987). Isolation and characterization of a mitochondrial RNA polymerase from *Drosophila melanogaster. Biochem Cell Biol* 65, 2: 173-82.

Goto, A., Matsushima, Y., Kadowaki, T. and Kitagawa, Y. (2001). *Drosophila* mitochondrial transcription factor A (d-TFAM) is dispensable for the transcription of mitochondrial DNA in Kc167 cells. *Biochem J* 354, Pt 2: 243-8.

Gray, M.W. (1992). The endosymbiont hypothesis revisited. *Int Rev Cytol* 141, 233-357.

Gray, M.W. (2003). Diversity and evolution of mitochondrial RNA editing systems. *IUBMB Life* 55, 4-5: 227-33.

Gray, M.W. (2005). Evolutionary biology: the hydrogenosome's murky past. *Nature* 434, 7029: 29-31.

Gray, M.W., Burger, G. and Lang, B.F. (1999). Mitochondrial Evolution. *Science* 283, 1476-1481.

Gray, M.W. and Doolittle, W.F. (1982). Has the endosymbiont hypothesis been proven? *Microbiol Rev* 46, 1: 1-42.

Gray, M.W., Lang, B.F. and Burger, G. (2004). Mitochondria of Protists. *Annu Rev Genet.*

Grohmann, K., Amairic, F., Crews, S. and Attardi, G. (1978). Failure to detect "cap" structures in mitochondrial DNA-coded poly(A)-containing RNA from HeLa cells. *Nucleic Acids Res* 5, 3: 637-51.

Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 41, 95-98.

Hamilton, K.G.A. (1982). "The Insects and Arachnids of Canada Part 10. The Spittlebugs of Canada. Homoptera: Cercopidae." The Insects and Arachnids of Canada, Biosystematics Research Institute, Ottawa, ON.

Hanada, T., Suzuki, T. and Watanabe, K. (2000). Translation activity of mitochondrial tRNA with unusual secondary structure. *Nucleic Acids Symp Ser* 44: 249-50.

Hanada, T., Suzuki, T., Yokogawa, T., Takemoto-Hori, C., Sprinzl, M. and Watanabe, K. (2001). Translation ability of mitochondrial tRNAsSer with unusual secondary structures in an in vitro translation system of bovine mitochondria. *Genes Cells* 6, 12: 1019-30.

Harlid, A., Janke, A. and Arnason, U. (1997). The mtDNA sequence of the ostrich and the divergence between paleognathous and neognathous birds. *Mol Biol Evol* 14, 7: 754-61.

Helfenbein, K.G., Fourcade, H.M., Vanjani, R.G. and Boore, J.L. (2004). The mitochondrial genome of *Paraspadella gotoi* is highly reduced and reveals that chaetognaths are a sister group to protostomes. *Proc Natl Acad Sci U S A* 101, 29: 10639-43.

Hendy, M.D. and Penny, D. (1989). A framework for the quantitive study of evolutionary trees. *Syst Zoo* 38, 297-309.

Hennig, W. (1981). "Insect phylogeny" Wiley-Interscience, New York.

Hoffmann, R.J., Boore, J.L. and Brown, W.M. (1992). A novel mitochondrial genome organization for the blue mussel, *Mytilus edulis*. *Genetics* 131, 2: 397-412.

Holt, I.J. and Jacobs, H.T. (2003). Response: The mitochondrial DNA replication bubble has not burst. *Trends Biochem Sci* 28, 7: 355-6.

Hrdy, I., Hirt, R.P., Dolezal, P., Bardonova, L., Foster, P.G., Tachezy, J. and Embley, T.M. (2004). *Trichomonas* hydrogenosomes contain the NADH dehydrogenase module of mitochondrial complex I. *Nature* 432, 7017: 618-22.

Huelsenbeck, J.P. (1997). Is the Felsenstein zone a fly trap? *Syst Biol* 46, 1: 69-74.

Huelsenbeck, J.P. (1998). Systematic bias in phylogenetic analysis: is the Strepsiptera problem solved? *Syst Biol* 47, 3: 519-37.

Hwang, U.W., Friedrich, M., Tautz, D., Park, C.J. and Kim, W. (2001). Mitochondrial protein phylogeny joins myriapods with chelicerates. *Nature* 413, 6852: 154-7.

Hwang, U.W., Kim, W. and Tautz, D. (1998). Molecular phylogenetics at the Felsenstein Zone: approaching the Strepsiptera problem using 5.8S abd 28S rDNA sequences. *Mol Phylogenet Evol* 9, 3: 470-480.

Inoue, J.G., Miya, M., Tsukamoto, K. and Nishida, M. (2001). A mitogenomic perspective on the basal teleostean phylogeny: resolving higher-level relatioships with longer DNA sequences. *Mol Phylogenet Evol* 20, 275-285.

Jacobs, H.T. (1991). Structural similarities between a mitochondrially encoded polypeptide and a family of prokaryotic respiratory toxins involved in plasmid maintenance suggest a novel mechanism for the evolutionary maintenance of mitochondrial DNA. *J Mol Evol* 32, 4: 333-9.

Jacobs, H.T., Herbert, E.R. and Rankine, J. (1989). Sea urchin egg mitochondrial DNA contains a short displacement loop (D-loop) in the replication origin region. *Nucleic Acids Res* 17, 22: 8949-65.

Jager, E. and Bass, R. (1975). Chloramphenicol/thiamphenicol and cycloheximide as tools for the measurement of mitochondrial protein synthesis in vitro during organogenesis of rat embryos. *Naunyn Schmiedebergs Arch Pharmacol* 290, 2-3: 161-73.

Janke, A., Feldmaier-Fuchs, G., Thomas, W.K., von Haeseler, A. and Paabo, S. (1994). The marsupial mitochondrial genome and the evolution of placental mammals. *Genetics* 137, 1: 243-56.

Janke, A. and Paabo, S. (1993). Editing of a tRNA anticodon in marsupial mitochondria changes its codon recognition. *Nucleic Acids Res* 21, 7: 1523-5.

Johnston, S.A., Anziano, P.Q., Shark, K., Sanford, J.C. and Butow, R.A. (1988). Mitochondrial transformation in yeast by bombardment with microprojectiles. *Science* 240, 4858: 1538-41.

Junqueira, A.C., Lessinger, A.C., Torres, T.T., da Silva, F.R., Vettore, A.L., Arruda, P. and Azeredo Espin, A.M. (2004). The mitochondrial genome of the blowfly *Chrysomya chloropyga* (Diptera: Calliphoridae). *Gene* 339, 7-15.

Kaguni, L.S. (2004). DNA polymerase gamma, the mitochondrial replicase. *Annu Rev Biochem*, 73: 293-320.

Kim, I., Young Cha, S., Hee Yoon, M., Sam Hwang, J., Mong Lee, S., Dae Sohn, H. and Rae Jin, B. (2005). The complete nucleotide sequence and gene organization of the mitochondrial genome of the oriental mole cricket, *Gryllotalpa orientalis* (Orthoptera: Gryllotalpidae). *Gene* 32, 2: 155.

Klass, K.D., Zompro, O., Kristensen, N.P. and Adis, J. (2002). Mantophasmatodea: a new insect order with extant members in the Afrotropics. *Science* 296, 5572: 1456-9.

Knoop, V. (2004). The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. *Curr Genet* 46, 3: 123-39.

Kocher, T.D., Thomas, W.K., Meyer, A., Edwards, S.V., Paabo, S., Villablanca, F.X. and Wilson, A.C. (1989). Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proc Natl Acad Sci U S A* 86, 16: 6196-200.

Korhonen, J.A., Pham, X.H., Pellegrini, M. and Falkenberg, M. (2004). Reconstitution of a minimal mtDNA replisome in vitro. *Embo J* 23, 12: 2423-9.

Koulintchenko, M., Konstantinov, Y. and Dietrich, A. (2003). Plant mitochondria actively import DNA via the permeability transition pore complex. *Embo J* 22, 6: 1245-1254.

Kristensen, N.P. (1991). "Phylogeny of Extant Hexapods." Pp. 125-142 *in* Organisation, C. S. I. R., ed The Insects of Australia, a textbook for Students and Research Workers. Melbourne University Press, Victoria.

Kruse, B., Narasimhan, N. and Attardi, G. (1989). Termination of transcription in human mitochondria: identification and purification of a DNA binding protein factor that promotes termination. *Cell* 58, 2: 391-7.

Kuhn, J. and Binder, S. (2002). RT-PCR analysis of 5' to 3'-end-ligated mRNAs identifies the extremities of cox2 transcripts in pea mitochondria. *Nucleic Acids Res* 30, 2: 439-46.

Kumar, S., Tamura, K., Jakobsen, I.B. and Nei, M. (2001). MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17, 12: 1244-5.

La Roche, J., Snyder, M., Cook, D.I., Fuller, K. and Zouros, E. (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, 1: 45-64.

L'Abbe, D., Duhaime, J.F., Lang, B.F. and Morais, R. (1991). The transcription of DNA in chicken mitochondria initiates from one major bidirectional promoter. *J Biol Chem* 266, 17: 10844-50.

Lang, B.F., Burger, G., O'Kelly, C.J., Cedergren, R., Golding, G.B., Lemieux, C., Sankoff, D., Turmel, M. and Gray, M.W. (1997). An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature* 387, 6632: 493-7.

Lavrov, D.V., Brown, W.M. and Boore, J.L. (2000). A novel type of RNA editing occurs in the mitochondrial tRNAs of the centipede *Lithobius forficatus*. *Proc Natl Acad Sci USA* 97, 25: 13738-13742.

Lavrov, D.V., Forget, L., Kelly, M. and Lang, B.F. (2005). Mitochondrial Genomes of Two Demosponges Provide Insights into an Early Stage of Animal Evolution. *Mol Biol Evol* 22, 5: 1231-1239.

Lavrov, D.V. and Lang, B.F. (2005). Transfer RNA gene recruitment in mitochondrial DNA. *Trends Genet* 21, 3: 129-33.

Le, T.H., Blair, D., Agatsuma, T., Humair, P.F., Campbell, N.J., Iwagami, M., Littlewood, D.T., Peacock, B., Johnston, D.A., Bartley, J., Rollinson, D., Herniou, E.A., Zarlenga, D.S. and McManus, D.P. (2000). Phylogenies inferred from mitochondrial gene orders-a cautionary tale from the parasitic flatworms. *Mol Biol Evol* 17, 7: 1123-5.

Leblanc, C., Richard, O., Kloareg, B., Viehmann, S., Zetsche, K. and Boyen, C. (1997). Origin and evolution of mitochondria: what have we learnt from red algae? *Curr Genet* 31, 3: 193-207.

Lefai, E., Calleja, M., Ruiz de Mena, I., Lagina, A.T., 3rd, Kaguni, L.S. and Garesse, R. (2000). Overexpression of the catalytic subunit of DNA polymerase gamma results in depletion of mitochondrial DNA in *Drosophila melanogaster*. *Mol Gen Genet* 264, 1-2: 37-46.

Lessinger, A.C., Junqueira, A.C., Conte, F.F. and Azeredo-Espin, A.M. (2004). Analysis of a conserved duplicated tRNA gene in the mitochondrial genome of blowflies. *Gene* 339, 1-6.

Lessinger, A.C., Junqueira, T.A., Lemos, T.A., Kemper, E.L., da Silva, F.R., Vettore, A.L., Arruda, P. and Azeredo-Espin, A.M.L. (2000). The mitochondrial genome

of the primary screwworm fly *Cochliomyia hominivorax* (Dipter: Calliphoridae). *Insect Molec Biol* 9, 5: 521-529.

Lewis, D.L., Farr, C.L., Farquhar, A.L. and Kaguni, L.S. (1994). Sequence, organization, and evolution of the A+T region of *Drosophila melanogaster* mitochondrial DNA. *Mol Biol Evol* 11, 3: 523-38.

Lewis, D.L., Farr, C.L. and Kaguni, L.S. (1995). *Drosophila melanogaster* mitochondrial DNA: completion of the nucleotide sequence and evolutionary comparisons. *Insect Mol Biol* 4, 4: 263-278.

Lin, S., Zhang, H., Spencer, D.F., Norman, J.E. and Gray, M.W. (2002). Widespread and extensive editing of mitochondrial mRNAS in dinoflagellates. *J Mol Biol* 320, 4: 727-39.

Lowe, T.D. and Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25, 955-964.

Lunt, D.H., Zhang, D.X., Szymura, J.M. and Hewitt, G.M. (1996). The insect cytochrome oxidase I gene: evolutionary patterns and conserved primers for phylogenetic studies. *Insect Mol Biol* 5, 3: 153-65.

Martin, G. and Keller, W. (1998). Tailing and 3'-end labeling of RNA with yeast poly(A) polymerase and various nucleotides. *RNA* 4, 2: 226-30.

Masta, S.E. (2000). Mitochondrial sequence evolution in spiders: intraspecific variation in tRNAs lacking the TPsiC Arm. *Mol Biol Evol* 17, 7: 1091-100.

Masta, S.E. and Boore, J.L. (2004). The complete mitochondrial genome sequence of the spider *Habronattus oregonensis* reveals rearranged and extremely truncated tRNAs. *Mol Biol Evol* 21, 5: 893-902.

Matsushima, Y., Garesse, R. and Kaguni, L.S. (2004). *Drosophila* mitochondrial transcription factor B2 regulates mitochondrial DNA copy number and transcription in Schneider cells. *J Biol Chem* 279, 26: 26900-5.

Matsushima, Y., Garesse, R. and Kaguni, L.S. (2005). *Drosophila* mitochondrial transcription factor B1 modulates mitochondrial translation but not transcription or DNA copy number in Schneider cells. *J Biol Chem.*

McCulloch, V. and Shadel, G.S. (2003). Human mitochondrial transcription factor B1 interacts with the C-terminal activation region of h-mtTFA and stimulates transcription independently of its RNA methyltransferase activity. *Mol Cell Biol* 23, 16: 5816-24.

Merten, S.H. and Pardue, M.L. (1981). Mitochondrial DNA in Drosophila. An analysis of genome organization and transcription in *Drosophila melanogaster* and *Drosophila virilis*. *J Mol Biol* 153, 1: 1-21.

Mili, S. and Pinol-Roma, S. 2003. LRP130, a pentatricopeptide motif protein with a noncanonical RNA-binding domain, is bound in vivo to mitochondrial and nuclear RNAs. Mol Cell Biol, **23**: 4972-82.

Mindell, D.P., Sorenson, M.D. and Dimcheff, D.E. (1998). An extra nucleotide is not translated in mitochondrial ND3 of some birds and turtles. *Mol Biol Evol* 15, 11: 1568-71.

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

Montoya, J., Christianson, T., Levens, D., Rabinowitz, M. and Attardi, G. (1982). Identification of initiation sites for heavy-strand and light-strand transcription in human mitochondrial DNA. *Proc Natl Acad Sci U S A* 79, 23: 7195-9.

Montoya, J., Gaines, G.L. and Attardi, G. (1983). The pattern of transcription of the human mitochondrial rRNA genes reveals two overlapping transcription units. *Cell* 34, 1: 151-9.

Montoya, J., Ojala, D. and Attardi, G. (1981). Distinctive features of the 5'-terminal sequences of the human mitochondrial mRNAs. *Nature* 290, 5806: 465-70.

Mootha, V.K., Lepage, P., Miller, K., Bunkenborg, J., Reich, M., Hjerrild, M., Delmonte, T., Villeneuve, A., Sladek, R., Xu, F., Mitchell, G.A., Morin, C., Mann, M., Hudson, T.J., Robinson, B., Rioux, J.D. and Lander, E.S. (2003). Identification of a gene causing human cytochrome c oxidase deficiency by integrative genomics. *Proc Natl Acad Sci U S A*, 100: 605-10.

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

Morl, M., Dorner, M. and Paabo, S. (1995). C to U editing and modifications during the maturation of the mitochondrial tRNA(Asp) in marsupials. *Nucleic Acids Res* 23, 17: 3380-4.

Muller, M. and Martin, W. (1999). The genome of *Rickettsia prowazekii* and some thoughts on the origin of mitochondria and hydrogenosomes. *Bioessays* 21, 5: 377-81.

Nagaike, T., Suzuki, T., Katoh, T. and Ueda, T. (2005). Human mitochondrial mRNAs are stabilized with polyadenylation regulated by mitochondria-specific poly(A) polymerase and polynucleotide phosphorylase. *J Biol Chem* 280, 20: 19721-7.

Naithani, S., Saracco, S.A., Butler, C.A. and Fox, T.D. (2003). Interactions among COX1, COX2, and COX3 mRNA-specific translational activator proteins on the

inner surface of the mitochondrial inner membrane of Saccharomyces cerevisiae. *Mol Biol Cell* 14: 324-33.

Nakamichi, N., Rhoads, D.D., Hayashi, J.I., Kagawa, Y. and Matsumura, T. (1998). Detection, localization, and sequence analyses of mitochondrial regulatory region RNAs in several mammalian species. *J Biochem (Tokyo)* 123, 3: 392-8.

Nardi, F., Carapelli, A., Dallai, R. and Frati, F. (2003a). The mitochondrial genome of the olive fly *Bactrocera oleae*: two haplotypes from distant geographical locations. *Insect Molec Biol* 12, 6: 605-611.

Nardi, F., Carapelli, A., Faniulli, P.P., Dallai, R. and Frati, F. (2001). The complete mitochondrial DNA sequence of the basal hexapod *Tetrodontophora bielanensis*: Evidence for heteroplasmy and tRNA translocations. *Mol Biol Evol* 18, 1293-1304.

Nardi, F., Spinsanti, G., Boore, J.L., Carapelli, A., Dallai, R. and Frati, F. (2003b). Hexapod origins: monophyletic or paraphyletic? *Science* 299, 5614: 1887-9.

Nass, M.M. and Nass, S. (1963a). Intramitochondrial Fibers with DNA Characteristics. I. Fixation and Electron Staining Reactions. *J Cell Biol* 19, 593-611.

Nass, S. and Nass, M.M. (1963b). Intramitochondrial Fibers with DNA Characteristics. II. Enzymatic and Other Hydrolytic Treatments. *J Cell Biol* 19, 613-29.

Nass, S., Nass, M.M. and Hennix, U. (1965). Deoxyribonucleic Acid in Isolated Rat-Liver Mitochondria. *Biochim Biophys Acta* 95, 426-35.

Nedelcu, A.M., Lee, R.W., Lemieux, C., Gray, M.W. and Burger, G. (2000). The complete mitochondrial DNA sequence of *Scenedesmus obliquus* reflects an intermediate stage in the evolution of the green algal mitochondrial genome. *Genome Res* 10, 6: 819-31.

Nirenberg, M.W. and Matthaei, J.H. (1961). The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proc Natl Acad Sci U S A* 47, 1588-602.

Ojala, D., Merkel, C., Gelfand, R. and Attardi, G. (1980a). The tRNA genes punctuate the reading of genetic information in human mitochondrial DNA. *Cell* 22, 2 Pt 2: 393-403.

Ojala, D., Montoya, J. and Attardi, G. (1981). tRNA punctuation model of RNA processing in human mitochondria. *Nature* 290, 5806: 470-4.

Ojala, D.K., Montoya, J. and Attardi, G. (1980b). The putative mRNA for subunit II of human cytochrome c oxidase starts directly at the translation initiator codon. *Nature* 287, 5777: 79-82.

Okimoto, R., Macfarlane, J.L., Clary, D.O. and Wolstenholme, D.R. (1992). The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. *Genetics* 130, 3: 471-98.

Olsen, G.J., Woese, C.R. and Overbeek, R. (1994). The winds of (evolutionary) change: breathing new life into microbiology. *J Bacteriol* 176, 1: 1-6.

Orr, A.T., Rabets, J.C., Horton, T.L. and Landweber, L.F. (1997). RNA editing missing in mitochondria. *RNA* 3, 4: 335-6.

Papillon, D., Perez, Y., Caubit, X. and Le Parco, Y. (2004). Identification of chaetognaths as protostomes is supported by the analysis of their mitochondrial genome. *Mol Biol Evol* 21, 11: 2122-9.

Peden, J.F. (1999). "Analysis of Codon Usage". Thesis for degree of Doctor of Philosophy. University of Nottingham, Nottingham, UK.

Perez-Martinez, X., Antaramian, A., Vazquez-Acevedo, M., Funes, S., Tolkunova, E., d'Alayer, J., Claros, M.G., Davidson, E., King, M.P. and Gonzalez-Halphen, D. (2001). Subunit II of cytochrome c oxidase in *Chlamydomonad* algae is a heterodimer encoded by two independent nuclear genes. *J Biol Chem* 276, 14: 11302-9.

Perna, N.T. and Kocher, T.D. (1995). Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *J Mol Evol* 41, 3: 353-8.

Pont-Kingdon, G., Okada, N.A., Macfarlane, J.L., Beagley, C.T., Watkins-Sims, C.D., Cavalier-Smith, T., Clark-Walker, G.D. and Wolstenholme, D.R. (1998). Mitochondrial DNA of the coral *Sarcophyton glaucum* contains a gene for a homologue of bacterial MutS: a possible case of gene transfer from the nucleus to the mitochondrion. *J Mol Evol* 46, 4: 419-31.

Pont-Kingdon, G.A., Okada, N.A., Macfarlane, J.L., Beagley, C.T., Wolstenholme, D.R., Cavalier-Smith, T. and Clark-Walker, G.D. (1995). A coral mitochondrial mutS gene. *Nature* 375, 6527: 109-11.

Popot, J.L. and de Vitry, C. (1990). On the microassembly of integral membrane proteins. *Annu Rev Biophys Biophys Chem* 19, 369-403.

Pruitt, K.D., Tatusova, T. and Maglott, D.R. (2005). NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 33, Database issue: D501-4.

Puranam, R.S., Attardi, G., Shadel, G.S., Clayton, D.A. and Dairaghi, D.J. (2001). The RNase P associated with HeLa cell mitochondria contains an essential RNA component identical in sequence to that of the nuclear RNase P. *Mol Cell Biol* 21, 2: 548-61.

Qiu, Y., Song, D., Zhou, K. and Sun, H. 2005. The Mitochondrial Sequences of Heptathela hangzhouensis and Ornithoctonus huwena Reveal Unique Gene Arrangements and Atypical tRNAs. *J Mol Evol*, 60: 57-71.

Race, H.L., Herrmann, R.G. and Martin, W. (1999). Why have organelles retained genomes? *Trends Genet* 15, 9: 364-70.

Ramon Valverde, J., Batuecas, B., Moratilla, C., Marco, R. and Garesse, R. (1994). The complete mitochondrial DNA sequence of the crustacean *Artemia franciscana*. *J Mol Evol* 39, 4: 400-8.

Rantanen, A., Gaspari, M., Falkenberg, M., Gustafsson, C.M. and Larsson, N.G. (2003). Characterization of the mouse genes for mitochondrial transcription factors B1 and B2. *Mamm Genome* 14, 1: 1-6.

Reyes, A., Gissi, C., Pesole, G. and Saccone, C. (1998). Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. *Mol Biol Evol* 15, 8: 957-66.

Reyes, A., Yang, M.Y., Bowmaker, M. and Holt, I.J. (2005). Bidirectional replication initiates at sites throughout the mitochondrial genome of birds. *J Biol Chem* 280, 5: 3242-50.

Roberti, M., Fernandez-Silva, P., Polosa, P.L., Fernandez-Vizarra, E., Bruni, F., Deceglie, S., Montoya, J., Gadaleta, M.N. and Cantatore, P. (2005). In vitro transcription termination activity of the *Drosophila* mitochondrial DNA-binding protein DmTTF. *Biochem Biophys Res Commun* 331, 1: 357-62.

Roberti, M., Polosa, P.L., Bruni, F., Musicco, C., Gadaleta, M.N. and Cantatore, P. (2003). DmTTF, a novel mitochondrial transcription termination factor that recognises two sequences of *Drosophila melanogaster* mitochondrial DNA. *Nucleic Acids Res* 31, 6: 1597-604.

Rocha, E.P. (2003). DNA repeats lead to the accelerated loss of gene order in bacteria. *Trends Genet* 19, 11: 600-3.

Rohrbaugh, M.L. and Hardison, R.C. (1983). Analysis of rabbit beta-like globin gene transcripts during development. *J Mol Biol* 164, 3: 395-417.

Rossmanith, W. (1997). Processing of human mitochondrial tRNA(Ser(AGY))GCU: a novel pathway in tRNA biosynthesis. *J Mol Biol* 265, 4: 365-71.

Rossmanith, W. and Karwan, R.M. (1998). Characterization of human mitochondrial RNase P: novel aspects in tRNA processing. *Biochem Biophys Res Commun* 247, 2: 234-41.

Rossmanith, W., Tullo, A., Potuschak, T., Karwan, R. and Sbis, E. (1995). Human Mitochondrial tRNA Processing. *J. Biol. Chem.* 270, 21: 12885-12891.

Saccone, C., De Giorgi, C., Gissi, C., Pesole, G. and Reyes, A. (1999). Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. *Gene* 238, 1: 195-209.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). "Book 1". Molecular Cloning. A Laboratory Manual. Second Edition.

Samuels, D.C., Schon, E.A. and Chinnery, P.F. (2004). Two direct repeats cause most human mtDNA deletions. *Trends in Genetics* 20, 9: 393-398.

Sarabhai, A. and Brenner, S. (1967). A mutant which reinitiates the polypeptide chain after chain termination. *J Mol Biol* 27, 1: 145-62.

Satta, Y., Ishiwa, H. and Chigusa, S.I. (1987). Analysis of nucleotide substitutions of mitochondrial DNAs in *Drosophila melanogaster* and its sibling species. *Mol Biol Evol* 4, 6: 638-50.

Schneider, I. (1972). Cell lines derived from late embryonic stages of *Drosophila melanogaster*. *J Embryol Exp Morphol* 27, 2: 353-65.

Schultheis, A.S., Weigt, L.A. and Hendricks, A.C. (2002). Arrangement and structural conservation of the mitochondrial control region of two species of Plecoptera: utility of tandem repeat-containing regions in studies of population genetics and evolutionary history. *Insect Mol Biol* 11, 6: 605-10.

Scouras, A., Beckenbach, K., Arndt, A. and Smith, M.J. (2004). Complete mitochondrial genome DNA sequence for two ophiuroids and a holothuroid: the utility of protein gene sequence and gene maps in the analyses of deep deuterostome phylogeny. *Mol Phylogenet Evol* 31, 1: 50-65.

Seidel-Rogol, B.L., McCulloch, V. and Shadel, G.S. (2003). Human mitochondrial transcription factor B1 methylates ribosomal RNA at a conserved stem-loop. *Nat Genet* 33, 1: 23-4.

Selwood, S.P., McGregor, A., Lightowlers, R.N. and Chrzanowska-Lightowlers, Z.M. (2001). Inhibition of mitochondrial protein synthesis promotes autonomous regulation of mtDNA expression and generation of a new mitochondrial RNA species. *FEBS Lett* 494, 3: 186-91.

Shao, R. and Barker, S.C. (2003). The highly rearranged mitochondrial genome of the Plague Thrips, *Thrips imaginis* (Insecta: Thysanoptera): convergence of two novel gene boundaries and an extraordinary arrangement of rRNA genes. *Mol Biol Evol* 20, 3: 362-370.

Shao, R., Campbell, N.J.H. and Barker, S.C. (2001). Numerous Gene Rearrangements in the mitochondrial genome of the Wallaby Louse, *Heterodoxus macropus* (Phthriptera). *Mol Biol Evol* 18, 5: 858-865.

Shao, R., Dowton, M., Murrell, A. and Barker, S.C. (2003). Rates of gene rearrangement and nucleotide substitution are correlated in the mitochondrial genomes of insects. *Mol Biol Evol* 20, 10: 1612-9.

Sharma, M.R., Koc, E.C., Datta, P.P., Booth, T.M., Spremulli, L.L. and Agrawal, R.K. (2003). Structure of the mammalian mitochondrial ribosome reveals an expanded functional role for its component proteins. *Cell* 115, 1: 97-108.

Shoubridge, E.A. (2002). The ABCs of mitochondrial transcription. *Nat Genet* 31, 3: 227-8.

Simpson, L., Aphasizhev, R., Gao, G. and Kang, X. (2004). Mitochondrial proteins and complexes in *Leishmania* and *Trypanosoma* involved in U-insertion/deletion RNA editing. *RNA* 10, 2: 159-70.

Snyder, M., Fraser, A.R., LaRoche, J., Gartner-Kepkay, K.E. and Zouros, E. (1987). Atypical Mitochondrial DNA from the Deep-Sea Scallop *Placopecten magellanicus*. *PNAS* 84, 21: 7595-7599.

Sonnhammer, E.L. and Durbin, R. (1995). A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* 167, 1-2: GC1-10.

Sonnhammer, E.L., Eddy, S.R. and Durbin, R. (1997). Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins* 28, 3: 405-20.

Spanos, L., Koutroumbas, G., Kotsyfakis, M. and Louis, C. (2000). The mitochondrial genome of the Mediterranean fruit fly, *Ceratitis capitata*. *Insect Molec Biol* 9, 2: 139-144.

Spelbrink, J.N., Li, F.Y., Tiranti, V., Nikali, K., Yuan, Q.P., Tariq, M., Wanrooij, S., Garrido, N., Comi, G., Morandi, L., Santoro, L., Toscano, A., Fabrizi, G.M., Somer, H., Croxen, R., Beeson, D., Poulton, J., Suomalainen, A., Jacobs, H.T., Zeviani, M. and Larsson, C. (2001). Human mitochondrial DNA deletions associated with mutations in the gene encoding Twinkle, a phage T7 gene 4-like protein localized in mitochondria. *Nat Genet* 28, 3: 223-31.

Spradling, A., Pardue, M.L. and Penman, S. (1977). Messenger RNA in heat-shocked *Drosophila* cells. *J Mol Biol* 109, 4: 559-87.

Stanton, D.J., Daehler, L.L., Moritz, C.C. and Brown, W.M. (1994). Sequences with the potential to form stem-and-loop structures are associated with coding-region duplications in animal mitochondrial DNA. *Genetics* 137, 1: 233-41.

Steinberg, S. and Cedergren, R. (1994). Structural compensation in atypical mitochondrial tRNAs. *Struct Biol* 1, 8: 507-510.

Steinberg, S., Gautheret, D. and Cedergren, R. (1994). Fitting structurally diverse animal mitochondrial $tRNA^{Ser}$ to common three-dimentional constraints. *J Mol Evol* 236, 982-989.

Stewart, J.B. and Beckenbach, A.T. (2003). Phylogenetic and genomic analysis of the complete mitochondrial DNA sequence of the spotted asparagus beetle *Crioceris duodecimpunctata*. *Mol Phylogenet Evol* 26, 3: 513-26.

Stewart, J.B. and Beckenbach, A.T. (2005). Insect mitochondrial genomics: the complete mitochondrial genome sequence of the meadow spittlebug *Philaenus spumarius* (Hemiptera: Auchenorrhyncha: Cercopoidae). *Genome* 48, 1: 46-54.

Stroumbakis, N.D., Li, Z. and Tolias, P.P. (1994). RNA- and single-stranded DNA-binding (SSB) proteins expressed during *Drosophila melanogaster* oogenesis: a homolog of bacterial and eukaryotic mitochondrial SSBs. *Gene* 143, 2: 171-7.

Stuart, K.D., Schnaufer, A., Ernst, N.L. and Panigrahi, A.K. (2005). Complex management: RNA editing in trypanosomes. *Trends Biochem Sci* 30, 2: 97-105.

Sunnucks, P. and Hales, D.F. (1996). Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). *Mol Biol Evol* 13, 3: 510-24.

Taanman, J.W. (1999). The mitochondrial genome: structure, transcription, translation and replication. *Biochim Biophys Acta* 1410, 2: 103-23.

Takata, K., Yoshida, H., Hirose, F., Yamaguchi, M., Kai, M., Oshige, M., Sakimoto, I., Koiwai, O. and Sakaguchi, K. (2001). Drosophila mitochondrial transcription factor A: characterization of its cDNA and expression pattern during development. *Biochem Biophys Res Commun* 287, 2: 474-83.

Temperley, R.J., Seneca, S.H., Tonska, K., Bartnik, E., Bindoff, L.A., Lightowlers, R.N. and Chrzanowska-Lightowlers, Z.M. (2003). Investigation of a pathogenic mtDNA microdeletion reveals a translation-dependent deadenylation decay pathway in human mitochondria. *Hum Mol Genet* 12, 18: 2341-8.

Tessier, D.C., Brousseau, R. and Vernet, T. (1986). Ligation of single-stranded oligodeoxyribonucleotides by T4 RNA ligase. *Anal Biochem* 158, 1: 171-8.

Thao, M.L., Baumann, L. and Baumann, P. (2004). Organization of the mitochondrial genomes of whiteflies, aphids, and psyllids (Hemiptera, Sternorrhyncha). *BMC Evol Biol* 4, 1: 25.

Thomas, M.A., Walsh, K.A., Wolf, M.R., McPheron, B.A. and Marden, J.H. (2000). Molecular phylogenetic analysis of evolutionary trends in stonefly wing structure and locomotor behavior. *Proc Natl Acad Sci U S A* 97, 24: 13178-83.

Thompson, J.A., Radonovich, M.F. and Salzman, N.P. (1979). Characterization of the 5'-terminal structure of simian virus 40 early mRNA's. *J Virol* 31, 2: 437-46.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res* 11, 4673-4680.

Tiranti, V., Savoia, A., Forti, F., D'Apolito, M.F., Centra, M., Rocchi, M. and Zeviani, M. (1997). Identification of the gene encoding the human mitochondrial RNA polymerase (h-mtRPOL) by cyberscreening of the Expressed Sequence Tags database. *Hum Mol Genet* 6, 4: 615-25.

Tomecki, R., Dmochowska, A., Gewartowski, K., Dziembowski, A. and Stepien, P.P. (2004). Identification of a novel human nuclear-encoded mitochondrial poly(A) polymerase. *Nucleic Acids Res* 32, 20: 6001-14.

Tomita, K., Ueda, T. and Watanabe, K. (1996). RNA editing in the acceptor stem of squid mitochondrial tRNA(Tyr). *Nucleic Acids Res* 24, 24: 4987-91.

Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K., Nakashima, R., Yaono, R. and Yoshikawa, S. (1996). The whole structure of the 13-subunit oxidized cytochrome c oxidase at 2.8 A. *Science* 272, 5265: 1136-44.

Tracy, R.L. and Stern, D.B. (1995). Mitochondrial transcription initiation: promoter structures and RNA polymerases. *Curr Genet* 28, 3: 205-16.

Tupy, J.L., Bailey, A.M., Dailey, G., Evans-Holm, M., Siebel, C.W., Misra, S., Celniker, S.E. and Rubin, G.M. (2005). Identification of putative noncoding polyadenylated transcripts in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A*.

Ueda, T., Watanabe, K. and Ohta, T. 1983. Structural analysis of bovine mitochondrial tRNASer (AGY). *Nucleic Acids Symp Ser,* 141-4.

Viale, A.M. and Arakaki, A.K. (1994). The chaperone connection to the origins of the eukaryotic organelles. *FEBS Lett* 341, 2-3: 146-51.

von Dohlen, C.D. and Moran, N.A. (1995). Molecular phylogeny of the Homoptera: a paraphyletic taxon. *J Mol Evol* 41, 2: 211-23.

von Heijne, G. (1986). Why mitochondria need a genome. *FEBS Lett* 198, 1: 1-4.

Watkins, R.F. and Beckenbach, A.T. (1999). Partial sequence of a sponge mitochondrial genome reveals sequence similarity to Cnidaria in cytochrome oxidase subunit II and the large ribosomal RNA subunit. *J Mol Evol* 48, 5: 542-54.

Wernette, C.M., Conway, M.C. and Kaguni, L.S. (1988). Mitochondrial DNA polymerase from *Drosophila melanogaster* embryos: kinetics, processivity, and fidelity of DNA polymerization. *Biochemistry* 27, 16: 6046-54.

Wernette, C.M. and Kaguni, L.S. (1986). A mitochondrial DNA polymerase from embryos of *Drosophila melanogaster*. Purification, subunit structure, and partial characterization. *J Biol Chem* 261, 31: 14764-70.

Wheeler, W.C., M.F., W., Wheeler, Q.D. and J., C. (2001). Phylogeny of the extant hexapod orders. *Cladistics* 17, 2: 113-210.

Whiting, M., Carpenter, J.C., Wheeler, Q.D. and Wheeler, W.C. (1997). The Strepsiptera problem: phylogeny of the holometabolous insect orders inferred from 18s and 28s ribosomal DNA sequences and morphology. *Syst Biol* 46, 1-68.

Whiting, M.F. (1998). Long-branch distraction and the Strepsiptera. *Syst Biol* 47, 1: 134-8.

Whiting, M.F. and Wheeler, W.C. (1994). Insect homeotic transformation. *Nature* 368, 6473: 696.

Wolstenholme, D.R. (1973). Replicating DNA molecules from eggs of *Drosophila melanogaster*. *Chromosoma* 43, 1: 1-18.

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

Wootton, R.J. (1981). Palaeozoic insects. *Ann. Rev. Entomol.* 36, 319-344.

Wrischnik, L.A., Higuchi, R.G., Stoneking, M., Erlich, H.A., Arnheim, N. and Wilson, A.C. (1987). Length mutations in human mitochondrial DNA: direct sequencing of enzymatically amplified DNA. *Nucleic Acids Res* 15, 2: 529-42.

Xu, F., Morin, C., Mitchell, G., Ackerley, C. and Robinson, B.H. (2004). The role of the LRPPRC (leucine-rich pentatricopeptide repeat cassette) gene in cytochrome oxidase assembly: mutation causes lowered levels of COX (cytochrome c oxidase) I and COX III mRNA. *Biochem J* 382: 331-6.

Yamauchi, M.M., Miya, M.U. and Nishida, M. (2004). Use of a PCR-based approach for sequencing whole mitochondrial genomes of insects: two examples (cockroach and dragonfly) based on the method developed for decapod crustaceans. *Insect Mol Biol* 13, 4: 435-42.

Yang, M.Y., Bowmaker, M., Reyes, A., Vergani, L., Angeli, P., Gringeri, E., Jacobs, H.T. and Holt, I.J. (2002). Biased incorporation of ribonucleotides on the mitochondrial L-strand accounts for apparent strand-asymmetric DNA replication. *Cell* 111, 4: 495-505.

Yasukawa, T., Yang, M.Y., Jacobs, H.T. and Holt, I.J. (2005). A bidirectional origin of replication maps to the major noncoding region of human mitochondrial DNA. *Mol Cell* 18, 6: 651-62.

Yokobori, S. and Paabo, S. (1995). Transfer RNA editing in land snail mitochondria. *Proc Natl Acad Sci U S A* 92, 22: 10432-5.

Yukuhiro, K., Sezutsu, H., Itoh, M., Shimizu, K. and Banno, Y. (2002). Significant levels of sequence divergence and gene rearrangements have occurred between the mitochondrial genomes of the wild mulberry silkmoth, *Bombyx mandarina*, and its close relative, the domesticated silkmoth, Bombyx mori. *Mol Biol Evol* 19, 8: 1385-9.

Zardoya, R. and Meyer, A. (1998). Complete mitochondrial genome suggests diapsid affinities of turtles. *Proc Natl Acad Sci U S A* 95, 24: 14226-31.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31, 13: 3406-15.

Zwick, P. (2000). Phylogenetic system and zoogeography of the Plecoptera. *Annu Rev Entomol* 45, 709-46.