

# University of Wollongong - Research Online

## Thesis Collection

Title: The evolution of mitochondrial genomes and phylogenetic relationships in the hymenoptera

Author: Lyda Raquel Castro

Year: 2006

Repository DOI:

### Copyright Warning

You may print or download ONE copy of this document for the purpose of your own research or study. The University does not authorise you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site.

You are reminded of the following: This work is copyright. Apart from any use permitted under the Copyright Act 1968, no part of this work may be reproduced by any process, nor may any other exclusive right be exercised, without the permission of the author. Copyright owners are entitled to take legal action against persons who infringe their copyright. A reproduction of material that is protected by copyright may be a copyright infringement. A court may impose penalties and award damages in relation to offences and infringements relating to copyright material.

Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.

**Unless otherwise indicated, the views expressed in this thesis are those of the author and do not necessarily represent the views of the University of Wollongong.**

Research Online is the open access repository for the University of Wollongong. For further information contact the UOW Library: [research-pubs@uow.edu.au](mailto:research-pubs@uow.edu.au)

*University of Wollongong Thesis Collections*

*University of Wollongong Thesis Collection*

---

*University of Wollongong*

*Year 2006*

---

The evolution of mitochondrial genomes  
and phylogenetic relationships in the  
hymenoptera

Lyda Raquel Castro  
University of Wollongong

Castro, Lyda Raquel, The evolution of mitochondrial genomes and phylogenetic relationships in the hymenoptera, PhD thesis, School of Biological Sciences, University of Wollongong, 2006. <http://ro.uow.edu.au/theses/660>

This paper is posted at Research Online.  
<http://ro.uow.edu.au/theses/660>

## **NOTE**

This online version of the thesis may have different page formatting and pagination from the paper copy held in the University of Wollongong Library.

## **UNIVERSITY OF WOLLONGONG**

### **COPYRIGHT WARNING**

You may print or download ONE copy of this document for the purpose of your own research or study. The University does not authorise you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site. You are reminded of the following:

Copyright owners are entitled to take legal action against persons who infringe their copyright. A reproduction of material that is protected by copyright may be a copyright infringement. A court may impose penalties and award damages in relation to offences and infringements relating to copyright material. Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.

**The evolution of mitochondrial genomes and phylogenetic  
relationships in the Hymenoptera**

A thesis submitted in fulfillment  
of the requirement for the award of the degree of

Doctorate of Philosophy

from the

University of Wollongong

by

Lyda Raquel Castro BSc, MSc

School of Biological Sciences

2006

### **Thesis Declaration**

I, Lyda Raquel Castro, declare that this thesis, submitted in partial fulfillment of the requirements for the award of Doctor of Philosophy, in the School of Biological Sciences, University of Wollongong, is wholly my own work unless otherwise referenced and acknowledged. This document has not been submitted for qualification at any other academic institution.

Lyda Raquel Castro

January 2006

## Acknowledgments

I am indebted to Dr. Mark Dowton from the University of Wollongong for supervising this project. I specially want to thank him for his continual support, guidance, encouragement, and knowledge over the past years. He has been always there for me and since the beginning has made my stay in Australia the best experience I could ever have had. I feel privileged to have worked in his lab. He has been the best example, not only as a professional, but as a person. I also want to thank all the members of his lab. for their support and friendship.

I would like to thank my family, especially my mum and dad. Their continual support, guidance and confidence made me achieve this goal. I also would like to thank Carlos, for his love, support and encouragement.

Thanks to Andy Austin, Dave Smith, and John Jennings for provision of the *Perga condei*, *Vanhornia eucnemidarum* and *Primeuchroeus* sp. specimens. For the kind donation of the other specimens, I want to thank Eveline Bartowsky, Robert Belshaw, Ferdinando Bin, Paul Dangerfield, John Early, F. Felipe, Scott Field, G. Fitt, Ian Gauld, Gary Gibson, E. Grissell, Paul Hanson, B. Hatami, John Heraty, Marianne Hellers, P. Horne, G. Jackson, John Jennings, Norm Johnson, Mike Keller, D. Kent, J. King, J. Kitt, M. Kulbars, Nina Laurence, Paul Lewis, Patrick Mardulyn, Lubomir Masner, Gwen Mayo, D. Murray, Ian Naumann, J. O'Hara, Donald Quicke, Nathan Schiff, Scott Shaw, Aguiar Sharkov, D. Smith, R. Storey, Gary Taylor, N. Tonkin, G. Tribe, G. Walter, Q. Wang, A. Wells, Bob Wharton, Jim Whitfield, N. Zareh.

I would like to thank Gary Gibson who generously shared his knowledge on apocritan morphology with us. Also Andy Austin for his suggestions and comments on preliminary versions of this thesis. Thanks to Rajkumar Kunaparaju for preliminary laboratory work on the sequencing of *Perga condei*, and Kalani Ruberu also for her help in cloning and sequencing *Vanhornia eucnemidarum*.

Finally, I want to thank the University of Wollongong, the Australian Museum and the Australian Research Council. Grants from these institutions supported this work.

## Abstract

This thesis studies the phylogenetic relationships in the Hymenoptera as well as the mitochondrial (mt) genomics of the group with a comparative approach. My principal purposes were to (1) reconstruct the phylogenetic relationships among the Apocrita, and (2) characterize the mt genome of the Hymenoptera and its utility as a phylogenetic marker both within the Apocrita and within the Holometabola. In order to achieve these aims: (1) 550 bp of the *18S* gene were sequenced in 87 apocritan taxa and analyzed using a Bayesian phylogenetic approach, including the sequences of two mitochondrial genes (*cox1*, *16S*) and another nuclear gene (*28S*) from Dowton and Austin (2001). Although the phylogeny of the Apocrita was not totally resolved, I was able to support some groups. In particular, the monophyly of the Proctotrupomorpha, and within this group the Chalcidoidea as sister taxon to the Diapriidae + Monomachidae + Maamingidae clade were consistently recovered and supported by high posterior probabilities. (2) Most of the mt genome of the sawfly *Perga condei* was sequenced. 12 protein coding genes, 16 *trn* genes, and the small and large rRNA genes, for a total of 13,416 bp. This mt genome has a conserved gene order, with the exception that *tnaL*<sup>CUN</sup> was not found in the position considered ancestral to insects and crustaceans (Boore et al. 1998, Flook et al. 1995). Apart from this rearrangement, the organization of the genes in *Perga condei* matches perfectly with distant species such as *Drosophila melanogaster* (Lewis et al. 1994) or *Triatoma dimidiata* (Dotson and Beard 2001). The base composition, the amino acid composition, and the codon usage of the mt genome of *P. condei* were reported. Similar to other insect mt genomes, this genome is A+T rich, and there is a correlation between the base composition and amino acid occurrence, with A+T rich codons predominating. (3) Two other mt



genomes of the Hymenoptera were also sequenced. The mt genome of *Vanhornia eucnemidarum* and of *Primeuchroeus* sp., both from the suborder Apocrita. Within the Apocrita, high rates of molecular evolution, compositional bias and gene rearrangements had been reported (Dowton and Austin 1997, Dowton et al. 2003). The mt genomes of *Vanhornia* and *Primeuchroeus* are further evidence of an increased rate of gene rearrangement within the Apocrita. In particular, there is a total of six *trn* genes rearranged in *Vanhornia eucnemidarum*. Additionally, several non-coding regions were found in the mt genome of *Vanhornia eucnemidarum*. One of these non-coding regions is around 600 bp long and has a high AT content, but does not seem to correspond to the typical A+T rich region present in other insect mt genomes. There are at least nine *trn* genes rearranged in the mt genome of *Primeuchroeus* sp. Further, the large and small rRNA genes are inverted. In both species, rearrangements of *trn* genes are the most common. The gene rearrangements found in the mt genomes of the hymenopteran taxa sequenced were characterized; however no synapomorphies were detected. Since the rate of gene rearrangement appears to be increased in this group of insects, only with increased taxon sampling will phylogenetically informative rearrangements be found. (4) Finally, the mt genome sequences previously described were tested as phylogenetic markers to reconstruct relationships both within the Holometabola and within the Hymenoptera. Results indicated that phylogenetic analyses using mt genomes were susceptible to outgroup and ingroup selection as well as analytical model. Analyses excluding 3<sup>rd</sup> codon positions were found to be the best model to analyze this type of data, but an increased taxonomic sampling within the Apocrita as well as within the outgroups is required to recover appropriate phylogenetic relationships.

## Table of Contents

Thesis Declaration.....	ii
Acknowledgments.....	iii
Abstract .....	v
List of Abbreviations .....	xi
List of Tables .....	xiv
List of Figures .....	xvi
CHAPTER 1 INTRODUCTION .....	1
1.1 MITOCHONDRIAL GENOME: EVOLUTION.....	2
1.1.1 Substitutional Rate Heterogeneity .....	2
1.1.2 Compositional Bias .....	5
1.1.3 Gene rearrangements.....	9
<i>1.1.3.1 Gene rearrangement mechanisms</i> .....	10
<i>1.1.3.2 Gene rearrangements as phylogenetic markers</i> .....	12
1.2 CONCATENATED MITOCHONDRIAL GENE SEQUENCES AND THEIR ...	14
USE FOR PHYLOGENETIC RECONSTRUCTION .....	14
<i>1.2.1 Cases in which concatenated mitochondrial gene sequences have been relatively</i>	
<i>successful in recovering phylogenetic relationships</i> .....	15
<i>1.2.2 Cases in which concatenated mitochondrial gene sequences have been</i>	
<i>problematic in recovering phylogenetic relationships</i> .....	16
1.3 EVOLUTION OF THE HYMENOPTERA.....	19
<i>1.3.1 Phylogenetic position of the Hymenoptera among other Holometabolan insects</i>	
.....	19
<i>1.3.2 Phylogenetic relationships within the Hymenoptera</i> .....	22
<i>1.3.2.1 Suborder Symphyta</i> .....	23
<i>1.3.2.2 Suborder Apocrita</i> .....	23
1.4 PARASITISM IN THE HYMENOPTERA.....	24
<i>1.4.1 Parasitoid Natural History</i> .....	25
<i>1.4.2 The Evolution of Parasitism</i> .....	28
1.5 AIMS.....	29

1.6 THESIS FORMAT .....	31
CHAPTER 2 The position of the Hymenoptera within the Holometabola as inferred from the mitochondrial genome of <i>Perga condei</i> (Hymenoptera:Symphyta:Pergidae) .....	34
2.1 Introduction.....	34
2.2 Materials and Methods.....	38
2.2.1 <i>Insects and DNA Extraction</i> .....	38
2.2.2 <i>PCR amplification and cloning</i> .....	39
2.2.3 <i>Identification of tRNAs</i> .....	40
2.2.4 <i>Alignments and phylogenetic analysis</i> .....	41
2.3 Results.....	43
2.3.1 <i>Genome structure and organization</i> .....	43
2.3.2 <i>Base composition, amino acid composition and codon usage</i> .....	45
2.3.3 <i>Phylogenetic analysis</i> .....	46
2.4 Discussion .....	48
2.4.1 <i>Perga condei mitochondrial genome structure</i> .....	48
2.4.2 <i>Phylogenetic analysis</i> .....	50
2.5 Conclusion .....	52
CHAPTER 3 Molecular Bayesian Analysis of the Apocrita (Insecta: Hymenoptera) suggests that the Chalcidoidea are sister to the (Monomachidae+Diapriidae+Maamingidae) .....	61
3.1 Introduction.....	61
3.2 Materials and Methods.....	63
3.2.1 <i>DNA Extraction and sequencing</i> .....	64
3.2.2 <i>Sequence alignment and phylogenetic analysis</i> .....	65
3.3 Results.....	67
3.3.1 <i>Molecular analyses</i> .....	68
3.3.1.1 <i>Unrooted analysis</i> .....	69
3.3.2 <i>Analysis including only the Proctotrupomorpha</i> .....	70
3.3.3 <i>Comparison between Bayesian runs</i> .....	71
3.4 Discussion .....	71
3.4.1 <i>Chalcidoidea+(Diapriidae+Monomachidae+Maamingidae)</i> .....	73
3.4.2 <i>Other hymenopteran relationships</i> .....	74
3.4.2.1 <i>Evaniiidae not sister to the Gasteruptiidae</i> .....	74
3.4.2.2 <i>Aculeata within the Evaniomorpha</i> .....	75

3.5 Conclusion .....	75
CHAPTER 4 Mitochondrial genomes of <i>Vanhornia eucnemidarum</i> and <i>Primeuchroeus</i> sp.: Evidence of rearranged mitochondrial genomes within the Apocrita (Hymenoptera) .....	84
4.1 Introduction .....	84
4.2 Materials and Methods.....	86
4.2.1 <i>Insects and DNA Extraction</i> .....	87
4.2.2 <i>PCR amplification and cloning</i> .....	87
4.2.3 <i>Identification of protein-coding, rrn, trn genes and gene rearrangements</i> .....	89
4.3 Results and Discussion.....	90
4.3.1 <i>Genome size and composition</i> .....	90
4.3.1.1 <i>Vanhornia eucnemidarum</i> .....	90
4.3.1.2 <i>Primeuchroeus</i> sp. ....	92
4.3.2 <i>Protein-coding genes and codon usage</i> .....	94
4.3.2.1 <i>Vanhornia eucnemidarum</i> .....	94
4.3.2.2 <i>Primeuchroeus</i> sp. ....	95
4.3.3 <i>Transfer RNA genes and ribosomal RNA genes</i> .....	95
4.3.3.1 <i>Vanhornia eucnemidarum</i> .....	95
4.3.3.2 <i>Primeuchroeus</i> sp. ....	96
4.3.4 <i>Genome organization</i> .....	97
4.3.4.1 <i>Vanhornia eucnemidarum</i> .....	97
4.3.4.2 <i>Primeuchroeus</i> sp. ....	98
4.3.5 <i>The ARNSEF region of Vanhornia eucnemidarum</i> .....	98
4.3.6 <i>Comparison of gene rearrangements</i> .....	99
4.4 Conclusion .....	100
CHAPTER 5 Mitochondrial genomes in the Hymenoptera and their utility as phylogenetic markers.....	114
5.1 Introduction .....	114
5.2 Materials and Methods.....	118
5.2.1 <i>Taxon Sampling</i> .....	118
5.2.2 <i>Alignment</i> .....	119
5.2.3 <i>Phylogenetic analysis</i> .....	119
5.3 Results and Discussion.....	120
5.3.1 <i>Model Specification and stationarity</i> .....	121
5.3.2 <i>Hymenopteran Relationships</i> .....	121
5.3.2.1 <i>All taxa included</i> .....	121
5.3.2.2 <i>Only Hymenoptera included</i> .....	122
5.3.3 <i>Holometabolan Relationships</i> .....	124

5.3.2.1 <i>Locusta</i> as outgroup.....	124
5.3.2.2 <i>Triatoma</i> as outgroup .....	125
5.3.2.3 <i>No outgroup</i> .....	125
5.4 Conclusion .....	126
Chapter 6 General Discussion .....	138
6.1 Significance of the study.....	138
6.2 Major results and implications.....	139
6.2.1 <i>An improved resolution of the Apocritan phylogeny</i> .....	139
6.2.2 <i>A more accurate estimate of the degree of gene rearrangements in</i> <i>hymenopteran mt genomes</i> .....	142
6.2.3 <i>The potential of mt genomes as phylogenetic markers</i> .....	144
6.3 Recommendations and future research .....	145
6.4 Conclusions.....	148
REFERENCES .....	151

## List of Abbreviations

$\mu\text{g}$	micro-gram
$\mu\text{L}$	micro-liter
$\mu\text{M}$	micro-molar
Ala [A]	Alanine (amino acid)
Arg [R]	Arginine (amino acid)
Asn [N]	Asparagine (amino acid)
Asp [D]	Aspartic acid (amino acid)
<i>atp6</i>	gene for ATP synthase subunit 6
<i>atp8</i>	gene for ATP synthase subunit 8
<i>cob</i>	cytochrome oxydase B
<i>cox1</i>	subunit 1 of cytochrome oxydase
<i>cox2</i>	subunit 2 of cytochrome oxydase
<i>cox3</i>	subunit 3 of cytochrome oxydase
Cys [C]	Cysteine (amino acid)
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide-triphosphate
EDTA	Ethylendiaminetetraacetic acid
Gln [Q]	Glutamine (amino acid)
Glu [E]	Glutamic acid (amino acid)
Gly [G]	Glycine (amino acid)
GTR	General time-reversible model
GTR + I + $\Gamma$	General time-reversible model with rates invgamma
His [H]	Histidine (amino acid)
Ile [I]	Isoleucine (amino acid)
Kb	Kilo-base
Leu [L]	Leucine (amino acid)
Lys [K]	Lysine (amino acid)
Met [M]	Metionine (amino acid)
ML	Maximum likelihood
mM	mili-molar

MP	Maximum parsimony
Mt	Mitochondrial
<i>nd1</i>	gene for NADH dehydrogenase subunit 1
<i>nd2</i>	gene for NADH dehydrogenase subunit 2
<i>nd3</i>	gene for NADH dehydrogenase subunit 3
<i>nd4</i>	gene for NADH dehydrogenase subunit 4
<i>nd4L</i>	gene for NADH dehydrogenase subunit 4L
<i>nd5</i>	gene for NADH dehydrogenase subunit 5
<i>nd6</i>	gene for NADH dehydrogenase subunit 6
ORF	Open reading frame
PCR	Polimerase chain reaction
PEG	Polyethylene glycol
Pfu	Pyrococcus furiosus
Phe [F]	Phenylalanine (amino acid)
PP	Posterior probability
Pro [P]	Proline (amino acid)
<i>rnl</i>	gene for ribosomal RNA large-subunit
<i>rns</i>	gene for ribosomal RNA small-subunit
RGR	rate of gene rearrangement
SDS	Sodium dodecyl sulfate
Ser [S]	Serine (amino acid)
Thr [T]	Threonine (amino acid)
<i>trn</i>	transfer RNA
<i>trnA</i>	transfer RNA gene for alanine
<i>trnC</i>	transfer RNA gene for cytosine
<i>trnD</i>	transfer RNA gene for aspartic acid
<i>trnE</i>	transfer RNA gene for glutamic acid
<i>trnF</i>	transfer RNA gene for phenylalanine
<i>trnH</i>	transfer RNA gene for histidine
<i>trnK</i>	transfer RNA gene for lysine
<i>trnL</i>	transfer RNA gene for leucine

<i>trnM</i>	transfer RNA gene for methionine
<i>trnN</i>	transfer RNA gene for asparagine
<i>trnP</i>	transfer RNA gene for proline
<i>trnR</i>	transfer RNA gene for arginine
<i>trnS</i>	transfer RNA gene for serine
<i>trnT</i>	transfer RNA gene for threonine
<i>trnV</i>	transfer RNA gene for valine
<i>trnW</i>	transfer RNA gene for tryptophan
<i>trnY</i>	transfer RNA gene for tyrosine
Trp [W]	Tryptophan (amino acid)
Tyr [Y]	Tyrosine (amino acid)
U	units
Val [V]	Valine (amino acid)



## List of Tables

<b>Table 2.1</b> Primer pairs used in PCR amplification of the <i>Perga condei</i> mitochondrial genome.....	54
<b>Table 2.2</b> Information of the taxa from which the whole mitochondrial genome have been sequenced, and were used in the phylogenetic analyses.....	55
<b>Table 2.3</b> AT content in different regions of the <i>Perga condei</i> mt genome.....	56
<b>Table 2.4</b> Codon usage in the <i>Perga condei</i> mt genome.....	57
<b>Table 3.1</b> List of taxonomic groups included in the phylogenetic analysis.....	77
<b>Table 3.2</b> Estimated model likelihood obtained in the different analyses.....	79
<b>Table 4.1</b> Primer pairs used in PCR amplification of the <i>V. eucnemidarum</i> and <i>Primeuchroeus</i> sp. mt genomes.....	101
<b>Table 4.2</b> Mt gene profile of <i>V. eucnemidarum</i> (Hymenoptera: Vanhnornidae).....	102
<b>Table 4.3</b> AT content in different regions of the <i>V. eucnemidarum</i> mt genome.....	103
<b>Table 4.4</b> Possible gene junctions in <i>Primeuchroeus</i> mt genome. Primers were carefully designed at the end of the genes and used in PCR reactions assuming all possible orientations.....	104

<b>Table 4.5</b> AT content in different regions of the <i>Primeuchroeus</i> sp. mt genome...	105
<b>Table 4.6</b> Comparison of the mt protein-encoding genes of <i>V. eucnemidarum</i> with those of other hymenopteran mt genomes available in GenBank.....	106
<b>Table 4.7</b> Codon usage in the <i>V. eucnemidarum</i> mt genome.....	107
<b>Table 4.8</b> Codon usage in the <i>Primeuchroeus</i> sp. mt genome.....	108
<b>Table 5.1</b> Information on the taxa whose whole mitochondrial genome sequence was used in the phylogenetic analyses.....	128
<b>Table 5.2</b> Estimated model likelihood obtained in the different analyses.....	129
<b>Table 5.3</b> Within hymenopteran relationships as recovered by the different phylogenetic analyses.....	130
<b>Table 5.4</b> Position of the Hymenoptera within the Holometabola as recovered by the different phylogenetic analyses.....	131
<b>Table 6.1</b> Assessment of the relative rate of genome rearrangement among hymenopteran mitochondrial genomes, using the RGR test.....	150

## List of Figures

<b>Figure 1.1</b> Previous hypotheses for holometabolan relationships.....	32
<b>Figure 1.2</b> Summary of what is known about symphytan relationships.....	33
<b>Figure 1.3</b> Major apocritan relationships proposed by Rasnitsyn (1988).....	33
<b>Figure 2.1</b> Holometabolan relationships obtained from the amino acid dataset.....	58
<b>Figure 2.2</b> Holometabolan relationships obtained from the nucleotide datasets.....	59
<b>Figure 3.1</b> Bayesian analysis including <i>28S</i> , <i>16S</i> , <i>COI-1</i> , <i>COI-2</i> (DA2001).....	80
<b>Figure 3.2</b> Bayesian analysis including <i>28S</i> , <i>16S</i> , <i>COI-1</i> , and <i>COI-2</i> partitions from DA2001, and <i>18S</i> partition. ....	81
<b>Figure 3.3</b> Bayesian analysis of the Proctotrupomorpha; including <i>28S</i> , <i>16S</i> , <i>COI-1</i> , and <i>COI-2</i> partitions from Downton and Austin (2001), and <i>18S</i> partition.....	82
<b>Figure 3.4</b> Generation plots of some of the substitution rates for the <i>COI-1</i> and <i>COI-2</i> partitions after 5,000,000 generation runs. A. Generation plot of the C-G rate parameter for the <i>COI-1</i> partition (diffuse posterior distribution). B. Generation plot of the C-G rate parameter for the <i>28S</i> partition (posterior remains focused). ....	83
<b>Figure 4.1</b> Linearized representation of the mt genome arrangement of <i>Vanhornia eucnemidarum</i> (Hymenoptera: Apocrita) and the putative ancestral arthropod organization	

represented by <i>Drosophila</i>	
<i>yacuba</i> .....	109

<b>Figure 4.2</b> Linearized representation of the mt genome arrangement of <i>Primeuchroeus sp.</i> (Hymenoptera: Apocrita) and the putative ancestral arthropod organization represented by <i>Drosophila yacuba</i> .....	110
--	-----

<b>Figure 4.3</b> Putative secondary structure for the 18 <i>trn</i> genes of the <i>V. eucnemidarum</i> mt genome.....	111
---	-----

<b>Figure 4.4</b> Inferred secondary structure of the <i>trn</i> genes of the mt genome of <i>Primeuchroeus sp.</i> .....	113
---	-----

<b>Figure 5.1</b> Accepted phylogenetic relationships between the five hymenopteran taxa included in this analysis.....	132
---	-----

<b>Figure 5.2</b> Phylogenetic relationships of the Holometabola. Bayesian analyses.....	134
--	-----

<b>Figure 5.3</b> Phylogenetic relationships of the Holometabola. Parsimony analyses..	136
--	-----

<b>Figure 5.4</b> Only Hymenopteran representatives included. A. Bayesian analyses of nucleotide sequences excluding 3 <sup>rd</sup> codon positions. B. Tree topology recovered in all other analyses (parsimony and Bayesian using proteins, parsimony and Bayesian including all codon positions and parsimony excluding 3 <sup>rd</sup> codon positions).....	137
---	-----

## **Chapter 1 . INTRODUCTION**

Since the late 1980's, mitochondrial DNA (mtDNA) analysis has become established as a powerful tool for evolutionary studies of animals. Information on factors that impact on the evolution of mitochondrial (mt) genomes provides critical background knowledge to improve the level of analytical sophistication in a range of studies, such as phylogeny estimation, population genetics, and mutation research (Moritz et al. 1987).

Features that affect the usefulness of mt genome sequences as markers for phylogenetic reconstruction include rates of nucleotide substitution (Felsenstein 1978, Hendy and Penny 1989), compositional bias (Conant and Lewis 2001, Foster and Hickey 1999, Hasegawa et al. 1993, Jermin et al. 1994), and gene rearrangements (Moritz et al. 1987, Boore and Brown 1998, Dowton et al. 2002b). The first section of the introduction will summarize the current state of knowledge regarding each of these phenomena. As one of the aims of this study was to use predominantly mt sequences to resolve the phylogeny of the Hymenoptera, a megadiverse order of Insecta (Kristensen 1999), a minor focus of the literature review will draw examples in relation to this particular group.

The second section of this introduction will review the current state of knowledge on concatenated mt gene sequences and their utility for phylogenetic reconstruction. I will also include current information concerning the phylogenetic position of the Hymenoptera as well as the relationships within the group. Finally, as a robust phylogeny of the Hymenoptera will allow us to trace the evolution of the parasitic life-style within the group,

in the last section of this introduction I will provide a review on parasitism in general and, specifically, on the evolution of parasitism in the Hymenoptera.

## ***1.1 MITOCHONDRIAL GENOME: EVOLUTION***

### **1.1.1 Substitutional Rate Heterogeneity**

Rates of nucleotide substitution are known to vary widely among phylogenetic groups (e.g. Adachi et al. 1993, Britten 1986, Cantatore et al. 1994, Martin et al. 1992, Martin 1995, Mindell and Thacker 1996, Wu and Li 1985). This has led to the rejection of a universal molecular clock (which implies that for any given macromolecule the rate of evolution is approximately constant over time in all evolutionary lineages [Zuckerkandl and Pauling 1965]) and has prompted investigators to search for possible causes of rate heterogeneity. Although it is unlikely that variation in the rate of molecular evolution is determined by a single factor, several individual factors have been associated with variation in rates of molecular evolution (Britten 1986, Adachi et al. 1993, Mindell and Thacker 1996). Some of the hypotheses that explain different molecular rates between taxa are:

*Differences in DNA repair efficiency among lineages*, proposed by Britten (1986) to explain the difference in substitution rate between the primate and rodent lineages. He suggested that a more efficient repair mechanism in primates could account for their slow molecular evolution.

*The generation time hypothesis* (Kohne 1970), which has gained widespread acceptance among evolutionary biologists, has been used to explain higher substitution rates for

monkeys than for humans (Li and Tanimura 1987, Chang et al. 1994) and higher rates for rodents than for primates (Wu and Li 1985, Chang et al. 1994). Additional evidence for generation time effects has also come from avian studies (Moore and Harvey 1994). The basic tenet of the generation time hypothesis is that organisms with shorter generation times should have a greater number of germ cell divisions per unit time and, therefore, a concomitantly higher mutation rate. This holds if the majority of mutations are the result of errors during DNA replication, if the number of germ cell divisions are roughly similar per generation in most organisms, and if the majority of mutations are neutral (Kimura 1979).

*The metabolic rate hypothesis*, proposed by Martin and Palumbi (1993), states that rates of molecular evolution should be positively correlated with metabolic rates. The logical basis of their hypothesis is twofold. First, increased rates of DNA replication and nucleotide replacement in organisms with higher metabolic rates should lead to higher mutation rates. Second, the increased concentrations of free oxygen radicals in cells with higher metabolic rates should be associated with a higher incidence of DNA damage. Support for this hypothesis comes from several observations. Rates of mitochondrial DNA synthesis are higher in tissues with higher metabolic rates (Gross et al. 1969). Similarly, small organisms have more mitochondria per unit body mass than large organisms, suggesting higher rates of mtDNA replication for smaller taxa. Finally, free oxygen radicals, the byproducts of metabolic respiration, are known mutagens (Shigenaga et al. 1989).

*The population size hypothesis*. The notion that population size can affect rates of molecular evolution derives from the nearly neutral theory (Ohta 1973, 1992). Specifically, this theory predicts that populations with small effective sizes will experience faster rates of evolution than populations with larger effective sizes. This follows because slightly deleterious mutations are more likely to be fixed in a small population. Thus, smaller

populations will experience a higher fixation rate for slightly deleterious mutations than will larger populations due to the increased influence of drift on selection.

Variation in evolutionary rates across taxa has been reported as having a significant effect in phylogenetic reconstruction. Felsenstein (1978) described the long-branch attraction (LBA) scenario, whereby parsimony and compatible methods are misled by pronounced branch length differences (Hendy and Penny 1989). The problem of long branches attracting is most likely to occur when rates of evolution show considerable variation among sequences, or where the sequences being analyzed are quite divergent. However, it has been found that the conditions of branch length differences necessary to create LBA can be provoked by unequal taxon sampling, without any need for unequal rates (Hendy and Penny 1989, Holland et al. 2003). It has been suggested that one strategy to reduce the effects of LBA could be to increase taxon sampling, by adding sequences that join onto those branches, thus breaking them up (Rannala et al. 1998). Many researchers recommend and employ maximum likelihood methods or other model-based methods such as Bayesian inference (Huelsenbeck et al. 2001), although the problem may be exacerbated by incorrect assumptions about the model of evolution (Chang 1996). Other workers have proposed modified versions of parsimony to ameliorate potential LBA problems (Steel et al. 1993, Willson 1999).

Siddall and Whiting (1999) and Pol and Siddall (2001) suggested a method to detect LBA. It is called “long-branch extraction” and involves removing the taxa which are potentially causing the long-branch attraction problem and re-running the analysis. If LBA is occurring, this test allows the long branches to find their correct position in two separate analyses. Bergsten (2005) found that the most common LBA artifact is related to ingroup



taxa being pulled towards a long branched outgroup. He suggested that all phylogenetic analyses should be run both with and without the outgroups to compare whether the outgroup only roots the ingroup tree, or if it simultaneously alters the ingroup topology. This technique proved useful in identifying LBA problems in his study (Bergsten 2005) as well as others (Lin et al. 2002 Hampl et al. 2004, Graham et al. 2002, Garcia-Moreno et al. 2003).

Interestingly, an increased rate of mt nucleotide substitution is coincident with the transition to the parasitic lifestyle in the Hymenoptera. A higher rate of mtDNA sequence divergence was found in the parasitic wasps (suborder Apocrita) compared with nonparasitic wasps (suborder Symphyta) (Dowton and Austin 1995b). Possible causes for this correlation are an increased rate of speciation (Page et al. 1998, Castro et al. 2002), an increased flux of mutagens, and a decreased DNA repair efficiency among the Apocrita (Dowton and Campbell 2001).

### **1.1.2 Compositional Bias**

A common feature of nucleotide sequences is compositional bias, the occurrence of the four bases A, G, C, and T in unequal proportions. The degree of compositional bias of mt genomes varies widely among genes and organisms. For example, the G+C content of third positions of codons from the protein–encoding genes in bacteria, vertebrates, and insects studied so far, ranges from 4% to 98% (Bernardi et al. 1985, Ikemura 1985, Jukes and Bhushan 1986, Sueoka 1988, Liu and Bechenbach 1992). When taxa under study exhibit a

similar pattern and degree of compositional bias, they are said to exhibit stationarity, sometimes referred to as base compositional equilibrium (Saccone et al. 1999). Variation in compositional bias among taxa is known as deviation from stationarity (Collins et al. 1994).

Two main evolutionary processes have been invoked to explain why patterns of base composition vary within and among species: biases in the process of mutation (such that the rates of change from G·C ↔ A·T are not constant in time or space) (Sueoka 1988), and natural selection (either in overall nucleotide content or on specific patterns of codon usage) (Akashi et al. 1998, Eyre-Walker 1999).

Animal mt genomes deviate from a random usage of nucleotides, particularly in the third positions of synonymous codons. These deviations are said to result from directional mutation pressure (Sueoka 1962, Asakawa et al. 1991, Jermini et al. 1994, 1996). Other potential causes for nucleotide biases have also been suggested, including reduced pools for specific nucleotides, a biased preference of the mt  $\gamma$  DNA polymerase for specific nucleotides, and a propensity for a specific mutational direction for relatively long term exposed H-strands during replication (Asakawa et al. 1991).

Compositionally biased codon usage has been found in Hymenoptera (Crozier and Crozier 1993, Jermini and Crozier 1994, Dowton and Austin 1995a), and has been associated with the evolution of parasitism (Dowton and Austin 1995a, 1997). Oxidative damage may be a significant source of mutations in mt genomes. This involves the oxidative or hydrolytic deamination of dC to dU, changing the base pairing from dG to dA during the next round of replication (Wagner et al. 1992). Consistent with this scenario, Dowton and Austin (1997)

found that the A-content of the mt *16S* fragment increased in parasitic wasps, at the expense of the G-content.

Compositional heterogeneity in sequence data has been identified as a problem that can mislead methods commonly used to infer phylogenetic trees (Conant and Lewis 2001, Jermini et al. 2004, Rosenberg and Kumar 2003). Conventional tree-building methods from amino acid and nucleotide sequences can be unreliable when the base composition of taxa varies between sequences since they tend to group sequences of similar nucleotide composition irrespective of the evolutionary history of the organisms (Lockhart et al. 1994). Although it has been suggested that amino acid sequences are more reliable (Hasegawa et al. 1993), it seems that amino acid composition is also affected by nucleotide compositional bias (Foster et al. 1997, Singer and Hickey 2000). In particular, compositional bias in sequence data was found to increase the difficulty with which short internal edges can be inferred using the maximum parsimony method, the maximum likelihood method with an F81 model of nucleotide substitution, and the neighbor-joining method using the Jukes-Cantor model of nucleotide substitution (Jermini et al. 2004).

In an attempt to correct for these biases, methods have been devised that take into account unequal base composition among sequences. One of these methods is the LogDet transformation of Lockhart et al. (1994), which allows tree-selection methods to consistently recover the correct tree when sequences evolve under simple asymmetric models that can vary between lineages. Some have suggested that the LogDet method does not consider rate variation among sites and that, similar to other distance methods, it performs poorly in analyses of taxa with moderate amount of substitution saturation (Mooers and Holmes 2000, Conant and Lewis 2001). Further, Jermini et al. (2004) showed

that the LogDet method is more likely than other phylogenetic methods to recover the true topology from compositionally heterogeneous sequences that have evolved at the same rate, but that the probability of inferring the true topology using this method, like other phylogenetic methods, decreased when internal edges in the true tree were very short.

Another method of distance estimation was presented by Galtier and Gouy (1995). They developed a non-homogeneous Markov model of nucleotide substitution that allowed equilibrium base composition to vary among lineages (Galtier and Gouy 1995). This model was also expanded to a maximum likelihood framework that allowed a higher number of sequences to be handled (Galtier and Gouy 1998). However, a major practical limitation of this new algorithm is running time. When the number of compared sequences is higher than seven or eight, only a small fraction of the tree space can be examined (Galtier and Gouy 1998).

Clearly, there is a need for more flexible methods to infer trees from compositionally heterogeneous data, particularly methods and programs that also allow a user to consider gene and site-specific differences in the nucleotide and amino acid content, the rates of change, and the distribution and proportion of invariable sites. It is also particularly important to assess the compositional heterogeneity of phylogenetic data as the number of sequences increases since as the number of ingroup taxa increases the lengths of the true internal edges become smaller (Jermiin et al. 2004).

### 1.1.3 Gene rearrangements

In animals, mt genomes are typically small (15-20kb), are circular, and encode 37 genes: two ribosomal RNAs (large and small ribosomal RNA), 22 *trn* genes, and 13 protein coding genes (Wolstenholme 1992, Boore 1999). There is one *trn* for every amino acid except leucine and serine, which have two genes. The protein coding genes consist of ATP synthase subunits 6 and 8 (*atp6* and *atp8*), cytochrome oxidase subunits 1-3 (*cox1-cox3*), cytochrome b (*cob*), and NADH dehydrogenase subunits 1-6 and 4L (*nd1-6* and *nd4L*). Typically, animal mt genomes also have a large noncoding region that contains elements that control the replication and transcription of the genome and is known to be A+T rich in insects (Wolstenholme 1992). With few exceptions, the gene content for animal mt genomes is well conserved (Boore 1999), but the gene order is more variable (Boore and Brown 1998).

Animal mtDNA is considerably more variable in size and gene organization than has generally been recognized (Moritz et al. 1987). The 37 mt genes can potentially be rearranged in an enormous number of combinations [ $1.9 \times 10^{54}$  different ways according to Dowton et al. (2002)], and the large number of different arrangements found among and within many metazoan phyla suggest that this character is relatively unconstrained (Boore and Brown 2000). Major rearrangements of genes, defined as translocations and/or inversions of one or more multigene tracts, appear to be infrequent on a geological timescale, although minor rearrangements, such as exchanges of position or polarity between neighboring tRNA genes, are encountered with greater frequency. With the possible exclusion of some minor rearrangements, convergent rearrangements in

independent lineages are highly unlikely, and rearrangements promise to be a reliable character for determining ancient relationships (Boore and Brown 1998).

The order of genes in the mt genome can be used to investigate the evolution of organisms and their genomes by providing information that can be used to develop models for the mechanisms involved in gene rearrangement, replication and regulation, and by providing characters that can be used in phylogenetic analysis (Serb and Lydeard 2003).

#### *1.1.3.1 Gene rearrangement mechanisms*

Most of the concepts concerning how mt genomes rearrange are based on comparisons of vertebrate mt genomes, due to the availability of more sequences for vertebrate taxa (Moritz et al. 1987, Macey et al. 1997, Macey et al. 1998). Based on these studies, gene rearrangements were first proposed to occur by tandem duplication of gene regions as a result of slipped-strand mispairing of the nascent DNA strand to a secondarily homologous region of the template strand during replication, and subsequent deletion of duplicated genes in a way that results in a different organization (Mortiz and Brown 1986, Mortiz et al. 1987, Kumazawa and Nishida 1995, Pääbo et al. 1991). Gene rearrangements observed in birds may well be explained by this mechanism (Desjardins and Morais 1990, Quinn and Wilson 1993), and its feasibility is also supported by the frequent polymorphic duplications of mtDNA sequences found in lizards (Mortiz and Brown 1986, Mortiz et al. 1987), snakes (Kumazawa and Nishida 1995), and marsupials (Pääbo et al. 1991).

Generally, *trn* genes appear to rearrange more frequently than protein-coding or ribosomal genes. One mechanism proposed to explain this, is the illicit priming of replication by *trn* genes and the resultant integration of *trn* genes around the control region (Cantatore et al. 1987). The proximity of *trn* genes to duplication boundaries, their apparently enhanced mobility relative to other mt genes, and their potential to form secondary structures suggest that they might facilitate rearrangements (Mortiz et al. 1987). Mt genomic rearrangements have also been explained by transposition events (Macey et al. 1997) or duplication and concerted evolution associated with replication of the heavy strand (Kumazawa et al. 1996). Rearrangements can also occur by inversions (in which the *trn* gene swaps strands) and translocations (in which *trn* genes move from a remote location into a *trn* gene cluster) that cannot be explained by the slipped-strand mispairing model. In this case, inversions may be produced by topoisomerases, which are known to mediate illegitimate recombination in prokaryotic systems (Boore and Brown 1998).

The mechanisms for rearrangements in invertebrates have not been studied as well as for vertebrates. Although inversions and translocations are rare in vertebrates, they seem to be common in invertebrate mt genomes (Dowton and Austin 1999b, Shao et al. 2001b). Dowton and Campbell (2001) proposed a model of intra mt recombination to explain inversions and convergent translocation events in invertebrates. The type of recombination that they proposed is similar to that described by Lunt and Hyman (1997). According to Lunt and Hyman (1997), a defining feature of recombination models is the breakage and rejoining of participating DNA strands. The products of intramolecular recombination in mtDNA would be reciprocal, double-stranded subgenomic circles closed by covalent bonds. Dowton and Campbell (2001) proposed that two proximate double-stranded breaks,

rather than being circularized to produce minicircles [as observed by Lunt and Hyman (1997)], might be reinserted back into the mt genome in the reverse orientation, producing the inversion. Although this model of gene rearrangement has not been experimentally evaluated, a model of inter-mtDNA recombination has also been proposed for mites (Shao et al. 2005).

#### *1.1.3.2 Gene rearrangements as phylogenetic markers*

Gene arrangement comparisons are emerging as powerful tools for resolving ancient phylogenetic relationships (Boore et al. 1995). For reconstructing specific metazoan branches, comparisons of gene arrangements in mtDNA seem to have exceptional advantages. First, gene content is nearly invariant. Second, the great number of potential arrangements makes convergence unlikely, therefore, gene arrangements are likely to be shared only as a result of common ancestry. Third, all of the genes commonly found in animal mtDNA have homologues in the mtDNAs of plants, fungi, and/or protists, making their homology near certain. Fourth, stability of gene arrangements is better explained by infrequency of rearrangement than by selection, also making convergence less likely. Fifth, gene arrangements commonly remain unchanged over long periods of evolutionary time, so they may retain the signal of ancient common ancestry (Boore and Brown 1998).

Gene rearrangements have been used to elucidate relationships among major clades of echinoderms (Smith et al. 1993), gastropods (Kurabayashi and Ueshima 2000), eutrozoans (Stechmann and Schlegel 1999), annelids (Boore and Brown 2000), and arthropods (Boore et al. 1995, Boore et al. 1998). However, recent attempts to use gene rearrangements to



reconstruct phylogenies have shown that some gene order comparisons are more problematic than others and that the methods to analyze gene rearrangements are still in their infancy. For example, Scouras et al. (2004) found that the increased number of gene rearrangements in ophiuroids is difficult to reconcile with other echinoderm mtDNAs with confidence. The diversity of mt DNA gene orders observed within the classes of Echinodermata and the possibility of independent origins of gene orders makes the analysis difficult (Scouras et al. 2004). Additionally, despite the claims that convergence of mt gene order is unlikely, there have been isolated cases of convergence (Flook et al. 1995, Mindell et al. 1998, Dowton and Austin 1999a).

The analysis of gene rearrangement characters is also a subject of current research. The lack of appropriate models of evolution, together with the extreme (at least in comparison with sequence data) mathematical complexity of gene order comparisons, creates major computational challenges (Gascuel, 2005). In general, gene organizational data has been coded and analyzed by standard parsimony methods (Dowton et al. 2002b). Likelihood methods are represented to date by a single effort (Larget et al. 2004). In Larget et al. (2004), Bayesian analysis of genome arrangements for 87 metazoan taxa gave inconclusive results and led the authors to doubt the usefulness of gene rearrangements alone to resolve evolutionary relationships among animals at the phylum level. However, a limitation of their analysis is that their Bayesian approach to estimate phylogeny and ancestral genome arrangements was based on a model in which gene inversion is the sole mechanism of change.

The use of more biologically realistic models of mt gene rearrangement, as well as the collection of mt gene rearrangements for more taxa, are necessary to assess the utility of these rearrangements as phylogenetic characters. Some groups present an amazing acceleration of the rearrangement rate: for example, the hemipteroid orders (Shao et al. 2001a, Shao et al. 2001b), decapod crustaceans (Morrison et al. 2002), and the parasitic Hymenoptera (Crozier and Crozier 1993, Dowton 1999, Dowton and Austin 1999b). These groups have been proposed as excellent models for examining the gene rearrangement mechanism and for assessing the phylogenetic utility of gene rearrangement characters (Dowton et al. 2002a).

## **1.2 CONCATENATED MITOCHONDRIAL GENE SEQUENCES AND THEIR USE FOR PHYLOGENETIC RECONSTRUCTION**

Mt DNA possesses a number of characteristics that have made it a popular marker in evolutionary studies (Avice et al. 1987). Because mt genes are effectively single copy, comparisons of paralogous genes is generally not a concern. The clonal pattern of maternal inheritance typical of most animals allows direct reconstruction of a bifurcating tree topology. Uniparental inheritance also reduces the effective population size of mt genes, which means that variants are fixed more quickly between speciation events (Kocher et al. 1989, Curole and Kocher 1999a).

The accumulation of complete mt DNA sequences in the gene databases has lead to an interest in the use of combined mt protein coding sequence data for resolving deep level

phylogenies (Arnason et al. 2002, Miya et al. 2001, Saccone et al. 1999). Several phylogenetic analyses have demonstrated that the use of complete mt genomes in phylogenetic studies significantly increases the confidence of the phylogenetic history inferred compared with phylogenetic hypotheses based on individual or partial mt genes (Cummings et al. 1995, Russo et al. 1996, Zardoya and Meyer 1996). The longer sequences obtained by sequencing complete genomes have encouraged attempts to reconstruct relationships among divergent lineages. However, complete mt genomes have sometimes failed to recover phylogenetic relationships supported with other markers (Curole and Kocher 1999b). This suggests that the limits and applicability of these data remain to be elucidated.

### ***1.2.1 Cases in which concatenated mitochondrial gene sequences have been relatively successful in recovering phylogenetic relationships***

Whole mt genome sequences have been successfully used to address questions concerning bat evolution (Lin and Penny 2001). This study supported bat monophyly and consistently placed bats close to the cetferungulates [whales (cetaceans) plus ferungulates (carnivores, ungulates, and perissodactyls)]. In a subsequent study, Lin et al. (2002) reported an increased stability of the evolutionary tree of mammals after improved taxon sampling. They collected mt genome sequences from 29 Laurasiatherians, 42 eutherians, and 47 mammals. A major finding was that additional taxa reduced LBA problems, and allowed them to obtain congruent information with morphological and nuclear datasets. In particular, analysis of their expanded dataset grouped the hedgehog with the mole, while in previous analyses the position of the hedgehog had been problematic (Krettek et al. 1995).

The correct position of the hedgehog, as well as eulipotyphalan monophyly, was also corroborated by Nikaido et al. (2003), who demonstrated the importance of using appropriate substitution models and species sampling when inferring phylogenetic relationships. In the same way, Phillips and Penny (2003) found that mt genomes favour Theria (marsupials plus placentals) over Marsiupionta (monotremes plus marsupials), in agreement with morphological and nuclear studies, after nucleotides are recoded as RY-characters, and maximum-likelihood analyses are partitioned among subsets of data.

Within the insects, mt genomes have been successful in recovering phylogenetic relationships concordant with traditional views of phylogeny and with convincing levels of support within the Diptera (Lessinger et al. 2000) and Coleoptera (Cameron et al. unpublished). Lessinger et al. (2000) performed phylogenetic analyses, both with nucleotides and amino acid data sets, and to minimize possibly inconsistent alignments they tested the effect of excluding the less conserved genes. All analyses recovered the same, well supported relationships, and were consistent with traditional groupings.

### ***1.2.2 Cases in which concatenated mitochondrial gene sequences have been problematic in recovering phylogenetic relationships***

Other studies that have used concatenated mt gene sequences for resolving phylogenies have produced results which are difficult to reconcile with trees produced using other markers, or that contradict traditionally accepted phylogenetic relationships. These studies have highlighted the need for sophisticated analysis of the different signals found within the mt genome (e.g. protein coding vs. ribosomal genes, first and second vs. third codon

positions, DNA vs. amino acid sequence data) and have highlighted the importance of evaluating the effect of other variables, such as outgroup and ingroup selection, data treatment, gene choice and optimality criteria when using mt genomes for phylogenetic reconstruction (Cameron et al. 2004).

Complete mt genomes failed to answer questions regarding vertebrate relationships. For example, with respect to tetrapod origins, different analyses supported different topologies with high bootstrap values, and the results were not sufficient to favor a lungfish + coelacanth clade or a lungfish + tetrapod clade (Zardoya and Meyer 1996, Zardoya et al. 1998). Another controversial question is the relationship of the two extant lineages of jawless fishes to the remaining vertebrates. One hypothesis predicts that hagfishes are the sister group to a clade containing the lamprey and Gnathostomata, while an alternative hypothesis predicts that the lampreys and hagfishes form a monophyletic clade that is sister to the Gnathostomata. Analysis of the complete mt genomes of a lamprey (*Petromyzon marinus*), a hagfish (*Myxine glutinosa*) and several teleost fishes, were not strong enough to support any hypothesis (Curole and Kocher 1999).

With reference to examination of insect relationships, some analyses using mt genomes have had to exclude taxa, probably due to their compositional bias or increased rates of nucleotide substitution. Several studies have excluded the honeybee mt genome in order to recover a monophyletic Holometabola (Black and Roehrdanz 1998, Friedrich and Muqim 2003, Nardi et al. 2001, Nardi et al. 2003, Stewart and Beckenbach 2003). Most studies have also had to exclude *Heterodoxus* (Insecta: Phthiraptera) and *Thrips* (Insecta: Thysanoptera) genome sequences due to evidence of unusual sequence evolution of these

two taxa (Foster and Hickey 1999). Additionally, Hassanin et al. (2005) have shown evidence of the dramatic consequences that mutational saturation and heterogeneity of nucleotide composition among taxa have on phylogenetic analyses using mt genomes.

Another recent and controversial example of using mt genomes for phylogenetic reconstruction was the study of Nardi et al. (2003). Nardi et al. (2003) contradict the generally accepted view of Hexapoda as a monophyletic clade and showed that Collembola, traditionally considered as basal to insects, appeared instead more closely related to one of the crustacean groups, the Brachiopoda. Delsuc et al. (2003) criticized the methodology used by Nardi et al. (2003), who drew their conclusions from maximum likelihood and Bayesian analyses of amino acids from four of the 13 mitochondrial proteins coding genes. Delsuc et al. (2003) argued that phylogenetic analyses of amino acids carry several pitfalls that might be avoided by analyzing nucleotide sequences for which more realistic models of sequence evolution and powerful reconstruction methods are available. Delsuc et al (2003) recoded the nucleotides as purines (R) and pyrimidines (Y), and their analyses recovered Collembola as sister to the insects, and the Hexapoda as a monophyletic clade. Cameron et al. (2004) re-examined the Nardi et al. (2003) analyses, not only based on the problems associated with the analytical methodology, but examining other areas that possibly affected the phylogenetic outcome (e.g. outgroup selection, ingroup taxon selection and alignment methodology, choices of genes). However, they were unable to confidently recover the sister-group of Collembola or make any conclusions regarding the monophyly of the Hexapoda. Instead they concluded that mt genome data alone were insufficient to resolve the issue and that the dataset appeared highly vulnerable to taxon selection, outgroup choice, data manipulation and gene selection.

The relatively limited number of species for which complete mt genome sequences are currently available make it difficult to deduce any firm conclusions about the usefulness of concatenated gene sequences in phylogenetic reconstruction. However, the apparently correct phylogenies recovered when a better taxon sampling is available (as is the case for the Diptera [Lessinger et al. 2000] and Coleoptera [Cameron et al. unpublished] within the insects, or the mammals [Lin et al. 2002]) indicates that whole mt genomes still represent promising candidates for resolving phylogenies.

### **1.3 EVOLUTION OF THE HYMENOPTERA**

#### ***1.3.1 Phylogenetic position of the Hymenoptera among other Holometabolan insects***

The Holometabola is the most diverse and successful group of terrestrial organisms, comprising 11 insect orders from which the Coleoptera, Hymenoptera, Diptera and Lepidoptera are the most diverse (over 99% of the total) (Kristensen 1999). The remaining seven orders include: caddisflies (order Trichoptera), lacewings (order Neuroptera), fleas (order Siphonaptera), twisted-winged parasites (order Strepsiptera), scorpionflies (order Mecoptera), and snakeflies (order Raphidioptera). There are good morphological characters to support the monophyly of most of these groups, and for more than a century, any newly described insect with complete metamorphosis could be easily assigned to one of these living lineages. What is not known, however, are the phylogenetic relationships among each of the 11 holometabolous insect orders (Whiting 2002a).

Many hypotheses have been presented for phylogenetic relationships among the holometabolous insect orders over the past century; these reflect the general difficulty of reconstructing the evolutionary history of this important insect group and the variety of opinions on the matter. Summaries of the current hypotheses are presented in figure 1.1.

Boudreaux (1979) and Hennig (1981) presented phylogenies based on different interpretations of morphological characters. Boudreaux placed Strepsiptera + Coleoptera as the most primitive holometabolan lineage and then argued for the placement of Hymenoptera at the base of the remaining orders. Hennig was uncertain as to the placement of Hymenoptera and Siphonaptera but argued for a sister group relationship between Strepsiptera and Coleoptera, and associated Trichoptera + Lepidoptera with Diptera + Mecoptera.

In his most recent summary, Kristensen (1999) divided the Holometabola into two main divisions: The Coleoptera + Neuropterida lineages (Neuroptera, Megaloptera, and Raphidioptera) form one division, and the remaining orders are placed in a second division (Hymenoptera + Mecopterida), with uncertainty as to the position of the Strepsiptera. Recently, Beutel and Gorb (2001) added a suite of morphological characters associated with the tarsi of insects and proposed a phylogeny that agrees with Kristensen (1999), but places the Strepsiptera as sister group to the Coleoptera.

Although a few attempts have been made from a molecular standpoint to decipher holometabolan phylogeny (Carmean et al. 1992, Chalwatzis et al. 1996, Pashley et al.



1993), Whiting et al. (1997) was the first formal analysis of morphological data in combination with extensive DNA sequence data. Their data consisted of 176 morphological characters coded across the Holometabola and outgroups, and portions of the 18S and 28S rDNA genes. Wheeler et al. (2001) expanded this study to include all hexapod orders and used a new analytical tool that obviates the need to generate a multiple alignment of DNA sequence data before phylogenetic reconstruction. Both studies largely concurred in their view of holometabolan phylogeny (Fig. 1.1). In contrast to morphology, these results suggested a sister-group relationship between Strepsiptera and Diptera, and as suggested by Kristensen (1999) demonstrated a close association of fleas with a family placed within the scorpionflies (Mecoptera). At the same time, their analyses indicated that many holometabolan interordinal relationships are not particularly well supported. Whiting (2002c) performed more extensive molecular analyses based on the entire *18S* rDNA gene for roughly three times more holometabolan species. Although this increased species sampling helped resolve some relationships (e.g. better support for Neuropterida), the general pattern of relationships provided by this single molecule is in some cases different than those found using morphology alone.

In summary, current DNA sequence data support the monophyly of most of the holometabolous insect orders, in agreement with morphology. DNA also supports the superordinal groups Amphiesmenoptera, Neuropterida and Halteria and the relationship between Mecoptera and Siphonaptera. DNA has not, however, been successful at confirming other relationships hypothesized by morphology, such as Mecopterida, Hymenoptera + Mecopterida, and Coleoptera + Neuropterida (Whiting et al. 1997, 2002b). The position of the Hymenoptera within the Holometabola appears to be especially

problematic and no characters support a firm placement of this group (Whiting 2002a).

Clearly, further work is needed to elucidate holometabolan relationships.

### ***1.3.2 Phylogenetic relationships within the Hymenoptera***

Königsmann (1978) and Rasnitsyn (1980, 1988) are two of the most influential works on the phylogenetics of the Hymenoptera, presenting the first phylogenetic hypotheses based on large morphological datasets (Whitfield 1998). However, the data used by Königsmann were based largely on the literature and were insufficient to resolve almost all apocritan relationships (Whitfield 1992). Rasnitsyn (1980) proposed an almost fully resolved hypothesis of higher-level relationships that included fossil taxa and substantial data from comparative morphology. In 1988, Rasnitsyn presented his slightly modified conclusions about hymenopteran relationships and these formed the framework for all subsequent investigations (Whitfield 1998).

The Hymenoptera are traditionally divided into the Symphyta (sawflies) and the Apocrita, the latter being further divided into the Parasitica (parasitic wasps) and the Aculeata (aculeate wasps) (Ronquist et al. 1999). There is a general consensus based on biology and the fossil record that the suborder Apocrita (which contains most of the parasitic wasps) has evolved from sawflies. This means that the traditional suborder Symphyta is a paraphyletic group (Quicke 1997). Although, there has been debate as to what group of sawflies is closest to the common ancestor of the Apocrita, a growing number of morphological studies have provided strong support for both a monophyletic Apocrita and for an Orussidae + Apocrita clade (Gibson 1985, Whitfield et al. 1989, Ronquist et al. 1999).

#### *1.3.2.1 Suborder Symphyta*

Comparative morphological work through the last decade has convincingly resolved many of the higher level groupings in the ‘Symphyta’ and their relationships to the Apocrita (Vilhelmsen 1997, Ronquist et al. 1999, Vilhelmsen 2000). Several major clades have bootstrap support exceeding 0.9 in phylogenetic analyses of morphological data, and although the analyses by Vilhelmsen (1997, 2000) and Ronquist et al. (1999) are based on different characters, the strict consensus trees from both analyses are perfectly congruent to each other (Fig. 1.2).

From these trees, well supported clades include the non-xyelid hymenopterans, core tenthredinoids (tenthredinoids excluding blasticotomids), the primitively wood- or stem-boring lineages (Cephidae, Anaxyelidae, Siricidae, Xiphydriidae, Orussidae, and Apocrita), and Orussidae + Apocrita.

#### *1.3.2.2 Suborder Apocrita*

Within the suborder Apocrita, Rasnitsyn (1988) suggested four major lineages; the Ichneumonomorpha, the Aculeata, the Proctotrupomorpha, and the Evaniomorpha (Fig. 3). However, a cladistic reanalysis of Rasnitsyn’s data (Ronquist et al. 1999) only supported the clades Ichneumonoidea and Aculeata.

Several studies have since generated different hypotheses concerning the phylogenetic relationships among apocritan groups, based on both morphological grounds (Gibson 1985, 1993, 1999, 2001, Whitfield 1992, Sharkey and Wahl 1992, Sharkey and Roy 2002, Masner 1993, Ronquist 1999, Quicke et al. 1992, 1994), or molecular data (Dowton and Austin 1994, 2001, Dowton et al. 1997). None of these provide a convincingly resolved tree based on measures of bootstrap support (Felsenstein 1985).

The most recent attempt to resolve the phylogeny of the Hymenoptera was the Dowton and Austin (2001) analysis, which surveyed 84 taxa and generated character information from three genes and morphology. Although this was a significant improvement over previous attempts, a range of relationships were still not stably recovered and were sensitive to the model of analysis. However, one group that was convincingly recovered was a monophyletic Proctotrupomorpha [as defined by Rasnitsyn (1988); including the superfamilies Proctotrupeoidea, Platygastroidea, Chalcidoidea, and Cynipoidea)]. Nevertheless, the relationships between each of the families of the Proctotrupomorpha were not resolved.

#### **1.4 PARASITISM IN THE HYMENOPTERA**

The Hymenoptera (sawflies, ants, bees, wasps) comprise an enormous number of species, more than all vertebrate species combined (Hammond 1992). Although in number of described species they are second only to the Coleoptera (Askew, 1971), it is actually the most species-rich in the most surveyed temperate regions (Whitfield 1998). Within the Hymenoptera, 75% of the species have been described as parasitoids (Whitfield 2003), and

have attracted considerable attention from biologists because of their importance in biological control (Strand and Obrycki 1996). Parasitic Hymenoptera also play an important role in maintaining the diversity of natural communities. LaSalle and Gauld (1993) emphasized their high position in food chains and have suggested that this makes them particularly liable to extinction, and also that they are likely to represent keystone species.

A parasitoid is “an organism which develops on or in another single host organism, extracts nourishment from it, and kills it as a direct or indirect result of that development” (Eggleton and Belshaw 1992). Parasitoid insects are dependent upon their hosts as larvae but are free-living as adults, and have also been referred to as protelean parasites by Askew (1971). Parasitoids use several strategies to alter the physiology, development and host immune responses of their host, mostly mediated by biochemical venoms or symbiotic viruses injected by the parent female at oviposition (Stolz and Whitfield 1992).

#### ***1.4.1 Parasitoid Natural History***

Godfray (1994) reviewed the general biology of hymenopteran parasitoids. They have highly specialized ovipositors, which are used both to manipulate eggs and to sting the host. The sting causes paralysis that may be permanent, or the host may recover and continue feeding. Parasitoids that attack concealed hosts often have long ovipositors which, when not in use, either extend beyond the end of the abdomen enclosed between protective valves, or are coiled inside the abdomen of the female. Cutting ridges at the end of the ovipositor allow wasps to drill through plant tissue and even wood to locate hidden hosts.

They possess special adaptations to allow the egg to pass down long and thin ovipositors; often the egg is very small and expands enormously within the host's body. In a number of hymenopteran parasitoids the whole abdomen is laterally or dorsoventrally compressed so that it can slide into narrow openings. In some cases, the adult female does not lay her eggs on the host but on the host's food plant. Parasitism occurs if the host eats the eggs. Within the Hymenoptera, this occurs in all members of the family Trigonalyidae. There are other parasitoids that lay their eggs away from the host but which have active free-living first instar larvae that are responsible for host location. This biology occurs in the two chalcidoid families Perilampidae and Eucharitidae, and the ichneumonoid subfamily Eucerotinae (Godfray 1994).

Parasitoids can be divided into two classes depending on the feeding behavior of the larvae. Some parasitoids develop within the body of the host, feeding from the inside, and are known as endoparasitoids. Ectoparasitoids, on the other hand, live externally, normally with their mouthparts buried in the body of their host (Godfray 1994). Apart from whether the parasitoid larva develops as an endo- or ectoparasitoid, another aspect of parasitoid biology involves whether the host insect continues to develop after it has been parasitized. Parasitoids whose hosts do not develop further are referred to as idiobionts and those whose hosts carry on their development post-parasitization are called koinobionts (Askew and Shaw 1986). Some idiobionts develop completely within a host egg or in a host pupa; others attack a mobile larval stage, but this is almost always paralysed permanently at the time of parasitization so that it can no longer move. Koinobionts typically attack larvae, often early instars, or eggs in the case of egg-larval and egg-pupal parasitoids (Quicke 1997)

Within the endo- ectoparasitoid and idio- koinobiont categories, there are different types of parasitoids displaying very different life-histories that provide interesting tests for evolutionary hypotheses. For instance, parasitoids that feed alone on a host are known as solitary parasitoids as opposed to gregarious parasitoids, where from two to several thousand individuals feed together on a single host. If further eggs are deposited on the host by the same species of parasitoid, superparasitism is said to occur. If a second female of a different species lays her eggs on the host, one of two things may happen: if the larvae of a second species compete with the resident larvae for host resources, multiparasitism occurs (Gauld and Bolton 1988). One of the most interesting examples of multiparasitism concerns the alder wood-wasp, *Xiphydria camelus*, and its two ichneumonid parasitoids, *Rhyssella curvipes* and *Pseudorhyssa alpestris*. *Rhyssella* lays its eggs upon the wood-wasp grub and it may be followed in this by *Pseudorhyssa*. The first instar larva of *Pseudorhyssa* kills that of *Rhyssella*. This relationship is obligatory for *Pseudorhyssa*, not because its larvae needs to feed upon a *Rhyssella* larva, but because it is unable to drill through wood with its own ovopositor to reach the wood-wasp larva (Askew 1971). However, if the larvae of the second species feed, not on the host, but on the parasitoid larvae already present, hyperparasitism occurs (Gauld and Bolton 1988). Hyperparasitism is generally of two kinds: facultative hyperparasitoids are able to attack unparasitized host individuals and only develop as hyperparasitoids when eggs are laid on a previously parasitized host; in contrast obligate hyperparasitoids are only able to develop as parasitoids of parasitoids (Godfray 1994). Hyperparasitism is quite frequent amongst the smaller parasitic Hymenoptera. For example, the larvae of the American chalcid *Perilampus hyalinus* penetrates the bodies of caterpillars (e.g. *Hyphantria*) and inside the host they search for a larva of the primary

parasite, *Ernestia* (Diptera: Tachinidae), which they in turn enter to complete their development (Askew 1971).

#### ***1.4.2 The Evolution of Parasitism***

The current phylogeny of the Hymenoptera suggests that parasitism arose once, in the common ancestor of the Orussoidea and Apocrita, whose wood-boring larva fed within tunnels of wood and at least partly upon fungi (Gibson 1985, Whitfield et al. 1989, Ronquist et al. 1999). The ancestral form of parasitism almost certainly was external (i.e. ectoparasitic); many of the basal clades within the Orussoidea, Stephanoidea, Ichneumonoidea, and Aculeata are ectoparasitoids of concealed hosts, often within galleries. This is also true for Megalyroidea and some members of the Evanioidea. Similarly, some members of the Chalcidoidea and Ceraphronoidea are also ectoparasitoids. However, some basal Cynipoidea attack hosts within galleries in wood, but are endoparasitic. In the Chalcidoidea, Evanioidea, Ceraphronoidea, and Cynipoidea, the ancestral form of parasitism is obscure, and although it seems more likely that the common ancestors of the “Proctotrupomorpha” (Proctotrupeoidea, Platygastroidea, Chalcidoidea, and Cynipoidea) and “Evaniomorpha” sensu Rasnitsyn (Evanioidea, Trigonalioidea, Megalyroidea, and Ceraphronoidea) lineages would have been ectoparasitic, there is no evidence for this conclusion (Whitfield 1992, Whitfield 1998). Instead, phylogenetic studies suggest that in some groups, particularly the Proctotrupomorpha, the basal lineages are generally endoparasitic (Dowton and Austin 2001).



In order to trace the evolutionary history of the forms of parasitism within the Hymenoptera, it is extremely important to perform comparative morphological and molecular research into the higher-level phylogenetic relationships for the suborder Apocrita, which contains all the parasitic Hymenoptera except for the primitive Orussoidea. Available phylogenetic analyses, seem to agree in that the earliest form of parasitism (ectoparasitism of wood-boring insects) has its origin in the common ancestor of Orussoidea+Apocrita (e.g. Rasnitsyn 1988, Dowton and Austin 2001). However, transitions from ecto- to endoparasitism within the Apocrita seem less clear, primarily because a robust apocritan phylogeny has yet to be recovered.

## 1.5 AIMS

***Aim 1.** To test the usefulness of concatenated complete mt gene sequences in resolving the phylogenetic position of the Hymenoptera within the holometabolan insects (Chapter 2).*

So far, attempts to resolve phylogenetic relationships among the Holometabola using complete mt genomes have had to exclude the honeybee (*Apis mellifera*) sequence. The honeybee appears to be problematic due to an increased rate of nucleotide substitution and compositional bias. In order to investigate the position of the Hymenoptera within the Holometabola, I will assess the impact of two analytical approaches, (1) sequence a hymenopteran mt genome suspected to have less compositional bias and slower rates of molecular evolution than the honeybee, and (2) use a range of analytical models for phylogenetic reconstruction. For this purpose, the complete mt genome of the sawfly *Perga condei* (Hymenoptera: Symphyta) will be sequenced.

*Aim 2. To improve the resolution of phylogenetic relationships within the Apocrita (Chapter 3).*

A well-supported phylogeny of the Apocrita will allow us to trace the evolution of parasitism as well as the evolution of the mt genome within the group. In the long term, this phylogeny will be used to test the usefulness of both concatenated mt genes as well as genome rearrangements in resolving phylogenetic relationships with confidence. The last attempt to recover phylogenetic relationships within the Apocrita was performed by Dowton and Austin (2001), but a range of relationships remain poorly supported. In an attempt to improve the resolution of apocritan relationships, I will analyze the Dowton and Austin (2001) dataset using a Bayesian phylogenetic approach. Part of the *18S* gene will be added to their dataset.

*Aim 3. To gain a more accurate estimate of the rate and nature of mt genome organization in apocritan wasps (Chapter 4).*

Current estimates of the rate and nature of mt genome reorganization are grossly underestimated in the Apocrita. This is likely due to the long divergence times between sampled taxa. Just two apocritan mt genomes have been entirely sequenced, and these are from relatively closely related taxa (*Apis* and *Melipona*, both from the family Apidae). Dowton et al. (1999, 2001) estimated the organization of a small number of tRNA genes from around 40 taxa, but their strategy was unlikely to detect protein-coding and ribosomal gene rearrangements. I will further characterize the degree of rearrangement and the evolution of the mt genome structure among the Apocrita by sequencing the complete mt genome of two other apocritan taxa, *Vanhornia eucnemidarum* and *Primeuchroeus* sp.

***Aim 4.** To explore the potential of mt genomes in resolving phylogenetic relationships within the Apocrita (Chapter 5).*

Since resolving relationships among the Apocrita has been difficult (Whitfield 2002), complete mt genomes might be useful candidates to improve the resolution of phylogenetic relationships of the group. However, sequencing complete genomes for this purpose, in such a large and diverse group, is a major undertaking. The phylogenetic position of the taxa whose mt genomes I will sequence under aim 3 is well established. I will use these genomes, together with those already available (*Apis* and *Melipona*) to explore the potential of concatenated mt genes in resolving phylogenetic relationships among the Apocrita. In doing this, I will also evaluate the effect of different variables such as DNA vs. amino acid sequence data or different models of phylogenetic reconstruction, in recovering the correct tree topology.

## **1.6 THESIS FORMAT**

Each chapter of this thesis has been written as a journal article, and the general format has been left this way except for a few modifications to make the format of all chapters the same. The title page of each chapter states where the chapter was submitted or accepted for publication. As a result there is some repetition between the general introduction and the introduction of each chapter. There is also some repetition in the methodology in the various chapters. The references from each journal article have been compiled together to form one reference chapter at the end of the thesis.

**Figure 1.1**

Previous hypotheses for holometabolan relationships. A. Boudreaux (1979), B. Hennig (1981), and C. Kristensen (1999) are based on morphological data. D. Whiting et al. (1997) is based on a combination of morphological and molecular data, and E. Whiting (2001) is the strict consensus tree from the 18S rDNA and improved taxonomic sampling.

**Figure 1.2**

Summary of what is known about symphytan relationships (Ronquist 1999).

**Figure 1.3**

Major apocritan relationships proposed by Rasnitsyn (1988). The Evaniomorpha comprise the Evanioidea, Ceraphronoidea, Megalyridae, Trigonalyidae and Stephanidae. The Proctotrupomorpha include the Proctotrupoidea, Platygastroidea, Chalcidoidea, and Cynipoidea.

## **Chapter 2 . The position of the Hymenoptera within the Holometabola as inferred from the mitochondrial genome of *Perga condei* (Hymenoptera:Symphyta:Pergidae)**

This chapter was slightly modified from a paper that was published in 2005 by Molecular Phylogenetics and Evolution.

Castro LR, 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, 469-479

### **2.1 Introduction**

The reconstruction of the phylogeny of insect orders has been a focus of several studies for more than a century (see discussion in Hennig 1981, Kristensen 1981, 1991, 1997, Whiting 2002a, Whiting et al. 1997). Despite this, controversies at all levels still exist, including evolutionary relationships among primitive insect groups and their allies, and the phylogenetic relationships of the most advanced insects: the Holometabola (Caterino et al. 2000, Whiting 2002a).

The Holometabola is the most diverse and successful group of terrestrial organisms, comprising 11 insect orders from which the Coleoptera, Hymenoptera, Diptera and Lepidoptera are the most diverse (over 99% of the total) (Kristensen 1999). Within the Holometabola, the Amphiesmenoptera (Lepidoptera + Trichoptera) is the only well-

established interordinal relationship (Hennig 1981, Kristensen 1997, Whiting et al. 1997); other relationships are not always recovered or not well supported. One of the most important examples is the highly controversial position of the Strepsiptera, which has been placed as sister group to the Coleoptera (Bourdreaux 1979, Hennig 1981) or the Diptera (Whiting 2002a, Whiting et al. 1997). The position of the Hymenoptera within the Holometabola appears to be especially unresolved. No characters support a firm placement of this group, which has been postulated as sister group to 'Meronida' (Mecopterida + Neuropterida) (Bourdreaux 1979) or to Mecopterida (Kristensen 1991, Whiting et al. 1997).

Of the studies addressing relationships among the major lineages of insects, Whiting et al. (1997), and Whiting (2002a) are the best approximations. They examined up to 147 taxa using *18S* and *28S* rDNA genes. From their study, with the exception of the Mecoptera and the Coleoptera, each holometabolous insect order is well supported as a monophyletic group, but relationships among most orders are still controversial. The greatest problem in their study appears to be the reliance on a single marker for phylogenetic inference (Whiting 2002a, 2002b).

In the past decade, mitochondrial DNA (mtDNA) analysis has become established as a powerful tool for evolutionary studies of animals (Boore and Brown, 1998). Further, several analyses have demonstrated recently that complete mitochondrial genomes provide higher levels of support than those based on individual or partial mt genes (Cummings et al. 1995, Russo et al. 1996, Zardoya and Meyer 1996).

Nevertheless, with respect to the insects, phylogenetic analyses employing entire mitochondrial DNA sequences have presented several inconsistencies, and have failed to generate a sufficiently robust phylogeny to make confident statements about insect evolution (Black and Roehrdanz 1998, Flook et al. 1995b, Friedrich and Muqim 2003, Nardi et al. 2001, 2003, Stewart and Beckenbach 2003, Wilson et al. 2000). The placement of the Hymenoptera has been especially problematic and has emphasized the difficulties of reconstructing phylogenies that include lineages with variable substitution rates and base compositional biases, as is evident in the honeybee mitochondrial genome (Flook et al. 1995b, Foster and Hickey 1999). Probably due to the highly divergent *Apis mellifera* mitochondrial DNA sequence, the Holometabola were not recovered as a monophyletic group in Flook et al. (1995b). Of even more concern, other studies have also had to exclude the honeybee sequence in order to recover a monophyletic clade for the Insecta and to avoid phylogenetic inconsistencies due to the significant amino acid substitution bias of this group (Black and Roehrdanz 1998, Friedrich and Muqim 2003, Nardi et al. 2001, 2003, Stewart and Beckenbach 2003, Wilson et al. 2000).

Some methods of phylogenetic analysis are more able to accommodate variation in substitution rates and compositional bias; it has been shown that under relatively simple or inadequate models of evolution, phylogenetic inference methods can actively mislead attempts to estimate evolutionary trees from molecular sequences, and that modelled analyses are more able to recover phylogeny when compositional heterogeneity or divergent rates are present (Swofford et al. 2001). Other studies have shown, that although usually associated with parsimony methods, long-branch attraction problems can also appear in maximum likelihood and distance analyses when the assumed substitution models



of these methods are strongly violated (Huelsenbeck 1995, Huelsenbeck and Hillis 1993, Sullivan and Swofford 1997).

It is highly desirable to resolve the placement of the Hymenoptera in the mtDNA phylogenetic tree of the insects, and especially to resolve its position among the Holometabola. In order to find the correct position of this group within the Holometabola, two approximations seem to be essential:

1. It would be appropriate to include in the analysis the sequence of a more basally derived hymenopteran than the honeybee. Increasing the number of taxa representing each insect order should generate a more accurate phylogeny (Lin et al. 2002). If there is a long-branch attraction problem (Hendy and Penny 1989) in relation to the honeybee, then, the addition of another less divergent hymenopteran mitochondrial genome has an improved chance of resolving the honeybee position. In this study, we sequenced the mitochondrial genome of *Perga condei* (Symphyta: Hymenoptera). Partial mt sequences of *Perga condei* display lower rates of molecular evolution and compositional bias compared with those of *Apis*, which belongs to the suborder Apocrita (Dowton and Austin 1995, 1997). Therefore, we expect *Perga* should be less prone to long-branch attraction.

2. The evolutionary model of phylogeny reconstruction also seems to affect the results considerably. When the substitution rates are considerably variable among different lineages, the use of a model that reflects the actual substitution probability is important for obtaining a correct tree topology (Lockhart et al. 1994, Yang 1996). Despite this problem, a thorough testing of different models of sequence evolution has not been done when

reconstructing the phylogeny of the insects using mt genomes. The honeybee misplacement represents a good model system for assessing the utility of different approaches to phylogenetic reconstruction.

The purpose of this research was to assess the impact of adopting both of these approaches (sequencing a hymenopteran suspected to have less compositional bias and slower rates of molecular evolution than the honeybee, and testing for different models of molecular evolution) on the reconstruction of the phylogeny of the Holometabola and to resolve the correct position of the Hymenoptera within this group.

## **2.2 Materials and Methods**

### ***2.2.1 Insects and DNA Extraction***

The spitting grub *Perga condei* (Hymenoptera: Symphyta: Pergidae) was collected by Andy Austin (University of Adelaide) into 100% ethanol and stored at 4°C before extraction.

Because this wasp is very big, flight muscle was sufficient for DNA extractions. Ethanol was removed by washing three times (30 min each) in 10 mM Tris-HCl (pH 8) containing 100 mM NaCl and 1 mM MgCl<sub>2</sub>. Tissue was homogenized in 400 µl of 10 mM Tris-HCl (pH 8), 10 mM EDTA, and 1% SDS containing 100 µg of proteinase K (Boehringer Mannheim) and incubated overnight at room temperature. DNA was separated from salt-insoluble material by the method of Sunnucks and Hales (1996). DNA was redissolved in 100 µl of sterile water and stored at 4°C. This DNA solution was used directly in the PCR reaction.

### ***2.2.2 PCR amplification and cloning***

A range of universal mitochondrial primers were tested and primers that generated the best amplifications were used for long PCRs. New primers were designed if necessary in order to generate perfectly matched primers. A total of 10 PCR fragments (500bp-3000bp) were the initial template for sequencing reactions or cloning (Table 2.1) and new primers were designed as sequence data accumulated. A 2 kb fragment from the same specimen was sequenced elsewhere (Dowton et al. 2003) and was obtained from GenBank.

PCR reactions were performed in a total volume of 20 µl. For the generation of short PCR fragments (<800 pb), reactions contained 10 mM Tris-HCl (pH 9), 50 mM KCl, 0.1% Triton X-100, 1.25–6.25 mM MgCl<sub>2</sub>, 0.4 µM of each primer, 25 µM of each dNTP, 0.75 U *Taq* DNA polymerase (PROMEGA) per reaction, and 0.5 µl of DNA extract. A negative control PCR tube was prepared with the same constituents but lacking DNA.

Amplifications were performed in a Hybaid Sprint PCR thermocycler or a Corbett Research thermocycler using the following program: 5 cycles (30 s at 94°C, 30 s at 45–55°C, and 1 min at 72°C), followed by 30 cycles (30 s at 94°C, 30 s at 55°C, and 1 min at 72°C). In addition, a 5-min extension at 72°C was added at the end of the 35 cycles in order to finish any incomplete amplification. PCR optimisation for each template involved the variation of possible primer combinations, MgCl<sub>2</sub> concentration, and annealing temperature. Longer amplifications (>800 bp) were performed as described above but using 0.75 U *Taq* and 2.5 mU *Pfu* DNA polymerase (*Pyrococcus furiosus*) and 100 µM of each dNTP per reaction. For these amplifications, we used an Eppendorf Gradient PCR cycler with a long PCR

program: an initial denaturation at 92°C for 2 min, followed by 35 cycles (denaturation at 92°C for 10 s, annealing at 45–65°C for 30 s, and extension at 68°C for 1-5 min).

In order to remove unincorporated primers and dNTPs before sequencing, double-stranded PCR products were purified using the PCR purification kit QIAquick (QIAGEN) or PEG (polyethylene glycol) precipitation (Maniatis et al., 1989) with some modifications (0.6 volumes of 30% PEG in 1.5 M NaCl was added to each PCR reaction). Cycle sequencing reactions were performed with the ABI Prism Dye Terminator cycle sequencing kit v.2 or v.3 (Perkin-Elmer) with *AmpliTaq* FS. Both strands of the PCR product were sequenced. Primer sequences were removed from the start and the end of the obtained sequence and sequence ambiguities were resolved by comparing the electropherograms using the program BioEdit v. 5.0.9 (Hall 1999).

PCR products that were difficult to sequence directly were cloned. For cloning we used the pGEM-T easy vector system from PROMEGA exactly as recommended by the manufacturer. Plasmids were extracted using the Wizard Plus SV Minipreps DNA purification system from PROMEGA and sequenced using both the original PCR primers and the M13-F and M13-R primers that anneal to the vector.

### ***2.2.3 Identification of trn genes***

Exact location of *trn* genes were found at the boundaries of the genes. The generated sequences were submitted for *trn* gene search using the program tRNA-Scan SE [v.1.1, <http://genome.wustl.edu/eddy/tRNAscan-SE>; Lowe and Eddy (1997)]. The parameters

for the tRNA scan program were set for mitochondrial-chloroplast DNA as the source and using the invertebrate mitochondrial genetic code. When long tracts of non coding sequence were apparent, the cove cut off score was reduced to 10 and the search repeated.

#### ***2.2.4 Alignments and phylogenetic analysis***

The edited sequences of *Perga condei* were aligned with other insect mt sequences using CLUSTAL V (Thompson et al. 1994), as distributed with the BIOEDIT program (Hall, 1999). Unequivocally alignable positions were imported into MEGA version 2.1 (Kumar et al. 2001) for calculating the nucleotide composition and amino acid composition.

The complete mtDNA coding sequences of other insects were obtained from GenBank. In order to avoid oversampling of some clades of Diptera or Lepidoptera we only included one representative of each of the available genera (Table 2.2). *Locusta migratoria* was used as the outgroup. *Triatoma dimidiata* (non-Holometabola) was also included in the ingroup in order to assess whether a particular analyses recovered a monophyletic Holometabola.

Individual protein coding genes were translated using the *Drosophila* mtDNA genetic code using the program Translation Machine ([www.2ebi.ac.uk/translate/](http://www.2ebi.ac.uk/translate/)). For each of the protein coding genes, amino acid sequences were initially aligned using CLUSTAL W (Thompson et al. 1994). The alignments were manually adjusted to avoid any ambiguous amino acid pairings. Nucleic acid sequences were aligned using PROTAL2DNA program (Schuerer, K. and Letondal, C. <http://bioweb.pasteur.fr/seqanal/interfaces/protal2dna.html>), which aligns DNA sequences based on the protein alignments. rRNA nucleotide sequences were also aligned using CLUSTAL W.

The individual alignments were combined to generate a single alignment. The genes were arranged in the order they occur on the majority strand of the *D. yakuba* sequence.

All phylogenetic analyses were conducted in PAUP\* version 4.0b10 (Swofford 1998) or MrBayes version 3.0b4 (Huelsenbeck and Ronquist 2001). Maximum parsimony (MP) (Fitch 1971) and Bayesian approaches were used for phylogenetic analysis using both amino acid and/or nucleotide sequences.

Maximum parsimony analysis was used with all characters weighted equally, and gaps treated as missing data. Non-parametric bootstrapping was performed using a full heuristic search with 100 replicates.

Bayesian analyses were conducted using the mtREV24 model of protein evolution (Adachi and Hasegawa 1996) when using amino acid sequences. The general time reversible model with some sites assumed to be invariable and with variable sites assumed to follow a discrete gamma distribution [GTR + I +  $\Gamma$ ; Yang (1994)] was selected as best-fit model of nucleotide substitution (ModelTest version 3.06, Posada and Crandall 1998). In these cases, we set the maximum likelihood parameters in MrBayes as follows: “lset nst=6” (GTR), “rates=invgamma” (I +  $\Gamma$ ), “covarion=yes” [to allow for rate heterogeneity between sites (Penny et al. 2001)]. The Markov chain Monte Carlo process was set so that four chains (three heated and one cold) ran simultaneously. We conducted runs for 100,000 generations, with trees being sampled every 100 generations, each of which started from a random tree. Independent analyses indicated that “stationarity” (or “burnin”: lack of improvement in ML scores) was reached at no later than 30,000 generations; thus, the first

300 trees were discarded from each analysis as the burnin, and the remaining trees were used to generate a 50% majority consensus tree, with the percentage of samples recovering any particular clade representing that clades' posterior probability (Huelsenbeck and Ronquist 2001).

## **2.3 Results**

### ***2.3.1 Genome structure and organization***

12 protein coding genes, 16 *trn* genes and the 2 small and large *rRNA* genes were sequenced for *Perga condei* (GenBank AY787816). The region comprising the *nd2* gene, 6 *trn* and the control region was not successfully amplified. The *nd2* gene is one of the most poorly conserved of the protein coding genes (Simon et al. 1994) and the absence of this gene in the analysis is not expected to change the final results. For the purpose of reconstructing the phylogeny of the Holometabola using mt genomes, the region sequenced gave us a total of 13416 bp. Except for the *trnL*<sup>CUN</sup> which was not found in the position considered ancestral to insects and crustaceans (Boore et al. 1998, Flook et al. 1995b), the arrangement of these genes matches perfectly with that of a number of distantly related insect species such as the fruitfly, *Drosophila melanogaster* (Lewis et al. 1994) and the true bug, *Triatoma dimidiata* (Dotson and Beard 2001).

The identification of ORFs in animal mtDNA, especially insects, is not conclusively established, and evidence on unusual initiation and termination codons is increasing as more mitochondrial genomes are described (Boore and Brown 2000; Campbell and Barker 1999, Spanos et al. 2000). Among dipterans, incomplete termination codons or even

absence of recognizable stop codons have been reported for a variety of protein-coding genes (Beard et al. 1993, Clary and Wolstenholme 1985a, Mitchell et al. 1993, Spanos et al. 2000). In our case, conventional stop codons could be assigned to most of the protein sequences. The *nd4* and *atp6* genes terminated with an incomplete (TA) codon. Identical situations have been described for other insect species where it has been proposed that the complete termination codon is created by polyadenylation as observed in other animal phyla (Nardi et al. 2001, Ojala et al. 1981).

Conventional ATA or ATG start codons could be assigned to five of the 12 protein coding genes. The other protein coding genes use ATC or ATT (codes for Ile) as start codon, which has also been reported for other species (Friedrich and Muqim 2003, Lessinger et al. 2000). Although the initiation of insect *coxI* genes does not perfectly agree with the invertebrate mt code (Lessinger et al. 2000), in our case an ATT start codon was identified for the *coxI* gene.

The regions coding for the small and large ribosomal subunit genes are 712 and 1359 bp long, respectively. Alignments of these regions with related insect sequences revealed numerous blocks of high sequence conservation, suggesting that the secondary structure elements were conserved (not shown). All *trn* gene sequences could be folded into cloverleaf secondary structures and are predicted to have secondary structures typical of previously published mitochondrial *trn* genes (data not shown). The *trnS*<sup>UCN</sup> gene was not identified by software, but was located due to its conserved relative genome position and sequence similarity.



Several non-coding and overlapping regions were identified in the mt genome of *Perga condei*. The largest non-coding region was 19 bp long (the AT rich region, which is generally the largest non-coding region, was not sequenced), and was located between the *cox2* and the *trnK* genes. Other non-coding regions of similar sizes (13, 12, 11bp) were located between the genes *nd5* and *trnH*, *cox3* and *trnG*, and *nd3* and *trnA* respectively. There are only two cases where genes overlap. *trnR* and *trnN* overlapped by 3 bp and there is an overlap of 6 bp between the *atp8* and *atp6* genes. Overlaps between these two coding genes are a common feature of metazoan mtDNA (Campbell and Barker 1999). In arthropods, these genes also overlap in *Locusta*, *Ceratis*, *Drosophila*, *Anopheles*, *Apis*, *Boophilus*, *Tribolium* and *Cochliomyia* (Campbell and Barker 1999, Clary and Wolstenholme 1985b, Crozier and Crozier 1993, Flook et al. 1995a, Lessinger et al. 2000, Mitchell et al. 1993, Spanos et al. 2000, Friedrich and Muqim 2003).

### ***2.3.2 Base composition, amino acid composition and codon usage***

As with previously published insect mitochondrial genomes, the *Perga condei* mt genome is A+T rich, with an average AT content of 78% (Table 2.3). This corresponds well with the AT bias ranging from 69.5% in *T. dimidiata* to up to 84.9% in *Apis mellifera* (Crozier and Crozier 1993, Dotson and Beard 2001). As expected from Dowton and Austin (1993), the total AT bias of *P. condei* genome is not as high as that of the honeybee. Third codon position nucleotides showed the highest A+T content (87.3%), while first and second codon position nucleotides had A+T content values less than the genome (72 and 69.7% respectively). These data agree with the suggestion that the AT bias is introduced by

mutational pressure, as has been found for other mitochondrial genomes (Foster et al. 1997).

The codon usage of the *P. condei* mitochondrial genome is shown in Table 2.4. A relationship between the base composition of codon families and amino acid occurrence was observed. This can be assessed by calculating the number G+C rich codons (Pro, Ala, Arg and Gly) and A+T rich codons (Phe, Ile, Met, Tyr, Asn and Lys) and then calculating their ratio (Crozier and Crozier 1993). In *P. condei*, this value was 0.30, which is lower than that found in dipteran insects (0.42-0.43) (Lessinger et al. 2000), but higher than the honeybee mitochondrial genome (0.18) (Crozier and Crozier 1993). In general, there is a significant correlation between codon usage and the nucleotide composition of *P. condei* mt genome, as is seen in other insect mt genomes (Foster et al. 1997).

### **2.3.3 Phylogenetic analysis**

Concatenated protein coding genes and ribosomal *12S* and *16S* genes produced a matrix of 12692 nucleotide characters used for parsimony and Bayesian analysis. The amino acid matrix consisted of 3587 characters. *Heterodoxus* and *Thrips* sequences were excluded from all analyses because of evidence of extreme A+T bias (Foster and Hickey 1999).

Preliminary analyses (not shown) always showed that, probably because of shared compositional bias and long branch attraction, these sequences were recovered as the sister to the *Apis mellifera* sequence independent of the type of analysis performed (data not shown). In the present study, the Hymenoptera were represented by three taxa; *Apis mellifera*, *Perga condei* and *Melipona bicolor*. Although the *Melipona* sequence has been

available for some time, no previous mitogenomic analyses have included it. *Melipona* is a relatively close relative of *Apis* (both belong to the family Apidae), and displays similarly accelerated rates of molecular evolution and composition bias (not shown). We performed various phylogenetic analyses with (i) all three taxa, (ii) just *Perga*, or (iii) just *Apis* in order to test the effect of the *Perga* sequence on the analysis.

Results from amino acid sequence analyses are shown in Fig. 2.1. In both parsimony and Bayesian analyses, the Hymenoptera is generally recovered as sister group to the Coleoptera. In some cases, the hemipteran (*T. dimidiata*) is obviously misplaced among the Holometabola (Fig. 2.1A). Bayesian and parsimony analyses with just *Apis* sequence recovered the same relationship (Fig. 2.1B). In Bayesian analysis with just *Perga* included, the Hymenoptera groups again with the Coleoptera (Fig. 2.1C), however, in parsimony analysis *Perga condei* is recovered as the sister taxon to *T. dimidiata* (Fig. 2.1D).

Phylogenetic trees based on nucleotide sequences are shown in Fig. 2.2. Different analyses were run in searching for an appropriate evolutionary model to analyze this type of data. In initial analyses, concatenated protein coding genes and ribosomal genes were treated uniformly, with all character changes weighted equally. Parsimony analysis recovered the hymenopterian clade (*Apis*, *Melipona* and *Perga* included) as sister group to the Lepidoptera (Fig. 2.2A). In Bayesian analyses, when all the hymenopterian representatives are included (*Perga*, *Apis* and *Melipona*), the Hymenoptera is recovered as sister group to the Mecopterida (Diptera-Lepidoptera) clade and the Coleoptera is recovered at the base of the tree (Fig. 2.2B). Parsimony analysis using *Apis* alone recovered the Hymenoptera as sister group to the Coleoptera (Fig. 2.2C). However, Bayesian analyses excluding *Perga* and *Melipona* sequences showed that *Apis mellifera* sequence is recovered as sister group to the

Lepidoptera (Fig. 2.2D). In both parsimony and Bayesian analyses excluding *Apis* and *Melipona* from the analysis, *Perga* places at the base of the Holometabola (Fig. 2.2E).

To avoid the possible saturation effect of third codon positions (Fitch, 1986), we performed parsimony analyses based on the first and second codon positions alone. However, the trees recovered were identical in topology to the parsimony trees obtained when all positions were included (data not shown). Bayesian analyses were run in which 1<sup>st</sup>, 2<sup>nd</sup>, and 3<sup>rd</sup> codon positions, and ribosomal (*12S* and *16S*) genes were treated as separate partitions. In these cases, MrBayes searched for the appropriate evolutionary model for each partition. Treating *12S* and *16S* as separate partitions did not make any difference (data not shown) so both *rns* and the *rnl* were treated as ‘ribosomal’ in a single partition. In these cases, when just *Apis* is included in the analysis, the Hymenoptera places again as sister group to the Lepidoptera (as shown in figure 2.2D). However, when *Perga*, *Apis* and *Melipona* are included, the hymenopteran clade is recovered as sister group to the Mecopterida (Diptera-Lepidoptera) and Coleoptera is at the base of the tree (as shown in Fig. 2.2B). *Perga* sequence alone recovered the same relationship (Fig. 2.2F).

## **2.4 Discussion**

### **2.4.1 *Perga condei* mitochondrial genome structure**

We have determined the first nearly complete mitochondrial genome sequence from a representative of the hymenopteran suborder Symphyta, the spitfire grub *Perga condei*.

Dowton and Austin (1995) had suggested that the rates of molecular evolution and

compositional bias within the Symphyta are lower than within the Apocrita. Our results confirm their findings and show that the *Perga* mitochondrial genome is actually similar to other conservative evolving insect mitochondrial genomes in terms of gene arrangement, base composition and amino acid composition. The A+T content of the mt genome of *Perga condei* was found to be within ranges observed for insect mitochondria and is lower than the total A+T content of the *Apis mellifera* mitochondrial genome. The same pattern is seen when the codon usage ratio is calculated. The greater base composition bias in *Apis* than in *Perga* seems to indicate that this bias affects amino acid use in the bee to a greater extent than in the sawfly. Based on these features, since apocritan sequences alone behave inadequately in phylogenetic analysis, the mitochondrial genomes of symphytan representatives seem more able to clarify the position of the Hymenoptera within the Holometabola.

We were able to identify a putative ATT start codon for the *coxI* gene of *Perga condei*, which is an unusual observation in insect mitochondria. Of the complete mitochondrial genomes available, only *Apis*, *Heterodoxus*, and *Triatoma* have been found to code for a putative inframe start codon for this gene (Crozier and Crozier 1993, Dotson and Beard 2001, Shao et al. 2001). The mitochondrial genome sequenced for *Perga condei* is generally compact. Although 13 non-coding regions were identified, these do not exceed more than 19 nucleotides and are not conserved in other mitochondrial genome sequences, implying absence of functional significance.

#### 2.4.2 Phylogenetic analysis

One of the first studies that used mitochondrial genomes to resolve the phylogeny of the insects was Flook et al. (1995b). However, in their study, the Hymenoptera were misplaced as sister group to the Orthoptera (*Locusta migratoria*), disrupting the monophyly of the Holometabola. Foster et al. (1997) and Foster and Hickey (1999) suggested that a combination of compositional bias and possibly a ‘long branch attraction problem’ was the reason for the misplacement of the honeybee in phylogenetic analyses. Thus, given its unexpected positioning with mitochondrial data, and its higher rates of molecular evolution, several subsequent studies (for example Black and Roehrdanz 1998, Friedrich and Muqim 2003, Nardi et al. 2001, 2003, Stewart and Beckenbach 2003, Wilson et al. 2000) have omitted the honeybee from their phylogenetic analyses using mitochondrial genomes, leaving the position of the Hymenoptera within the Holometabola unclear.

The mt genome of *Melipona bicolor* became available in 2003, but being closely related to *Apis mellifera* (both from the family Apidae) and sharing similar molecular features (high rates of molecular evolution and compositional bias), this genome was likely to be as problematic as the honeybee for phylogenetic reconstruction. Alternatively, our characterization of *Perga condei* mitochondrial genome suggests it has evolutionary properties more consistent with other insect mitochondrial genomes, and thus, it seems less prone to compositional bias or ‘long branch attraction’ problems.

However, amino acid analyses utilizing the *Perga* genome did not recover relationships consistent with established insect phylogeny; the Holometabola were not monophyletic in some of the analyses, while the Mecoptera were not recovered in others. The

Hymenoptera tended to fall out as sister group to the Coleoptera, possibly because these two have the longest branches. It has been suggested that phylogenetic analyses of amino acid sequences may be problematic (Delsuc et al. 2003), the currently available models of mitochondrial amino acid substitution are based on empirically deduced matrices from mammalian sequence databases and may not be appropriate when analyzing insect datasets (Adachi and Hasegawa 1996).

Due to the failure of previous analyses to resolve a range of insect relationships, we expected that the position of the Hymenoptera would only be resolved when appropriate models of analysis were used in conjunction with expanded taxonomic sampling. Although an enormous range of models can be applied to such a dataset (e.g. distinct models for each codon position within each gene would yield 36 partitions, each with its own model), it has been shown that there are distinct problems with the unnecessary over-parameterization of Bayesian analyses (Rannala 2002). Thus, we considered that, due to the differences in selective constraints, a minimum of four partitions required distinct models in this analysis: each of the first, second and third codon positions of the protein coding genes, and the ribosomal genes were treated as separate partitions.

Consistent with this expectation, we found that simple models and/or inadequate taxon sampling (maximum parsimony and bayesian analyses where all positions were treated uniformly, and analyses including just *Apis*) did not recover relationships consistent with established insect phylogeny. The Hymenoptera either disrupted the monophyly of the Coleoptera (as seen in Fig. 2.2C) or the Mecopterida (Diptera + Lepidoptera) (as in Fig. 2.2A and 2.2D).

When more realistic models of analysis were employed, Bayesian analysis recovered the Hymenoptera as sister group to the Mecopterida (Fig. 2.2B and 2.2F). This was the case whether *Perga* alone, or all three hymenopteran taxa were included, but not when only *Apis* was included.

Previous morphological and molecular studies (Kristensen 1991, 1999, Whiting et al. 1997) have postulated a placement of the Hymenoptera as sister group to the Mecopterida, although not with substantial support. Our analyses tend to most consistently support this relationship. The only plausible alternative placement is at the base of the Holometabola (as obtained in Fig 2.2E), however, this grouping presented a lower likelihood (-110397) compared to the tree in which all the taxa are included, and more appropriate models of molecular evolution are specified (-100155).

In general, our study indicates that with appropriate models and appropriate taxon sampling, complete mitochondrial genome sequences adequately resolve relationships among the insect orders. However, the mitochondrial genome of groups such as Mecoptera, Trichoptera, Neuroptera or Strepsiptera would need to be included in order to make confident conclusions about the position of the Hymenoptera within the Holometabola and about relationships between the other holometabolan orders.

## **2.5 Conclusion**

In this chapter I sequenced most of the mitochondrial genome of the sawfly *Perga condei* (Insecta: Hymenoptera: Symphyta: Pergidae) and tested different models of phylogenetic reconstruction in order to resolve the position of the Hymenoptera within the



Holometabola, using mitochondrial genomes. The mitochondrial genome sequenced for *Perga condei* had less compositional bias and slower rates of molecular evolution than the honeybee, as well as a less rearranged genome organization. Phylogenetic analyses showed that, when using mitochondrial genomes, both adequate taxon sampling and more realistic models of analysis are necessary to resolve relationships among insect orders. Both parsimony and Bayesian analyses performed better when nucleotide instead of amino acid sequences were used. In particular, this study supports the placement of the Hymenoptera as sister group to the Mecoptera.

**Table 2.1**

Primer pairs used in PCR amplification of the *Perga condei* mitochondrial genome

**Table 2.2**

Information of the taxa from which the whole mitochondrial genome have been sequenced, and were used in the phylogenetic analyses.

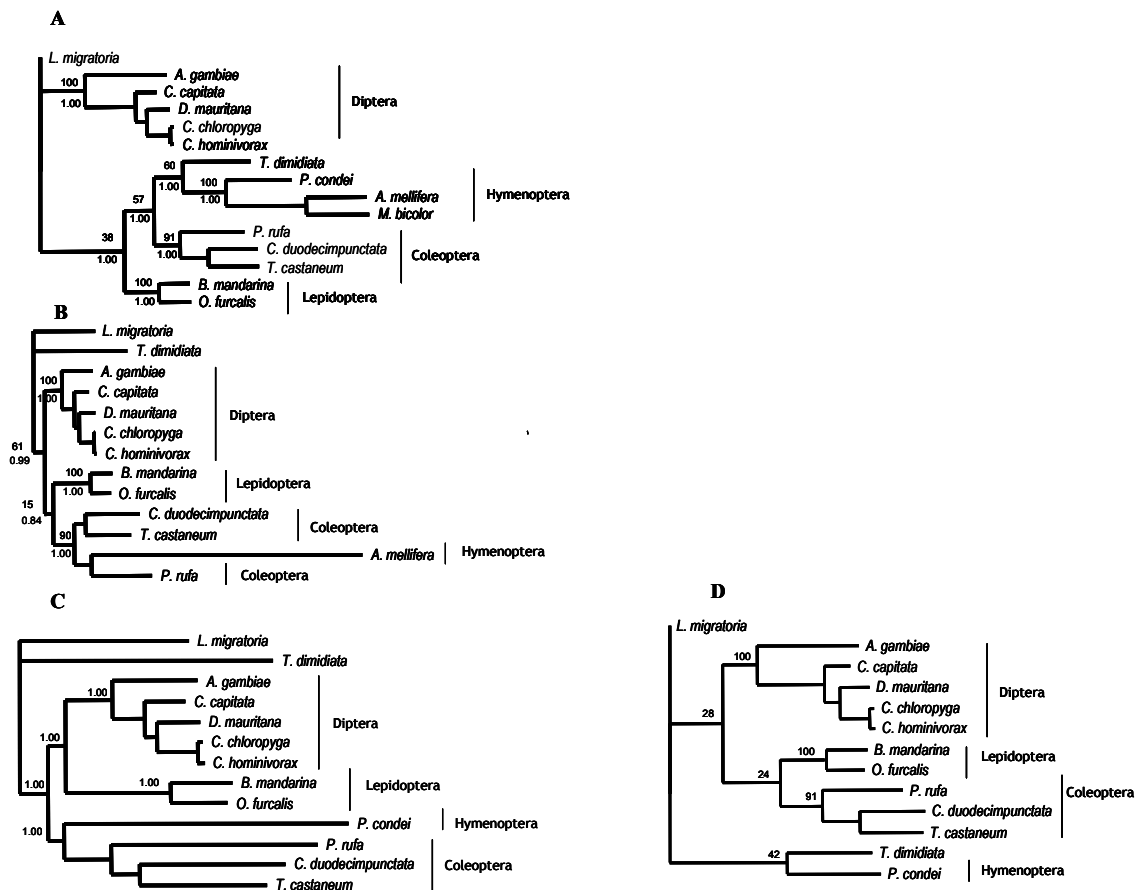
**Table 2.3**

AT content in different regions of the *Perga condei* mitochondrial genome

	<b>Total</b>	<b>%T</b>	<b>%C</b>	<b>%A</b>	<b>%G</b>	<b>%AT</b>
<b>All sites</b>	13413	35.2	13.8	42.8	8.3	78
<i>cox1</i>	1538	37.5	16.1	34.1	12.4	71.6
<i>cox2</i>	696	32.5	15.1	42.4	10.1	74.9
<i>atp8</i>	165	43.6	11.5	41.8	3	85.4
<i>atp6</i>	666	39.9	14.3	37.1	8.7	77
<i>cox3</i>	780	36	16.9	34.9	12.2	70.9
<i>nd3</i>	354	44.6	13.6	33.6	8.2	78.2
<i>nd5</i>	1641	48.9	7.3	30.3	13.5	79.2
<i>nd4</i>	1344	50.7	6.5	28.3	14.4	79
<i>nd4L</i>	294	54.4	4.4	25.5	15.6	79.9
<i>nd6</i>	532	41.9	9.6	42.5	6	84.4
<i>cytB</i>	1137	37.2	16.4	35.8	10.6	73
<i>nd1</i>	942	49.4	7.6	26.9	16.1	76.3
<b>Protein coding total</b>	10089	43	11.7	33.4	12	76.4
<b>First codon positions</b>	3363	35.3	11.1	36.7	16.9	72
<b>Second codon positions</b>	3363	48.4	17	21.3	13.4	69.7
<b>Third codon positions</b>	3363	45.2	7	42.1	5.8	87.3
<b>Ribosomal RNA</b>	2048	38.4	11.8	43.8	6.1	82.2

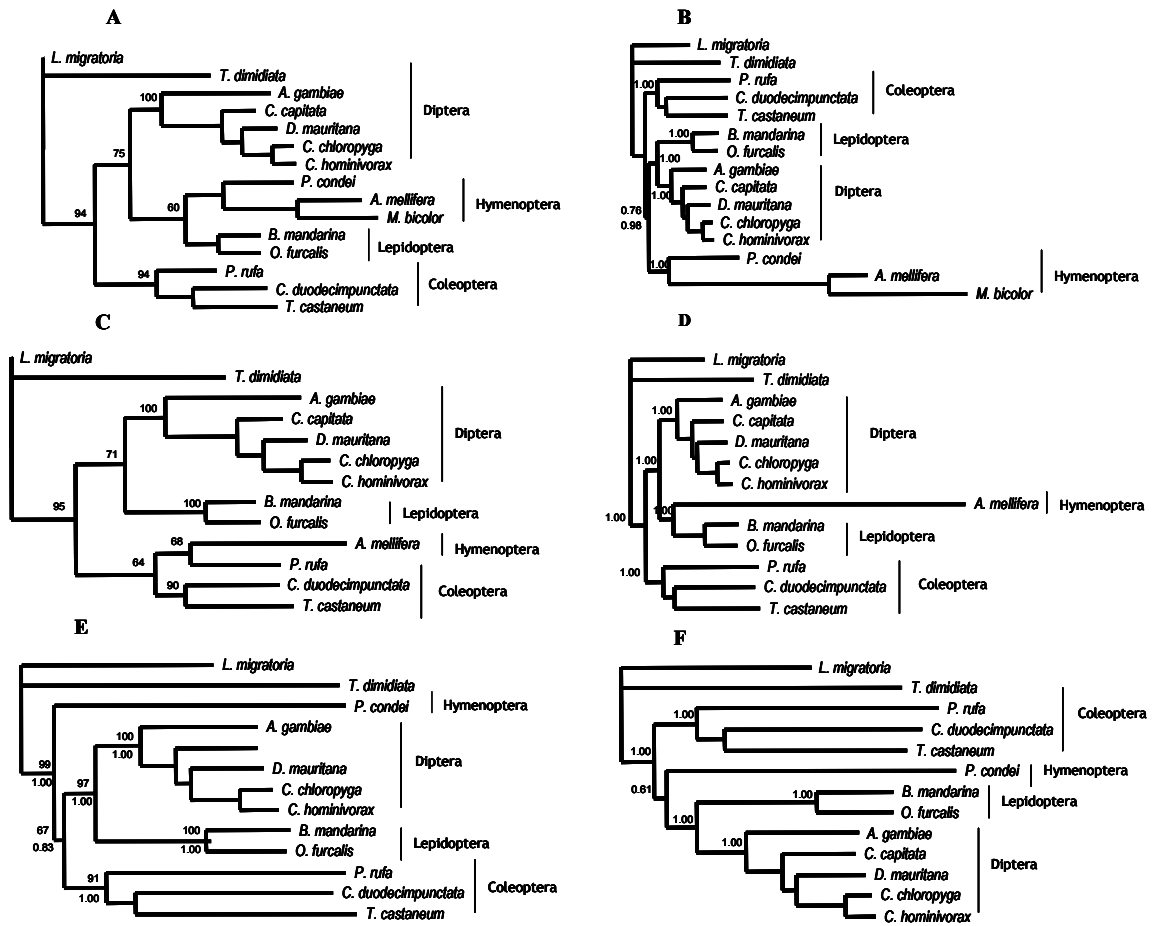
**Table 2.4**Codon usage in the *Perga condei* mitochondrial genome

Amino Acid	Codon	n	RSCU	Amino Acid	Codon	n	RSCU
Phe [F]	UUU	286	1.74	Tyr [Y]	UAU	135	1.79
	UUC	42	0.26		UAC	16	0.21
Leu [L]	UUA	343	4.27	Ter [end]	UAA	0	0
	UUG	40	0.5		UAG	0	0
	CUU	43	0.54	His [H]	CAU	45	1.38
	CUC	6	0.07		CAC	20	0.62
	CUA	49	0.61	Gln [Q]	CAA	38	1.62
	CUG	1	0.01		CAG	9	0.38
Ile [I]	AUU	320	1.73	Asn [N]	AAU	166	1.69
	AUC	49	0.27		AAC	30	0.31
Met [M]	AUA	262	1.78	Lys [K]	AAA	84	1.56
	AUG	32	0.22		AAG	24	0.44
Val [V]	GUU	79	2.08	Asp [D]	GAU	57	1.73
	GUC	6	0.16		GAC	9	0.27
	GUA	61	1.61	Glu [E]	GAA	51	1.48
	GUG	6	0.16		GAG	18	0.52
Ser [S]	UCU	105	2.63	Cys [C]	UGU	23	1.84
	UCC	4	0.1		UGC	2	0.16
	UCA	92	2.31	Trp [W]	UGA	71	1.73
	UCG	6	0.15		UGG	11	0.27
Pro [P]	CCU	49	1.72	Arg [R]	CGU	7	0.61
	CCC	22	0.77		CGC	1	0.09
	CCA	42	1.47		CGA	33	2.87
	CCG	1	0.04		CGG	5	0.43
Thr [T]	ACU	73	1.92	Ser [S]	AGU	29	0.73
	ACC	15	0.39		AGC	1	0.03
	ACA	63	1.66		AGA	74	1.86
	ACG	1	0.03		AGG	8	0.2
Ala [A]	GCU	51	2.1	Gly [G]	GGU	49	1.07
	GCC	8	0.33		GGC	2	0.04
	GCA	36	1.48		GGA	103	2.24
	GCG	2	0.08		GGG	30	0.65



**Figure 2.1**

Phylogenetic trees obtained from amino acid dataset. (A) Parsimony analysis, all three hymenopteran taxa included. Bayesian analysis recovers the same relationships. Numbers over the branches indicate bootstrap values calculated in PAUP. Values under the branches are posterior probabilities obtained in MrBayes. (B) Bayesian analysis, only *Apis mellifera* included. Posterior probability values indicated under the nodes. Parsimony analysis recovers the same relationships. Numbers above the nodes represent bootstrap values. (C) Bayesian analysis, only *Perga condei* included. Posterior probability values indicated above the nodes. (D) Parsimony analysis, just *Perga condei*. Numbers over the branches represent bootstrap values.



## Figure 2.2

Phylogenetic trees obtained from nucleotide datasets. (A) Parsimony analysis, all character changes weighted equally. All Hymenoptera included. (B) Bayesian analysis. *Apis mellifera*, *Melipona bicolor* and *Perga condei* included, all characters treated uniformly. The same relationships are recovered when partitions are considered. When different, posterior probability values are indicated over and under the branches for both analyses respectively. (C) Parsimony analysis all character changes weighted equally, only *Apis mellifera* included. (D) Bayesian analysis including just *Apis mellifera* sequence. The same relationships are recovered when all characters are treated as a single partition or when partitions are considered. Posterior probability values are indicated over the branches and were the same for both analyses. (E) Parsimony analysis, all character changes weighted equally, only *Perga condei* included. Bayesian analysis with all characters treated uniformly recovered this tree topology. Bootstrap values and posterior probabilities are indicated over and under the branches respectively. (F) Bayesian analysis. 1st, 2nd, 3rd codon positions and ribosomal genes treated as separate partitions. Only *Perga* sequence included. Posterior probabilities are shown above the nodes.



### **Chapter 3 . Molecular Bayesian Analysis of the Apocrita (Insecta: Hymenoptera) suggests that the Chalcidoidea are sister to the (Monomachidae+Diapriidae+Maamingidae)**

This chapter was slightly modified from a paper that is pending for acceptance after minor revisions to the journal *Invertebrate Systematics*.

#### **3.1 Introduction**

Despite recent efforts, evolutionary relationships among the apocritan wasps (Insecta: Hymenoptera: Apocrita) remain elusive. Neither morphology nor molecular analyses have yet provided well-resolved phylogenies for this large and diverse insect suborder. Based on comparative morphology and an intuitive approach, Rasnitsyn (1980, 1988) proposed an almost fully resolved phylogenetic hypothesis of higher level relationships. However, explicit cladistic analyses of those characters did not generally support that hypothesis (Ronquist et al. 1999). Moreover, when reductional wing characters were removed [which have been argued to be prone to convergence (Sharkey and Roy 2002)], many nodes of the apocritan tree collapsed, with the remaining relationships mostly those that are already generally recognized (Sharkey and Roy 2002). Inclusion of a set of additional wing venation characters did not increase resolution. The authors of this analysis concluded that the current morphological data matrix is best considered as a starting point for further research (Sharkey and Roy 2002). Fortunately, such morphological data appears imminent

through the Hymenoptera: Assembling the Tree of Life program

(<http://www.hymatol.org/>).

Similarly, molecular data have not yielded well-supported phylogenetic hypotheses for the Apocrita. In the most recent study (Dowton and Austin 2001), most relationships were sensitive to the model of analysis employed, despite the inclusion of three genes [representing both the nuclear (28S) and mitochondrial (mt) genomes (*16S* and *cox1*)] and, in some cases, morphology. The only consistently resolved relationships (excluding those that are already generally recognized) were the monophyly of the Proctotrupomorpha (included Chalcidoidea, Platygastroidea, Cynipoidea, Proctotrupidae, Diapriidae, Maamingidae, Monomachidae, Pelecinidae, Roproniidae, and Heloridae); and within the Proctotrupomorpha, a clade comprising Monomachidae, Diapriidae and Maamingidae. The Ceraphronoidea also fell within the Proctotrupomorpha, but only when morphological characters were included. Although this large and diverse suborder clearly demands a more extensive sampling of both characters and taxa, the lack of resolution may be in some part due to methodological limitations (at least at the time of the study) when simultaneously analyzing genes while applying different models to each partition (see below).

The simultaneous analysis of different genes may be particularly problematic in the Apocrita (Dowton and Austin 2002), given the disparate rates and base composition of mitochondrial and nuclear genes (Dowton and Austin 1995, 1997). Both simulation and empirical studies support the notion that modeling the process of molecular evolution increases phylogenetic accuracy (Cunningham 1997, Hillis et al. 1994, Whilfield and Cameron 1998), whether maximum parsimony or maximum likelihood is used. However,

although different models can easily be applied to different genes within a parsimony analysis, Dowton and Austin (2002) showed that this can arbitrarily influence the phylogeny by increasing the contribution of one partition over another to the total tree length (i.e. the shortest tree found after simultaneous analysis). They also showed that metrics used by others, based on the incongruence length difference test (Wheeler and Hayashi 1998), do not circumvent this problem (Dowton and Austin 2002).

It has only recently become computationally tractable to analyze heterogeneous datasets within a likelihood framework, by applying different models to different partitions (Nylander et al. 2004). Previously, only a single, general likelihood model could be applied to all partitions of a molecular dataset, but with the availability of MrBayes (v3), it is now possible to apply different models to each partition (Ronquist and Huelsenbeck 2003).

For this reason, we considered it timely to reanalyze the molecular dataset of Dowton and Austin (2001) (DA2001 from here on), using a mixed-model Bayesian phylogenetic approach. In addition, we sequenced 550 bp of the nuclear *18S* rRNA gene, which is generally more conserved than any of the genes included in previous analyses of apocritan relationships, and has been proposed to be a good candidate for resolving Mesozoic divergences (65-250 million years ago) (Wiegmann et al. 2000). Our molecular analyses do not confidently resolve the phylogeny of the Apocrita, but they do consistently, and with reasonable support, recover the Chalcidoidea as the sister group to the (Monomachidae + Diapriidae + Maamingidae), in contrast to DA2001.

### **3.2 Materials and Methods**

### 3.2.1 DNA Extraction and sequencing

The same taxa sampled in the DA2001 study were sampled for the *18S* gene. In this way, we were able to add the newly collected data directly to the previous molecular data set containing 3 genes. Vouchers are available at the Hymenopteran library in Dowton's laboratory. DNA was extracted following the method of Sunnucks and Hales (1996). 550 bp of the *18S* gene were amplified using primers 3F (5' GTTCGATTCCGGAGAGGA 3') and 5R (5' CTTGGCAAATGCTTTCGC 3') from Giribet *et al.* (1996) for 79 aprocritan taxa. We chose to sequence this region because it overlaps a region already sequenced in a range of holometabolan taxa, and includes most of regions A and D of the *18S* gene, as described in Whiting *et al.* (1997). Amplifications were not successful for 8 of the taxa used in Dowton and Austin (2001). These were coded as missing data (Table 1).

PCR reactions were performed in a total volume of 20 µl containing 10 mM Tris-HCl (pH 9), 50 mM KCl, 0.1% Triton X-100, 2.4 mM MgCl<sub>2</sub>, 0.4 µM each primer, 25 µM of each dNTP, 0.75 U Taq DNA polymerase (PROMEGA) per reaction, and 0.5 µl of DNA extract. Amplifications were performed in a Corbett Research thermocycler using the following program: an initial denaturation at 94°C for 2 min, followed by 35 cycles (denaturation at 94°C for 30 s, annealing at 50°C for 30 s, and extension at 72°C for 1 min). In addition, a 5 min extension at 72°C was added at the end of the 35 cycles in order to finish any incomplete amplification.

In order to remove unincorporated primers and dNTPs before sequencing, double-stranded PCR products were purified using PEG (polyethylene glycol) precipitation (Maniatis *et al.*

1989, p.1.40) with some modifications (0.6 volumes of 30% PEG in 1.5 M NaCl was added to each PCR reaction). Cycle sequencing reactions were performed with the ABI Prism Dye Terminator cycle sequencing kit v.3.1 (Perkin-Elmer) with AmpliTaq FS. Both strands of the PCR product were sequenced and ambiguities were resolved by comparing the electropherograms using the program BioEdit v. 5.0.9 (Hall 1999). Newly generated sequences were deposited in GenBank ([AY918964-AY919042](#)).

### ***3.2.2 Sequence alignment and phylogenetic analysis***

Edited *18S* sequences were aligned using CLUSTAL W (Thompson et al. 1994), with the program BioEdit (Hall 1999), and manually adjusted. Alignments for *16S*, *28S*, and *COI* were as described previously (Dowton and Austin 2001). The final data set thus comprised 4 gene fragments, for 87 taxa (available as supplementary material).

*Orussus* (Symphyta: Orussidae) was specified as the outgroup. Inclusion of a broader collection of symphytans (*Xiphydria*, *Hartigia*) in preliminary bayesian analyses led to clearly erroneous groupings [such as *Vanhornia* pulled towards the outgroup (data not shown)]. This may be due to long-branch attraction artifacts. Bergsten (2005) found that the most common long-branch attraction artifact is related to ingroup taxa being pulled towards a long branched outgroup. Further, due to the different rates of molecular evolution evident in symphytan and apocritan mitochondrial genes (Dowton and Austin, 1995), it is possible that model misspecification becomes a problem when both groups of taxa are included. For these reasons, we included a single symphytan representative as the

outgroup (*Orussus*), the taxon most broadly accepted as the closest symphytan relative of the Apocrita.

Molecular phylogenetic analyses were conducted in MrBayes v. 3.1 (Ronquist and Huelsenbeck 2003, Huelsenbeck and Ronquist 2005). Due to previous analyses in which nuclear and mitochondrial hymenopteran gene phylogenies were generally judged incongruent (Dowton and Austin 1998, Belshaw et al. 2000, Chen et al. 2004), six molecular partitions were considered: *18S*, *16S*, *28S*, *COI* 1<sup>st</sup> and 2<sup>nd</sup> codon positions. Saturation analysis from Dowton and Austin (2001, 2002) indicated strong saturation of the *COI* 3<sup>rd</sup> codon partition, with corrected pairwise distances departing widely from uncorrected distances. In order to avoid noise due to saturation of 3<sup>rd</sup> codon positions, we excluded this data partition. A total of 3 different analyses were run: an unpartitioned Bayesian analysis of the DA2001 data set (28S16SCOInopart), a partitioned Bayesian analysis of the DA2001 data set (28S16SCOIpart), and a partitioned Bayesian analysis including the *18S* gene (18S28S16SCOIpart). We calculated Bayes factors to compare the predicted likelihoods between analyses, as described in Nylander et al. (2004).

Mr Modeltest (Nylander 2004) selected the general time reversible model with some sites assumed to be invariable and with variable sites assumed to follow a discrete gamma distribution [GTR + I +  $\Gamma$ ; Yang (1994)], as a best fit model of nucleotide substitution for all partitions. As the results from Mr Modeltest do not indicate whether or not the covarion option should be included, we did preliminary analyses with and without the covarion option. According to Posada and Crandall (1998) the use of a model is only justifiable if it

significantly increases the likelihood of the analysis; so the likelihood values both with and without the covarion option were compared (Table 2). We found few differences between the topologies of the covarion=yes and covarion=no analyses. Additionally, Bayes factors favored the simpler model. For these reasons, we present only “covarion=no” results. Accordingly, we set the maximum likelihood parameters in MrBayes as follows: “lset nst=6” (GTR), “rates=invgamma”, and “covarion=no”.

The Markov chain Monte Carlo process was set so that four chains (three heated and one cold) ran simultaneously. For each analysis we conducted four runs of 5,000,000 generations, with trees being sampled every 100 generations, each of which started from a random tree. Independent analyses indicated that “stationarity” (or “burn-in”: lack of improvement in ML scores) was reached at no later than 70,000 generations; thus, the first 700 trees were discarded from each analysis as the burnin, and the remaining trees were used to generate a 50% majority rule consensus tree but including all compatible groupings, with the percentage of samples recovering any particular clade representing that clades’ posterior probability (Huelsenbeck and Ronquist 2001). To provide additional confirmation of convergence and appropriate mixing, we compared the harmonic mean of the ML scores after burnin from the independent runs (Nylander et al. 2004). We also compared posterior distributions and generation plots for all model parameters, to check if the runs were producing similar marginal posterior distributions and were mixing appropriately over the distributions.

### **3.3 Results**

### ***3.3.1 Molecular analyses***

Our molecular Bayesian analyses of apocritan relationships displayed many similarities with those reported previously (Dowton and Austin 2001), but unfortunately not a well resolved phylogeny of the group. Several nodes have low posterior probabilities and some taxa are not consistently recovered throughout the analyses. However, a consistent difference with DA2001 and also a well supported group according to the Bayesian analysis was the clade (Chalcidoidea, (Monomachidae, Maamingidae, Diapriidae)). For this reason, we will focus on this group in the present study. Nevertheless, in order not to bias our analyses (by excluding potentially disrupting families/superfamilies), we included representatives from across the Apocrita.

The most important drawback of the DA2001 study was that all analyses were performed using only maximum parsimony (MP). At that time, computational limitations precluded maximum likelihood (ML) analysis. MrBayes 3.1 allowed us to perform such an analysis. The first purpose of this research was to reanalyze DA2001 molecular data set using MrBayes. An initial Bayesian analysis was run using a simple model, with all genes treated as a single partition. A second Bayesian analysis was then performed in which *16S*, *28S* and the *COI* 1<sup>st</sup> and 2<sup>nd</sup> codon positions were considered as separate partitions. Both analyses recovered a tree with very similar topologies. However, in the second analysis we obtained a tree with an increase of more than 500 log likelihood units, and comparing the predicted likelihoods of the mixed model with the partitioned model, the Bayes factor value was very strong against the simpler model (Bayes factor=2032, Table 2). As a result, the tree from the unpartitioned analysis is not presented. The Bayesian analysis of the DA2001 molecular



dataset recovered a monophyletic Proctotrupomorpha (pp= 0.98), and within the Proctotrupomorpha the Chalcidoidea as sister group to the (Diapriidae+Monomachidae+Maamingidae) (pp= 0.99) (Fig. 1). The Heloridae were recovered inside the Proctotrupomorpha, as expected, but in contrast to their apparently erroneous placement in most of the analyses presented in the DA2001 study.

There were only minor differences when the *18S* data was included in the analysis (Fig. 2). Here again, we consistently recovered the Proctotrupomorpha forming one clade (pp= 0.98), and the Chalcidoidea as sister group to the (Diapriidae+Monomachidae+Maamingidae) (pp= 0.91).

#### *3.3.1.1 Unrooted analysis*

It has been suggested that erroneous rooting can cause misplacement of some taxa due to long branch attraction to the outgroup (Bergersten 2005). Holland et al.'s (2003) study showed that the inclusion of an outgroup can frequently disrupt the ingroup tree. For this reason they suggested trees should always be constructed both with and without the outgroup. Due to the apparent misplacement of (Ichneumonoidea+Braconidae) in some of the analyses (being pulled towards the outgroup), we sought to evaluate the effect of outgroup choice to the ingroup topology by running an unrooted analysis. After removing *Orussus* from the analysis, ingroup relationships were consistent with previous analyses. The monophyly of the Proctotrupomorpha and other ingroup relationships such as the (Chalcidoidea (Diapriidae+Monomachidae+Maamingidae)) were still supported (data not shown but available upon request).

### 3.3.2 Analysis including only the *Proctotrupomorpha*

In the Dowton and Austin (2001) study, a range of relationships were sensitive to the model of analysis. However, a group that was particularly stable was the monophyletic Proctotrupomorpha; nevertheless, the relationships between each of the proctotrupomorph families were not identically recovered. Since Bayesian analyses also supported a monophyletic Proctotrupomorpha, we attempted an additional analysis including only the proctotrupomorph families. We included the Ceraphronoidea, since it has been proposed that this group falls inside the Proctotrupomorpha. *Venturia* and *Schlettererius* were included as outgroups. The Bayesian analysis including all molecular partitions supported a monophyletic Proctotrupomorpha, excluding the Ceraphronoidea. The (Chalcidoidea (Diapriidae+Monomachidae+Maamingidae)) group was also recovered with high support (pp= 0.98) (Fig. 3).

In order to assess whether our evidence for (Chalcidoidea (Diapriidae+Monomachidae+Maamingidae)) was better supported than a (Chalcidoidea+Platygastroidea) sister group relationship, we performed a Bayesian analysis with topological constraints. In this latter analysis, the search was restricted to those trees that contained all chalcidoids and platygastroids as a monophyletic group; this has only become recently possible with MrBayes (v. 3.1.2). The most likely tree that retained the Chalcidoidea and Platygastroidea as sister groups had a likelihood of -18,798, more than 1500 log<sub>e</sub> units worse than the most likely tree in which the Chalcidoidea were sister to the (Diapriidae+Monomachidae+Maamingidae) (likelihood = -17,212). Using Bayes factors to compare these two values, our data are strongly in favour of the latter relationship.

### ***3.3.3 Comparison between Bayesian runs***

When independent analyses were performed on the same dataset, the likelihoods were very similar, suggesting that stationarity had been reached. Similarly, the average standard deviation of split frequencies was less than 0.01, reflecting the fact that the tree samples from the different runs became increasingly similar and suggesting convergence between the four runs. After 5,000,000 generations, the marginal posterior distributions of gamma and proportion of invariant sites remained focused, and the chain rapidly mixed over them. However, we found some examples of parameter sets for which the marginal posterior distributions are diffuse with a higher density throughout the parameter space. This is particularly the case for some of the substitution rates on the *COI-1* and *COI-2* partitions (Fig. 4). This seems to happen in all analyses (both partitioned and unpartitioned). Similar observations were reported by Nylander et al. (2004); it seems that these parameters will continue to fluctuate throughout the run, but since the posterior remained focused for other parameters, the bayesian run seems robust despite the inclusion of some “weak” parameters in the model.

## **3.4 Discussion**

The most recent attempt to resolve the phylogeny of the Hymenoptera was the DA2001 analysis, in which they sampled 84 taxa and generated character information from three genes. A morphological dataset was also included in some analyses. This was a significant

improvement over previous attempts; however, a range of relationships were not stably recovered and were sensitive to the model of analysis. From their analysis, a group that was particularly stably recovered was the monophyletic Proctotrupomorpha (except for the dubious placement of the Ceraphronoidea in some trees); nevertheless, the relationships between each of the proctotrupomorph families were not identically recovered. The present analysis supported the monophyly of the Proctotrupomorpha and confirmed the placement of the Ceraphronoidea outside of the Proctotrupomorpha.

In contrast to the DA2001 study, we decided not to include morphological dataset generated by Ronquist et al. (1999). Sharkey and Roy (2002) recently described a number of shortcomings in Ronquist et al.'s matrix, due to erroneous coding of wing characters (most of which are reductional). They generated a new morphological matrix with the original 39 wing characters deleted and replaced with new wing characters. Although they generated an improved morphological matrix, in both cases the coding of the morphological characters was an intuitive estimation of the groundplan character states of a taxon (in most cases at the family level). This method can be equivocal and has been criticized (Yeates 1995, Prendini 2001). With this method, the analysis of higher-level relationships involves estimating the plesiomorphic states of the higher taxa under consideration, from information of a sample of their constituent species (or the extraction of character state information from the literature), and then summarizing the character state information into supraspecific terminals (Prendini 2001). According to Yeates (1995) this approach may result in non-parsimonious ancestral state assignments, even in simple cases, and there are conceptual issues to be dealt with regarding how to infer the ancestral states in the best possible way.

### **3.4.1 Chalcidoidea+(Diapriidae+Monomachidae+Maamingidae)**

Very little is known about the phylogenetic relationships within the Proctotrupomorpha.

Within this group, one of the best-recovered associations is the

Platygastroidea+Chalcidoidea sister-group relationship, which since Rasnitsyn (1988) has been supported by several studies (Whitfield 1992, Ronquist et al. 1999, Sharkey 2001, Dowton et al. 1997, Dowton and Austin 2001). However, Gibson (1999) presented an alternate hypothesis to the Platygastroidea+Chalcidoidea relationship based on two character systems: the structure of the mesotrochanteral depressor muscle and the relative structure and position of the pronotum, prepectus, and mesothoracic spiracle. According to Gibson (1999), neither character supports a Platygastroidea+Chalcidoidea sister-group relationship. Rather, both characters indicate that Platygastroidea is most closely related to Pelecinidae+Poctotrupidae+Vanhorniidae; and Chalcidoidea+Mymarommatoidea were indicated as possibly being more closely related to Diapriidae or Ceraphronoidea based on other characters (Gibson 1999).

Similarly, little is known about the sister group to the Diapriidae, which is likely the recently described Maamingidae (Early et al. 2001), Monomachidae or Austroniidae (Sharkey and Roy 2001). According to Ronquist (1999), some characters suggested the Diapriidae as sister group to the Cynipoidea. The expanded dataset of Dowton and Austin (2001) supported a monophyletic group including Diapriidae, Monomachidae and Maamingidae, but again, the position of this group within the Proctotrupomorpha was not consistently recovered. In some cases it was recovered as sister group to the

Chalcidoidea+Platygastroidea [as in Figs 3, 7, and 11 from Dowton and Austin (2001)], at the base of the Proctotrupomorpha (Figs 4, 5, 8, 12), as sister group to the Chalcidoidea+Cynipoidea (Fig. 10) or just to Chalcidoidea (Fig. 2).

Within the Proctotrupomorpha, our molecular analyses strongly supported the Chalcidoidea+ (Diapriidae+Monomachidae+Maamingidae) clade. Most of the analyses also recovered a Platygastroidea+Cynipoidea association, but with poor support. The inclusion of other critical taxa such as Mymarommatidae would be essential to further test this relationship.

### ***3.4.2 Other hymenopteran relationships***

#### ***3.4.2.1 Evaniidae not sister to the Gasteruptiidae***

Despite conflicting evidence, several recent phylogenetic and other studies predict an Evaniidae+Gasteruptiidae clade. Ronquist et al. (1999) using morphological data recovered these two families as sister taxa with low support. Dowton and Austin (2001) also recovered Evaniidae+Gasteruptiidae, however their placement of Evaniidae depended on the tree building methods and models employed. Our Bayesian analysis did not support an Evaniidae+Gasteruptiidae clade, instead we recovered an Evaniidae+(Gasteruptiidae+Ceraphronoidea) group. The most apparent evidence suggesting a non-monophyletic Evanioidea (includes Evaniidae, Gasteruptiidae, Aulacidae), is the differences in host biology of the families (Ronquist 1999). The evaniids oviposit into cockroach oothecae buried in substrate, loose in leaf litter, or attached to female cockroaches. The gasteruptiids are predator-inquilines that lay eggs inside the cells

of solitary bees and wasps nesting in plant stems or in underground nests, with the subsequent larvae developing on the food stores and/or nest inhabitants.

#### *3.4.2.2 Aculeata within the Evaniomorpha*

Rasnitsyn (1988) suggested a monophyletic Evaniomorpha combining groups as diverse as the Ceraphronoidea, Evanioidea, Megalyridae, Trigonalysidae and Stephanidae. The cladistic reanalysis of Rasnitsyn's data indicated that the evaniomorphs form a grade of basal apocritan lineages rather than a monophyletic group (Ronquist et al. 1999). Dowton et al. (1997) supported again the Evaniomorpha as a monophyletic group, but Dowton and Austin (2001) suggested the Evaniomorpha as a grade including the Ichneumonoidea, Aculeata and Stephanidae. Our molecular analyses gave strong support to the association of Aculeata with Stephanidae, Megalyridae and Trigonalysidae in one group, but contradicts the traditional Aculeata+Ichneumonoidea relationship suggested by morphology (Ronquist et al. 1999, Whitfield 1998), and previous molecular analyses (Dowton and Austin 1994, 2001).

### **3.5 Conclusion**

In this chapter I performed Bayesian analyses to reanalyze the Dowton and Austin (2001) molecular dataset. Partial sequences of the *18S* gene were also generated and added to the original dataset in an attempt to improve the resolution of the phylogenetic relationships within the Apocrita. All analyses consistently recovered the Proctotrupomorpha as a monophyletic clade and, as in Dowton and Austin (2001), the Diapriidae+Monomachidae+Maamingidae was always recovered with very strong support ( $pp > 0.90$  in all analyses). In contrast to Dowton and Austin (2001), in this study the

Chalcidoidea and Platygastroidea were not recovered as sister groups. The Chalcidoidea was instead consistently recovered as sister taxon to the (Diapriidae+Monomachidae+Maamingidae) clade.



**Table 3.1** List of taxonomic groups included.

Sub-order Superfamily Family	Taxon	16S	28S	COI	18S
Symphyla					
Cephoidea					
Cephidae	<i>Hartigia trimaculata</i> (Say)	√	√	√	√
Siricoidea					
Xiphydriidae	<i>Xiphydria mellipes</i> (Harris)	√	√	√	√
Orussoidea					
Orussidae	<i>Orussus terminalis</i> (Newman)	√	√	√	√
Apocrita					
Ceraphronoidea					
Megaspilidae	<i>Conostigmus</i> sp.	√	√	√	√
	<i>Dendrocerus carpenteri</i> (Curtis)	√	√	√	√
Ceraphronidae	<i>Aphanogmus</i> sp.	√	√	√	√
	<i>Ceraphron</i> sp. 1	√	√	√	√
	<i>Ceraphron</i> sp. 2	X	√	√	X
Chalcidoidea					
Aphelinidae	<i>Encarsia formosa</i> (Gahan)	√	√	√	√
	<i>Aphytis melinus</i> (De Bach)	√	√	√	√
Chalcididae	<i>Brachymeria phya</i> (Walker)	√	√	√	X
Encyrtidae	<i>Leptomastix dactylopii</i> (Howard)	√	√	√	√
Eulophidae	<i>Melittobia australica</i> (Girault)	√	√	√	√
Eupelmidae	<i>Eusandalum</i> sp.	√	√	√	√
Mymaridae	<i>Gonatocerus</i> sp.	X	√	√	√
Pteromalidae	<i>Trichilogaster</i> sp.	X	√	√	√
	<i>Pteromalus puparum</i> (L.)	√	√	√	√
	<i>Megastigmus</i> sp.	√	√	√	√
Torymidae					
Cynipoidea					
Cynipidae	<i>Xestophanes</i> sp.	X	√	√	X
Figitidae	<i>Anacharis zealandica</i> Ashmead	√	√	√	√
Ibaliidae	<i>Ibalia leucospoides</i> (Hochenwarth)	√	√	√	√
Evanioidea					
Evaniidae	<i>Evania</i> sp. 1	√	√	√	√
	<i>Evania</i> sp. 2	√	√	√	√
	<i>Evania</i> sp. 3	√	X	√	X
Gasteruptiidae	<i>Gasteruption</i> sp. 1	√	√	√	√
	<i>Gasteruption</i> sp. 2	√	√	√	√
	<i>Gasteruption</i> sp. 3	√	X	√	√
	<i>Eufoenus</i> sp.	√	√	√	√
Ichneumonoidea					
Braconidae	<i>Ascogaster</i> sp.	√	√	√	X
	<i>Diospilus</i> sp.	√	√	√	√
	<i>Dolopsidea</i> sp.	√	√	√	√
	<i>Jarra maculipennis</i>	√	√	√	√
	<i>Megalohelcon ichneumonoides</i>	√	√	√	√
	<i>Miropotes</i> sp.	√	√	√	√
	<i>Neoneurus mantis</i> Shaw	√	√	√	√
	<i>Sigalphus</i> sp.	√	√	√	√
	<i>Toxoneuron abdominalis</i> Cresson	√	√	√	√
Ichneumonidae	<i>Ichneumon promissorius</i> (Erichson)	√	√	√	√
	<i>Venturia canescens</i> (Gravenhorst)	√	√	√	√
	<i>Xorides praecatorius</i> (F.)	√	√	√	√

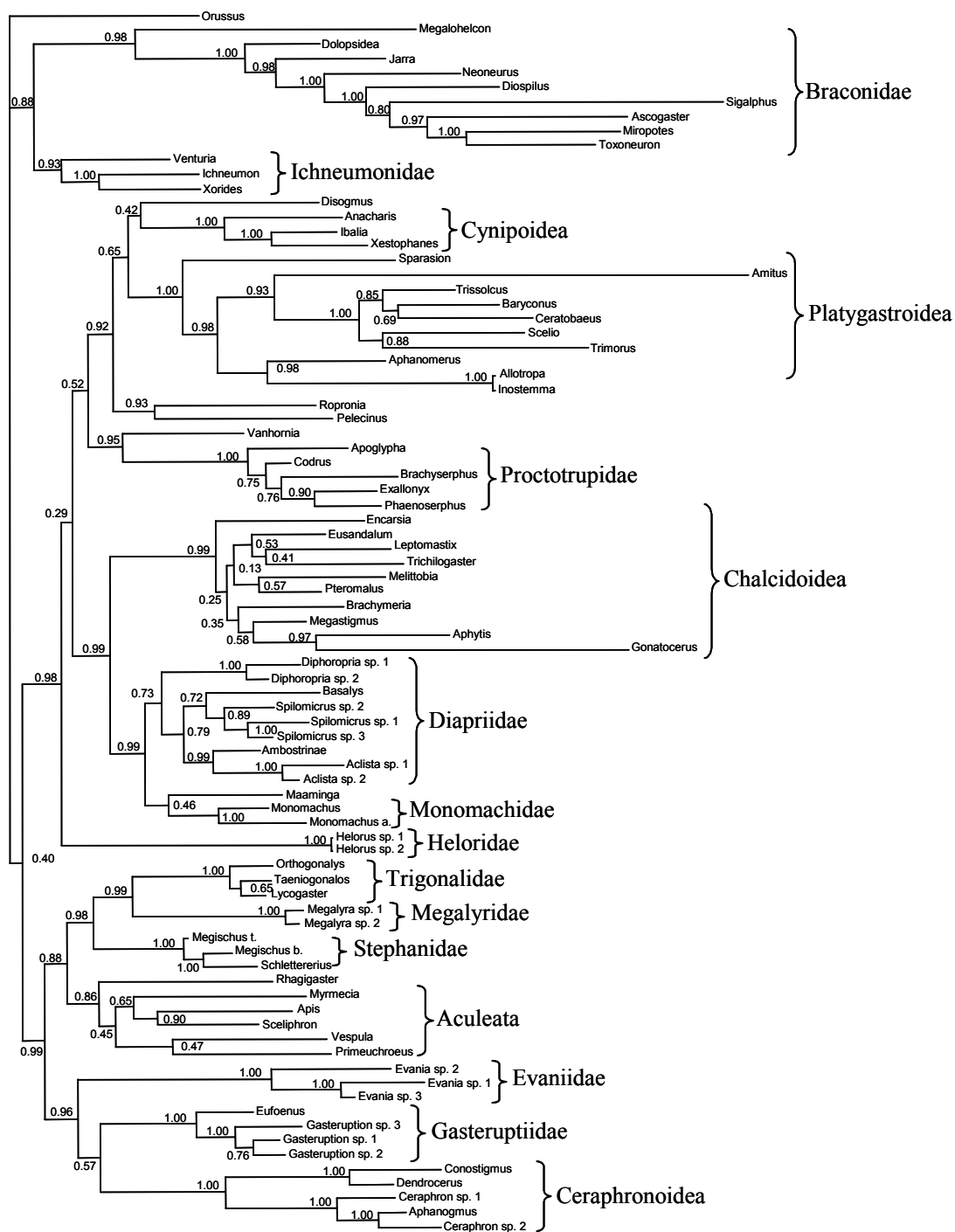
<i>Megalyroidea</i>					
Megalyridae	<i>Megalyra sp. 1</i>	√	√	√	√
	<i>Megalyra sp. 2</i>	X	√	√	√
<i>Platygaстроidea</i>					
Scelionidae	<i>Baryconus sp.</i>	√	√	√	√
	<i>Ceratobaeus sp.</i>	√	√	√	√
	<i>Scelio fulgidus</i> (Crawford)	√	√	√	√
	<i>Sparasion sp.</i>	√	√	√	√
	<i>Trimorus sp.</i>	√	√	X	√
	<i>Trissolcus basalis</i> (Wollaston)	√	√	√	√
Platygastridae	<i>Allotropa sp.</i>	X	√	√	√
	<i>Aphanomerus sp.</i>	X	√	√	√
	<i>Amitus sp.</i>	√	√	X	√
	<i>Inostemma sp.</i>	√	√	X	√
<i>Proctotrupoidea</i>					
Diapriidae	<i>Aclista sp. 1</i>	√	√	√	√
	<i>Diphoropria sp. 1</i>	√	√	√	√
	<i>Spilomicrus sp. 1</i>	√	√	√	X
	<i>Spilomicrus sp. 2</i>	√	√	√	X
	<i>Spilomicrus sp. 3</i>	√	√	√	√
	<i>Aclista sp. 2</i>	√	√	X	√
	<i>Basalys sp.</i>	X	√	√	√
	Genus indet. (Ambostrinae)	√	√	X	√
	<i>Diphoropria sp. 2</i>	√	√	X	√
Heloridae	<i>Helorus sp. 1</i>	√	√	√	√
	<i>Helorus sp. 2</i>	√	√	√	√
Maamingidae	<i>Maaminga rangi</i>	√	√	√	√
Monomachidae	<i>Monomachus antipodalis</i>	√	√	√	√
	<i>Monomachus sp.</i> (Chile)	√	√	√	√
Pelecniidae	<i>Pelecinius polyturator</i> (Drury)	√	√	√	√
Proctotrupidae	<i>Apoglypha sp.</i>	√	√	√	√
	<i>Brachyserphus abruptus</i> (Say)	√	√	√	√
	<i>Exallonyx obsoletus</i> (Say)	√	√	√	√
	<i>Codrus sp.</i>	√	√	X	√
	<i>Disogmus areolator</i> (Haliday)	√	√	X	√
	<i>Phaenoserphus viator</i> (Haliday)	√	X	√	√
Roproniidae	<i>Repronis garmani</i> (Ashmead)	√	√	√	√
Vanhorniidae	<i>Vanhornia eucnemidarum</i>	√	√	√	√
<i>Stephanoidea</i>					
Stephanidae	<i>Megischus bicolor</i> (Westwood)	X	√	√	√
	<i>Megischus texanus</i> (Cresson)	X	√	√	√
	<i>Schelettererius cinctipes</i> (Cresson)	X	√	√	√
<i>Trigonalioidea</i>					
Trigonalidae	<i>Taeniogonalos gundlachii</i> (Cresson)	√	√	√	√
	<i>Orthogonalys pulchella</i> (Cresson)	√	√	√	√
	<i>Lycogaster sp.</i>	√	√	X	X
<i>Apoidea</i>					
Apidae	<i>Apis mellifera</i> (L.)	√	√	√	√
<i>Sphecoidea</i>	<i>Sceliphron sp.</i>	√	√	√	√
<i>Chrysidoidea</i>					
Chrysididae	<i>Primeuchroeus sp.</i>	√	X	√	√
<i>Vespoidea</i>					
Formicidae	<i>Myrmecia forficata</i> (F.)	√	√	√	√
Vespidae	<i>Vespula germanica</i> (Fabricius)	X	√	√	√
Tiphiidae	<i>Rhagigaster sp.</i>	√	√	√	√

**Table 3.2**

Estimated model likelihood obtained in the different phylogenetic analysis

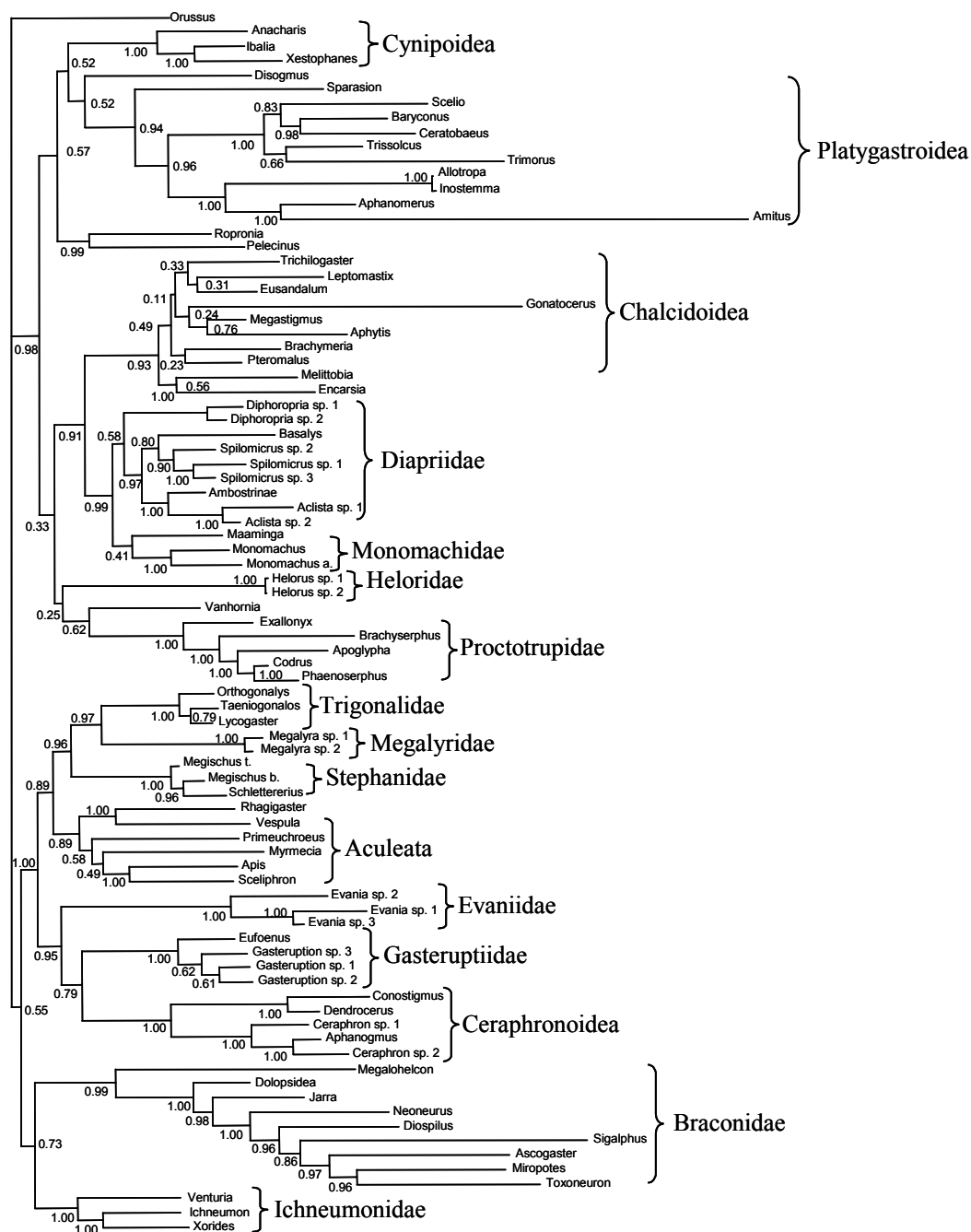
Analysis		Predicted likelihood		Bayes factor
28S16SCOI-nopart. (covarion=no)	28S16SCOI-nopart. (covarion=yes)	-32620	-32663	86
28S16SCOI-part. (covarion=no)	28S16SCOI-part. (covarion=yes)	-31604	-31873	538
28S16SCOI-part. (covarion=no)	28S16SCOI-nopart. (covarion=no)	-31604	-32620	2032
18S28S16SCOI-part. (covarion=no)	18S28S16SCOI-part. (covarion=yes)	-37254	-37798	1088

part= Genes treated as different partitions, nopart.= all genes treated as a single partition



**Figure 3.1**

Bayesian analysis of the *28S*, *16S*, *COI-1*, *COI-2* partitions from DA2001. All genes treated as a single partition. Model selected according to MrModeltest.

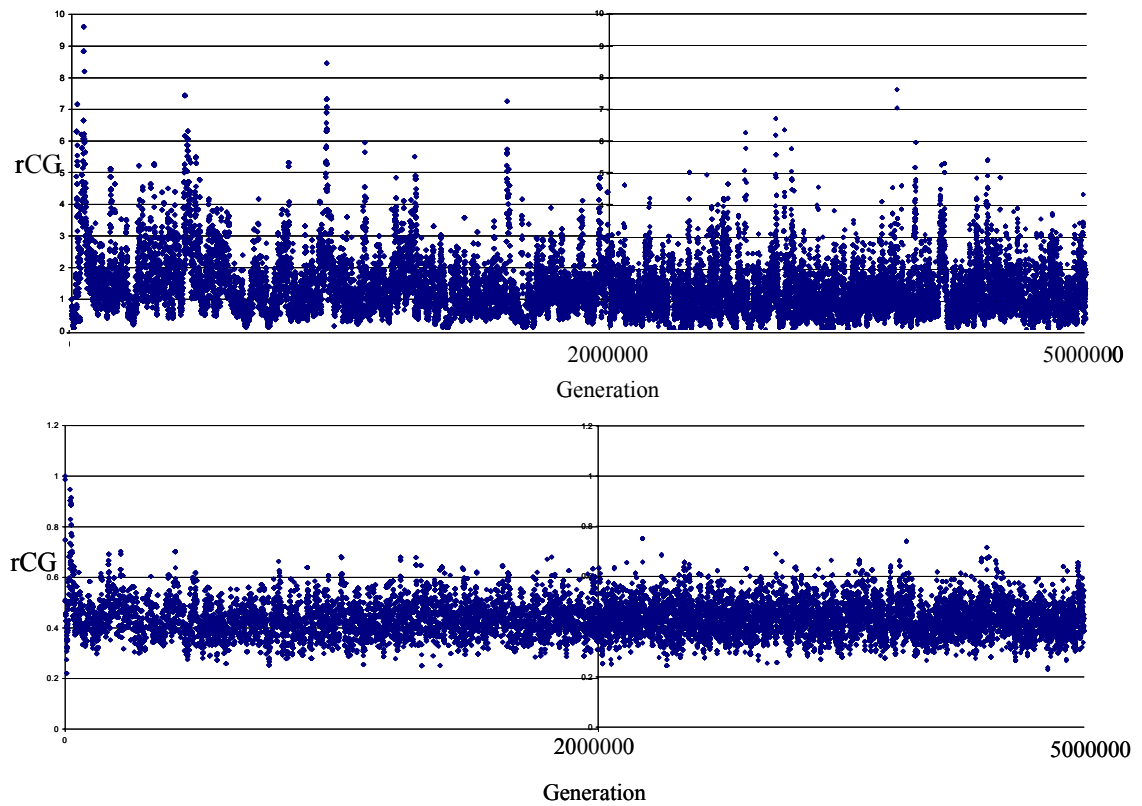


**Figure 3.2**

Bayesian analysis of the *28S*, *16S*, *COI-1*, and *COI-2* partitions from DA2001, and *18S* partition. Model selected according to MrModeltest.

**Figure 3.3**

Bayesian analysis of the Proctotrupomorpha; including *28S*, *16S*, *COI-1*, and *COI-2* partitions from Dowton and Austin (2001), and *18S* partition.



**Figure 3.4**

Generation plots of some of the substitution rates for the *COI-1* and *COI-2* partitions after 5,000,000 generation runs. A. Generation plot of the C-G rate parameter for the *COI-1* partition (diffuse posterior distribution). B. Generation plot of the C-G rate parameter for the *28S* partition (posterior remains focused).

**Chapter 4 . Mitochondrial genomes of *Vanhornia eucnemidarum* and  
*Primeuchroeus* sp.: Evidence of rearranged mitochondrial genomes within  
the Apocrita (Insecta: Hymenoptera)**

This chapter was slightly modified from a manuscript that has been accepted in the journal *Genome*.

#### **4.1 Introduction**

Mitochondrial (mt) genomes have been broadly studied because of the ease of recovering genetic information that may be useful for investigating molecular and organismal evolution. In general, metazoan mtDNA contains thirteen protein-coding genes, two ribosomal RNA genes (rRNA), and twenty-two transfer RNA genes (*trn*) (Boore 1999). In addition, the mtDNA molecule contains a major noncoding (A+T rich in insects) region that is thought to play a role in the initiation of transcription and replication (Wolstenholme 1992). With an average size of 16 kb, it is a very compact circular genome. The predominance of maternal inheritance, lack of extensive recombination, and accelerated rates of nucleotide substitution are features that have favoured the use of mtDNA as an evolutionary marker (Lessinger et al. 2000). It has been extensively used for the investigation of population structures and phylogenetic relationships at various taxonomic levels (Avisé 1995).



As more mt genomes are completely sequenced, rearrangements appear more common than originally thought (Boore and Brown 1998). For example, among arthropods, highly rearranged mt genomes have been found for the wallaby louse (Shao et al. 2001), the hermit crab (Hickerson and Cunningham 2000), and metastrate ticks (Black and Roehrdanz 1998, Campbell and Baker 1998). High levels of rearrangement have also been noted for nematodes and brachiopods (Boore 1999, Boore and Brown 2000). Gene rearrangements have been shown to be very powerful characters for reconstructing evolutionary relationships (Smith et al. 1993, Boore et al. 1995, Boore and Brown 1998, Curole and Kocher 1999, Le et al. 2000, Roehrdanz et al. 2002), and the rapidity of rearrangement within a lineage determines the level at which rearrangements are likely to be phylogenetically informative (Boore et al. 1998).

The growing interest in mt genomes for phylogenetic reconstruction has triggered a rapid increase of published complete mt genome sequences (Curole and Kocher 1999). However, most of these sequences are from vertebrates while the numbers in other animal phyla are lagging behind (Boore 1999). This is also true for insects, despite the fact that this group represents the largest animal taxon on earth with many phylogenetic questions unresolved. 40 complete mt insect genomes have been sequenced with 13 of them from the order Diptera and 11 from the Hemiptera. Only three hymenopteran mt genomes are available in GenBank, *Apis mellifera*, *Melipona bicolor*, and *Perga condei* (see chapter 2). Only two of them belong to the megadiverse suborder Apocrita (*Apis* and *Melipona*), and with both of them belonging to the superfamily (Apoidea), leaving 13 of the 14 apocritan superfamilies unrepresented.

Additionally, sequences of two mt multigenic regions within the Hymenoptera have indicated that a number of apocritan lineages have distinctly rearranged genomes compared with the ancestral hexapod organization, and that these rearrangements are restricted to the Apocrita (Dowton and Austin 1999, Dowton et al. 2003). A number of molecular and biological transitions also map to this node of the hymenopteran tree, among them the parasitic lifestyle and an increased rate of mt genetic divergence (Dowton and Campbell 2001). However, since the available hymenopteran data is extremely limited, an expanded sample of complete hymenopteran mt genomes would considerably enhance our understanding of the evolutionary biology of the mt genome in this group.

Here we report the almost complete mt genome of two other hymenopteran taxa, *Vanhornia eucnemidarum* (Hymenoptera:Apocrita:Vanhorniidae), and *Primeuchroeus* sp. (Hymenoptera:Apocrita:Chrysididae). Our goal in sequencing the mt genome of other apocritan taxa was to further characterize the evolution of the mt genome structure and to gain a more accurate estimate of the nature of mt genome organization among the Apocrita. Additionally, Dowton et al. (2003) had found a major rearrangement in the mt genome of *V. eucnemidarum*, this rearrangement involves a five fold repetition of three of the *trns* from the ‘*ARNSEF*’ region, with some copies having mutations that would likely render them non-functional (Dowton et al. 2003). We sequenced this region in another *V. eucnemidarum* population, with the purpose of further characterizing the process of gene rearrangement in this region.

## 4.2 Materials and Methods

#### **4.2.1 Insects and DNA Extraction**

*Vanhornia eucnemidarum* and *Primeuchroeus* specimens, provided by Andy Austin (University of Adelaide), were collected into 100% ethanol and stored at 4°C before extraction. Ethanol was removed by washing three times (30 min each) in 10 mM Tris-HCl (pH 8) containing 100 mM NaCl and 1 mM MgCl<sub>2</sub>. Tissue was homogenized in 400 µl of 10 mM Tris-HCl (pH 8), 10 mM EDTA, and 1% SDS containing 100 µg of proteinase K (Boehringer Mannheim) and incubated overnight at room temperature. DNA was separated from salt-insoluble material by the method of Sunnucks and Hales (1996). DNA was redissolved in 100 µl of sterile water and stored at 4°C. This DNA solution was used directly in PCR reactions.

#### **4.2.2 PCR amplification and cloning**

A range of universal insect mt primers (Simon et al. 1994) were tested and primers that generated the best amplifications were used for long PCRs. New primers were designed where necessary in order to generate perfectly matched primers. A total of 9 overlapping PCR fragments (*V. eucnemidarum*) and 8 overlapping fragments (*Primeuchroeus*) between 500 bp and 3000 bp were the initial template for sequencing reactions or cloning (Table 4.1), and new primers were designed as sequence data accumulated. A ~2 kb fragment containing part of *nd3*, 6 *trns* and part of *nd5* genes was sequenced elsewhere for both taxa (Dowton et al. 2003), and obtained from GenBank.

PCR reactions were performed in a total volume of 20  $\mu$ l. For the generation of short PCR fragments (<800 bp), reactions contained 10 mM Tris-HCl (pH 9), 50 mM KCl, 0.1% Triton X-100, 1.25–6.25 mM MgCl<sub>2</sub>, 0.4  $\mu$ M of each primer, 25  $\mu$ M of each dNTP, 0.75 U *Taq* DNA polymerase (PROMEGA) per reaction, and 0.5  $\mu$ l of DNA extract. A negative control PCR tube was prepared with the same constituents but lacking DNA.

Amplifications were performed in a Hybaid Sprint PCR thermocycler or a Corbett Research thermocycler using the following program: 5 cycles (30 s at 94°C, 30 s at 45–55°C, and 1 min at 72°C), followed by 30 cycles (30 s at 94°C, 30 s at 55°C, and 1 min at 72°C). In addition, a 5-min extension at 72°C was added at the end of the 35 cycles in order to finish any incomplete amplification. PCR optimization for each template involved the variation of possible primer combinations, MgCl<sub>2</sub> concentration, and annealing temperature. Longer amplifications (>800 bp) were performed as described above but using 0.75 U *Taq* and 2.5 mU *Pfu* DNA polymerase (*Pyrococcus furiosus*) and 100  $\mu$ M of each dNTP per reaction, or alternatively 1.2 U of BIO-X-ACT Long DNA polymerase from BIOLINE and all PCR conditions as suggested by the manufacturer. For these amplifications, we used the following long PCR program: an initial denaturation at 92°C for 2 min, followed by 35 cycles (denaturation at 92°C for 10 s, annealing at 45–65°C for 30 s, and extension at 68°C for 1–5 min) and a final extension step of 5 min.

After optimization of the amplification reactions, unincorporated primers and dNTPs were removed prior to sequencing. PCR products were purified either using the QIAquick PCR purification kit (QIAGEN) or PEG (polyethylene glycol) precipitation (Maniatis 1989) with some modifications (0.6 volumes of 30% PEG in 1.5 M NaCl was added to each PCR reaction). Cycle sequencing reactions were performed with the ABI Prism Dye Terminator

cycle sequencing kit v.2 or v.3 (Perkin-Elmer) with *AmpliTaq* FS. Both strands of the PCR product were sequenced. Primer sequences were removed from the start and the end of the obtained sequence and sequence ambiguities were resolved by comparing the electropherograms using the program BioEdit v. 5.0.9 (Hall 1999).

PCR products that were difficult to sequence directly were cloned. For cloning we used the pGEM-T easy vector system from PROMEGA exactly as recommended by the manufacturer. Plasmids were extracted using the Wizard Plus SV Minipreps DNA purification system from PROMEGA and sequenced using both the original PCR primers and the M13-F and M13-R primers that anneal to the vector.

#### ***4.2.3 Identification of protein-coding, rrn, trn genes and gene rearrangements***

Mt gene names are abbreviated according to Boore (1999). Protein-coding and rrn gene sequences were initially identified using BLAST searches in GenBank, and then subsequently by alignment with other complete mt DNA and amino acid sequences from other closely related insects (*Drosophila yakuba* NC001322, *Anopheles gambiae* L20934, *Bombyx mandarina* AB070263 *Apis mellifera* L06178, *Melipona bicolor* NC004529, *Perga condei* AY787816). Protein coding genes were translated using the *Drosophila* mtDNA genetic code using the program Translation Machine ([www.2ebi.ac.uk/translate/](http://www.2ebi.ac.uk/translate/)). Edited sequences were subsequently imported into MEGA version 2.1 (Kumar et al. 2001) for calculating the nucleotide and amino acid composition.

After mapping the position of the protein-coding and rrn genes, sequences were screened for trn genes. These sequences were submitted for trn gene search using the program tRNA-

Scan SE [v. 1.1, <http://genome.wustl.edu/eddy/tRNAscan-SE>; Lowe and Eddy (1997)]. The parameters for the trn scan program were set for mt-chloroplast DNA as the source and using the invertebrate mt genetic code. When long tracts of non-coding sequence were apparent, the cove cut off score was reduced to 10 and the search repeated.

Finally, possible gene rearrangements were detected by comparing the complete organization of the mt genome of *V. eucnemidarum* and *Primeuchroeus* sp. with that of *D. yakuba*, which has a genome organization that corresponds to the ancestral organization for insects and crustaceans (Boore et al. 1998, Flook et al. 1995).

## **4.3 Results and Discussion**

### ***4.3.1 Genome size and composition***

#### ***4.3.1.1 Vanhornia eucnemidarum***

The almost complete mt genome of *V. eucnemidarum* was sequenced (total of 16610 bp). We were not able to amplify the region comprising part of the *rns*, the control region, and four *trns*. This region has proven difficult to sequence in other insects, including other hymenopteran mt genomes [*Perga condei* (Castro and Dowton 2005) and *Melipona bicolor* (Silvestre and Arias, unpublished)]. Analysis of the sequence obtained revealed the typical gene content found in other metazoan mt genomes. All 13 protein-coding genes were entirely sequenced, the *rnl*, and 18 of the 22 *trns*. In addition, part of the *rns* subunit was also sequenced.

Several non-coding and overlapping regions were identified in the mt genome of *V. eucnemidarum*. The largest non-coding region was 663 bp long and is located between the *trnL<sup>UAA</sup>* and the *trnK* genes (Fig. 4.1). We initially thought that this region corresponded to the control region; however, it has a low AT content (76%) when compared with control regions of other closely related insects (between 86.9% and 96% in Diptera, and 96% in the honeybee). Other major non-coding regions ranged in size from 2 bp to 236 bp and were found throughout the genome, giving a total of 1199 noncoding nucleotides (Table 4.2). This is considerably more than other insect mt genomes, in which noncoding nucleotides outside the AT rich region can number less than 50 bp, and usually are limited to only 1 or 2 bp. In only a few other insects have long spacer sequences been reported. For example, *Pyrocoelia rufa* (Coleoptera) contains a 1724 bp long intergenic spacer sequence composed of twelve 134 bp tandem repeats plus one incomplete 116 bp repeat (Bae et al. 2004). Also, *Triatoma dimidiata* (Hemiptera) contains a 314 bp intergenic spacer sequence (Dotson and Beard 2001), while *Apis mellifera* (Hymenoptera) contains a 193 bp long intergenic sequence, which was thought to function as an additional origin of replication (Crozier and Crozier 1993). Cournuet et al. (1991) found that the 193 bp region of *Apis*, located between the *trnL* and the *cox2* genes, occurs as longer variants in other honeybee strains, with sizes ranging from 200 bp to 650 bp, and that it probably arose as a duplication of the *trnL* gene and the 3' end of the *coxI* gene. In our case, there is evidence of a *trn*-like structure at the 5' end of this non-coding region. However this *trn* does not resemble any of the 22 *trns*. Our attempts to find similar sequences elsewhere in the genome, or in GenBank, were unsuccessful, and there is no evidence that this intergenic spacer arose through duplication and subsequent degeneration of another gene.

Some *V. eucnemidarum* mt genes overlap, for a total of 30 bp in 4 locations (Table 4.2).

The longest overlap is 13 bp, involving the *atp8* and *atp6* genes. Overlaps between these two coding genes are a common feature of metazoan mt genomes (e.g. Campbell and Barker 1999), and had been reported in other hymenopteran taxa (Crozier and Crozier 1993, Castro and Dowton 2005).

As widely reported in other insect mtDNA sequences, the nucleotide composition of the *V. eucnemidarum* mt genome is biased towards adenine and thymine, with an average AT content of 80.1% (Table 4.3). The total AT bias of the *V. eunemidarun* mt genome is not as high as that of the honeybee [84.3% excluding the AT rich region (Crozier and Crozier 1993)], but higher than for *Perga condei* [78% (Castro and Dowton 2005)]. Third codon position nucleotides showed the highest AT content (86.4%), while first and second codon position nucleotides had AT content values less than the genome (79.1 and 72.6% respectively). As in other insect mt genomes, the observed base composition differences are reflected in the different composition of amino acids and codon usage. The total number of codons in the *V. eunemidarun* protein coding genes was 3644, excluding termination codons. This also falls within the range found in other insects, where values range from 3624 in *H. macropus* to 3746 in *Ceratis capitata* (Kim et al. 2005).

#### 4.3.1.2 *Primeuchroeus* sp.

Three portions of the mt genome comprising a total of 12341 bp were sequenced (Fig. 4.2).

We had difficulties amplifying the junctions between these regions and thus, some of the protein coding genes are slightly incomplete. Others have found it difficult to sequence sections of the mt genome due to the presence of repeat regions which interfere with the



sequencing process, and have suggested the presence of additional large repeat units (Lavrov and Brown 2001). In our case, the presence of this type of large repeat unit might be the reason for the failure to amplify some gene junctions. We systematically attempted amplifications of all possible gene junctions (60 in total; Table 4.4). In total, 12 of the 13 protein coding genes were sequenced, the large and the small *rrn* genes and 19 of the 22 *trn* genes, which is preliminary evidence that the gene content of the *Primeuchroeus* sp. mt genome is conserved when compared to the typical mt genome.

A total of 693 non-coding nucleotides are evident, with 616 bp in a single non-coding region. We proposed this might represent the control region (CR) due to its high AT content (86.9%), and proximity to the *rrn* genes; however, other characteristics of insect control regions such as T-stretch sections and a stem-loop (Zhang and Hewitt 1997) were not evident. The size of the control region in other insect genomes ranges from 440 in *Thrips imagines* (Shao and Baker 2003) to 4061 in *Drosophila melanogaster* (Lewis et al. 1994). The non-coding region of *Primeuchroeus* falls within this range. Other large intergenic regions were 38 bp (located between the two *trnL* genes), 16 bp (between the genes *cox1* and *cox2*), and 10 bp (located between the *cob* and *nd1* genes). There are only two cases where genes overlap. The largest overlap is between the *atp8* and the *atp6* genes. Again, this overlap is shared with a range of other metazoan mt genomes and was also seen in *Vanhornia*, *Perga* and *Apis* mt genomes. As in *V. eucnemidarum*, and other insect mt DNA sequences, the nucleotide composition of the *Primeuchroeus* genome is also A+T rich, with an average A+T content of 78.2% (Table 4.5).

### 4.3.2 Protein-coding genes and codon usage

#### 4.3.2.1 *Vanhornia eucnemidarum*

The size of the protein-coding genes in the *V. eucnemidarum* mt genome is very similar to that of the corresponding orthologous genes of *A. mellifera*, *M. bicolor*, and *P. condei* (Table 4.6). The more conserved gene is *cox1* with an average amino acid identity of 60.12%, while the less conserved genes are *nd2* and *nd6* with average amino acid identities of only 20.63% and 18.75% respectively (Table 4.6). This amino acid identity pattern is in agreement with those observed in other arthropods. However, even the most conserved gene, *cox1*, is generally less conserved among Hymenoptera than in other invertebrates (e.g., Lavrov et al. 2000, Negrisolo et al. 2004).

Conventional ATA or ATG start codons could be assigned to most of the protein-coding genes (Table 4.2). The *nd5* gene uses ATT (codes for Ile) as a start codon, which has also been reported in other insect mt genomes (Friedrich and Muqim 2003; Lessinger et al. 2000). Only the *cox1* gene has an unconventional TTG start codon. Unconventional start codons for *cox1* have been extensively reported in several arthropod species, including insects (Clary and Wolstenholme 1985, Beard et al. 1993, Spanos et al. 2000, Lessinger et al. 2000, among others).

Among insects, incomplete or even absence of recognizable termination codons have been reported (Clary and Wolstenholme 1985, Beard et al. 1993, Kim et al. 2005). In *V. eucnemidarum*, conventional stop codons were assigned to most of the protein genes (table 4.2). Only the *nd4* gene terminated with an incomplete (TA) codon. In these cases, it has

been proposed that the complete termination codon could be created by post-transcriptional polyadenylation, as observed in other animal phyla (Nardi et al. 2001, Ojala et al. 1981).

The codon usage of the *V. eucnemidarum* mt genome is shown in Table 4.7. There is a correlation between codon usage and nucleotide composition, as seen in other insect mt genomes. Leu, Phe, Ile, Met and Lys are the most frequently used amino acids as well as those with AT rich codons.

#### 4.3.2.2 *Primeuchroeus* sp.

Most of the protein-coding genes have ATG initiation codons; however, the *atp8* gene uses ATC, while *nd4L* and *nd6* use ATT. In contrast to *V. eucnemidarum*, the *Primeuchroeus* *cox1* gene has a traditional ATG start codon. As for *V. eucnemidarum*, conventional stop codons were assigned to most of the protein coding genes, except for *cox2* and *nd6*, that had incomplete stop codons (T and TA, respectively). Again, the A+T rich codons are the most frequently used (Tables 4.5, 4.8).

### 4.3.3 *Transfer RNA genes and ribosomal RNA genes*

#### 4.3.3.1 *Vanhornia eucnemidarum*

The predicted structures of the *V. eucnemidarum* *trn* genes are shown in Fig. 4.3. All *trn* genes have the typical cloverleaf structure except for *trnS*<sup>UCU</sup>. In this case, the dihydrouridine (DHU) arm forms a simple loop, as in several other metazoan species (including insects) (Wolstenholme 1992). The *V. eucnemidarum* *trn* genes are 64 to 71 bp

long, and the anticodons are identical to those observed in *Drosophila yakuba* (Clary and Wolstenholme 1985) and *Apis mellifera* (Crozier and Crozier 1993).

Both ribosomal genes are located on the minority strand and separated by the *trnV* gene, a common pattern in the mt genome of arthropods. Alignment of these regions with related insect sequences revealed numerous blocks of high sequence conservation, suggesting that the secondary structure elements were conserved (data not shown). The A + T content in the *rns* and *rnl* genes is 80.8% and 82.8% respectively. These percentages correspond to that found in other hymenopteran mt genomes, in which the AT content of the *rrn* genes are generally higher than the genome (Crozier and Crozier 1993, Castro and Dowton 2005).

#### 4.3.3.2 *Primeuchroeus* sp.

The *Primeuchroeus* sp. *trn* genes are shown in Fig. 4.4. 19 *trn* genes were identified on the basis of their respective anticodons and secondary structures. The length of the *trn* genes range from 66 to 74 bp and anticodon nucleotides were also the same as those described for other related insect species such as *Apis mellifera* and *Perga condei*.

Although *rrn* genes in arthropods and other reported hymenopteran mt genomes are usually encoded on the minority strand, in the *Primeuchroeus* genome the *rrn* genes were found together separated by the *trnV* gene as in the ancestral organization, but inverted to the opposite strand. Among insects, this situation has only been reported for *Thrips imaginis*, for which the *rrn* genes have undergone translocation, with both encoded on the majority strand (Shao and Barker 2003). In *H. sapiens*, the proximity of the *rrn* genes to the transcription promoter site (within the control region) ensures that the *rrn* genes are

expressed at a much higher rate than other mt genes (Montoya et al. 1982). Shao and Barker (2003) suggested that, in the *Thrips imaginis* case in which the *rrn* genes were not located close to each other nor to the control region, two other sets of promoters and termination elements may exist. In our case, there is a major non-coding region close to the *rnl* gene (with three *trn* genes in between), but sequencing of the region next to the *rns* gene is required in order to make further conclusions about the mechanisms of *rrn* transcription.

#### **4.3.4 Genome organization**

##### **4.3.4.1 Vanhornia eucnemidarum**

The mt genome of *D. yakuba* has the organization considered ancestral for hexapods. This inference is based on the observation that its mt genome has the same organization as the crustacean *Daphnia pulex*. Comparison of the *V. eucnemidarum* mtDNA arrangement with the corresponding arrangement in *D. yakuba* reveals a number of rearrangements, mostly involving *trn* genes (Fig. 4.1). Probably due to their small size, translocation and/or inversion of *trn* genes seem to be more common than positional changes involving protein-coding genes and ribosomal genes (Negrisolo et al. 2004). The only other unusual feature in the *V. eucnemidarum* mt genome is the presence of a non-coding region between the *trnL<sup>UAA</sup>* and the *trnK* genes, not present in the ancestral organization.

According to Dowton et al. (2003), *trn* gene rearrangements can be classified as translocations, local inversions or shuffling. A translocation is a movement of a gene to another position across a protein-coding gene. A rearrangement is classified as an inversion

when the *trn* is found on the opposite strand, and as shuffling when the *trn* gene is on the same mt strand, but in a different position compared to the ancestral organization (without movement across a protein coding gene). As shown in Fig. 4.1, comparison to the *Drosophila* mt genome reveals at least one translocation (*trnL*<sup>UAA</sup>), 2 translocation + inversions (*trnM*, *trnL*<sup>UAG</sup>), 2 inversions (*trnY*, *trnR*), and 2 shuffling events (*trnN*, *trnS*). Translocation and shuffling events probably took place by a duplication of a large region followed by loss of all but one or a few genes in a single or several deletion events (Lavrov et al. 2002), while the inversion events probably arose by intra-mt recombination as proposed by Dowton and Austin (1999), involving breakage and rejoining of the mt genome.

#### 4.3.4.2 *Primeuchroeus* sp.

The organization of the *Primeuchroeus* sp. genome is also unique compared to other hymenopteran or insect mt genomes (Fig. 4.2). Comparison of the *Primeuchroeus* sp. mtDNA arrangement with the corresponding arrangement in *D. yakuba* reveals at least 3 translocations (*trnI*, *trnM*, *trnL*<sup>UAA</sup>), 2 translocation + inversion events (*trnQ*, *trnH*), and 2 inversions (*trnY*, *trnR*). Additionally, the most remarkable rearrangement in *Primeuchroeus* appears to be an inversion involving the large and small *rrn* genes and including the *trnL*<sup>UAG</sup> and *trnV*. None of the protein coding genes appeared rearranged, although the *nd2* gene was not found. At least 9 *trn* genes in total are in a different position when compared to the ancestral organization.

#### 4.3.5 The ARNSEF region of *Vanhornia eucnemidarum*

The size of an animal mt genome is usually minimized over time (Rand 1993). It has been proposed that when duplications of a genomic region occur, they are generally followed by a rapid elimination of redundant copies, which may or may not result in a change of gene order (Macey et al. 1997). However, direct observations of this process are rarely documented. The evolution of mt genomes with multiple copies of some genes may provide a direct way to observe the evolution of mt genome rearrangements. The most remarkable rearrangement in the mt genome of *V. eucnemidarum* was reported by Dowton et al. (2003), and involves multiple, tandemly repeated copies of three *trn* genes (*trnS*, *trnN*, *trnA*) in the junction between the *nd3* and *nd5* genes, with some of the copies having point mutations that seem to have changed a functional gene into a pseudogene. By sequencing the same region in other *V. eucnemidarum* populations, we hoped to directly observe mutational changes that describe this rearrangement event. However the organization of genes in a second *V. eucnemidarum* population was exactly the same as that originally found in the first population sequenced. We found the 5-fold repetition of *trnS*, *trnN* and *trnA* genes and the same pseudogenes. Only one nucleotide mutation (C to T) in the *trnN* of the fifth repetition and one nucleotide deletion (of a T) in the non-coding region between the *trnN* and the *trnA* pseudogene of repeat 4 were found. Although this might suggest conservation of these pseudogenes, the sampled populations were only separated by 75 miles. Sampling of more remote populations of *Vanhornia* is beyond the scope of the present study, but should shed light on how long such pseudogenes remain in the mt genome.

#### **4.3.6 Comparison of gene rearrangements**

The phylogenetic relationships among the apocritan wasps remain to be fully described (Dowton and Austin 2001). One of the aims of our study was to examine whether mt gene rearrangements might provide informative phylogenetic characters to resolve hymenopteran relationships. However, none of the rearrangements reported here are shared among the present sample of Hymenoptera whose mt genome has been sequenced. The present study indicates that taxonomic sampling of apocritan mt genomes remains insufficient to infer phylogeny or to examine the mechanism of gene rearrangement. The degree of change between sequenced genomes is presently too great to reconstruct these events.

#### **4.4 Conclusion**

I sequenced most of the mitochondrial (mt) genomes of two apocritan taxa: *Vanhornia eucnemidarum* and *Primeuchroeus* sp. These mt genomes have similar nucleotide composition and codon usage to mt genomes reported for other Hymenoptera (Crozier and Crozier 1993, Castro and Dowton 2005), with a total A+T content of 80.1% and 78.2% respectively. Gene content corresponds to other metazoan mt genomes, however, gene organization is not conserved. There are a total of six *trn* genes rearranged in *Vanhornia eucnemidarum* and nine in *Primeuchroeus* sp. relative to the ancestral genome organization. Additionally, several non-coding regions were found in the mt genome of *Vanhornia eucnemidarum*, as well as evidence of a sustained gene duplication involving three *trn* genes. I also report an inversion of the large and small *rrn* genes in *Primeuchroeus* sp. mt genome. However, none of the rearrangements reported are phylogenetically informative with respect to the current taxon sample.



**Table 4.1**

Primer pairs used in PCR amplification of the *V. eucnemidarum* mt genome

Primer pairs used in PCR amplification of the *Primeuchroeus* mt genome

**Table 4.2**Mt gene profile of *V. eucnemidarum* (Hymenoptera: Vanhnormidae)

Feature	Position number		Size	Strand	Codon	Stop	Intergenic
	From	To	(bp)		start		nucleotides
<i>trnL</i> <sup>(UAG)</sup>	1	69	69	+			
<i>nd2</i>	132	1154	1023	+	ATA	TAA	62
<i>trnW</i>	1153	1219	67	+			-2
<i>trnC</i>	1212	1275	64	-			-8
<i>trnY</i>	1286	1351	66	-			12
<i>cox1</i>	1355	2926	1572	+	TTG	TAA	3
<i>cox2</i>	3034	3717	684	+	ATA	TAA	107
<i>trnL</i> <sup>UAA</sup>	3728	3794	67	+			10
<i>trnK</i>	4458	4525	68	+			663
<i>trnD</i>	4543	4610	68	+			17
<i>atp8</i>	4770	4931	162	+	ATA	TAA	159
<i>atp6</i>	4919	5602	684	+	ATG	TAA	-13
<i>cox3</i>	5609	6403	795	+	ATG	TAA	6
<i>nd3</i>	6475	6876	402	+	ATG	TAA	71
<i>trnS</i> <sup>UCU</sup>	6895	6962	68	-			18
<i>trnN</i>	6972	7035	64	+			9
<i>trnA</i>	7049	7115	67	+			13
<i>trnS</i> <sup>UCU</sup>	7142	7206	65	-			26
<i>trnN</i>	7215	7279	65	+			8
<i>trnA</i>	7295	7360	66	+			15
<i>trnS</i> <sup>UCU</sup>	7387	7451	65	-			26
<i>trnN</i>	7460	7524	65	+			8
<i>trnA</i>	7540	7603	64	+			15
<i>trnS</i> <sup>UCU</sup>	7631	7695	65	-			27
<i>trnN</i>	7705	7769	64	+			9
<i>trnA</i>	7785	7849	65	+			15
<i>trnS</i> <sup>UCU</sup>	7877	7941	65	-			27
<i>trnN</i>	7951	8015	65	+			9
<i>trnA</i>	8030	8094	65	+			14
<i>trnS</i> <sup>UCU</sup>	8121	8165	45	-			26
<i>trnE</i>	8226	8293	68	-			60
<i>trnF</i>	8350	8415	66	-			56
<i>nd5</i>	8446	10126	1558	-	ATT	TAA	153
<i>trnH</i>	10129	10195	67	-			2
<i>nd4</i>	10196	11539	1345	-	ATG	TA*	0
<i>nd4L</i>	11533	11827	295	-	ATA	TAA	-7
<i>trnT</i>	11828	11895	68	+			0
<i>trnP</i>	11895	11961	67	-			-1
<i>nd6</i>	11964	12299	328	+	ATG	TAA	2
<i>cob</i>	12528	13667	1140	+	ATG	TAA	236
<i>trnS</i> <sup>(UGA)</sup>	13680	13750	71	+			12
<i>nd1</i>	13763	14713	956	-	ATA	TAA	9
<i>trnM</i>	14722	14788	67	-			6
<i>rnl</i>	14789	16115	1325	-			
<i>trnV</i>	16116	16179	64	-			
<i>rns</i>	16180	16568	388	-			

**Table 4.3**AT content in different regions of the *V. eucnemidarum* mt genome

	TOTAL	%T	%C	%A	%G	%AT
All sites	16568	36.6	13.2	23.3	6.7	80.1
<i>nd2</i>	1023	41.4	12.0	42.2	4.3	83.6
<i>cox1</i>	1569	39.5	17.3	35.6	11.1	71.6
<i>cox2</i>	684	37.1	17.0	36.8	9.1	73.9
<i>atp8</i>	162	36.4	13.6	46.3	3.7	82.7
<i>atp6</i>	684	38.9	17.3	37.4	6.4	76.3
<i>cox3</i>	795	39.0	15.7	38.0	7.3	77.0
<i>nd3</i>	399	38.3	13.0	41.1	7.5	79.4
<i>nd5</i>	1555	50.5	6.0	29.1	14.4	79.6
<i>nd4</i>	1349	47.2	5.9	33.3	13.6	80.5
<i>nd4L</i>	300	51.0	4.0	33.7	11.5	84.7
<i>nd6</i>	325	45.2	12.9	36.9	4.9	82.1
<i>cob</i>	1137	39.7	15.7	36.0	8.6	75.5
<i>nd1</i>	950	47.9	6.9	31.2	14.0	79.1
Protein coding total	10932	42.6	11.9	35.4	10.1	78
First codon positions	3644	36.0	11.1	40.1	12.8	76.1
Second codon positions	3644	49.2	16.3	22.4	12.1	71.6
Third codon positions	3644	42.7	8.3	43.7	5.4	86.4
Ribosomal RNA	5666	43.2	5.5	39.2	12.1	82.4

**Table 4.4**

Possible gene junctions in *Primeuchroeus* mt genome. Primers were carefully designed at the end of the genes and used in PCR reactions assuming all possible orientations

**Possible gene junctions in *Primeuchroeus* mt genome**

<i>cox3-nd3</i>	<i>cox3-nd5</i>
<i>cox3-nd4</i>	<i>cox3-nd1</i>
<i>cox3-rns</i>	<i>nd3-rns</i>
<i>nd3-nd4</i>	<i>nd3-nd1</i>
<i>nd3-nd5</i>	<i>nd5-nd4</i>
<i>nd5-nd1</i>	<i>nd5-rns</i>
<i>nd1-rns</i>	<i>nd4-rns</i>
<i>nd1-nd4</i>	

**Table 4.5**AT content in different regions of the *Primeuchroeus* sp. mt genome

	TOTAL	%T	%C	%A	%G	%AT	Start codon	Stop codon	
All sites	12341	39.0	10.6	39.2	11.2	78.2			
<i>cox1</i>	1602	39.5	10.5	32.1	17.9	71.6	ATG	TAA	<i>cox2</i>
	678	40	9.9	33.0	17.1	73	ATG	T	
<i>atp8</i>	159	39.6	5.7	43.4	11.3	83	ATC	TAA	
<i>atp6</i>	657	39.1	14.3	36.5	10	75.6	ATG	TAA	
<i>cox3</i> *	680	35.0	17.8	35.4	11.8	70.4	ATG		
<i>nd3</i> *	141	48.9	2.1	27.7	21.3	76.6		TAA	
<i>nd5</i> *	594	36.9	16.7	42.9	3.5	79.8		TAA	
<i>nd4</i> *	648	48.8	3.4	32.3	15.6	81.1	ATG		
<i>nd4L</i>	318	55.7	0.6	30.2	13.5	86.5	ATT		
<i>nd6</i>	570	42.6	16.8	37.9	2.6	80.5	ATT	TA	
<i>cob</i>	1143	40.2	18.9	32.8	8.0	73	ATG	TAG	
<i>nd1</i> *	807	43.5	5.9	34.3	16.2	77.8		TAA	
Protein coding total	7997	41.2	11.8	34.5	12.5	75.7			
First codon positions	2666	34.2	11.8	38.0	15.9	72.2			
Second codon positions	2666	47.7	15.0	24.3	12.9	72			
Third codon positions	2666	41.7	8.6	41.0	8.6	82.7			
Ribosomal RNA*	2275	43.1	13.4	38.7	4.8	81.8			

\*incomplete gene

**Table 4.6**

Comparison of the mt protein-encoding genes of *V. eucnemidarum* with those of other hymenopteran mt genomes available in GenBank. Protein lengths were inferred from the original sequences obtained from GenBank. Numbers in parentheses indicate percentage identity with *V. eucnemidarum* (number of identical amino acid in the pairwise alignment performed with ClustalW)

**Table 4.7**Codon Usage in the *V. eucnemidarum* mt genome

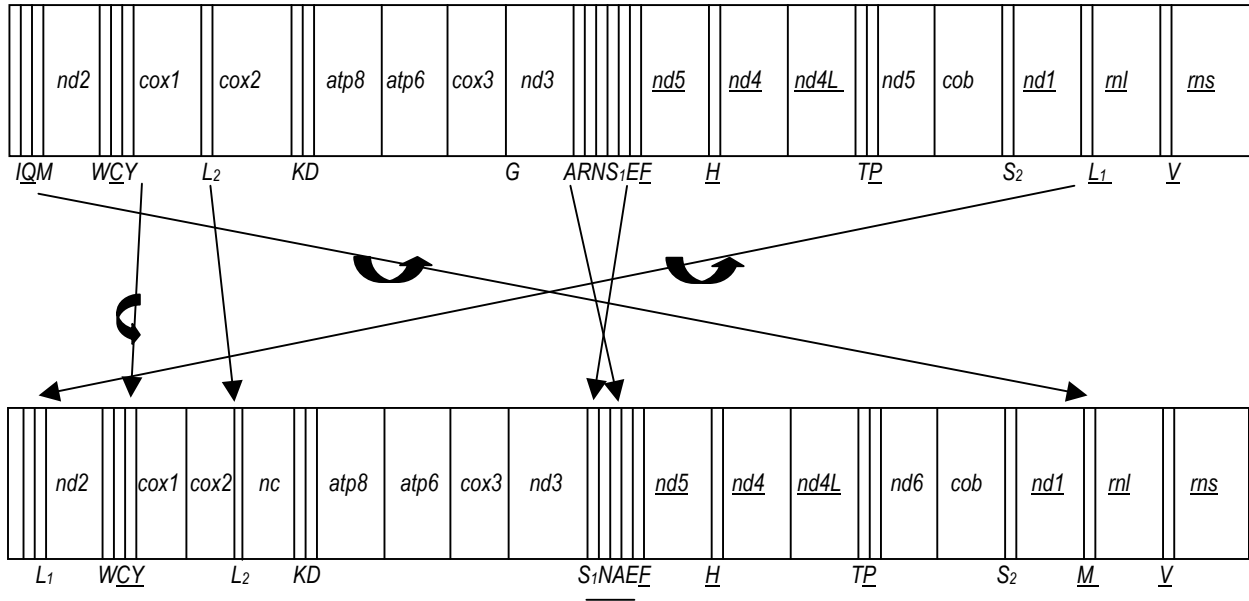
Amino acid	Codon	n	RSCU
Phe [F]	UUU	268	1.59
	UUC	69	0.41
Leu [L]	UUA	330	3.76
	UUG	59	0.67
	CUU	48	0.55
	CUC	8	0.09
	CUA	82	0.93
	CUG	0	0
Ile [I]	AUU	358	1.69
	AUC	66	0.31
Met [M]	AUA	326	1.72
	AUG	54	0.28
Val [V]	GUU	69	2.21
	GUC	5	0.16
	GUA	48	1.54
	GUG	3	0.10
Ser [S]	UCU	100	2.01
	UCC	21	0.42
	UCA	148	2.97
	UCG	4	0.08
Pro [P]	CCU	42	1.49
	CCC	10	0.35
	CCA	60	2.12
	CCG	1	0.04
Thr [T]	ACU	67	1.87
	ACC	13	0.36
	ACA	61	1.71
	ACG	2	0.06
Ala [A]	GCU	34	2.13
	GCC	7	0.44
	GCA	23	1.44
	GCG	0	0
Tyr [Y]	UAU	165	1.75
	UAC	24	0.25
His [H]	CAU	45	1.58
	CAC	12	0.42
Gln [Q]	CAA	45	1.73
	CAG	7	0.27
Asn [N]	AAU	205	1.59
	AAC	53	0.41
Lys [K]	AAA	106	1.64
	AAG	23	0.36
Asp [D]	GAU	50	1.79
	GAC	6	0.21
Glu [E]	GAA	52	1.55
	GAG	15	0.45
Cys [C]	UGU	26	1.86
	UGC	2	0.14
Trp [W]	UGA	85	1.91
	UGG	4	0.09
Arg [R]	CGU	11	1.05
	CGC	0	0
	CGA	31	2.95
	CGG	0	0
Ser [S]	AGU	23	0.46
	AGC	1	0.02
	AGA	92	1.85
	AGG	9	0.18
Gly [G]	GGU	39	1.01
	GGC	4	0.10
	GGA	96	2.48
	GGG	16	0.41

**Table 4.8**Codon Usage in the *Primeuchroeus* sp. mt genome

Amino acid	Codon	n	RSCU
Phe [F]	UUU	202	1.70
	UUC	35	0.30
Leu [L]	UUA	240	3.90
	UUG	50	0.81
	CUU	30	0.49
	CUC	11	0.18
	CUA	38	0.62
	CUG	0	0
Ile [I]	AUU	254	1.68
	AUC	49	0.32
Met [M]	AUA	186	1.65
	AUG	40	0.35
Val [V]	GUU	70	2.041
	GUC	4	0.12
	GUA	56	1.64
	GUG	7	0.20
Ser [S]	UCU	43	1.42
	UCC	14	0.46
	UCA	81	2.67
	UCG	6	0.20
Pro [P]	CCU	40	1.52
	CCC	20	0.76
	CCA	41	1.56
	CCG	4	0.15
Thr [T]	ACU	34	1.43
	ACC	12	0.51
	ACA	46	1.94
	ACG	3	0.13
Ala [A]	GCU	21	1.50
	GCC	7	0.50
	GCA	26	1.86
	GCG	2	0.14
Tyr [Y]	UAU	119	1.72
	UAC	19	0.28
His [H]	CAU	49	1.58
	CAC	13	0.42
Gln [Q]	CAA	31	1.63
	CAG	7	0.37
Asn [N]	AAU	156	1.69
	AAC	29	0.31
Lys [K]	AAA	84	1.58
	AAG	22	0.42
Asp [D]	GAU	40	1.74
	GAC	6	0.26
Glu [E]	GAA	39	1.32
	GAG	20	0.68
Cys [C]	UGU	19	1.65
	UGC	4	0.35
Trp [W]	UGA	54	1.66
	UGG	11	0.34
Arg [R]	CGU	3	0.39
	CGC	2	0.26
	CGA	22	2.84
	CGG	4	0.52
Ser [S]	AGU	7	0.23
	AGC	0	0
	AGA	77	2.53
	AGG	15	0.49
Gly [G]	GGU	26	0.82
	GGC	4	0.13
	GGA	59	1.86
	GGG	38	1.20



*Drosophila yakuba* mt genome arrangement



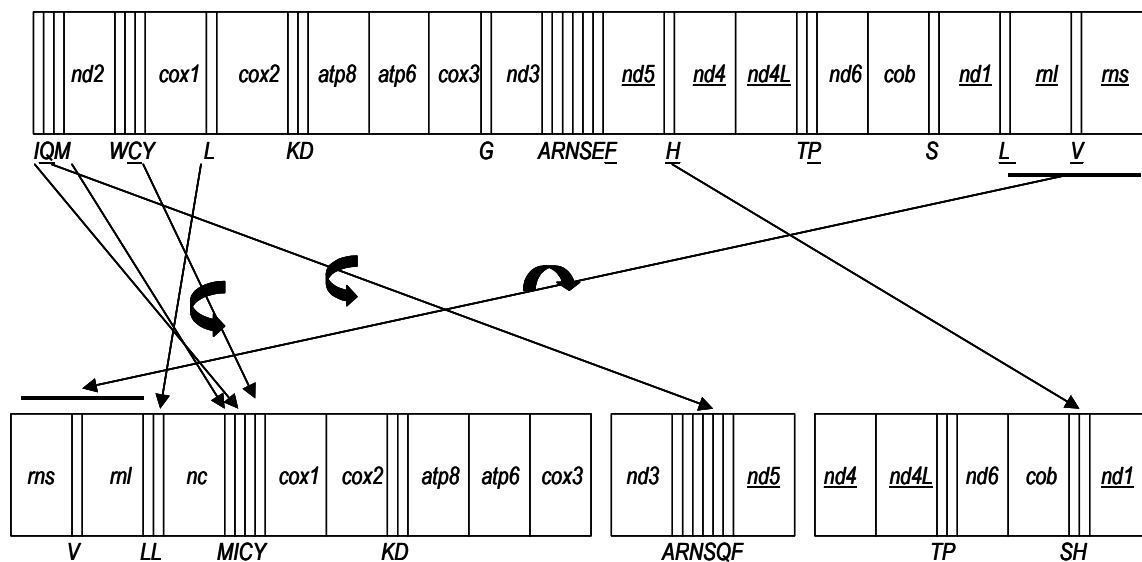
*Vanhornia eucnemidarum* mt genome arrangement

**Figure 4.1**

Linearized representation of the mt genome arrangement of *Vanhornia eucnemidarum*

(Hymenoptera: Apocrita) and the putative ancestral arthropod organization represented by *Drosophila yakuba*. Genes are transcribed from left to right except those indicated by underlining, which are transcribed from right to left. *trn* genes are designated by single-letter amino acid codes except those encoding leucine and serine, which are labeled  $L_1$  *trnL* (UAG),  $L_2$  *trnL* (UAA),  $S_1$  *trnS* (UCU),  $S_2$  *trnS* (UGA). Arrows indicate differences in gene locations between *D. yakuba* and *V. eucnemidarum*. The circling arrows indicate inversions.

*Drosophila yakuba* mt genome arrangement



*Primeuchroeus sp.* mt genome arrangement

**Figure 4.2**

Linearized representation of the mt genome arrangement of *Primeuchroeus sp.*

(Hymenoptera: Apocrita) and the putative ancestral arthropod organization represented by

*Drosophila yakuba*. Genes are transcribed from left to right except those indicated by

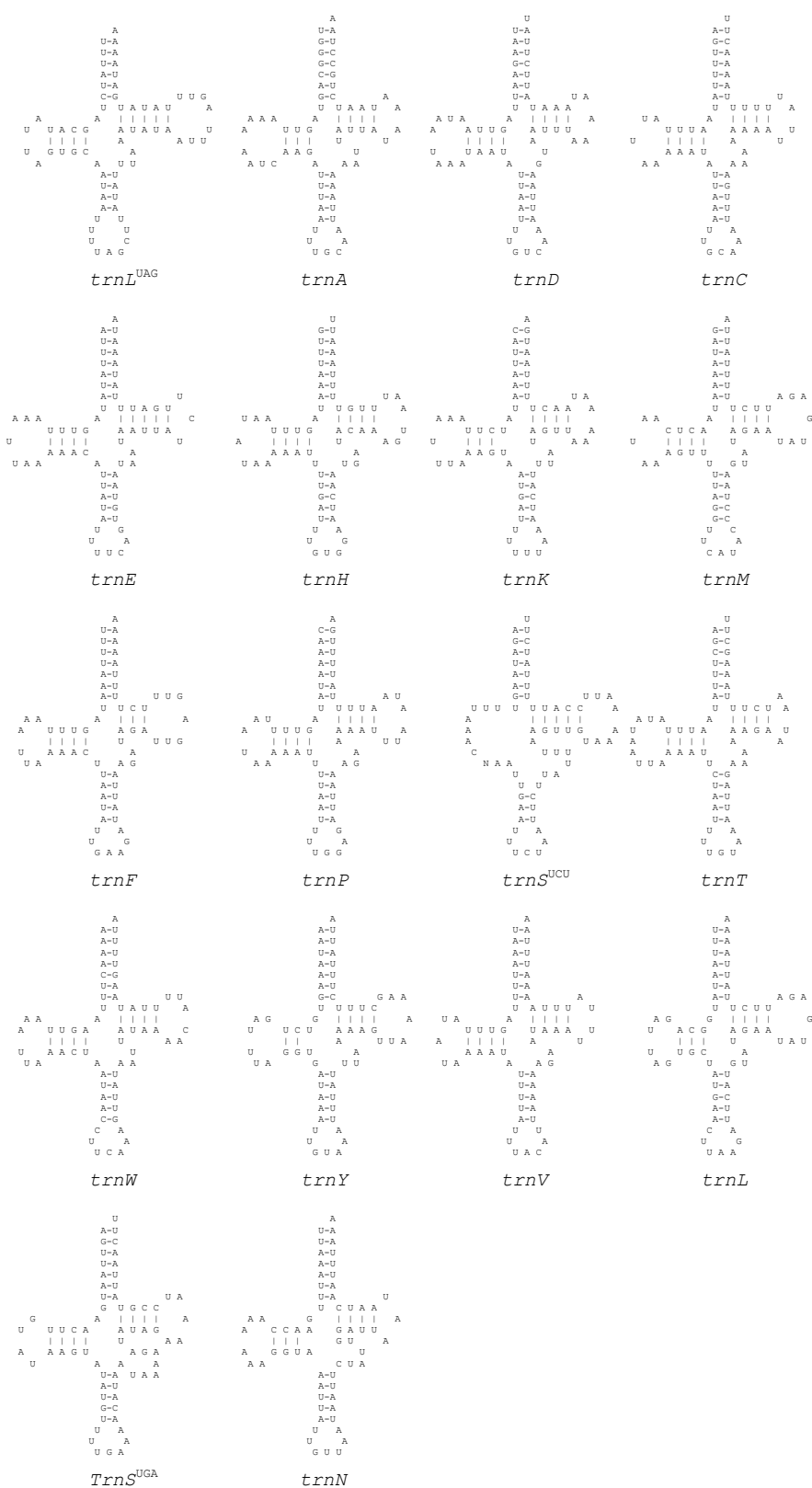
underlining, which are transcribed from right to left. *trn* genes are designated by single-

letter amino acid codes except those encoding leucine and serine, which are labeled *L<sub>1</sub> trnL*

(UAG), *L<sub>2</sub> trnL* (UAA), *S<sub>1</sub> trnS* (UCU), *S<sub>2</sub> trnS* (UGA). Arrows indicate differences in gene

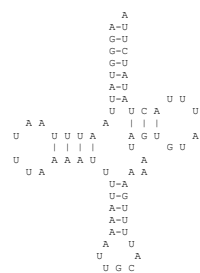
locations between *D. yakuba* and *Primeuchroeus sp.*. The circling arrows indicate

inversions.

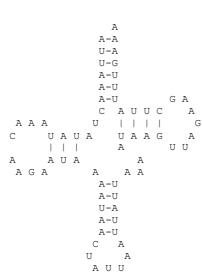


**Figure 4.3**

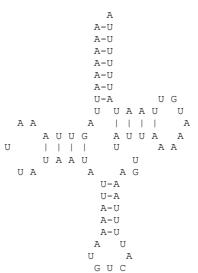
Putative secondary structure for the 18 *trn* genes of the *V. eucnemidarum* mt genome.



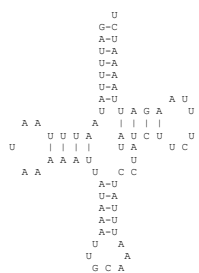
*trnA*



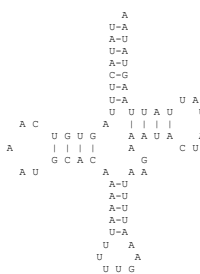
*trnN*



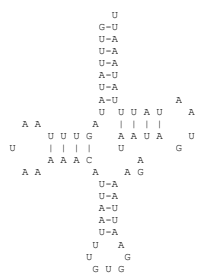
*trnD*



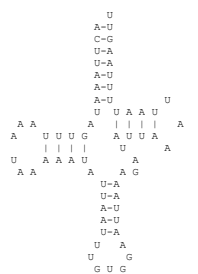
*trnC*



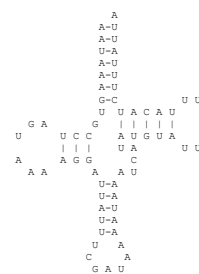
*trnQ*



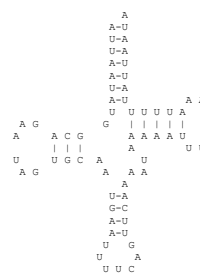
*trnE*



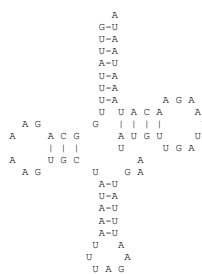
*trnH*



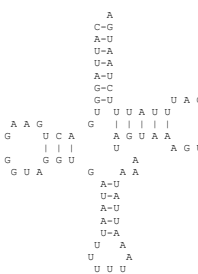
*trnI*



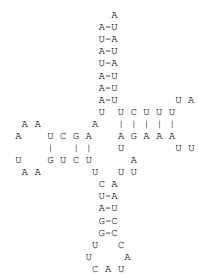
*TrnL<sup>UAA</sup>*



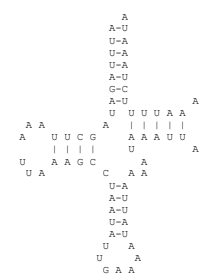
*trnL<sup>UAG</sup>*



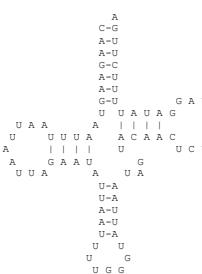
*trnK*



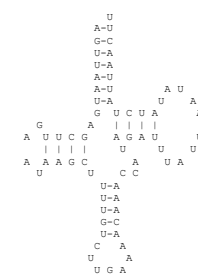
*trnM*



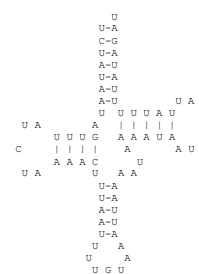
*trnF*



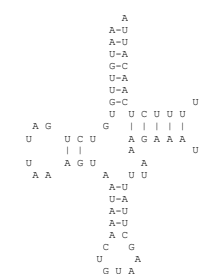
*trnP*



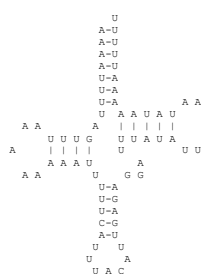
*trnS<sup>UGA</sup>*



*trnT*



*trnY*



*trnV*

**Figure 4.4**

Inferred secondary structure of the *trn* genes of the mt genome of *Primeuchroeus sp.*. The *trn* genes are labeled with the abbreviations of their corresponding amino acids.

## **Chapter 5 . Mitochondrial genomes in the Hymenoptera and their utility as phylogenetic markers**

This chapter was slightly modified from a paper accepted in the journal *Systematic Entomology*.

### **5.1 Introduction**

The accumulation of complete mitochondrial (mt) DNA sequences in the gene databases has lead to an interest in the use of entire mt coding sequence data for resolving deep level phylogenies (Arnason et al. 2002, Miya et al. 2001, Saccone et al. 1999). Several phylogenetic analyses have demonstrated that the use of complete mt genomes in phylogenetic studies significantly increases the confidence of the phylogenetic history inferred compared with phylogenetic hypotheses based on individual or partial mt genes (Cummings et al. 1995, Russo et al. 1996, Zardoya and Meyer 1996). The longer sequences obtained by sequencing complete mt genomes have encouraged attempts to reconstruct relationships among divergent lineages. However, complete mt genomes have sometimes failed to recover phylogenetic relationships supported with other markers (Curole and Kocher 1999). This suggests that the limits and applicability of these data remain to be elucidated.

For example, whole mt genome sequences were successfully used to address questions concerning bat evolution (Lin and Penny 2001). Lin et al. (2002) also reported an increased stability of the evolutionary tree of mammals after improved taxon sampling. A major finding of their study was that additional taxa reduced long branch attraction (LBA) problems, and allowed them to obtain congruent information with morphological and nuclear datasets. The utility of mt genomes in the phylogeny of mammals was also corroborated by Nikaido et al. (2003), who demonstrated the importance of using appropriate substitution models and species sampling when inferring phylogenetic relationships. In the same way, Phillips and Penny (2003) found that mt genomes favoured Theria (marsupials plus placentals) over Marsupionta (monotremes plus marsupials), in agreement with morphological and nuclear studies, after nucleotides were recoded as RY-characters, and maximum-likelihood analyses were partitioned among subsets of data.

Within the insects, mt genomes have been successful in recovering phylogenetic relationships concordant with traditional views of phylogeny and with convincing levels of support within the Diptera (Lessinger et al. 2000, Junqueira et al. 2004, Cameron et al. in review). Lessinger et al. (2000) performed phylogenetic analyses, both with nucleotides and amino acid data sets, and with the inclusion or exclusion of some genes to minimize possibly inconsistent alignments. All analyses recovered the same, well supported relationships, and were consistent with traditional groupings. After extending the dataset to eight dipteran taxa Junqueira et al. (2004) also supported the monophyly of the Nematocera (mosquitoes) and the Brachycera (flies), and within the Brachycera the monophyly of the Calyptratae and the Acalyptratae, consistent with Lessinger et al. (2000) and the classic interpretation of dipteran phylogeny (Junqueira et al. 2004).

On the other hand, some studies that have used concatenated mt gene sequences for resolving phylogenies have produced results which are difficult to reconcile with trees produced using other markers, or that contradict traditionally accepted phylogenetic relationships. For example, complete mt genomes have failed to answer a number of questions regarding vertebrate relationships. With respect to tetrapod origins, different analyses supported different topologies with high bootstrap values, and the results were not sufficient to distinguish a lungfish + coelacanth clade from a lungfish + tetrapod clade (Zardoya and Meyer 1996, Zardoya et al. 1998). Another controversial question is the relationship of the two extant lineages of jawless fishes to the remaining vertebrates, which after including the mt genomes of a lamprey (*Petromyzon marinus*), a hagfish (*Myxine glutinosa*) and several teleost fishes, was still not resolved (Curole and Kocher 1999).

With reference to examination of insect relationships, some analyses using mt genomes have had to exclude taxa, probably due to their compositional bias or increased rates of nucleotide substitution. Several studies have excluded the honeybee mt genome in order to recover a monophyletic Holometabola (Black and Roehrdanz 1998, Friedrich and Muqim 2003, Nardi et al. 2001, Nardi et al. 2003, Stewart and Beckenbach 2003). Most studies have also had to exclude the *Heterodoxus* (Insecta:Phthiraptera) and *Thrips* (Insecta:Phthiraptera) genome sequences due to evidence of unusual sequence evolution of these two taxa (Foster and Hickey 1999). Additionally, Hassanin et al. (2005) have shown evidence of the dramatic consequences that mutational saturation and heterogeneity of nucleotide composition among taxa have on phylogenetic analyses using mt genomes.



With respect to the phylogeny of the hexapods, Cameron et al. (2004) examined a number of variables that affect the phylogenetic outcome when using entire mt genomes (e.g. outgroup selection, ingroup taxon selection and alignment methodology, choice of genes). However, they were unable to confidently recover the sister-group of Collembola or make any conclusions regarding the monophyly of the Hexapoda. Instead they concluded that mt genome data alone were insufficient to resolve the phylogeny and that the dataset appeared highly vulnerable to taxon selection, outgroup choice, data manipulation and gene selection.

The relatively limited number of species for which complete mt genome sequences are currently available make it difficult to make any firm conclusions about the usefulness of concatenated gene sequences in phylogenetic reconstruction. However, the apparently correct phylogenies recovered when a better taxon sampling is available (as is the case for the Diptera within the insects, or the mammals) indicates that whole mt genomes still represent promising candidates for resolving phylogenies.

Resolving relationships among the Apocrita (wasps, bees, ants) has proven difficult (Whitfield 2002, Dowton and Austin 2001). In the most recent analysis, Dowton and Austin (2001) sampled 84 taxa and generated character information from three genes, and included morphology. Although this was a significant improvement over previous attempts, a range of relationships were still not stably recovered and were sensitive to the model of analysis. Complete mt genomes seem useful candidates to improve the resolution of phylogenetic relationships of the group, although sequencing complete genomes in such a large and diverse group is a major undertaking. There are five hymenopteran species whose

mt genomes have been sequenced. The phylogenetic relationships of these five species in uncontroversial and well supported by morphology and other molecular studies (Rasnitsyn 1988, Ronquist et al. 1999, Dowton and Austin 1994, 2001, Carpenter 1999, Brothers and Carpenter 1999) (Fig. 5.1). In this study we used these mt genomes to assess the utility of complete mt genomes to resolve phylogenetic relationships among the Hymenoptera. In addition, we assessed the utility of these genomes to recover the position of the Hymenoptera among other holometabolan insects.

## **5.2 Materials and Methods**

### ***5.2.1 Taxon Sampling***

The sequences of *Vanhornia eunemidarum* and *Primeuchroeus* sp., as well as the complete mtDNA coding sequences of other insects obtained from GenBank were used in this analysis. In order to avoid over-sampling of some clades of Diptera or Lepidoptera, we only included one representative of each of the available genera (Table 5.1). *Locusta migratoria* was used as the outgroup in most analyses including. Other analyses were conducted with *Triatoma dimidiata* as outgroup, and a final unrooted tree was generated with the purpose of testing whether ingroup topology was susceptible to outgroup selection. Additional analyses were performed including only the hymenopteran representatives. In these cases either *Drosophila melanogaster* (Diptera) or *Crioceris duodecimpunctata* (Coleoptera) was selected as the outgroup.

### **5.2.2 Alignment**

Individual protein coding genes were translated using the *Drosophila* mtDNA genetic code using the program Translation Machine ([www.2ebi.ac.uk/translate/](http://www.2ebi.ac.uk/translate/)). For each of the protein coding genes, amino acid sequences were initially aligned using CLUSTAL W (Thompson et al. 1994). The alignments were manually adjusted to avoid any ambiguous amino acid pairings. Nucleic acid sequences were aligned using PROTAL2DNA program (Schuerer, K. and Letondal, C. <http://bioweb.pasteur.fr/seqanal/interfaces/protal2dna.html>), which aligns DNA sequences based on the protein alignments. rRNA nucleotide sequences were also aligned using CLUSTAL W. The individual alignments were combined to generate a single alignment. The genes were arranged in the order they occur on the first strand of the *D. yakuba* sequence.

### **5.2.3 Phylogenetic analysis**

All phylogenetic analyses were conducted in PAUP\* version 4.0b10 (Swofford 1998) or MrBayes version 3.1 (Huelsenbeck and Ronquist 2001). Maximum parsimony (MP) (Fitch 1971) and Bayesian approaches were used for phylogenetic analysis using both amino acid and nucleotide sequences. MP analysis was used with all characters weighted equally, and gaps treated as missing data. Non-parametric bootstrapping was performed using a full heuristic search with 100 replicates. Bayesian analyses were conducted using the mtREV24 model of protein evolution (Adachi and Hasegawa 1996) when using amino acid sequences. For the nucleotide dataset, the general time reversible model with some sites assumed to be invariable and with variable sites assumed to follow a discrete gamma distribution [GTR + I +  $\Gamma$ ; Yang (1994)] was selected as best-fit model of nucleotide

substitution [MrModeltest, Nylander (2004)]. Previous analyses had suggested that partitioned analyses were superior to unpartitioned analyses (Castro and Dowton 2005), therefore, the dataset was divided into five partitions (1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup> codon positions, *rrnS*, and *rrnL*). To avoid the possible saturation effect of third codon positions (Fitch 1986), we performed an additional analysis excluding 3<sup>rd</sup> codon positions. In these cases, we set the maximum likelihood parameters in MrBayes as follows: “lset nst=6” (GTR), “rates=invgamma” (I +  $\Gamma$ ). The Markov chain Monte Carlo process was set so that four chains (three heated and one cold) ran simultaneously. We conducted two independent runs for 1,000,000 generations, with trees being sampled every 100 generations, each of which started from a random tree. Independent analyses indicated that “stationarity” (or “burnin”: lack of improvement in ML scores) was reached at no later than 70,000 generations; thus, the first 700 trees were discarded from each analysis as the burnin, and the remaining trees were used to generate a 50% majority consensus tree, with the percentage of samples recovering any particular clade representing that clades’ posterior probability (Huelsenbeck and Ronquist 2001). As the results from MrModeltest do not indicate whether the covarion option should be included, we conducted analyses with and without the covarion option. According to Posada & Crandall (1998) the use of a model is only justifiable if it significantly increases the likelihood of the analysis; the likelihoods of the analyses both with and without the covarion option were compared using Bayes factors as describe by Nylander et al. (2004).

### 5.3 Results and Discussion

### ***5.3.1 Model Specification and Stationarity***

The choice of models is relatively straightforward for Bayesian analyses with the use of MrModeltest. However, MrModeltest does not distinguish whether a covarion model, (which allows rate variation across sites and also allows those rates to change through time [Penny et al. 2001]) should be applied. This may be particularly important in analyses utilizing hymenopteran mt genomes; however our results showed that there are no differences in the tree topologies of the analyses when covarion is set to YES.

Interpretation of Bayes factors indicated that in general there is a significant difference in the likelihoods between the two runs, and in all the analyses the simplest model (covarion=NO) tended to be favored (Table 5.2). When independent analyses were performed on the same dataset, the likelihoods were very similar, suggesting that stationarity had been reached. Similarly, the average standard deviation of split frequencies was less than 0.01, reflecting that the tree samples from the different runs became increasingly similar and suggesting convergence between the four runs. After 1,000,000 generations, the marginal posterior distributions of gamma and proportion of invariant sites remained focused, and the chain rapidly mixed over them.

### ***5.3.2 Hymenopteran Relationships***

#### ***5.3.2.1 All taxa included***

*Bayesian analyses:* Bayesian phylogenetic analyses including the 21 taxa, both with all codons included or 3<sup>rd</sup> codon positions excluded, recovered the same within-hymenopteran relationships with posterior probability values exceeding 0.95 (Fig. 5.2A, B, C). In these

cases, the Apoidea (represented by *Apis mellifera* and *Melipona bicolor*) is recovered as monophyletic, as is the Aculeata, (Apoidea + *Primeuchroeus sp.*). The Symphyta (*Perga condei*) is recovered at the base of the Hymenoptera. In general, the hymenopteran phylogenetic relationships obtained in all Bayesian analyses using nucleotide sequences conform to the test phylogeny. However, Bayesian analyses using amino acids did not recover the test hymenopteran phylogeny (Fig. 5.2D, E), with *Vanhornia* disrupting the Aculeata. Holometabolan relationships recovered with these analyses are discussed below.

*Parsimony analyses:* Parsimony analyses (of both proteins and nucleotides) including the 21 taxa recovered different within-hymenopteran relationships, with *Vanhornia* (a member of the Proctotrupomorpha) as sister to the Apoidea, disrupting the Aculeata (as in Fig. 5.3A-F). In this case, the hymenopteran relationships recovered are not supported by the traditional view shown in Fig 5.1.

#### 5.3.2.2 Only Hymenoptera included

*With single outgroup (Coleoptera or Diptera):* When analyses were performed including only the hymenopteran representatives, with a single coleopteran or a dipteran as outgroup, both bayesian analyses (proteins and nucleotides including all codon positions and rRNA genes) and parsimony analyses (proteins and nucleotides) recovered the same topology, in which *Vanhornia* disrupted the monophyly of the Aculeata (Figure 5.4A). In this case, only Bayesian analyses excluding the 3<sup>rd</sup> codon positions recovered again the Aculeata as monophyletic (Figure 5.4B, Table 5.3).

*Unrooted analysis:* An unrooted analysis including just the hymenoptera representatives recovered the same relationships, in which *Vanhornia* is more closely related to the Apoidea than *Primeuchroeus* (data not shown). This result contradicts morphological and previous molecular views.

Clearly, the utility of mt genomes to resolve phylogenetic relationships within the Hymenoptera is variable (Table 5.3 shows the results for each analysis). At this level, the phylogenetic analyses were susceptible to analytical model. Hymenopteran relationships were reliably recovered only when all 21 taxa were included, and only Bayesian analyses excluding 3<sup>rd</sup> codon positions were able to recover the test phylogeny when a single taxon was selected as the outgroup to the Hymenoptera. This suggests that exclusion of 3<sup>rd</sup> codon positions is in fact a better way of analyzing this type of data, and that a complete taxon sampling, not only of ingroup but of outgroup taxa, is crucial.

These results contrast with Lessinger et al. (2000) and Junqueira et al. (2002) who performed a phylogenetic analyses of dipteran mt genomes, and independent of the type of model of analyses always recovered very strong well supported relationships. However, it has been reported that mtDNA evolution in the Apocrita is particularly rapid and that these taxa have extreme compositional bias (Dowton and Austin 1995, 1997). These factors may make phylogenetic recovery more challenging within the Hymenoptera, compared to the Diptera. Our results suggest that an extensive taxon sampling will be required to confidently resolve hymenopteran relationships, and that, as more Apocritan taxa are sequenced, inclusion of other Holometabolan taxa might be necessary in recovering an appropriate hymenopteran phylogeny.

### 5.3.3 Holometabolan Relationships

Currently, morphological and molecular data support the monophyly of most of the holometabolous insect orders, however, the most recent expanded molecular data sets of holometabolous insects including *18S* and *28S* genes failed to confirm some relationships hypothesized by morphology, such as Mecopterida, Hymenoptera + Mecopterida, and Coleoptera + Neuropterida (Whiting et al. 1997, 2002b). The position of the Hymenoptera within the Holometabola appeared to be especially unresolved and no characters support a firm placement of this group (Whiting 2002a). Analyses using complete mt genomes had suggested the Hymenoptera as sister taxa to the Mecopterida, however these analyses were susceptible to both ingroup selection and analytical model (Castro and Dowton 2005). In the present study, we expanded the hymenopteran dataset by including two other mt genomes (*Primeuchroeus* sp. and *Vanhornia eucnemidarum*). We tested the robustness of the ingroup topology to outgroup selection and analytical model.

#### 5.3.2.1 *Locusta* as outgroup

*Bayesian analyses:* When *Locusta migratoria* was used as the outgroup, all Bayesian analyses (with both the protein and all nucleotide dataset) recovered the Hymenoptera as sister group to the Coleoptera, with posterior probabilities above 0.95. Bayesian analyses excluding 3<sup>rd</sup> codon positions recovered the same relationships (Fig. 5.2A, 5.2D).

*Parsimony analyses:* Parsimony analyses using protein sequences recovered the Hymenoptera as sister group to the Coleoptera (as in Fig. 5.3A), while the analyses of



nucleotide sequences supported the Hymenoptera-Mecopterida association (Fig. 5.3B).

Parsimony analyses excluding 3<sup>rd</sup> codon positions recovered the Hymenoptera at the base of the Holometabola, however with *Philaenus* (Hemiptera) included within the Hymenoptera, which disrupts the monophyly of the Holometabola (Fig. 5.3C).

#### 5.3.2.2 *Triatoma* as outgroup

*Bayesian analyses:* Bayesian analysis including all codon positions and ribosomal genes recovered the Hymenoptera as sister group to the Mecopterida (Fig. 5.2B). However, analyses excluding 3<sup>rd</sup> codon positions again recovered the Hymenoptera as sister group to the Coleoptera (Fig. 5.2C), and using the amino acids Hymenoptera places at the base (Fig. 5.2E).

*Parsimony analyses:* When *Triatoma* was used as the outgroup, parsimony analysis including all codon positions and ribosomal genes also recovered the Hymenoptera as sister group to the Mecopterida (Fig. 5.3E). Parsimony analyses of the protein dataset, as well as parsimony analyses excluding 3<sup>rd</sup> codon positions, recovered the Hymenoptera at the base of the Holometabola (Fig. 5.3D, 5.3F).

#### 5.3.2.3 *No outgroup*

Holland et al.'s (2003) study showed that the inclusion of an outgroup can frequently disrupt the ingroup tree. For this reason they suggested trees should always be constructed both with and without the outgroup. Since some of the analyses seemed susceptible to outgroup selection, we performed an unrooted analysis in order to test if any ingroup

topology was maintained. In these cases, parsimony and Bayesian analyses recovered the Hymenoptera as closest relative to the Coleoptera (data not shown).

Again, mt genomes failed to resolve phylogenetic relationships within the Holometabola with confidence. Most of the analyses recovered the Hymenoptera as sister group to the Coleoptera which does not agree with previous morphological studies (Kristensen 1999, Beutel and Gorb 2001). In particular, in the analyses in which the test hymenopteran phylogeny is recovered, the Hymenoptera seems to be supported as sister group to the Coleoptera (Figure 5.2A). Other studies also supported this relationship (Cameron et al. 2006). However, there is also strong support for a Hymenoptera+Mecopterida association as well as Hymenoptera at the base of the Holometabola (Table 5.4).

## **5.4 Conclusion**

In this chapter I assessed the ability of mitochondrial (mt) genome sequences to recover a test phylogeny of five hymenopteran taxa from which phylogenetic relationships are well accepted. Our analyses indicated that the test phylogeny is well recovered in all nucleotide Bayesian analyses when all the available holometabolan taxa were included, but only in Bayesian analyses excluding 3<sup>rd</sup> codon positions when only the hymenopteran representatives and a single outgroup were included. This result suggests that taxon sampling of the outgroup might be as important as taxon sampling of the ingroup when recovering hymenopteran phylogenetic relationships using whole mt genomes. Parsimony analyses were more sensitive to both taxon sampling and analytical model than Bayesian analyses, and analyses using the protein dataset did not recover the test phylogeny. In

general, mt genomes did not resolve the position of the Hymenoptera within the Holometabola with confidence, suggesting that an increased taxon sampling, both within the Holometabola and among outgroups, may be necessary.

**Table 5.1**

Information on the taxa whose whole mitochondrial genome sequence was used in the phylogenetic analyses

ORDER	TAXA	AUTHOR	YEAR	ACCESSION No.
Diptera	<i>Drosophila melanogaster</i>	Clary et al.	1983	U37541
	<i>Ceratis capitata</i>	Spanos et al.	2000	AJ242872
	<i>Cochliomyia hominivorax</i>	Lessinger et al.	2000	AF260826
	<i>Chysomya chloropyga</i>	Junqueira, A.	2004	AF352790
	<i>Anopheles gambiae</i>	Beard et al.	1993	L20934
	<i>Bactrocera oleae</i>	Nardi et al.	2003	NC005333
	<i>Haematobia irritans</i>	Lessinger et al.	Unpublished	NC007102
Lepidoptera	<i>Bombyx mandarina</i>	Yukuhiro, K. et al.	2000	AB070263
	<i>Ostrinia furnacalis</i>	Coates, B. et al	2005	AF467260
	<i>Antheraea pernyi</i>	Liu, Y. et al.	Unpublished	NC004622
Hymenoptera	<i>Apis mellifera</i>	Crozier and Crozier	1993	L06178
	<i>Melipona bicolor</i>	Silvestre and Arias	Unpublished	NC004529
	<i>Perga condei</i>	Castro and Dowton	2005	AY787816
	<i>Vanhornia eucnemidarum</i>	Castro and Dowton	Unpublished	DQ302100
	<i>Primeuchroeus</i> sp.	Castro and Dowton	Unpublished	DQ302101/02
Coleoptera	<i>Tribolium castaneum</i>	Friedrich and Muqim	2003	AJ312413
	<i>Crioceris duodecimpunctata</i>	Stewart and Beckenbach	2003	AF467886
	<i>Pyrocoelia rufa</i>	Bae et al	2004	AF452048
	<i>Locusta migratoria</i>	Flook et al	1995	X80245
Hemiptera	<i>Triatoma dimidiata</i>	Dotson and Beard	2001	AF301594
	<i>Philaenus spumarius</i>	Stewart and Beckenbach	2005	NC005944

**Table 5.2**

Estimated model likelihood obtained in the different phylogenetic analyses

Analysis		Likelihood		Bayes factor
allcodonsRNA-outLocusta (covarion=no)	allcodonsRNA- outLocusta (covarion=yes)	-172418	-173238	1640
no3rdcodons- outLocusta (covarion=no)	no3rdcodons- outLocusta (covarion=yes)	-107845	-107947	204
allcodonsRNA-outTriatoma (covarion=no)	allcodonsRNA- outTriatoma (covarion=yes)	-154636	-155320	1370
no3rdcodons- outTriatoma (covarion=no)	no3rdcodons- outTriatoma (covarion=yes)	-96425	-96446	42
allcodonsRNA-Hymenoptera (covarion=no)	allcodonsRNA-Hymenoptera (covarion=yes)	-64470	-64545	150
no3rdcodons-Hymenoptera (covarion=no)	no3rdcodons-Hymenoptera (covarion=yes)	-45453	-45465	24

Bayes factors > 10 indicates that there is strong support in favor of the analysis with higher likelihood score.

Allcodons=1<sup>st</sup>, 2<sup>nd</sup>, and 3<sup>rd</sup> codon positions included. No3rdcodons=analyses in which 3<sup>rd</sup> codon positions are excluded. In all cases, covarion=NO returned higher likelihoods.

**Table 5.3**

Within hymenopteran relationships as recovered by the different phylogenetic analyses

All taxa included	Test phylogeny recovered
Bayesian analyses	
Proteins/rRNA	X
Allcodons/rRNA	√
No3rdcodons/rRNA	√
Parsimony	
Proteins/rRNA	X
Allcodons/rRNA	X
No3rdcodons/rRNA	X
Only Hymenoptera and single outgroup	
Bayesian analyses	
Proteins/rRNA	X
Allcodons/rRNA	X
No3rdcodons/rRNA	√
Parsimony	
Proteins/rRNA	X
Allcodons/rRNA	X
No3rdcodons/rRNA	X
Only Hymenoptera no outgroup-all analyses	X

**Table 5.4**

Position of the Hymenoptera within the Holometabola as recovered by the different phylogenetic analyses

---

<i>Locusta migratoria</i> as outgroup		
Bayesian analyses		
Proteins/rRNA		(Hymenoptera + Coleoptera)
Allcodons/rRNA		(Hymenoptera + Coleoptera)
No3rdcodons/rRNA		(Hymenoptera + Coleoptera)
Parsimony analyses		
Proteins/rRNA		(Hymenoptera + Coleoptera)
Allcodons/rRNA		(Hymenoptera + Mecopterida)
No3rdcodons/rRNA		(Hymenoptera at base of Holometabola)
 <i>Triatoma dimidiata</i> as outgroup		
Bayesian analyses		
Proteins/rRNA		(Hymenoptera base Holometabola)
Allcodons/rRNA		(Hymenoptera + Mecopterida)
No3rdcodons/rRNA		(Hymenoptera + Coleoptera)
Parsimony analyses		
Proteins/rRNA		(Hymenoptera at base of Holometabola)
Allcodons/rRNA		(Hymenoptera + Mecopterida)
No3rdcodons/rRNA		(Hymenoptera at base of Holometabola)
No outgroup-all analyses		(Hymenoptera + Coleoptera)

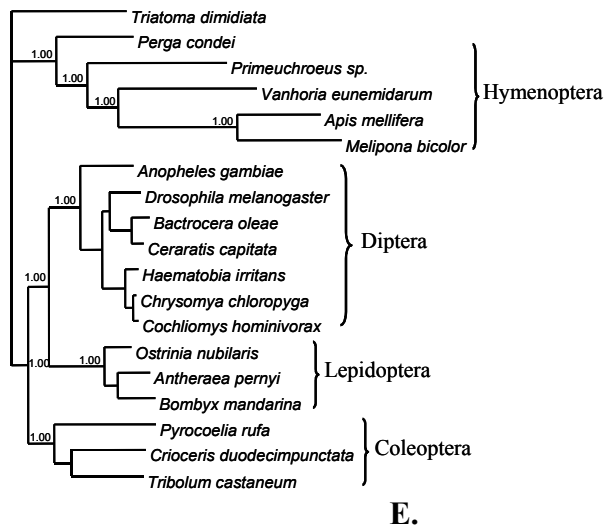
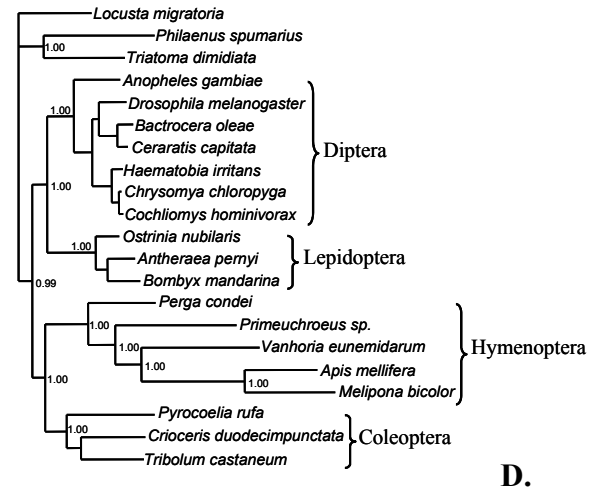
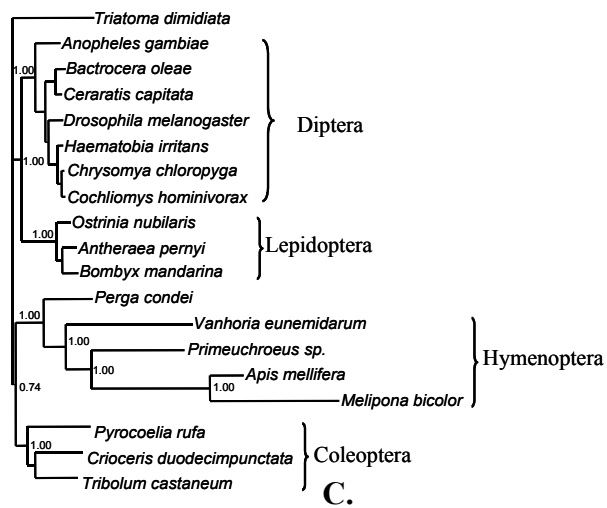
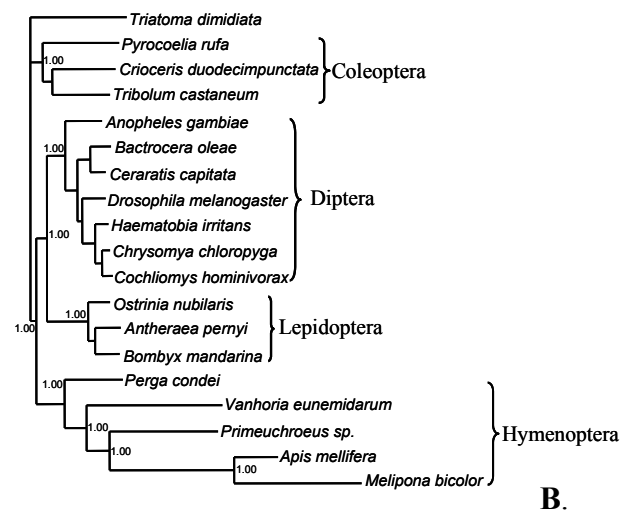
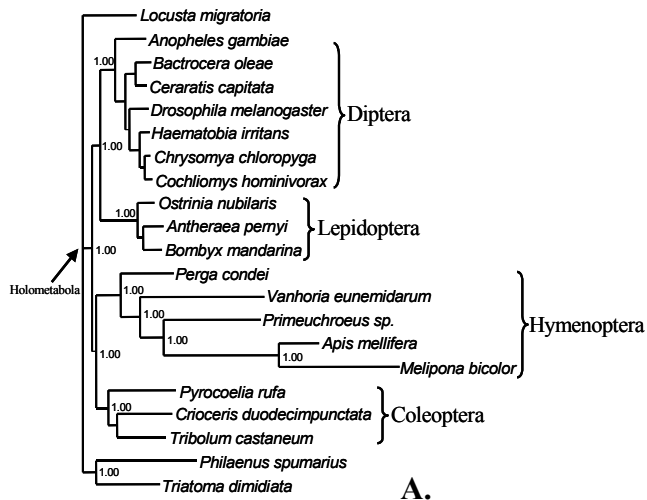
---

### **Figure 5.1**

Accepted phylogenetic relationships between the five hymenopteran taxa included in this analysis. Monophyly of the Aculeata is well supported by morphological and molecular studies (Rasnitsyn 1988, Ronquist et al. 1999, Downton and Austin 1994, 2001).

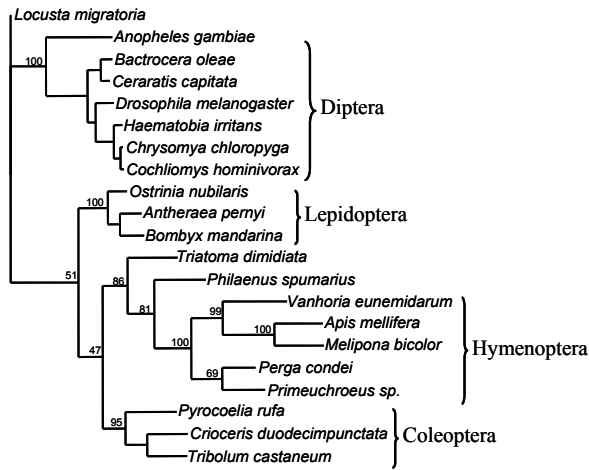
Chrysidoidea (represented here by *Primeuchroeus sp.*) is well supported within the Aculeata as sister taxa to Aculeata *sensu stricto* (here represented by *Apis* and *Melipona*) (Carpenter 1999, Brothers and Carpenter 1993, Brothers 1999). There is general consensus based on biology and the fossil record that the sub order Apocrita evolved from sawflies (Symphyta, represented here by *Perga condei*) (Gibson 1985, Ronquist et al. 1999, Rasnitsyn 1980, 1988).



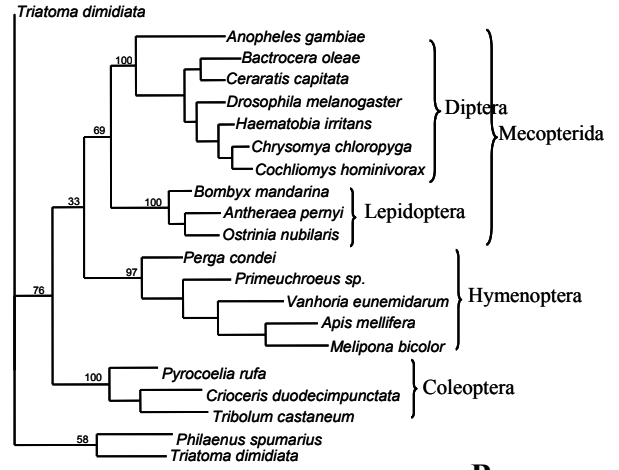


## Figure 5.2

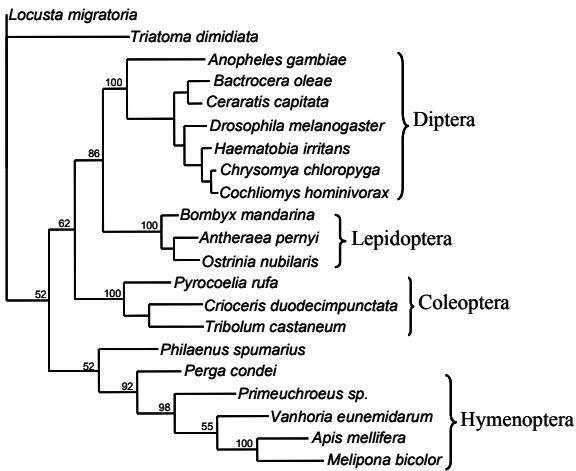
Phylogenetic relationships of the Holometabola. Bayesian analyses. A. Bayesian analysis including all the 21 taxa and 5 molecular partitions (1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup> codon positions, 12S and 16S genes). Analyses excluding 3<sup>rd</sup> codon positions recovered the same relationships. All posterior probabilities above 0.95. B. Bayesian analysis including all codon positions and rRNA genes, *Triatoma dimidiata* as outgroup. C. Bayesian analyses excluding 3<sup>rd</sup> codon positions, *Triatoma dimidiata* as outgroup D. Bayesian analyses using proteins, *Locusta* as outgroup. E. Proteins, *Triatoma* as outgroup.



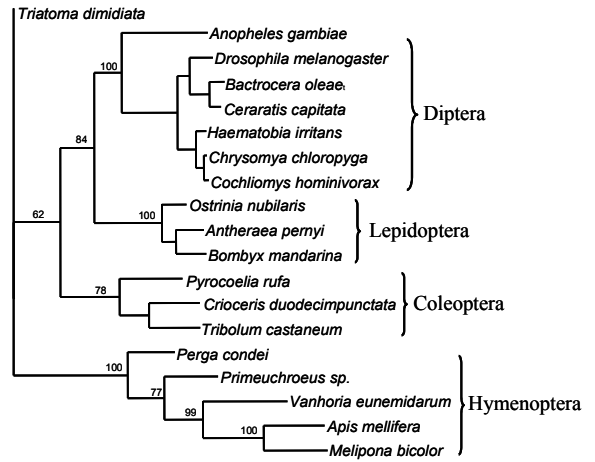
A.



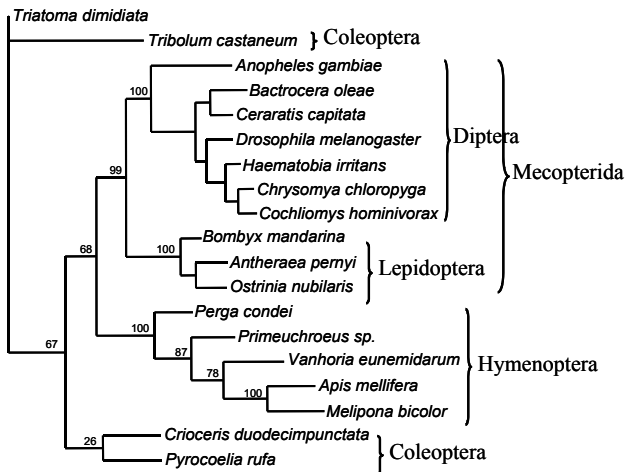
B.



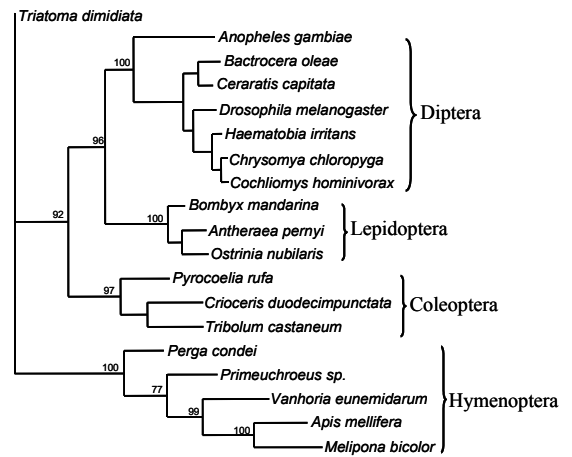
C.



D.



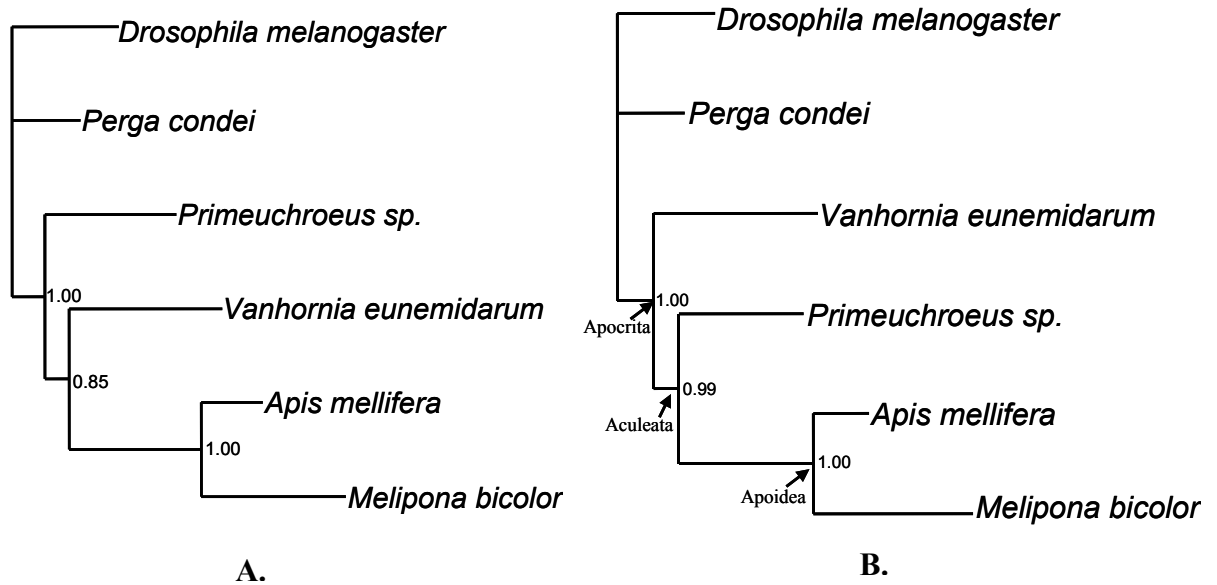
E.



F.

### Figure 5.3

Phylogenetic relationships of the Holometabola. Parsimony analyses. A. Parsimony analyses using proteins. B. Analysis including all the 21 taxa and 5 molecular partitions (1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup> codon positions, 12S and 16S genes). C. Parsimony analysis of nucleotide sequences excluding 3<sup>rd</sup> codon positions. D. Parsimony analysis of proteins with *Triatoma dimidiata* as outgroup. E. Including all codon positions and rRNA genes, *Triatoma dimidiata* as outgroup. F. Parsimony analyses excluding 3<sup>rd</sup> codon positions, *Triatoma dimidiata* as outgroup.



**Figure 5.4**

Only Hymenopteran representatives included. A. Bayesian analyses of nucleotide sequences excluding 3<sup>rd</sup> codon positions. B. Tree topology recovered in all other analyses (parsimony and Bayesian using proteins, parsimony and Bayesian including all codon positions and parsimony excluding 3<sup>rd</sup> codon positions).

## **Chapter 6 . General Discussion**

In this chapter, I bring together the preceding results, and discuss how they link together to form a broader perspective of the major issues explored in this thesis. I first outline the wider significance of the study, then give an overview of the major results and their implications. Finally I give recommendations and future research.

### **6.1 Significance of the study**

This study provides evolutionary information on a very important group of insects, the parasitic wasps, which are a significant source of biological control agents (Quicke 1997). In particular, the molecular dataset supports phylogenetic relationships not previously recovered and provides a preliminary phylogeny from which to start tracing the evolution of some parasitic biologies. This phylogeny contributes to the worldwide effort of assembling the tree of life, which will form the critical infrastructure on which all comparative biology rests, and which seeks the resolution of different nodes in the tree of life as a starting point to understand the evolution of biodiversity (Cracraft and Donoghue 2004).

Additionally, although mitochondrial (mt) genes have been widely used as phylogenetic markers (Avice 1997), studies of the evolution of this molecule are still in their infancy, especially for the invertebrate phyla. This study provides a greater understanding of the evolution of hymenopteran mt genomes and their utility as phylogenetic markers.

## 6.2 Major results and implications

### 6.2.1 *An improved resolution of the Apocritan phylogeny*

One of the aims (aim 2 from introduction) of this study was to improve the resolution of the phylogenetic relationships within the Apocrita. The purpose of doing so was to understand the evolutionary history of the Hymenoptera, and to reveal both lifestyle changes and mitochondrial genome changes during the evolution of the group (both require a robust phylogeny in order to interpret the direction and nature of changes during evolution). Due to the immense diversity of this insect order, deciphering the entire phylogeny of the Apocrita is a very difficult task, but some progress is being made. In the following section, the family and superfamily relationships that are becoming either broadly accepted, or that were recently recovered by molecular analyses, are outlined. Although these account for only a small portion of the hymenopteran evolutionary tree, they represent an important starting point for the interpretation of evolutionary change. In particular, these were some of the relationships used in testing mitochondrial genomes as phylogenetic markers, as outlined later in this chapter.

*-Symphyta (sawflies) as the basally divergent group:* There is now a general consensus based on biology and the fossil record that the Apocrita have evolved from sawflies, although there has been some debate as to which group of sawflies is closest to the common ancestor of the Apocrita [with both the stem sawflies, Cephioidea (e.g. Konigsmann 1977) and Orussidae as the stronger options (Quicke 1997)]. A growing number of morphological studies have now provided strong support for both a monophyletic Apocrita and for an

Orussidae + Apocrita clade (e.g. Gibson 1985, Whitfield et al. 1989, Basibuyuk and Quicke 1995, Vilhelmsen 1997, 2001). Additionally, simultaneous analyses of molecular and morphological data of basal Hymenoptera also supported this relationship (Schulmeister et al. 2002).

*-Monophyly of the Ichneumonoidea (Braconidae + Ichneumonidae):* Ichneumonoidea was proposed as a major monophyletic lineage by Rasnitsyn (1988) and then supported in a cladistic reanalysis of the same dataset by Ronquist et al. (1999). Sharkey and Wahl (1992) described six morphological synapomorphies for the superfamily, making this one of the most strongly supported family level relationships among the Apocrita, based on morphology. Although the expanded molecular dataset of Dowton and Austin (2001) did not recover the Ichneumonoidea with convincing support, the present research provides strong support for the monophyly of the group.

*-Monophyly of the Aculeata, and the arrangement (Chrysidoidea, (Apoidea, Vespoidea)):* The Aculeata is a well supported monophyletic group based on morphology (Brothers 1975, Brothers and Carpenter 1993). The family level relationships among the parasitic Aculeata have been also intensively investigated, since the work of Brothers (1975) who put together a great amount of data for phylogenetic analysis. Carpenter (1986) provided a detailed analysis of the families of the Chrysidoidea, and Brothers and Carpenter (1993) studied the Chrysidoidea and Vespoidea. These works supported essentially the same relationships, with the Chrysidoidea being a well-supported monophyletic taxon basal to the other aculeates, and with the Vespoidea also monophyletic and the sister group of the



Apoidea (including Sphecoidea). Aculeata is also recovered as monophyletic with strong support in all the molecular analyses performed in this thesis.

*-Monophyly of the Proctotrupomorpha:* The Proctotrupomorpha has been considered a monophyletic group since Rasnitsyn (1988). It was also generally recovered as monophyletic in the Dowton and Austin (2001) expanded analysis, however, the position of the Ceraphronoidea inside the Proctotrupomorpha (i.e. as resolved by Ronquist et al. 1999) was sensitive to the model of analysis. This research supported the monophyly of the Proctotrupomorpha, with the Ceraphronoidea outside this group. Relationships of the groups inside the Proctotrupomorpha are less clear. Dowton and Austin (2001) strongly suggested the Diapriidae, Monomachidae and Maamingidae as a natural group; the analyses performed in this thesis also supported this relationship. Additionally, my data supports the (Maamingidae + Diapriidae + Monomachidae) as sister group to the Chalcidoidea with strong support. Both these groups are endoparasitoids. This relationship could suggest as a possible scenario that the ancestor of the Chalcidoidea, Mymaridae, Maamingidae, Diapriidae and Monomachidae were parasitoids of concealed, possibly leaf-mining hosts. The Chalcidoidea then spread into a range of hosts, while Maamingidae, Diapriidae, and Monomachidae specialized on Diptera. Other relationships within the Proctotrupomorpha are not well supported. For example, in this study the Platygastroidea is recovered as sister group to the Cynipoidea but with low support. Other analyses based on molecular data supported a sister group relationship of Platygastroidea with Chalcidoidea (Dowton and Austin 1997, 2001), and available morphological data indicated a relationship of Platygastroidea with Ceraphronoidea (Ronquist et al. 1999). Gibson (1999) also supported a clade comprising Platygastroidea+Pelecinidae+Proctotrupidae+Vanhorniidae. On the other

hand, although the position of Heloridae is not well supported in this thesis, it was always included inside the Proctotrupomorpha, contrary to Dowton and Austin (2001). These results suggest that it would be useful and more tractable to concentrate on deducing the phylogenetic relationships within the Proctotrupomorpha prior to the Apocrita.

### ***6.2.2 A more accurate estimate of the degree of gene rearrangements in hymenopteran mt genomes***

Despite the general lack of resolution of the hymenopteran phylogeny, knowledge concerning some nodes can be used to direct taxon sampling of mt genomes. With Aim 3 of this thesis I wanted to gain a more accurate estimate of the rate and nature of mitochondrial genome rearrangements in the Hymenoptera. In order to do so, the mt genome of one symphytan representative, *Perga condei*, and two apocritan representatives, *Vanhornia eucnemidarum* and *Primeuchroeus* sp. were sequenced.

Examination of genome organization shows that the Symphyta are hardly rearranged. The mt genome of *Perga condei* resembles that considered ancestral for insects and crustaceans, with only one gene not found in the ancestral position. This result agrees with Dowton and Austin (1999) and Dowton et al. (2003), who had sequenced two different portions of the mt genomes, the *cox2-atp8* gene junction and the *nd3-nd5* gene junction. However, the rate of genomic change was accelerated in Apocrita, with both *Vanhornia* and *Primeuchroeus* genomes very rearranged when compared to the ancestral organization. Rates of gene rearrangements for the Hymenoptera can be calculated using the rate of gene rearrangement (RGR) test from Dowton (2004), which works similarly to Tajima's relative rate test

(Tajima 1993). In order to do this I tabulated the number of genes rearranged in each of the hymenopteran taxa for which genomes are available: 1 in *Perga*, 7 in *Vanhornia*, 11 in *Primeuchroeus*, 8 in *Apis*, 7 in *Melipona*. I used the number of gene rearrangements rather than breakpoint distances since, as suggested by Dowton (2004), the number of breakpoint distances may inflate the significance value when making comparisons between relatively unrearranged genomes. The RGR test requires three genomes to be compared and it measures the relative rate of gene rearrangement between genomes 1 and 2, by comparing them with genome 3, a reference genome. In this case, *Daphnia* (Crustacea) was used as the reference mt genome (i.e. genome 3), and *Drosophila* as genome 2. The significance of the calculated value is assessed by reference to chi-square tables, using 1 degree of freedom after Bonferroni correction. Within the Hymenoptera clearly there is an increased rate of gene rearrangement in the Apocrita (*Vanhornia* [P=0.0082], *Primeuchroeus* [P=0.0009], *Apis* [P=0.0047] and *Melipona* [P=0.0082]) but not in the Symphyta (*Perga* [P=0.3173]), as shown in table 6.1.

Dowton and Austin (1997a, 1997b) also had found an increased rate of compositional bias and nucleotide substitutions in the Apocrita with respect to the Symphyta. Accordingly, this thesis supports a correlation in the rate of nucleotide substitution and the rate of gene rearrangement in the Apocrita. This result is in line with Shao et al. (2003) and other studies that have indicated that there might be a correlation between the rates of nucleotide substitution and the rates of gene rearrangements. Other lineages with both a high rate of mt gene rearrangement and a high rate of nucleotide substitution include the lineages of the blue mussel, *Mytilus edulis* (Hoffmann et al. 1992), the akamata snake, *Dinodon*

*semicarinatus* (Kumazawa et al. 1998), the ascidian, *Halocynthia roretzi* (Yokobori et al. 1999), and some hemipteroids (Page et al. 1998, Shao et al. 2001, Shao and Baker 2003).

This study indicates that within the Apocrita, *trn* gene rearrangements predominate. None of the protein coding genes was found in a rearranged position, and the two *rRNA* genes were inverted in the mt genome of *Primeuchroeus*. It is expected that due to their small size, *trn* genes would translocate and/or invert easily (Negrisola et al. 2004).

Translocation, shuffling and inversion types of rearrangement were identified in the mt genomes of *Vanhornia* and *Primeuchroeus*. Translocation and shuffling events probably took place by a duplication of a large region followed by loss of all but one or a few genes in a single or several deletions events as suggested in Lavrov et al. 2002, while inversion events probably arose by intra-mitochondrial recombination as proposed by Dowton and Austin (1999), involving breakage and rejoining of the mt genome. An extended taxon sampling will be required in order to test the different models and mechanisms of gene rearrangements in the group.

### **6.2.3 The potential of mt genomes as phylogenetic markers**

Finally, an aim of this project, as outlined by aims 1 and 4, was to explore the potential of mt genomes in resolving phylogenetic relationships within the Apocrita, and between the Hymenoptera and other Holometabola. This is difficult to assess in the absence of a well-resolved phylogeny, but I was able to use the nodes in the hymenopteran tree that are well supported as a test case. Broad assessment of analytical conditions appeared to identify those conditions that reliably recovered the test phylogeny, with the partitioned Bayesian

analysis of nucleotide sequences excluding 3<sup>rd</sup> codon positions as the preferred model. One pitfall of this approach is the very small taxon sampling, with just five representatives of the Hymenoptera included. As suggested by various studies, with a small taxon sampling the phylogeny may be difficult to recover accurately (Hillis 1996, 1998, Hillis et al. 2003, Murphy et al. 2001, Pollock and Bruno 2000, Pollock et al. 2000, 2002). However, it is important to keep in mind that, such a small taxon size also means that there are few alternate phylogenies possible (with 5 taxa there are only 15 possible trees), making the test phylogeny relatively easy to recover. The present analyses indicated that when the appropriate model of phylogenetic reconstruction was applied, the test phylogeny was recovered. Further assessment of this will rely on a broader sampling of mt genomes, which will make any test phylogeny more difficult to recover.

### **6.3 Recommendations and future research**

Despite the accomplishments of this research, it is apparent that our understanding of the phylogeny of apocritan relationships still requires increased taxonomic sampling and genomic sampling, as well as a better morphological matrix that could be analyzed in combination with the molecular data. In particular, protein-coding nuclear genes would be excellent candidates. For example, the nuclear gene encoding the long-wavelength rhodopsin (LW Rh) has been shown in several studies to be a promising gene to determine higher-level phylogenies, as well as intrafamily relationships (Carulli et al 1994, Mardulyn and Cameron 1999, Cameron and Mardulyn 2001, 2003, Lockhart and Cameron 2001). Elongation factor 1 alpha (EF-1-alpha) is a highly conserved ubiquitous protein involved in translation that has also been suggested to have desirable properties for phylogenetic

inference and as a marker in resolving deep phylogenetic relationships (eg. Cho et al. 1995, Roger et al. 1999).

The results of this work suggest that complete mt genomes are promising candidates for resolving phylogenetic relationships of the group. However, selection of appropriate taxa to be sequenced, based on availability and taxonomic position, is crucial. A good option would be to start by sampling the Proctotrupomorpha more fully, particularly as some nodes in the phylogeny appear to be resolved. A diapriid, a monomachid, and a maamingid would be a good starting point, since relationships among these taxa are well resolved. Inclusion of a mymarommatid would be beneficial, as this is well-supported as the sister group to the Chalcidoidea (Gibson 1986). Future taxonomic sampling should aim at sequencing at least a single representative from each family/superfamily of the Proctotrupomorpha. This will allow us to test further the Chalcidoidea + (Diapriidae+Monomachidae+Maamingidae) sister-group relationship, as well as the position of other families. Additionally, this research has demonstrated the importance of using appropriate models of phylogenetic reconstruction when using mt genomes; however, sequencing more mt genomes will be crucial in understanding the evolution of this molecule.

Finally, gene rearrangements are abundant among the apocritan taxa and might also be good candidates to resolve the phylogenetic relationships. However, as concluded in chapter four of this thesis, an increased taxonomic sampling seems to be essential for detecting synapomorphies. In order to achieve this, it is particularly important to accelerate the rate of sequencing of mt genomes in this group. The long-PCR (polymerase chain

reaction) amplification of complete mt genomes represents the most widely used tool in the purification of arthropodan mtDNA. Of the 93 complete mt genomes sequenced for this phylum, 46 were amplified by long PCR, 23 were isolated and recovered based on other technical approaches (CsCl gradient, multiple PCR, etc), and 24 have not yet been published (Barau et al. 2005). There are a number of described conserved primers specially designed for the long-PCR amplification of complete arthropodan mt genomes (Hwang et al. 2001, Barau et al. 2005), and the PCR conditions used by the different researchers are very similar, with different types of specific long-PCR *Taq* polymerases available in the market. In some cases complete mt genomes have been amplified in two pieces of about 8 kb each. However, successful long PCR amplifications within the Hymenoptera have been difficult to achieve (the longest PCR amplification I obtained was about 3.5 kb), probably due to the extensive amount of gene rearrangements and compositional bias. In my case, other *Taq* polymerases remain to be tested [(e.g. Elongase enzyme mix (Invitrogen), Expand Long Template PCR system (Boehringer Mannheim Co.)], and with five available mt genomes, specific hymenopteran long-PCR primers need to be designed and tested. Additionally, I also had some difficulties during the primer-walking process when using both amplified and cloned fragments, while this also is a slow process. The Erase-a-base or transposon systems could be used to overcome these problems. With these methods, there is no need to either deal with primer walking nor order any template specific primers. The entire DNA template can be sequenced using the universal sequencing primers included in the kit, and all individual clones can be sequenced at the same time, reducing the time and probably the cost of the sequencing process.

## 6.4 Conclusion

Phylogenetic relationships within the order Apocrita have proven difficult to resolve (Whitfield 1998, Dowton and Austin 2001). Both molecular and morphological studies have advanced in the understanding of phylogenetic relationships within the group, however, several nodes are still not well supported. This study is the first Bayesian analysis of an increased molecular dataset that includes two mitochondrial and two nuclear genes. Although the molecular Bayesian analysis did not provide the expected well supported phylogeny of the Apocrita, it provided support for some nodes and generated a new hypothesis to work on. In particular, this study supported an association of Chalcidoidea with the (Diapriidae + Monomachidae + Maamingidae). Future research needs to focus on smaller groups, the Proctotrupomorpha, for example.

This study is the first comparative study of complete mitochondrial genomes within the order Hymenoptera. It was able to describe the degree of gene rearrangement and to address the utility of whole mt genomes as phylogenetic markers within the group. The complete mt genomes of *Perga*, *Vahnornia* and *Primeuchroeus* are evidence of increased rates of mt gene rearrangement in the Apocrita with respect to the Symphyta. However, an increased taxon sampling is required to study the evolution of these gene rearrangements, with no clear synapomorphies found among the five complete mt genomes available for the Hymenoptera. Whole mt genomes were proven usefull as markers to resolve phylogenetic relationships in the group; however only under specific models of molecular evolution. Different models of phylogenetic recontruction seemed to affect the results considerably, as well as taxonomic sampling. These findings contribute to other studies on mt genomes,



both within the Hymenoptera and among insects.

**Table 6.1**

Assesment of the relative rate of genome rearrangment among hymenopteran mitochondrial genomes, using the RGR test

Genome	GR	<i>P</i>
<i>Perga condei</i> (Hymenoptera: Symphyta)	1	0.3173
<i>Vanhornia eucnemidarum</i> (Hymenoptera: Apocrita)	7	0.0082*
<i>Primeuchroeus</i> sp. (Hymenoptera: Apocrita)	11	0.0009**
<i>Apis mellifera</i> (Hymenoptera: Apocrita)	11	0.0047*
<i>Melipona bicolor</i> (Hymenoptera: Apocrita)	7	0.0082*

For all comparisons using the RGR test genomes 2 and 3 are *Drosophila* and *Daphnia*. According to the test  $\chi^2_{1 \text{ d.f.}} = [\text{GR}_{(1,3)} - \text{GR}_{(2,3)}]^2 / \text{GR}_{(1,3)} + \text{GR}_{(2,3)}$ , GR= number of gene rearrangments. The tabulated *P* values are uncorrected values; asterisk indicate the level of sifnificance after Bonferroni correction (\**P* < 0.05, \*\* *P* < 0.01, \*\*\* *P* < 0.001).

## Chapter 7 . REFERENCES

- Adachi J, Cao Y, Hasegawa M (1993) Tempo and mode of mitochondrial DNA evolution in vertebrates at the amino acid sequence level: rapid evolution in warm-blooded vertebrates. *Journal of Molecular Evolution*, 36, 270-281
- Adachi J, Hasegawa M (1996) Model of amino acid substitution in proteins encoded by mitochondrial DNA. *Journal of Molecular Evolution*, 42, 459-468
- Akashi H, Kliman RM, Eyre-Walker A (1998) Mutation pressure, natural selection, and the evolution of base composition in *Drosophila*. *Genetica*, 102/103, 49-60
- Arnason U, Adegoke JA, Bodin K, Born EW, Esa YB, Gullberg A, Nilsson M, Short RV, Xu XF, Janke A (2002) Mammalian mitogenomic relationships and the root of the eutherian tree. *Proceedings of the National Academy of Science of the United States of America*, 99, 8151-8156.
- Asakawa S, Kumazawa Y, Araki T, Himeno H, Miura K-I, Watanabe K (1991) Strand-specific nucleotide composition bias in echinoderm and vertebrate mitochondrial genomes. *Journal of Molecular Evolution*, 32, 511-520
- Askew RR (1971) *Parasitic Insects*. Heinemann Educational Books Ltd. London
- Askew RR, Shaw MR (1986) Parasitoid communities: their size, structure and development. In: *Insect Parasitoids* (Waage, J., D. Greathead eds.) Academic Press, London.
- Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Evolution and Systematics*, 18, 489-522

- Avise JC (1995) Mitochondrial DNA polymorphism and a connection between genetics and demography of relevance to conservation. *Conservation Biology*, 9, 686-690
- Bae JS, Kim I, Sohn HD, Jin BR (2004) The mitochondrial genome of the firefly, *Pyrocoelia rufa*: complete DNA sequence, genome organization, and phylogenetic analysis with other insects. *Molecular Phylogenetics and Evolution*, 32, 978-985
- Ballard JW (2000) When one is not enough: Introgression of mitochondrial DNA in *Drosophila*. *Molecular Biology and Evolution*, 17, 1126-1130
- Beard DB, Hamm DM, Collins FH (1993) The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects. *Insect Molecular Biology*, 2, 103-124
- Belshaw R, Quicke DLJ (2002) Robustness of ancestral state estimates: evolution of life history strategy in ichneumonoid parasitoids. *Systematic Biology*, 51, 450-477
- Belshaw R, Dowton M, Quicke DLJ, Austin AD (2000) Estimating ancestral geographical distributions: a Gandwanan origin for aphid parasitoids? *Proceedings of the Royal Society of London Series B*, 267, 491-496
- Bergsten J (2005) A review of long-branch attraction. *Cladistics*, 21, 163-193
- Bernardi G, Olofsson B, Filipski J, Zerial M, Salinas J, Cuny G, Meunier-Rotival M, Rodier F (1985) The mosaic genome of warm blooded vertebrates. *Science*, 228, 953-957
- Beutel RG, Gorb SN (2001) Ultrastructure of attachment specializations of hexapods (Arthropoda): evolutionary patterns inferred from a revised ordinal phylogeny. *Journal of Zoological Systematics and Evolutionary Research*, 39, 177-207

- Black WC, Roehrdanz RL (1998) Mitochondrial gene order is not conserved in Arthropods: Prostriate and Metastriate tick mitochondrial genomes. *Molecular Biology and Evolution*, 15, 1772-1785.
- Boore JL (1999) Animal mitochondrial genomes. *Nucleic Acids Research*, 27, 1767-1780
- Boore JL, Brown WM (1998) Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. *Current Opinion in Genetics and Development*, 8, 668-674
- Boore JL, Brown WM (2000) Mitochondrial genomes of *Galathealinum*, *Helobdella*, and *Platynereis*: sequence and gene arrangement comparisons indicate that Pogonophora is not a phylum and Annelida and Arthropoda are not sister taxa. *Molecular Biology and Evolution*, 17, 87-106
- Boore JL, Collins TM, Stanton D, Daehler LL, Brown WM (1995) Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements. *Nature*, 376, 163-165
- Boore JL, Lavrov DV, Brown WM (1998) Gene translocation links insects and crustaceans. *Nature*, 392, 667-668
- Bourdreaux HB (1979) Arthropod phylogeny with special reference to insects. John Wiley and Sons, New York
- Britten RJ (1986) Rates of DNA sequence evolution differ between taxonomic groups. *Science*, 231, 1393-1398
- Cameron SL, Miller KB, D'Haese CA, Whiting MF, Barker SC (2004) Mitochondrial genome data alone are not enough to unambiguously resolve the relationships of Entognatha, Insecta and Crustacea *sensu lato* (Arthropoda). *Cladistics*, 20, 534-557

- Campbell NJH, Barker, SC (1998) An unprecedented major rearrangement in an arthropod mitochondrial genome. *Molecular Biology and Evolution*, 15, 1786-1787
- Campbell NJH, Barker SC (1999) The novel mitochondrial gene arrangement of the cattle tick, *Boophilus microplus*: fivefold tandem repetition of a coding region. *Molecular Biology and Evolution*, 16, 732-740
- Cantatore P, Gadaleta MN, Roberti M, Saccone C, Wilson AC (1987) Duplication and remoulding of tRNA genes during the evolutionary rearrangement of mitochondrial genomes. *Nature*, 329, 853-855
- Cantatore P, Roberti M, Pesole G, Ludovico A, Milella F, Gadaleta MN, Saccone C (1994) Evolutionary analysis of cytochrome b sequences in some perciformes - evidence for a slower rate of evolution than in mammals. *Journal of Molecular Evolution*, 39, 589-597
- Carnean D, Kimsey LS, Berbee ML (1992) 18S rDNA sequences and the holometabolous insects. *Molecular Phylogenetics and Evolution*, 1, 270-278
- Castro LR, Austin AD, Dowton M (2002) Contrasting rates of mitochondrial molecular evolution in parasitic Diptera and Hymenoptera. *Molecular Biology and Evolution*, 19, 1100-1113
- Castro LR, 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, 469-479.
- Caterino MS, Cho S, Sperling FAH (2000) The current state of insect molecular systematics: a thriving tower of Babel. *Annual Review of Entomology*, 45, 1-54

- Chalwatzis N, Hauf J, van de Peer Y, Kinzelbach R, Zimmermann FK (1996) 18S ribosomal RNA genes of insects: primary structure of the genes and molecular phylogeny of the Holometabola. *Annals of the Entomological Society of America*, 89, 788-803
- Chang BHJ, Shimmin LC, Shyue SK, Hewett-Emmett D (1994) Weak male-driven molecular evolution in rodents. *Proceedings of the Natural Academy of Science USA*, 91, 827-831
- Chang JT (1996) Inconsistency of evolutionary tree topology reconstruction methods when substitution rates vary across characters. *Mathematical Biosciences*, 134, 189-215
- Chen Y, Xiao H, Fu J, Huang D-W (2004) A molecular phylogeny of eurytomid wasps inferred from DNA sequence data of 28S, 18S, 16S and COX1 genes. *Molecular Phylogenetics and Evolution*, 31, 300-307
- Clary DO, Wolstenholme DR (1985a) The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. *Journal of Molecular Evolution*, 22, 252-271
- Clary DO, Wolstenholme DR (1985b) The ribosomal RNA genes of *Drosophila* mitochondrial DNA. *Nucleic Acids Research*, 13, 4029-4045
- Coates BS, Sumerford DV, Hellmich RL, Lewis LC (2005) Partial mitochondrial genome sequences of *Ostrinia nubilalis* and *Ostrinia furnicalis*. *International Journal of Biological Sciences*, 1, 13-18
- Collins TM, Wimberger PH, Naylor GJP (1994) Compositional bias, character-state bias, and character-state reconstruction using parsimony. *Systematic Biology*, 43, 482-496

- Conant GC, Lewis PO (2001) Effects of nucleotide composition bias on the success of the parsimony criterion in phylogenetic inference. *Molecular Biology and Evolution*, 18, 1024-1033
- Cracraft J, Donoghue MJ (2004) *Assembling the Tree of Life*. Oxford University Press. New York.
- Crozier RH, Crozier YC (1993) The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. *Genetics*, 133, 97-117
- Cummings MP, Otto SP, Wakeley J (1995) Sampling properties of DNA sequence data in phylogenetic analysis. *Molecular Biology and Evolution*, 12, 814-822
- Cunningham CW (1997) Is congruence between data partitions a reliable predictor of phylogenetic accuracy? Empirically testing and iterative procedure for choosing among phylogenetic methods. *Systematic Biology*, 46, 464-478
- Curole JP, Kocher TD (1999) Mitogenomics: digging deeper with complete mitochondrial genomes. *Trends in Ecology and Evolution*, 14, 394-398
- Delsuc F, Phillips MJ, Penny D (2003) Comment on "hexapod origins: monophyletic or paraphyletic? *Science*, 201, 1482
- Desjardins P, Morais R (1990) Sequence and gene organization of the chicken mitochondrial genome. *Journal of Molecular Evolution*, 212, 599-634
- Dotson EM, Beard CD (2001) Sequence and organization of the mitochondrial genome of the Chagas disease vector, *Triatoma dimidiata*. *Insect Molecular Biology* 10: 205-215.



- Dowton M (1999) Relationships among the cyclostome braconid (Hymenoptera: Braconidae) subfamilies inferred from a mitochondrial tRNA gene rearrangement. *Molecular Phylogenetics and Evolution*, 11, 283-287
- Dowton M, Austin AD (1993) Direct sequencing of double-stranded PCR products without intermediate fragment purification; digestion with mung bean nuclease. *Nucleic Acids Research*, 21, 3599-3600
- Dowton M, Austin AD (1994) Molecular phylogeny of the insect order Hymenoptera: Apocritan relationships. *Proceedings of the National Academy of Science of the United States of America*, 91, 9911-9915.
- Dowton M, Austin AD (1995) Increased genetic diversity in mitochondrial genes is correlated with the evolution of parasitism in the Hymenoptera. *Journal of Molecular Evolution*, 41, 958-965.
- Dowton M, Austin AD (1997a) The evolution of strand-specific compositional bias. A case study in the hymenopteran mitochondrial 16S rRNA gene. *Molecular Biology and Evolution*, 14, 109-112.
- Dowton M, Austin AD (1997b) Evidence for AT-transversion bias in wasp (Hymenoptera: Symphyta) mitochondrial genes and its implications for the origin of parasitism. *Journal of Molecular Evolution*, 44, 398-405
- Dowton M, Austin AD (1998) Phylogenetic relationships among the microgastroid wasps (Hymenoptera: Braconidae): combined analysis of 16S and 28S rDNA genes, and morphological data. *Molecular Phylogenetics and Evolution*, 10, 354-366

- Dowton M, Austin AD (1999) Evolutionary dynamics of a mitochondrial rearrangement "Hot Spot" in the Hymenoptera. *Molecular Biology and Evolution*, 16:298-309
- Dowton M, Austin AD (2001) Simultaneous analysis of 16S, 28S, COI and morphology in the Hymenoptera: Apocrita - evolutionary transitions among parasitic wasps. *Biological Journal of the Linnean Society*, 74, 87-111.
- Dowton M, Austin AD (2002) Increased congruence does not necessarily indicate increased phylogenetic accuracy - the behavior of the incongruence length difference test in mixed-model analyses. *Systematic Biology*, 51, 19-31
- Dowton M, Campbell NJH (2001) Intramitochondrial recombination - is it why some mitochondrial genes sleep around? *Trends in Ecology and Evolution*, 16, 269-271
- Dowton M, Castro LR, Austin AD (2002) Mitochondrial gene rearrangements as phylogenetic characters in the invertebrates: the examination of genome 'morphology'. *Invertebrate Systematics*, 16, 345-356
- Dowton M, Castro LR, Campbell SL, Bargon SD, Austin AD (2003) Frequent mitochondrial gene rearrangements at the Hymenopteran nd3-nd5 junction. *Journal of Molecular Evolution*, 56, 1-10
- Early JW, Masner L, Naumann ID, Austin AD (2001) Maamingidae, a new family of proctotrupoid wasp (Insecta : Hymenoptera) from New Zealand. *Invertebrate Taxonomy*, 15, 341-352
- Eggleton P, Belshaw R (1992) Insect parasitoids: an evolutionary overview. *Philosophical Transactions of the Royal Society of London Series B*, 337, 1-20
- Eyre-Walker A (1999) Evidence of selection on silent site base composition in mammals: potential implications for the evolution of isochores and junk DNA. *Genetics*, 152, 675-683

- Felsenstein J (1985) Phylogenies and the comparative method. *American Naturalist*, 125, 1-15
- Felsenstein JF (1978) Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology*, 27, 401-410
- Fitch WM (1971) Towards defining the course of evolution: Minimum change for a specific tree topology. *Systematic Zoology*, 20, 406-416.
- Fitch WM (1986) The estimate of total nucleotide substitutions from pairwise differences is biased. *Philosophical Transactions of the Royal Society of London Series B*, 312, 317-324
- Flook PK, Rowell CHF, Gellissen G (1995a) The sequence, organization, and evolution of the *Locusta migratoria* mitochondrial genome. *Journal of Molecular Evolution*, 41, 928-941
- Flook P, Rowell H, Gellissen G (1995b) Homoplastic rearrangements of insect mitochondrial tRNA genes. *Naturwissenschaften*, 83, 336-337
- Foster PG, Hickey DA (1999) Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. *Journal of Molecular Evolution*, 48, 284-290
- Foster PG, Jermiin LS, Hickey DA (1997) Nucleotide composition bias affects amino acid content in proteins coded by animal mitochondria. *Journal of Molecular Evolution*, 44, 282-288
- Friedrich M, Muqim N (2003) Sequence and phylogenetic analysis of the complete mitochondrial genomes of the flour beetle *Tribolium castaneum*. *Molecular Phylogenetics and Evolution*, 26, 502-512.

- Galtier N, Gouy M (1995) Inferring phylogenies from DNA sequences of unequal base compositions. *Proceedings of the Natural Academy of Science USA*, 92, 11317-11321
- Galtier N, Gouy M (1998) Inferring pattern and process: maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. *Molecular Biology and Evolution*, 15, 871-879
- Garcia-Moreno J, Sorenson MD, Mindell DP (2003) Congruent avian phylogenies inferred from mitochondrial and nuclear DNA sequences. *Journal of Molecular Evolution*, 57, 27-37
- Gauld ID, Bolton B (1988) *The Hymenoptera*. Oxford University Press, Oxford
- Gauld ID, Bolton B (1996) *The Hymenoptera*. Oxford University Press, Oxford
- Gauld ID, Bolton B (1988) Introducing the Hymenoptera. In: Gauld ID, Bolton B (eds) *The Hymenoptera*. (pp. 1-39). New York: Oxford University Press
- Gauld ID (1988) Evolutionary patterns of host utilization by ichneumonoid parasitoids (Hymenoptera: Ichneumonidae and Braconidae). *Biological Journal of Linnean Society*, 35, 351-377
- Gibson GAP (1993) Superfamilies Mymarommatoidea and Chalcidoidea. In: Goulet H, Huber JT (eds) *Hymenoptera of the world: an identification guide to families*. Research Branch, Agriculture Canada, Ottawa, Ontario
- Gibson GAP (1985) Some pro- and mesothoracic structures important for phylogenetic analysis of Hymenoptera, with a review of terms used for the structures. *Canadian Entomologist*, 117, 1395-1443

- Gibson GAP (1999) Sister-group relationships of the Platygastroidea and Chalcidoidea (Hymenoptera) - an alternative hypothesis to Rasnitsyn (1988). *Zoologica Scripta*, 28, 125-138
- Giribet G, Carranza S, Baguna J, Riutort M, Ribera C (1996) First molecular evidence for the existence of a Tardigrada + Arthropoda clade. *Molecular Biology and Evolution*, 13(1), 76-84
- Godfray HCJ (1994) *Parasitoids - behavior and evolutionary ecology*. Princeton University Press, New Jersey
- Goldman N, Anderson JP, Rodrigo AG (2000) Likelihood-based tests of topologies in phylogenetics. *Systematic Biology*, 49, 652-670
- Graham SW, Olmstead RG, Barrett S (2002) Rooting phylogenetic trees with distant outgroups: a case study from the commelinoid monocots. *Molecular Biology and Evolution*, 19, 1769-1781
- Gross NJ, Getz GS, Rabinowitz M (1969) Apparent turnover of mitochondrial deoxyribonucleic acid mitochondrial phospholipids in the tissues of the rat. *Journal of Biological Chemistry*, 244, 1552-1562
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 41, 95-98
- Hammond PM (1992) Species Inventory. In global biodiversity, status of the earth's living resources (B. Groombridge ed.) pp. 17-39. London. Chapman and Hall
- Hampl V, Cepicka I, Flegel J, Tachezy J, Kulda J (2004) Critical analysis of the topology and rooting of the parabasal 16S rRNA tree. *Molecular Phylogenetics and Evolution*, 32, 711-723

- Hasegawa M, Hashimoto T, Adachi J, Iwabe N, Miyata T (1993) Early branchings in the evolution of eukaryotes: ancient divergence of *Entamoeba* that lacks mitochondria revealed by protein sequence data. *Journal of Molecular Evolution*, 36, 380–388
- Hassanin A, Leger N, Deutsch J (2005) Evidence for multiple reversals of asymmetric mutational constraints during the evolution of the mitochondrial genome of Metazoa, and consequences for phylogenetic inferences. *Systematic Biology*, 54, 277–298
- Hendy MD, Penny D (1989) A framework for the quantitative study of evolutionary trees. *Systematic Zoology*, 38, 297–309
- Hennig W (1981) *Insect phylogeny*. Academic Press, New York
- Hickerson MJ, Cunningham CW (2000) Dramatic mitochondrial gene rearrangements in the hermit crab *Pagurus longicarpus* (Crustacea, Anomura). *Molecular Biology and Evolution*, 17, 639–644
- Hillis DM, Huelsenbeck JP, Cunningham CW (1994) Application and accuracy of molecular phylogenies. *Science*, 264, 671–677
- Hoffmann RJ, Boore JL, Brown WM (1992) A novel mitochondrial genome organization for the blue mussel, *Mytilus edulis*. *Genetics*, 131, 397–412
- Holland BR, Penny D, Hendy MD (2003) Outgroup misplacement and phylogenetic inaccuracy under a molecular clock—a simulation study. *Systematic Biology*, 52, 229–238
- Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP (2001) Bayesian inference of phylogeny and its impact on evolutionary biology. *Science*, 294, 2310–2314

- Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17, 754-755
- Huelsenbeck JP (1995) The robustness of two phylogenetic methods: four-taxon simulations reveal a slight superiority of maximum likelihood over neighbor joining. *Molecular Biology and Evolution*, 12, 843-849
- Huelsenbeck JP, Bull JJ, Cunningham CW (1996) Combining data in phylogenetic analysis. *Trends in Ecology and Evolution*, 11, 152-158
- Huelsenbeck JP, Hillis DM (1993) Success of phylogenetic methods in the four-taxon case. *Systematic Biology*, 42, 247-265
- Ikemura (1985) Codon usage and tRNA content in unicellular and multicellular organisms. *Molecular Biology and Evolution*, 2, 13-24
- Jermiin LS, Crozier RH (1994) The cytochrome *b* region in the mitochondrial DNA of the ant *Tetraponera rufoniger*: sequence divergence in Hymenoptera may be associated with nucleotide content. *Journal of Molecular Evolution*, 38, 282-294
- Jermiin LS, Foster PG, Graur D, Lowe RM, Crozier RH (1996) Unbiased estimation of symmetrical directional mutation pressure from protein-coding DNA. *Journal of Molecular Evolution*, 42, 476-480
- Jermiin LS, Graur D, Lowe RM, Crozier RH (1994) Analysis of directional mutation pressure and nucleotide content in mitochondrial cytochrome *b* genes. *Journal of Molecular Evolution*, 39, 160-173
- Jermiin LS, Ho SYW, Ababneh F, Robinson J, Larkum AW (2004) The biasing effect of compositional heterogeneity on phylogenetic estimates may be underestimated. *Systematic Biology*, 53, 638-643

- Jukes TH, Bhushan V (1986) Silent nucleotide substitutions and G + C content of some mitochondrial and bacterial genes. *Journal of Molecular Evolution*, 24, 39-44
- Junqueira ACM, Lessinger AC, Torres TT, Rodriguez da Silva F, Vettore AL, Arruda P, Azeredo-Espin AML (2004) The mitochondrial genome of the blowfly *Chrysomya chloropyga* (Diptera: Calliphoridae). *Gene*, 339, 7-15
- Kim I, Cha SY, Hwang JS, Lee SM, Sohn HD, Jin BR (2005) The complete nucleotide sequence and gene organization of the mitochondrial genome of the oriental mole cricket, *Gryllotalpa orientalis* (Orthoptera: Gryllotalpidae). *Gene*, 353, 166-168
- Kimura, M (1979) The neutral theory of molecular evolution. *Scientific American*, 241, 94-104
- Kocher TD, Thomas WK, Meyer A, Edwards SV, Paabo S, Villablanca FX, Wilson AC (1989) Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proceedings of the National Academy of Science USA*, 86, 6196-6200
- Kohne DE (1970) Evolution of higher organism DNA. *Quarterly Reviews of Biophysics*, 33, 327-375
- Königsmann E (1978) Das phylogenetische system der Hymenoptera. Teil 3: Terebrantes (Unterordnung Apocrita). *Deutsche Entomologische Zeitschrift*, 25, 1-55
- Krettek A, Gullberg A, Arnason U (1995) Sequence analysis of the complete mitochondrial DNA molecule of the hedgehog, *Erinaceus europaeus*, and the phylogenetic position of the Lipotyphla. *Journal of Molecular Evolution*, 41, 952-957
- Kristensen NP (1999) Phylogeny of endopterygote insects, the most successful lineages of living organisms. *European Journal of Entomology*, 96, 237-253



- Kristensen NP (1981) Phylogeny of insect orders. *Annual Review of Entomology*, 26, 135-157
- Kristensen NP (1991) Phylogeny of the extant hexapods. In: CSIRO (ed) The insects of Australia: a textbook for students and research workers. CSIRO and Melbourne University Press, Melbourne, p 125-140
- Kristensen NP (1997) The groundplan and basal diversification of the hexapods. Arthropod relationships. Chapman and Hall, London, p 281-293
- Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics*, 17, 1244-1245
- Kumazawa Y, Nishida M (1995) Variations in mitochondrial tRNA gene organization of reptiles as phylogenetic markers. *Molecular Biology and Evolution*, 12, 759-772
- Kumazawa Y, Ota H, Nishida M, Ozawa T (1996) Gene rearrangements in snake mitochondrial genomes: highly concerted evolution of control-region-like sequences duplicated and inserted into a tRNA gene cluster. *Molecular Biology and Evolution*, 13, 1242-1254
- Kumazawa Y, Ota H, Nishida M, Ozawa T (1998) The complete nucleotide sequence of a snake (*Dinodon semicarinatus*) mitochondrial genome with two identical control regions. *Genetics*, 150, 313-329
- Kurabayashi A, Ueshima R (2000) Complete sequence of the mitochondrial DNA of the primitive opisthobranch gastropod *Pupa strigosa*: systematic implication of the genome organization. *Molecular Biology and Evolution*, 17, 266-277
- Larget B, Simon DL, Kadame JB, Sweet D (2004) A bayesian analysis of metazoan mitochondrial genome arrangements. *Molecular Biology and Evolution*, 22, 486-495

- LaSalle J, Gauld ID (1993) Hymenoptera: Their diversity, and their impact on the diversity of other organisms. In: LaSalle J, Gauld ID (eds) Hymenoptera and Biodiversity. CAB International, Wallingford, U.K., p 1-26
- Lavrov DV, Boore JL, Brown WM (2000) The complete mitochondrial DNA sequence of the horseshoe crab *Limulus polyphemus*. *Molecular Biology and Evolution*, 17, 813-824
- Lavrov DV, Brown WM (2001) *Trichinella spiralis* mtDNA: a nematode mitochondrial genome that encodes a putative ATP8 and normally structured tRNAs and has a gene arrangement relatable to those of coelomate metazoans. *Genetics*, 157, 621-637
- Le TH, Blair D, Agatsuma T, Humair P-F, Campbell NJH, Iwagami M, Littlewood TJ, Peacock B, Johnston DA, Bartley J, Rollinson D, Herniou EA, Zarlenga DS, McManus DP (2000) Phylogenies inferred from mitochondrial gene order - a cautionary tale from the parasitic flatworms. *Molecular Biology and Evolution*, 17, 1123-1125
- Lessinger AC, Junqueira ACM, Lemos TA, Kemper EL, da silva FR, Vettore AL, Arruda P, Azeredo-Espin AML (2000) The mitochondrial genome of the primary screwworm fly *Cochliomyia hominivorax* (Diptera: Calliphoridae). *Insect Molecular Biology*, 9, 521
- Lewis DL, Farr CL, Farquhar AL, Kaguni LS (1994) Sequence, organization, and evolution of the A+T region of *Drosophila melanogaster* mitochondrial DNA. *Molecular Biology and Evolution*, 11, 523-538
- Li W-H, Tanimura M (1987) The molecular clock runs more slowly in man than in apes and monkeys. *Nature*, 326, 93-96

- Lin Y-H, McLenachan PA, Gore AR, Phillips MJ, Ota R, Hendy MD, Penny D (2002) Four new mitochondrial genomes and the increased stability of evolutionary trees of mammals from improved taxon sampling. *Molecular Biology and Evolution*, 19, 2060-2070
- Lin Y-H, Penny D (2001) Implications for bat evolution from two new complete mitochondrial genomes. *Molecular Biology and Evolution*, 18, 684-688
- Liu H, Bechenbach AT (1992) Evolution of the mitochondrial cytochrome oxidase II gene among 10 orders of insects. *Molecular Phylogenetics and Evolution*, 1, 41-52
- Lockhart PJ, Steel MA, Hendy MD, Penny D (1994) Recovering evolutionary trees under a more realistic model of sequence evolution. *Molecular Biology and Evolution*, 11, 605-612
- Lowe TM, Eddy SR (1997) tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, 25, 955-964
- Lunt DH, Hyman BC (1997) Animal mitochondrial DNA recombination. *Nature*, 387, 247
- Macey JR, Larson A, Ananjeva NB, Papenfuss TJ (1997) Replication slippage may cause parallel evolution in the secondary structures of mitochondrial transfer RNAs. *Molecular Biology and Evolution*, 14, 30-39
- Macey JR, Schulte JA, II, Larson A, Papenfuss TJ (1998) Tandem duplication via light-strand synthesis may provide a precursor for mitochondrial genomic rearrangement. *Molecular Biology and Evolution*, 15, 71-75

- Macey JR, Larson A, Ananjeva NB, Fang Z, Papenfuss TJ (1997) Two novel gene orders and the role of light-strand replication in rearrangement of the vertebrate mitochondrial genome. *Molecular Biology and Evolution*, 14, 91-104
- Maniatis T, Sambrook J, Fritsch EE (1989) *Molecular Cloning*. Laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Martin AP (1995) Mitochondrial DNA sequence evolution in sharks: rates, patterns, and phylogenetic inferences. *Molecular Biology and Evolution*, 12, 1114-1123
- Martin AP, Naylor GJP, Palumbi SR (1992) Rates of mitochondrial DNA evolution in sharks are slow compared with mammals. *Nature*, 357, 153-155
- Martin AP, Palumbi SR (1993) Body size, metabolic rate, generation time, and the molecular clock. *Proceedings of the National Academy of Science USA*, 90, 4087-4091
- Masner L (1993) Superfamily Proctotrupoidea. In: Goulet H, Huber JT (eds) *Hymenoptera of the World: An Identification guide to families*. pp. 537-557 Research Branch, Agriculture Canada, Ottawa, Ontario.
- Mindell DP, Soreson MD, Dimcheff DE (1998) Multiple independent origins of mitochondrial gene order in birds. *Proceedings of the National Academy of Science USA*, 95, 10693-10697
- Mindell DP, Thacker CE (1996) Rates of molecular evolution: Phylogenetic issues and applications. *Annual Review of Ecology and Systematics*, 27, 279-303
- Mitchell SE, Cockburn AF, Seawright JA (1993) The mitochondrial genome of *Anopheles quadrimaculatus*, species A: complete nucleotide sequence and gene organization. *Genome*, 36, 1058-1073

- Miya MA, Kawaguchi A, Nishida M (2001) Mitogenomic exploration of higher teleostean phylogenies: a case study of moderate-scale evolutionary genomics with 38 newly determined complete mitochondrial DNA sequences. *Molecular Biology and Evolution*, 18, 1993-2009
- Montoya J, Christianson T, Levens D, Rabinowitz M, Attardi G (1982) Identification of initiation sites for heavy-strand and light-strand transcription in human mitochondrial DNA. *Proceedings of the Natural Academy of Science USA*, 79, 7195-7199
- Mooers AO, Holmes EC (2000) The evolution of base composition and phylogenetic inference. *Trends of Ecology and Evolution*, 15, 365-369
- Moores AO, Harvey PH (1994) Metabolic rate, generation time, and rate of molecular evolution in birds. *Molecular Phylogenetics and Evolution*, 3, 344-350
- Moritz C, Brown WM (1986) Tandem duplication of D-loop and ribosomal RNA sequences in lizard mitochondrial DNA. *Science*, 233, 1425-1427
- Moritz C, Dowling TE, Brown WM (1987) Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annual Review of Ecology and Systematics*, 18, 269-292
- Morrison CL, Harvey AW, Lavery S, Tieu K, Huang Y, Cunningham CW (2002) Mitochondrial gene rearrangements confirm the parallel evolution of the crab-like form. *Proceedings of the Royal Society of London Series B*, 269, 345-350
- Nardi F, Carapelli A, Fanciulli PP, Dallai R, Frati F (2001) The complete mitochondrial DNA sequence of the basal hexapod *Tetradontophora bielensis*: evidence for heteroplasmy and tRNA translocations. *Molecular Biology and Evolution*, 18, 1293-1304

- Nardi F, Spinsanti G, Boore JL, Carapelli A, Dallai R, Frati F (2003) Hexapod origins: monophyletic or paraphyletic? *Science*, 299, 1887-1889
- Negrisol E, Minelli A, Valle G (2004) Extensive gene order rearrangement in the mitochondrial genome of the centipede *Scutigera coleoptrata*. *Journal of Molecular Evolution*, 58, 413-423
- Nikaido M, Cao Y, Harada M, Okada N, Hasegawa M (2003) Mitochondrial phylogeny of hedgehogs and monophyly of Eulipotyphla. *Molecular Phylogenetics and Evolution*, 28, 276-284
- Nylander JAA, Ronquist F, Huelsenbeck JP, J.L. N-A (2004) Bayesian phylogenetic analysis of combined data. *Systematic Biology*, 53, 47-67
- Nylander JAA (2004) *MrModeltest 2.0*. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.
- Ohta T (1973) Slightly deleterious mutant substitutions in evolution. *Nature*, 246, 96-98
- Ohta T (1992) The nearly neutral theory of molecular evolution. *Annual Review of Ecology and Systematics*, 23, 263-286
- Ojala D, Montoya J, Attardi G (1981) tRNA punctuation model of RNA processing in human mitochondria. *Nature*, 290, 470-474
- Pääbo S, Thomas WK, Whitfield KM, Kumazawa Y, Wilson AC (1991) Rearrangements of mitochondrial transfer RNA genes in marsupials. *Journal of Molecular Evolution*, 33, 426-430

- Page RDM, Lee PLM, Becher SA, Griffiths R, Clayton DH (1998) A different tempo of mitochondrial DNA evolution in birds and their parasitic lice. *Molecular Phylogenetics and Evolution*, 9, 276-293
- Pashley DP, McPherson BA, Zimmer EA (1993) Systematics of holometabolous insect orders based on 18S ribosomal RNA. *Molecular Phylogenetics and Evolution*, 2, 132-142
- Penny D, McComish BJ, Charleston MA, Hendy MD (2001) Mathematical elegance with biochemical realism: the covarion model of molecular evolution. *Journal of Molecular Evolution*, 53, 711-723
- Phillips MJ, Penny D (2003) The root of the mammalian tree inferred from whole mitochondrial genomes. *Molecular Phylogenetics and Evolution*, 28, 171-185
- Pol D, Siddall ME (2001) Biases in maximum likelihood and parsimony: a simulation approach to a 10-taxon case. *Cladistics*, 17, 266-281
- Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics*, 14, 817-818
- Prendini L (2001) Species or supraspecific taxa as terminals in cladistic analysis? Groundplans versus exemplars revisited. *Systematic Biology*, 50, 290-300
- Quicke DLJ (1997) *Parasitic Wasps*. Chapman and Hall, London
- Quicke DLJ, Fitton MG, Ingram S (1992) Phylogenetic implications of the structure and distribution of ovipositor valvelli in the Hymenoptera (Insecta). *Journal of Natural History*, 26, 587-608

- Quicke DLJ, Fitton MG, Tunstead JR, Ingram SN, Gaitens PV (1994) Ovipositor structure and relationships within the Hymenoptera, with special reference to the Ichneumonoidea. *Journal of Natural History*, 28, 635-682
- Quinn TW, Wilson AC (1993) Sequence evolution in and around the control region in birds. *Journal of Molecular Evolution*, 37, 417-425
- Rand DM (1993) Endotherms, ectotherms, and mitochondrial genome-size variation. *Journal of Molecular Evolution*, 37, 281-295
- Rannala B, Huelsenbeck J, Yang Z, Nielsen R (1998) Taxon sampling and the accuracy of large phylogenies. *Systematic Biology*, 47, 702-710
- Rannala B (2002) Identifiability of parameters in MCMC bayesian inference of phylogeny. *Systematic Biology*, 51, 754-760
- Rasnitsyn AP (1980) The origin and evolution of the Hymenoptera. Trudy Paleontologicheskogo Instituta. *Akademiya Nauk SSSR* 174:1-192 (In Russian)
- Rasnitsyn AP (1988). An outline of evolution of the hymenopterous insects (Order Vespida). *Oriental Insects*, 22, 115-145
- Roehrdanz RL, Degrugillier ME, Black WC IV (2002) Novel rearrangements of arthropod mitochondrial DNA detected with long-PCR: applications to arthropod phylogeny and evolution. *Molecular Biology and Evolution*, 19, 841-849
- Ronquist, F., Rasnitsyn, A.P., Roy, A., Eriksson, K. & Lindgren, M. (1999). Phylogeny of the Hymenoptera: a cladistic reanalysis of Rasnitsyn's (1988) data. *Zoologica Scripta*, 28, 13-50
- Rosenberg MS, Kumar S (2003) Taxon sampling, bioinformatics, and phylogenomics. *Systematic Biology*, 52, 119-124



- Russo CA, Takezaki N, Nei M (1996) Efficiencies of different genes and different tree-building methods in recovering a known phylogeny. *Molecular Biology and Evolution*, 13, 525-536
- Saccone C, De Giorgi C, Gissi C, Pesole G, Reyes A (1999) Evolutionary genomics in metazoa: the mitochondrial DNA as a model. *Gene*, 238, 195-209
- Schulmeister S, Wheeler WC, Carpenter JC (2002) Simultaneous analysis of the basal lineages of Hymenoptera (Insecta) using sensitivity analysis. *Cladistics*, 18, 455-484
- Scouras A, Beckenbach K, Arndt A, Smith MJ (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. *Molecular Phylogenetics and Evolution*, 31, 50-65
- Serb JM, Lydeard C (2003) Complete mtDNA sequence of the North American freshwater mussel, *Lampsilis ornata* (Unionidae): an examination of the evolution and phylogenetic utility of mitochondrial genome organization in Bivalvia (Mollusca). *Molecular Biology and Evolution*, 20, 1854-1866
- Shao R, Campbell NJH, Barker SC (2001a) Numerous gene rearrangements in the mitochondrial genome of the wallaby louse, *Heterodoxus macropus* (Phthiraptera). *Molecular Biology and Evolution*, 18, 858-865
- Shao R, Campbell NJH, Schmidt ER, Barker SC (2001b) Increased rate of gene rearrangement in the mitochondrial genomes of insects in three hemipteroid orders. *Molecular Biology and Evolution*, 18:1828-1832

- Shao R, Mitani H, Barker SC, Takahashi M, Fukunaga M (2005) Novel mitochondrial gene content and gene arrangement indicate illegitimate inter-mtDNA recombination in the Chigger Mite, *Leptotrombidium pallidum*. *Journal of Molecular Evolution*, 39, 1-10
- Shao R, Barker C (2003) The highly rearrange mitochondrial genome of the plague thrips, *Thrips imaginis* (Insecta:Thysanoptara): convergence of two novel gene boundaries and an extraordinary arrangement of rRNA genes. *Molecular Biology and Evolution*, 20, 362-370
- Shao R, Dowton M, Murrel A, Barker SC (2003) Rates of Gene Rearrangement and Nucleotide Substitution are correlated in the Mitochondrial Genomes of Insects. *Molecular Biology and Evolution*, 20,1612-1619
- Sharkey MJ, Wahl DB (1992) Cladistics of the Ichneumonoidea (Hymenoptera). *Journal of Hymenopteran Research*, 1, 15-24
- Sharkey M J, Roy A (2002) Phylogeny of the Hymenoptera: a reanalysis of the Ronquist et al. (1999) reanalysis, emphasizing wing venation and apocritan relationships. *Zoologica Scripta*, 31, 57-66
- Shaw M (1983). On[e] evolution of endoparasitism: the biology of some genera of Rogadinae (Braconidae). *Contributions of the American Entomological Institute*, 20, 307-328
- Shigenaga MK, Gimeno CJ, Ames BN (1989) Urinary 8 hydroxy-2'-deoxyguanosine as a biological marker of in-vivo oxidative DNA damage. *Proceedings of the National Academy of Science USA*, 86, 9697-9701
- Siddall ME, Whiting MF (1999) Long-branch abstractions. *Cladistics*, 15, 9-24

- Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P (1994). Evolution, weighting , and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaccion primers. *Annals of the Entomological Society of America*, 87, 653-701
- Singer GAC, Hickey DA (2000) Nucleotide bias causes a genome wide bias in the amino acid composition of proteins. *Molecular Biology and Evolution*, 17, 1581-1588
- Smith MJ, Arndt A, Gorski S, Fajber E (1993) The phylogeny of echinoderm classes based on mitochondrial gene arrangements. *Journal of Molecular Evolution*, 36, 545-554
- Spanos L, Koutroumbas G, Kotsyfakis M, Louis C (2000) The mitochondrial genome of the Mediterranean fruit fly, *Ceratitis capitata*. *Insect Molecular Biology*, 9, 139-144
- Stechmann A, Schlegel M (1999) Analysis of the complete mitochondrial DNA sequence of the brachiopod *Terebratulina retusa* places Brachiopoda within the protostomes. *Proceedings of the Royal Society of London Series B*, 266, 2043-2052
- Steel M, Lockhart PJ, Penny D (1993) Confidence in evolutionary trees from biological sequence data. *Nature*, 364, 440-442
- Stewart JB, Beckenbach AT (2003) Phylogenetic and genomic analysis of the complete mitochondrial DNA sequences of the spotted asparagus beetle *Crioceris duodecimpunctata*. *Molecular Phylogenetics and Evolution*, 26, 513-526
- Stolz D, Whitfield JB (1992) Viruses and virus-like entities in the parasitic Hymenoptera. *Journal of Hymenopteran Research*, 1, 125-139
- Strand MR, Obrycki JJ (1996) Host specificity of insect parasitoids and predators. *BioScience*, 46, 422-429

- Sueoka N (1962) On the genetic basis of variation and heterogeneity of DNA base composition. *Proceedings of the Natural Academy of Science USA*, 48, 582-592
- Sueoka N (1988) Directional mutation pressure and neutral molecular evolution. *Proceedings of the Natural Academy of Science USA*, 85, 2653-2657
- Sullivan J, Swofford DL (1997) Are guinea pigs rodents? The importance of adequate models in molecular phylogenetics. *Journal of Mammalian Evolution*, 4, 77-86
- Sunnucks P, Hales DF (1996) Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). *Molecular Biology and Evolution*, 13, 510-524
- Swofford DL (1998) PAUP\*. Phylogenetic Analysis using Parsimony (\*and other methods). Version 4. Sinauer Associates, Sunderland, MA.
- Swofford DL, Olsen GJ, Waddell PJ, Hillis DM (1996) Phylogenetic inference. in D.M. Hillis, C. Mortiz, and B.K. Mable (Eds.) *Molecular Systematics* (pp. 407-514), Sinauer Associates, Sunderland, Massachusetts.
- Swofford DL, Waddell PJ, Huelsenbeck JP, Foster PG, Lewis PO, Rogers JS (2001) Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. *Systematic Biology*, 50, 525-539
- Tajima F (1993) Simple methods for testing the molecular clock hypothesis. *Genetics*, 135, 599-607
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22, 4673-4680

- Vilhelmsen L (1997) The phylogeny of lower Hymenoptera (Insecta), with a summary of the early evolutionary history of the order. *Journal of Zoological Systematics and Evolutionary Research*, 35, 49-70
- Vilhelmsen L (2000) The ovipositor apparatus of basal Hymenoptera (Insecta): phylogenetic implications and functional morphology. *Zoologica Scripta*, 29, 319-345
- Vilhelmsen L (2001) Phylogeny and classification of the extant basal lineages of the Hymenoptera. *Zoological journal of the Linnean Society*, 29, 319-345
- Wagner JR, Hu C-C, Ames BN (1992) Endogenous oxidative damage of deoxycytidine in DNA. *Proceedings of the Natural Academy of Science USA*, 89, 3380-3384
- Wheeler WC, Whiting M, Wheeler QD, Carpenter JM (2001) The phylogeny of the extant hexapod orders. *Cladistics*, 17, 113-169
- Wheeler WC, Hayashi CY (1998) The phylogeny of the extant chelicerate orders. *Cladistics*, 14, 173-192
- Whitfield JB (2003) Phylogenetic insights into the evolution of parasitism in Hymenoptera. *Advances in Parasitology*, 54, 69-100
- Whitfield JB (1992) Phylogeny of the non-aculeate Apocrita and the evolution of parasitism in the Hymenoptera. *Journal of Hymenopteran Research*, 1, 3-14
- Whitfield JB (1998). Phylogeny and evolution of host-parasitoid interactions in Hymenoptera. *Annual Review of Entomology*, 43, 129-151

- Whiting MF (2002a) Mecoptera is paraphyletic: multiple genes and phylogeny of Mecoptera and Siphonaptera. *Zoologica Scripta*, 31, 93-104
- Whiting MF (2002b) Phylogeny of the holometabolous insect orders: molecular evidence. *Zoologica Scripta*, 31, 3-15
- Whiting MF (2002c) XXI International Congress of Entomology, Iguassu Falls, Brazil, August, 2000. *Zoologica Scripta*, 31, 1-2
- Whiting MF, Carpenter JC, Wheeler Q, Wheeler WC (1997) The Strepsiptera problem: phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. *Systematic Biology*, 46, 1-68
- Whiting MF, Carpenter JC, Wheeler Q, Wheeler WC (1997) The Strepsiptera problem: phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. *Systematic Biology*, 46, 1-68
- Wiegmann BM, Mitter C, Regier JC, Friedlander TP, Wagner DM, Nielsen ES (2000) Nuclear genes resolve Mesozoic-aged divergences in the insect order Lepidoptera. *Molecular Phylogenetics and Evolution*, 15, 242-259
- Willson (1999) A higher order parsimony method to reduce long-branch attraction. *Molecular Biology and Evolution*, 16, 694-705
- Wilson K, Cahill V, Ballment E, Benzie J (2000) The complete sequence of the mitochondrial genome of the crustacean *Penaeus monodon*: are malacostracan crustaceans more closely related to insects than to branchiopods? *Molecular Biology and Evolution*, 17, 863-874
- Wolstenholme DR (1992) Animal mitochondrial DNA: structure and evolution. *International Review of Cytology-A Survey of Cell Biology*, 141, 173-216
- Wu C-I, Li W-H (1985) Evidence for higher rates of nucleotide substitution in rodents than in man. *Proceedings of the Natural Academy of Science USA*, 82, 1741-1745

- Yang Z (1994). Maximum Likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular Evolution*, 39, 306-314
- Yang Z (1996) Among-site rate variation and its impact on phylogenetic analysis. *Trends in Ecology and Evolution*, 11, 367-372
- Yeates DK (1995) Groundplans and exemplars: paths to the tree of life. *Cladistics*, 11, 343-357
- Yokobori S, Ueda T, Feldmaier-Fuchs G, Paabo S, Ueshima R, Kondow A, Nishikawa K, Watanabe K (1999) Complete DNA sequence of the mitochondrial genome of the ascidian *Halocynthia roretzi* (Chordata, Urochordata). *Genetics*, 153,1851-1862
- Yukuhiro K, Sezutsu H, Itoh M, Shimizu K, Banno Y (2000) Rearrangements have occurred between the mitochondrial genomes of the wild Mulberry Silkmoth, *Bombyx mandarina*, and its close relative, the Domesticated Silkmoth, *Bombyx mori*. *Molecular Biology and Evolution*, 19, 1385-1389
- Zardoya R, Cao Y, Hasegawa M, Meyer A (1998) Searching for the closest living relative(s) of tetrapods through evolutionary analyses of mitochondrial and nuclear data. *Molecular Biology and Evolution*, 15, 506-517
- Zardoya R, Meyer A (1996) Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. *Molecular Biology and Evolution*, 13, 933-942
- Zhang D-X, Hewitt GM (1997) Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. *Biochemical Systematics and Ecology*, 25, 99-120

Zuckerlandl E, Pauling L (1965) Evolutionary divergence and convergence in proteins.  
Pp. 97-166. In: Bryson, V., and H.J. Vogel (eds). *Evolving Genes and Proteins*.  
Academic Press, New York.