University of Kentucky
UKnowledge

University of Kentucky Doctoral Dissertations
Graduate School

2009

HYMENOPTERAN MOLECULAR PHYLOGENETICS: FROM APOCRITA TO BRACONIDAE (ICHNEUMONOIDEA)

Barbara J. Sharanowski
University of Kentucky, barb.sharanowski@gmail.com

Right click to open a feedback form in a new tab to let us know how this document benefits you.

Recommended Citation

Sharanowski, Barbara J., "HYMENOPTERAN MOLECULAR PHYLOGENETICS: FROM APOCRITA TO BRACONIDAE (ICHNEUMONOIDEA)" (2009). *University of Kentucky Doctoral Dissertations*. 695.
https://uknowledge.uky.edu/gradschool_diss/695

This Dissertation is brought to you for free and open access by the Graduate School at UKnowledge. It has been accepted for inclusion in University of Kentucky Doctoral Dissertations by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

ABSTRACT OF DISSERTATION

Barbara J. Sharanowski

The Graduate School

University of Kentucky

2009

HYMENOPTERAN MOLECULAR PHYLOGENETICS: FROM APOCRITA TO BRACONIDAE (ICHNEUMONOIDEA)

ABSTRACT OF DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for degree of Doctor of Philosophy in the Department of Entomology, College of Agriculture at the University of Kentucky

By
Barbara J. Sharanowski
Lexington, Kentucky

Director: Dr. Michael J. Sharkey, Professor of Entomology
Lexington, Kentucky
2009

Copyright © Barbara J. Sharanowski 2009

ABSTRACT OF DISSERTATION

HYMENOPTERAN MOLECULAR PHYLOGENETICS: FROM APOCRITA TO BRACONIDAE (ICHNEUMONOIDEA)

Two separate phylogenetic studies were performed for two different taxonomic levels within Hymenoptera. The first study examined the utility of expressed sequence tags for resolving relationships among hymenopteran superfamilies. Transcripts were assembled from 14,000 sequenced clones for 6 disparate Hymenopteran taxa, averaging over 660 unique contigs per species. Orthology and gene determination were performed using modifications to a previously developed computerized pipeline and compared against annotated insect genomes. Sequences from additional taxa were added from public databases with a final dataset of 24 genes for 16 taxa.

The concatenated dataset recovered a robust and well-supported topology; however, there was extreme incongruity among individual gene trees. Analyses of sequences indicated strong compositional and transition biases, particularly in the third codon positions. The use of filtered supernetworks aided visualization of the existing congruent phylogenetic signal that existed across the individual gene trees. Additionally, treeness triangle plots indicated a strong residual signal in several gene trees and across codon positions in the concatenated dataset. However, most analyses of the concatenated dataset recovered expected relationships, known from other independent analyses. Thus, ESTs provide a powerful source of information for phylogenetic analysis, but results are sensitive to low taxonomic sampling and missing data.

The second study examined subfamilial relationships within the parasitoid family Braconidae, using over 4kb of sequence data for 139 taxa. Bayesian inference of the concatenated dataset recovered a robust phylogeny, particularly for early divergences within the family. There was strong evidence supporting two independent lineages within the family: one leading to the non-cyclostomes and one leading to the cyclostomes. Ancestral state reconstructions were performed to test the theory of ectoparasitism as the ancestral condition for all taxa within the family. Results indicated an endoparasitic ancestor for the family and for the non-cyclostome lineage, with an early transition to ectoparasitism for the cyclostome lineage. However, reconstructions of some nodes were sensitive to outgroup coding and will also be impacted with increased biological knowledge.

KEYWORDS: Molecular Phylogenetic Systematics, Hymenopteran Phylogenomics, Evolution of Parasitism, Braconidae, Expressed Sequence Tags (ESTs)

MULTIMEDIA ELEMENTS USED: JPEG (.jpg);

Barbara J. Sharanowski

May 6, 2009

HYMENOPTERAN MOLECULAR PHYLOGENETICS: FROM APOCRITA TO BRACONIDAE (ICHNEUMONOIDEA)

By

Barbara J. Sharanowski

Dr. Michael Sharkey
Director of Dissertation

Dr. Kenneth Yeargan
Director of Graduate Studies

May 6, 2009
Date

FOR THE USE OF DISSERTATIONS

Unpublished dissertations submitted for the Doctor's degree and deposited in the University of Kentucky Library are as a rule open for inspection, but are to be used only with due regard to the rights of the authors. Bibliographical references may be noted, but quotations or summaries of parts may be published only with the permission of the author, and with the usual scholarly acknowledgments.

Extensive copying or publication of the dissertation in whole or in part also requires the consent of the Dean of the Graduate School of the University of Kentucky.

A library that borrows this dissertation for use by its patrons is expected to secure the signature of each user.

Name	Date

DISSERTATION

Barbara J. Sharanowski

The Graduate School

University of Kentucky

2009

HYMENOPTERAN MOLECULAR PHYLOGENETICS: FROM APOCRITA TO BRACONIDAE (ICHNEUMONOIDEA)

DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for degree of Doctor of Philosophy in the Department of Entomology, College of Agriculture at the University of Kentucky

By
Barbara J. Sharanowski
Lexington, Kentucky

Director: Dr. Michael J. Sharkey, Professor of Entomology
Lexington, Kentucky
2009

Copyright © Barbara J. Sharanowski 2009

This thesis is dedicated to my wonderful and loving husband Terry, who supported me in every way imaginable. Thanks for the moving, job changing, becoming my own personal maid while writing this thesis, and lifting my spirits every day for the past 11 years.

(Now it's time, L.L.M.E!).

ACKNOWLEDGEMENTS

First and foremost I would like to thank my graduate advisor, Dr. Michael Sharkey, who has been incredibly supportive throughout my tenure at the University of Kentucky. He has allowed me the flexibility to learn and grow as a researcher, inspired me to improve my work and go the extra mile, and has aided me to attend numerous conferences, collecting and museum trips, and educational workshops. I would also like to thank the following labmates and colleagues, who have taught me invaluable skills and my fields, constructively challenged my ideas, collaborated with me, and have helped me in countless ways to complete my projects on time: Andy Boring, Kevin Pitz, Dicky Yu, Katja Seltmann, Ashley Dowling, Carlos Sarmiento, Kacie Johansen, Ray Fisher, Adam Kesheimer, Cheryl Lindsay, Jonathan Roden, Khartik, Jennie Edelen, Stephanie Clutts, Martha Potts, Debra Murray, and Tom Dodson. I would especially like to thank Andy Boring for his friendship and understanding throughout the years. I would like to extend a special thanks to all of my committee members who have helped me in different ways. My gratitude extends to: Dr. Dan Potter for inspiring me to be a great teacher; to Dr. Randall Voss for challenging me to new heights in my research and allowing me to utilize his lab and personnel to learn genomics; and to Dr. John Obrycki helping me think about my research in a larger perspective and for all the administrative help along the way. I would also like to thank Dr. Chris Schardl, Dr. Eric Chapman, and Dr. Wiesrock for insightful comments on my thesis and helping me learn a variety of new analytical techniques. Within the Entomology Department, I would like to thank Kelly McHone, Karissa Arnold, and Darlene Thorpe for making my life as a student easier with continual administrative support. I would also like to thank Dr. Yeargan, who was an excellent DGS, always helpful and with an open door. Additionally, I extend my gratitude to Dr. Reddi Palli, Dr. Bruce Webb, Dr. James Harwood, and Dr. Stephen Dobson for allowing me to use equipment within their labs. In Dr. Webb's lab, I would like to thank Tonja Fisher for helping me with bioinformatics software. I would also like to thank Dr. Randall Voss's entire lab for countless help, but I would especially like to thank Dr. John Walker for all of the time he devoted to teaching me genomics bench work. I would like to thank Catherin Linnen, Dr. Bruce Webb, and Kimberley Ferrero for providing specimens from colonies for my phylogenomic research. I would like to thank all of my family and friends who have supported and inspired my academic pursuits, especially my husband Terry Sharanowski, who has often

put his own career on hold to accommodate mine. I want to warmly thank all of the Hymenopterists who have inspired me, welcomed me, sent specimens, and taught me over the years, particularly Dr, Andy Bennett, Dr. Bob Wharton, and Dr. Donald Quicke. I also extend sincere gratitude to all of the organizations that have provided funding for research and travel, including the University of Kentucky Graduate School, The University of Kentucky Women's Club, The Department of Entomology, and the National Science Foundation.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS……iii
List of Tables……vii
List of Figures……viii
CHAPTER 1: INTRODUCTION……1
CHAPTER 2: HYMENOPTERAN PHYLOGENOMICS……3

2.1 Introduction……3
2.2 Taxonomic Background……4
2.3 Materials and Methods……6

2.3.1 Insect Specimens……6
2.3.2 RNA Extraction and Construction of cDNA libraries ……7
2.3.3 Contig Assembly and Identification of Orthologs……9
2.3.4 Genetic Statistics……11
2.3.5 Phylogenetic inference……11
2.3.6 Filtered Supernetworks and Treeness Triangles……12

2.4 Results……12

2.4.1 Concatenated Datasets……12
2.4.2 Nucleotide Composition Bias……20
2.4.3 Individual Gene Analyses ……25
2.4.4 Visualizing Conflict and Compatibility with Filtered Supernetworks……25
2.4.5 Visualizing Phylogenetic Signal Using Treeness Triangles……30

2.5 Discussion……32

2.5.1 Phylogenetic Implications……35

2.6 Conclusions……37

CHAPTER 3: MOLECULAR SYSTEMATICS OF THE BRACONIDAE (HYMENOPTERA: ICHNEUMONOIDEA) ……38

3.1 Introduction……38
3.2 Taxonomic Background……39
3.3 Materials and Methods……44

3.3.1 Taxonomic Sampling……44
3.3.2 DNA protocols……48
3.3.3 Multiple Sequence Alignment……50
3.3.4 Phylogenetic Analyses……51
3.3.5 Ancestral State Reconstructions……52

3.4 Results……………………………………………………………………………………………………52

3.4.1 Gene statistics………………………………………………………………………………52
3.4.2 Individual Gene Anlayses…………………………………………………………………55

3.4.2.1 Ribosomal DNA……………………………………………………………………55
3.4.2.2 Protein-coding genes………………………………………………………………59

3.4.3 Concatenated Analysis……………………………………………………………………66
3.4.4 Ancestral State Reconstruction……………………………………………………………68

3.5 Discussion…………………………………………………………………………………………71

3.5.1 Utility of protein-coding markers…………………………………………………………71
3.5.2 Phylogenetic implications…………………………………………………………………72

3.5.2.1 Orthostoma…………………………………………………………………………76

3.5.2.1.1 Microgastroda……………………………………………………………76
3.5.2.1.2 Sigalphoda…………………………………………………………………77
3.5.2.1.3 Euphoroda…………………………………………………………………78
3.5.2.1.4 Helconoda…………………………………………………………………78

3.5.2.1.4.1 Helconopia……………………………………………………79
3.5.2.1.4.2 Macrocentropia………………………………………………80

3.5.2.2 Cyclostoma…………………………………………………………………………80

3.5.2.2.1 Aphidioda…………………………………………………………………81
3.5.2.2.2 Braconoda…………………………………………………………………81

3.5.3 Evolution of Life History among the Braconidae…………………………………………83

3.6 Conclusions………………………………………………………………………………………84

CHAPTER 4: OVERALL SUMMARY AND FUTURE DIRECTIONS……………………………………87
Appendix A…………………………………………………………………………………………………89
Appendix B…………………………………………………………………………………………………95
References…………………………………………………………………………………………………101
VITA……112

List of Tables

Table 2.1. List of taxa used in phylogenetic analyses and the number of unique contigs generated from each cDNA library sequenced 8

Table 2.2. List of genes used in analyses and the taxa represented for each gene 14

Table 2.3. List of which genes supported the clades recovered in Figure 2.1 16

Table 3.1. Exemplars utilized in this study, including location of collection and the genes that were amplified for each taxon 45

Table 3.2. Primer pairs used to amplify gene regions employed in this study and associated references. 49

Table 3.3. Gene regions utilized for individual and concatenated analyses, with the number of included taxa, aligned length, and percentage of parsimony informative sites. 53

Table 3.4. Proposed classification scheme for Braconidae. 75

List of Figures

Figure 2.1. Bayesian phylogram inferred from the concatenated dataset of 24 genes. 15

Figure 2.2. A-C. Bayesian phylograms inferred from the concatenated 24-gene dataset under varying subsets of outgroup taxa with posterior probabilities.. 18

Figure 2.3. A-B. Parsimony and maximum likelihood analyses of the concatenated 24-gene dataset 19

Figure 2.4. A-H. Bayesian phylograms inferred from various data partitioning schemes of the concatenated 24-gene dataset 22

Figure 2.5. Bayesian phylograms inferred from individual gene datasets with all data included 26

Figure 2.6. Bayesian phylograms inferred from individual gene datasets with the third position removed from genes demonstrating third position heterogeneity. 27

Figure 2.7. Filtered supernetworks of the 24-gene dataset 28

Figure 2. 8. Treeness triangle with data points calculated using the closest tree algorithm based on LogDet distance matrices of the individual genes and concatenated data for the 1st and 2nd codon positions. 31

Figure 3.1. Alternative hypotheses on the relationships among the major lineages within Braconidae. 40

Figure 3.2. Tree summarizing current knowledge of braconid subfamilial relationships based on previous molecular and morphological analyses.. 43

Figure 3.3. Nucleotide composition for all gene partitions. 54

Figure 3.4. Inferred topology from the Bayesian analysis of 28S (D1-D3) rDNA. 56

Figure 3.5. Inferred topology from the Bayesian analysis of 28S (D3-D5) rDNA. 58

Figure 3.6. Inferred topology from the Bayesian analysis of 18S rDNA....................................... 60

Figure 3.7. Inferred topology from the Bayesian analysis of CAD (54-405)................................. 61

Figure 3.8. Inferred topology from the Bayesian analysis of CAD (apmod)................................. 63

Figure 3.9. Inferred topology from the Bayesian analysis of ACC... 65

Figure 3.10. Inferred topology from the Bayesian analysis of all genes concatenated 67

Figure 3.11. Ancestral state reconstruction under the Asymmetrical Markov k-state 2 parameter model (AsymmMK) using maximum likelihood 70

Figure 3.12. Summary tree of well-supported relationships among Braconidae with proposed taxonomic changes 74

CHAPTER 1: INTRODUCTION

Phylogenetic systematics is a vital field of study that aims to establish the evolutionary relationships between organisms at all levels of classification, from intra-specific populations to the relationships between all organisms on the planet. The importance of taxonomy and systematics cannot be overstated. These fundamental disciplines provide the foundation for applied biological research by provisioning the framework for accurate taxonomic identification and host association, and for understanding evolutionary patterns and biogeographic distribution. One of the central goals of phylogenetic systematics is to understand the evolutionary patterns and processes that have given rise to the immense and magnificent diversity of organisms throughout time. Darwin (1872) advanced this sentiment when he stated, "I believe it has been with the great Tree of Life, which fills with its dead and broken branches the crust of the earth, and covers the surface with its ever-branching and beautiful ramifications (p.105)". The focus of this dissertation is on one of largest and most diverse branches on the tree of life: Hymenoptera (Insecta).

Hymenoptera is one of the most speciose groups of organisms on earth, with the number of estimated species rivaling that of the mega-diverse beetles (Austin & Dowton, 2000; LaSalle & Gauld, 1993). Beyond greatness in numbers, Hymenoptera also have an immense impact on the world's ecosystems, either as pollinators of plants (e.g. bees) (Fontaine et al., 2006), cornerstone species for ecosystem function (e.g. ants) (Hölldobler & Wilson, 1990), or as regulators of arthropod populations (e.g. parasitoids) (Austin & Dowton, 2000). Furthermore, for decades Hymenoptera have been utilized as a model system for evolutionary studies on haplo-diploid sex determination (Bull, 1981; Trivers & Hare, 1976), the development of complex social behavior (Alexander, 1974), and co-evolution in a variety of ecological settings (Janzen, 1966; Kiester et al., 1984; Wiebes, 1979). More recently, researchers have utilized advanced genetic techniques to investigate evolutionary patterns within Hymenoptera (Bezier et al., 2009; Toth et al., 2007; Wertheim et al., 2005; Wilfert et al., 2007). Similarly, the two main studies in this dissertation are attempts to utilize molecular data to explore the evolutionary relationships at different taxonomic levels within Hymenoptera, albeit with very different phylogenetic approaches.

In Chapter 2, a phylogenomic approach is utilized to examine apocritan evolutionary relationships. This relatively new approach focuses more heavily on genetic sampling by

examining substitutions in expressed sequence tags (ESTs). Since most transcripts that are found in common across a diverse set of taxa are relatively conserved, this technique offers an excellent approach for examining higher level relationships in insects at the ordinal or superfamilal level (Theodorides et al., 2002). Thus, the main premise of the study in Chapter 2 is to test the utility of using genomic information for understanding higher level hymenopteran relationships that have typically been unresolved using more conventional techniques (Castro & Dowton, 2006; Dowton & Austin, 2001; Ronquist *et al.*, 1999).

Alternatively, the study outlined in Chapter 3 utilizes a more traditional molecular phylogenetic approach to examine relationships among subfamilies within the Braconidae, with increased taxonomic sampling for a handful of molecular markers. The study in Chapter 3 builds on previous research (Belshaw et al., 1998; Dowton et al., 2002) by the addition of nuclear protein coding genes, and increased taxonomic sampling in previously unresolved lineages. Additionally, transitional patterns between endo- and ectoparasitism are examined for early diverging lineages within the Braconidae, to test long-standing assumptions (Čapek, 1970; Gauld, 1988; Tobias, 1967) on the evolution of modes of parasitism within the family.

Copyright © Barbara J. Sharanowski 2009

CHAPTER 2: HYMENOPTERAN PHYLOGENOMICS

2.1. Introduction

The Apocrita (Hymenoptera: Insecta), including bees, ants, and parasitoid wasps, constitutes one of the most important and diverse groups of organisms on earth from both an anthropogenic and environmental perspective (Austin & Dowton, 2000; Gauld & Bolton, 1988; Whitfield, 1998). Members of Apocrita are invaluable insects to humans, working as efficient parasitoids of destructive pests, as important pollinators of plants, and as keystone species in ecosystem function. Unfortunately, there is little understanding of the phylogenetic relationships among superfamilies, particularly among the highly diverse parasitic lineages. Several studies have attempted to resolve higher-level Hymenopteran relationships using morphological data (Königsmann, 1976; Königsmann, 1978a; Königsmann, 1978b; Rasnitsyn, 1988; Ronquist et al., 1999), molecular data (Castro & Dowton, 2006; Castro & Dowton, 2007; Dowton & Austin, 1994; Dowton *et al.*, 1997), or a combination of both (Carpenter & Wheeler, 1999; Dowton & Austin, 2001). Morphological datasets have been hampered by convergent homoplastic characters typical among parasitoids, as unrelated organisms have been shown to possess the same phenotypic adaptations to a similar host (Whitfield, 1992a). To date, molecular datasets have been restricted to mitochondrial and ribosomal DNA markers that are relatively easy to amplify across a wide range of taxa. While taxonomic sampling has been relatively robust, the limited number of genetic loci has failed to resolve most superfamilial relationships (Sharkey, 2007). This lack of knowledge prevents understanding of the mode and pattern of evolutionary traits, such as the evolution of parasitism strategies, social behavior, complex venoms, and polydna viruses (Whitfield, 1998; Whitfield et al., 2003).

Recent studies demonstrate the power of utilizing genomic information for phylogenetic reconstruction (Dunn et al., 2008; Philippe et al., 2005; Rokas et al., 2003; Savard et al., 2006). Expressed sequence tags (ESTs), which are fragments of coding sequence, offer a particularly abundant and efficient source of new genetic markers for phylogenetic analysis (Hughes et al., 2006). Utilizing ESTs also allows for amplification of a wider range of taxa than just those species involved in whole genome sequencing projects. Datasets based on ESTs utilize significantly more genetic information than traditional polymerase chain reaction (PCR) approaches. Increasing the number of independent molecular markers often causes gene trees to converge upon a more accurate species tree (Rokas et al., 2003; Savard et al., 2006).

However, this phylogenomic approach is often weakened by limited taxon sampling, which may increase systematic error (Baurain *et al.*, 2007; Dávalos & Perkins, 2008; Zwickl & Hillis, 2002).

The main purpose of this paper is to test the utility of using ESTs for phylogenetic analysis of Hymenoptera at the superfamilal level. The dataset includes ten hymenopteran taxa, with six of these newly sequenced for representative transcripts. Taxon sampling includes representatives of superfamilies that have been historically unresolved, and although the taxon sampling is limited, this paper presents the first attempt to reconstruct hymenopteran evolutionary relationships utilizing a phylogenomics approach.

2.2. Taxonomic Background

Hymenoptera has traditionally been divided into 2 suborders, Symphyta, or sawflies, and Apocrita, or wasp-waisted hymenopterans. While the monophyly of Apocrita has long been recognized (Königsmann, 1978a; Rasnitsyn, 1988; Ronquist et al., 1999), Symphyta is now acknowledged as a paraphyletic basal grade (Rasnitsyn, 1988; Schulmeister et al., 2002; Vilhelmsen, 2001). Apocrita has been further subdivided into two groups: Aculeata, containing the bees, ants, and stinging wasps; and Parasitica, most members of which are parasitoids of insects and arachnids. Parasitica, containing the majority of the diversity of the order, is the least understood group and is likely paraphyletic with respect to Aculeata (Brothers, 1975). Rather than utilizing these two unnatural but traditional subdivisions, Rasnitsyn (1988) proposed a new infraorder system for the extant apocritan lineages (= suborder Vespina), including Orussomorpha, Evaniomorpha, Proctotrupomorpha, Ichneumonomorpha, and Vespomorpha (more traditionally known as Aculeata). Although Rasnitsyn placed the Orussoidea within Apocrita, it is generally accepted that the parasitic Orussoidea is the sister group to Apocrita (Sharkey, 2007; Vilhelmsen, 2003). Consistent with Brothers (1975), Rasnitsyn proposed a sister relationship between Ichneumonomorpha (=Ichneumonoidea) and Vespomorpha (=Aculeata). Additionally, he suggested that the Evaniomorpha and Proctotrupomorpha were sister groups. The erection of the Proctotrupomorpha (including Cynipoidea, Proctotrupoidea, Platygastroidea, and Chalcidoidea) was a novel hypothesis from his earlier work (Rasnitsyn, 1980b) that had placed these superfamilies within the Ichneumonomorpha. Interestingly, when Ronquist et al. (1999) reanalyzed Rasnitsyn's (1988) morphological data using cladistic techniques, the relationships proposed by Rasnitsyn were not recovered (with most conflicting resolution attributable to reductional characters, see Sharkey & Roy, 2002). However, Rasnitsyn's (1988)

proposed classification of the Hymenoptera remains the most widely accepted tested hypothesis (for a full review, see Sharkey, 2007; and Whitfield, 1992a).

Dowton and Austin (1994) performed one of the first molecular analyses of Hymenoptera based on one mitochondrial gene (16S rRNA). Whereas most relationships were unresolved, they did recover a sister relationship between Ichneumonoidea and Aculeata, as proposed by Rasnitsyn (1988), albeit with very low nodal support. Additionally they recovered a clade consistent with Rasnitsyn's (1988) Proctotrupomorpha, but again with little support. Carpenter and Wheeler (1999) performed an analysis of 36 hymenopteran taxa for three genes (18S rDNA, 28S rDNA, and two regions of cytochrome oxidase I (COI)) and the morphological dataset of Ronquist (1999). Although the combined analyses recovered a monophyletic Apocrita and Aculeata, all other clades demonstrated paraphyletic relationships and paraphyletic superfamilies. Dowton and Austin (2001) expanded their dataset in 2001 to include three genes (28S rDNA, 16S rDNA, and COI), 87 taxa and the morphological dataset from Ronquist (1999). They performed multiple analyses under variable weighting schemes; however, the dataset was sensitive to analytical technique and the inclusion of morphology. Under at least one model, Dowton and Austin (2001: Figure 5., pg. 98) recovered a sister relationship between the Ichneumonoidea and Aculeata, as well as a monophyletic Proctotrupomorpha, but again these clades had relatively weak support. More recently, Castro and Dowton (2006) conducted Bayesian and parsimony analyses on the Dowton and Austin (2001) dataset with the addition of 18S rRNA sequences. They recovered a strongly supported Proctotrupomorpha in most analyses, however most other relationships were sensitive to outgroup selection, method of analysis, and gene inclusion. Unlike previous analyses, they usually recovered the Aculeata within Evaniomorpha with variable levels of support.

Molecular analyses of hymenopteran relationships have never incorporated nuclear protein coding genes, as all previous analyses have been based on ribosomal DNA or mitochondrial genes. Ribosomal DNA markers typically have great utility for lower level relationships among Hymenoptera, such as family, subfamily or tribal relationships (Deans *et al.*, 2006; Mardulyn & Whitfield, 1999; Zaldivar-Riverón *et al.*, 2008). Alternatively, mitochondrial genes appear to be too saturated for higher level relationships among Hymenoptera (Castro & Dowton, 2007). Nuclear protein-coding genes, particularly regulatory genes critical to cell function, tend to be relatively conserved across distant organisms and may offer a potent source

of genetic information that may help resolve higher-level relationships among Hymenoptera. Additionally, nuclear protein-coding genes are relatively easy to align, thereby diminishing uncertainty with respect to homology statements across alignments. Here, expressed sequence tags are used as a source of molecular characters for a small subset of hymenopteran taxa (10) representing 8 of the 15 apocritan superfamilies as recognized by Sharkey (2007). Obviously the 10 hymenopteran taxa utilized here do not represent a comprehensive sample of the taxonomic diversity within the order. However, this approach contrasts with the higher taxonomic, but low genetic sampling of previous analyses. Even with the low taxonomic sampling, it is possible to test the relationships proposed by Rasnitsyn (1988) which have been variably supported with molecular data, including: the monophyly of the Proctotrupomorpha; the placement of Aculeata, and to a limited extent, the monophyly of Evaniomorpha, and the placement of Ceraphronoidea within this putative clade.

2.3. Materials and Methods

2.3.1 Insect Specimens

The extraction of RNA necessary for developing CDNA libraries requires extremely fresh and properly preserved specimens. The main motivation for taxon selection was to sample specimens that represented apocritan superfamilies that have been historically unresolved. In particular, attempts were made to obtain representative taxa from at least one symphytan and the following apocritan superfamilies: Ichneumonoidea, Proctotrupoidea, Ceraphronoidea, Evanioidea, and Cynipoidea. However, taxon selection was limited by the availability of extremely fresh material and available funds for sequencing. Where possible, organisms were obtained from established colonies. Additional material was obtained by collecting live material from the field, although it was not always possible to obtain multiple specimens for extraction or to establish exact identifications due to the limited number of specimens and the need to keep available specimens fresh while taxonomically identifying the organisms.

Of the six species of Hymenoptera sequenced for this experiment, two were obtained from existing colonies from colleagues as follows: the symphytan, *Neodiprion sertifer* (Hymenoptera: Diprionoidea: Diprionidae, (ten males, ten females, Catherine Linnen, Harvard University); *Campoletis sonorensis* (Hymenoptera: Ichneumonoidea: Ichneumonidae) (ten males, ten females, Bruce Webb, University of Kentucky). The other four apocritan specimens were collected in Kentucky by the author (BJS) with a sweep net, including: *Pelecinus polyturator*

(Hymenoptera: Proctotrupoidea: Pelecinidae) (2 females); *Pristaulacus strangliae* (Hymenoptera: Evanioidea: Aulacidae) (3 females); an unidentified ceraphronid (Hymenoptera: Ceraphronidae) (1 female); and an unidentified eucoiliine (Hymenoptera: Figitidae) (2 females). Specimens were stored whole at -80°C until used. Table 2.1 lists all taxa in the analyses, including those whose sequences were mined from public databases, and the higher taxonomic names that are employed in all phylogenetic figures. Hymenopteran sequences mined for taxa from public databases were chosen based on availability. Outgroup sequences were chosen based on availability with an attempt to sample a broad range of taxa in which the relationships among outgroups have been well supported in other datasets. Additionally, annotated model genomes were utilized where possible to enhance the ability to determine orthology among loci.

2.3.2 RNA Extraction and Construction of cDNA Libraries

Total RNA was extracted from all available specimens using TRIzol reagent (Invitrogen) (Chomczynski & Sacchi, 1987) according to the manufacturer's instructions and further cleaned using the RNeasy Mini Kit (Qiagen). The integrity of RNA of each species was analyzed on denaturing formaldehyde/agarose gel and quantified in a spectrometer to ensure a minimum of 50 ng starting material in a maximum of 3 µL. Additionally, RNA quality assessments were performed on a bioanalyzer at the University of Kentucky MicroArray Core Facility.

Libraries were constructed using SMART™ cDNA Library Construction kit (Protocol PT3000-1, CLONTECH Laboratories), using the long-distance PCR method (Barnes, 1994; Chenchik et al., 1998). First strand cDNA synthesis was achieved using 1-3 µL of RNA sample, 1 µL Superscript II reverse transcriptase (Life Technologies), 1 µL SMART IV Oligonucleotide (10 µM) (5′-AAG CAG TGG TAT CAA CGC AGA GTG GCC ATT ACG GCC GGG-3′), 1 µL CDS III/3′ PCR primer (10 µM) (5′-ATT CTA GAG GCC GAG GCG GCC GAC ATG-$d(T)_{30}$ (A/G/C)N-3′), 1 µL dNTP (10 mM), 1 µL dithiothreitol (DTT) (20 mM) in 2 µL of buffer (250 mM Tris (pH 8.3), and 30 mM $MgCl_2$, 375 mM KCl). Amplification of cDNA by PCR was performed in a GeneAmp 480 thermocycler using 5′PCR Primer (5′-AAG CAG TGG TAT CAA CGC AGA GT-3′) and CDS III/3′ PCR primer with the Advantage PCR kit (CLONTECH Laboratories) following the manufacturer's instructions. Thermocycler conditions were as follows: 1 min at 95°C followed by 18 to 24 cycles of 15 s at 95°C and 6 min at 68°C. Subsequently, the double stranded cDNA was digested with proteinase K (20 µg/µL), digested with a *Sfi*I restriction enzyme and size fractioned following the manufacturer's instructions (CLONTECH Laboratories).

Table 2.1. List of taxa used in phylogenetic analyses and the number of unique contigs generated from each cDNA library sequenced. Abbreviated names are used in some tables for brevity, but all figures use the names listed in the 5th column to demonstrate the higher level relationships. N/A is not applicable as sequences for these taxa were mined from public databases

Species	Abbr. name	Family	Superfamily	Taxon name used in phylogenies	No. of clones sampled from library	No. of unique contigs
Neodiprion sertifer	Ns	Diprionidae	Tenthredinoidea	Symphyta	2000	795
Campoletis sonorensis	Cs	Ichneumonidae	Ichneumonoidea	Ichneumonidae	2000	761
Lysiphlebus testacipes	Lt	Braconidae	Ichneumonoidea	Braconidae	n/a	n/a
Pristaulacus strangliae	Ps	Aulacidae	Evanioidea	Evanioidea	2000	581
Pelecinus polyturator	Pp	Pelecinidae	Proctotrupoidea	Proctotrupoidea	3000	842
Eucoiliinae sp.	Fe	Figitidae	Cynipoidea	Cynipoidea	2500	536
Nasonia vitripennis	Nv	Pteromalidae	Chalcidoidea	Chalcidoidea	n/a	n/a
Ceraphronidae sp.	Ce	Ceraphronidae	Ceraphronoidea	Ceraphronoidea	2500	492
Apis mellifera	Am	Apidae	Apoidea	Apoidea	n/a	n/a
Solenopsis invicta	Si	Formicidae	Vespoidea	Vespoidea	n/a	n/a
Tribolium castaneum	Tc	Tenebrionidae	Tenebrionoidea	Coleoptera	n/a	n/a
Bombyx mori	Bm	Bombycidae	Bombycoidea	Lepidoptera	n/a	n/a
Drosophila melanogaster	Dm	Drosophilidae	Ephydroidea	Diptera	n/a	n/a
Acyrthosiphon pisum	Ap	Aphididae	Aphidoidea	Hemiptera	n/a	n/a
Myzus persicae	Am	Aphididae	Aphidoidea	Hemiptera	n/a	n/a
Locusta migratoria	Lm	Acrididae	Acridoidea	Orthoptera	n/a	n/a

The cDNA libraries were ligated to λ TriplEx2™ vector in a packaging reaction using PhageMaker® System (Novagen), following the manufacturer's instructions. Phage transductions were performed for 2 hours at 31°C using the BM25.8 *E.coli* host strain in LB broth with 10 mM $MgSO_4$. The converted library was then plated on LB agar plates containing carbenicillin (50 µg/ml) and grown overnight at 37°C. Isolated plaques were sampled and placed into 96-well PCR plates containing 50 µL of LB broth with 8% glycerol and carbenicillin (50 µg/ml) and grown overnight at 37°C. The individual colonies were then sampled and picked into 20 µL of water and heated at 95°C for 2 minutes. This mixture (2 µL) was then used as template in a 25 µL PCR reaction with 1 µL (50 nM) of TripleX 5LD (5′-CTC GGG AAG CGC GCC ATT GTG TTG GT-3′), 1 µL (50 nM) of Triplex 3LD (5′-TAA TAC GAC TCA CTA TAG GGC GAA TT-3′), 2.5 µL

dNTP (1.25mM), 0.4 of Taq, 2.5 µL of buffer (500mM KCl, 100mM Tris-HCl (pH 9.0), and 1% Triton-X-100), and 1.2 µL $MgCl_2$ (25 mM). Thermocycler conditions were as follows: 3 min at 94°C followed by 32 cycles of 30 s at 94°C, 30 s at 60°C, and 1 min at 72°C, with a final extension of 7 min at 72°C. Amplified samples were electrophoresed in 1% agarose gel alongside a 1 kb ladder and all reactions demonstrating single bands above 200 bp were sent to the Advanced Genetic Technologies Center, University of Kentucky, for sequencing. Both product purification and sequencing were performed using Agencourt CleanSEQ magnetic beads and an Applied Biosystems 3730xl DNA Analyzer, respectively.

2.3.3 Contig Assembly and Identification of Orthologs

Sequences and chromatograms were subject to vector and low-quality sequence removal and assembled into contigs using high stringency settings in SeqMan (DNASTAR Inc., Madison, WI, USA). As an initial search to identify orthologs, modifications were made to a pre-developed automated software program, entitled Hal, designed for identifying orthologs of proteomes (Robbertse et al., 2006). In addition to the 6 species of Hymenoptera analyzed here, the predicted genes of 3 annotated model genomes were utilized, including: *Drosophila melanogaster* (Diptera), *Bombyx mori* (Lepidoptera), and *Apis mellifera* (Hymenoptera). These coding sequences were downloaded from the following resources: Flybase (The FlyBase Consortium, 2008; Tweedie et al., 2009), SilkDB (Beijing Genomics Institute, 200; Wang et al., 2005), and BeeBase (Elsik et al., 2006; The Honeybee Genome Sequencing Consortium, 2008), respectively.

All sequences were translated into amino acid sequences and run through the Hal pipeline which included: an all versus all blastp search (Altschul et al., 1990) with a cutoff e-value of 1e-1; clustering with MCL (http://micans.org/mcl/) across several inflation parameters (Enright et al., 2002); and cluster filtering. Filtering involved selecting clusters containing proteins that had best hits to other proteins within that same cluster. Additionally, clusters were excluded if it contained more than one protein per species. At minimum, 5 of the 9 taxa had to be included in each cluster. A total of 76 clusters were identified using the pipeline.

When there are hundreds of proteins for each taxon, this high-throughput method of identifying orthologs is extremely efficient. If paralogous sequences seep into the dataset, the conflicting phylogenetic signal is likely to be swamped out by the hundreds of orthologous genes. However, when there are fewer sequences for each taxon, paralogy can contribute

significant noise to the dataset and potentially affect the outcome. Thus, to further prevent paralogs, the sequences from each cluster were filtered through another set of criteria. Each nucleotide sequence from each cluster was subject to a tBlastX search with a higher cutoff e-value of 1e-25 (Altschul et al., 1990). To be included in the final list of genes for phylogenetic analysis, each sequence had to have the same annotated gene be the best hit for 3 different model genomes: *D. melanogaster*, *B. mori*, and *A. mellifera*. To prevent the inclusion of gene family members with conserved domains, genes were excluded if multiple genes hit below an e-value of 1e-25 for any of these taxa. Additionally, the best hits had to have an identity of greater than 50% over a minimum of 60 amino acids. For these genes, additional sequences were assigned to the cluster from the following 7 taxa if they also met the above criteria: *Nasonia vitripennis* (Hymenoptera), *Solenopsis invicta* (Hymenoptera: Formicidae), *Lysiphlebus testacipes* (Hymenoptera: Braconidae) *Tribolium castaneum* (Coleoptera), *Myzus persicae* (Hemiptera: Aphididae), *Acyrthosiphon pisum* (Hemiptera: Aphididae), and *Locusta migratoria* (Orthoptera). The 'non-redundant nucleotide' and 'EST Others' database (NCBI) were used as the source of sequences for the database searches. These taxa increased sampling within the ingroup and provided multiple outgroups for the analysis. Although these criteria are somewhat arbitrary, they are fairly conservative compared to other studies (Mushegian et al., 1998; Remm et al., 2001). To minimize the amount of missing data, clusters were included only if they contained representative sequences from at least 3 of the 6 hymenopteran taxa sequenced for this experiment. While 29 of the 76 clusters met the stricter search criterion, only 12 of these contained at least 3 of the sequenced hymenopteran taxa.

Since only one sequence per species can be contained in the cluster, the pipeline eliminates potentially useful genes, as some taxa possess multiple transcript variants. Transcript variants often do not vary across the coding sequence or differ only in one or a few sites that will likely will not affect the overall phylogenetic analysis (Goodstadt & Ponting, 2006). To increase the number of genes available for analysis, all sequences from the 6 hymenopteran libraries were again examined using an all versus all blastn search (Altschul et al., 1990) with a cutoff e-value of 1e-25 using the stand alone blastall program (NCBI). All hits that were not identified with the Hal pipeline were filtered using the same criteria and methodology mentioned previously. An additional 12 genes were identified, all with at least one taxon having multiple transcript variants. All sequences were aligned using MUSCLE (Edgar, 2004), and hand edited to ensure a proper reading frame. To test whether genes with multiple transcripts were

useful and did not represent out-paralogs (Remm et al., 2001), all transcripts for all taxa were tested phylogentically. If the transcripts for a given taxon clumped together on the tree, they were considered homologous and therefore included within the dataset, provided they met all other criteria. The final dataset consisted of 24 genes, 12 identified from the pipeline and 12 identified through the method just described.

2.3.4. Genetic Statistics

The number of informative sites and tests for base composition homogeneity were performed in Paup* 4.0b10 (Swofford, 2000). Disparity index tests for substitution homogeneity across lineages and substitution pattern calculations were compiled in MEGA (Tamura et al., 2007). A Monte Carlo test with 1000 replicates was used to estimate the *p*-values for the disparity index tests (Kumar & Gadagkar, 2001a). A maximum composite likelihood estimate was used to calculate the substitution patterns for each codon position, with a heterogeneous pattern among lineages and variable rates among sites estimated with a gamma distribution (Tamura et al., 2004). Estimates for the gamma parameter were obtained using Paup* 4.0b10 (Swofford, 2000) with 500 random addition sequences with tree bisection and reconnection (TBR), using the Sullivan et al. (1995) estimate. Phylogenetic trees were viewed and manipulated using Dendroscope (Huson et al., 2007).

2.3.5 Phylogenetic Inference

Phylogenetic assessments of taxa with multiple transcript variants were performed using maximum composite likelihood distances (Tamura et al., 2004) and the neighbor-joining method with MEGA 4.0.2 (Tamura et al., 2007). All analyses performed in Paup* 4.0b10 (Swofford, 2000) were aided with the PaupUp graphical interface (Calendini & Martin, 2005). MrModeltest v2.3 (Nylander, 2004; Posada & Crandall, 1998) was used with Paup* 4.0b10 (Swofford, 2000) to test for the best evolutionary model applicable to individual gene datasets using hierarchical likelihood ratio tests. For all genes, the general time reversible model had the highest likelihood with a parameter for invariant sites and among-site rate variation modeled with a gamma distribution (GTR +I+G). Peptide alignments were analyzed using the amino acid general time reversible model in MrBayes. Bayesian inference was used to analyze all concatenated and individual gene datasets with MrBayes v3.1.2 (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003). All analyses were run with 4 chains and 2 independent runs until stationarity was reached. Stationarity of the independent runs was determined using

convergence diagnostics and plots of generation versus the log probability of the data as guidelines. All datasets were partitioned for each gene and codon position. The maximum likelihood analysis was performed on the concatenated dataset with all data included,, using RAxML VI-HPC (Stamatakis, 2006) on the CIPRES Portal v. 1.14 (CIPRES Collaborative Group, 2005-2008), the general time reversible model with a parameter for invariant sites and among-site rate variation modeled with a gamma distribution (GTRGAMMAI) with rapid bootstrapping (under GTRCAT model) and automatic determination of the number of replications required (Stamatakis et al., 2008). Parsimony analyses were also performed on the full concatenated dataset using Paup* 4.0b10 (Swofford, 2000) with a heuristic search, 1000 random additions sequences, TBR, holding 5 trees per rep, and multiple states treated as polymorphisms. Standard bootstrap resampling was performed with the same heuristic search settings with 1000 replications.

2.3.6 Filtered Supernetworks and Treeness Triangles

Evolutionary networks were constructed using SplitsTree v.4.0 (Huson & Bryant, 2006), with filtered supernetworks performed using the Z-closure method (see Huson *et al.*, 2004 for a detailed explanation). Treeness triangles were generated using the Treeness Triangle program described in White et al. (2007), using the distance matrix option with trees estimated using the closet tree algorithm (Hendy, 1991). Distance matrices for treeness triangles were calculated using the LogDet method (Tamura & Kumar, 2002) in MEGA (Tamura et al., 2007), which accounts for multiple substitutions. Additionally, coordinates for the triangles were normalized using a Perl script kindly provided by the author of the program (Tim White).

2.4. Results

2.4.1 Concatenated Datasets

The final concatenated dataset contained 24 genes with an aligned length of 10917 base pairs of which 48.6 percent were parsimony informative (not including apomorphic sites). Table 2.2 lists which genes were included in the dataset and which taxa were represented in the individual gene datasets. All of the individual gene datasets had a minimum of 12 taxa with a representative transcript. Under a Bayesian framework, the 24-gene dataset recovered several expected relationships consistent with other molecular and morphological phylogenetic studies of Hymenoptera, including: a monophyletic Aculeata, a monophyletic Ichneumonoidea, Symphyta as sister to all apocritan taxa, and a sister relationship between the two most closely

related putative proctotrupomorphs (Cynipoidea + Proctotrupoidea) (Figure 2.1). Additionally, the phylogenetic positions of all outgroups were consistent with previously recovered relationships, including a monophyletic Holometabola, Hymenoptera as sister to all other Holometabola, and a sister relationship between the two included Panorpoid orders (Savard et al., 2006; Wheeler et al., 2001; Whiting, 2002).

Within the ingroup, Ceraphronoidea and Evanioidea were recovered as sister taxa, consistent with Rasnitsyn's (1988) proposed Evaniomorpha. This Evaniomorpha clade was recovered as sister to all other apocritan taxa. The Proctotrupomorpha was proposed by Rasnitsyn (1988) to include Cynipoidea, Proctotrupoidea, Chalcidoidea, and Platygastroidea. Although the platygastroids were not represented in this analysis, Chalcidoidea was not recovered with the other putative proctotrupomorphs. Rather, Proctotrupoidea + Cynipoidea were recovered as sister to Aculeata (Apoidea + Vespoidea) and Chalcidoidea was recovered as sister to Ichneumonoidea (Aculeata (Cynipoidea + Proctotrupoidea)).

Figure 2.1 depicts the number of genes that recovered a given node, determined from examining the recovered clades from individual gene analyses. Bayesian posterior probabilities are depicted above the node. Although there was high support over most of the tree, there was relatively low nodal support for Evaniomorpha and for the node containing the remaining apocritan lineages. It is possible that sampling error affected node recovery and support , particularly for the Evaniomorpha clade, as both the ceraphronoid and evanioid had 66.2 and 51.7 percent missing data (including gaps), respectively (Table 2.2). The symphytan and cynipoid also had relatively high levels of missing data, at 51.0 and 52.1 percent, respectively.

Table 2.2. List of genes used in analyses and the taxa represented for each gene. Gene numbers and symbols are referenced to FlyBase (The FlyBase Consortium, 2008). See Table 2.1 for the key to abbreviated taxon names. (P.I. = Parsimony informative).

FlyBase Gene number	FlyBase Gene symbol	Aligned Length	No. P.I. sites	Outgroups						Hymenopteran taxa									
				Lm	Bm	Dm	Tc	Ap	Mp	Nv	Am	Si	Lt	Cs	Ce	Ns	Fe	Pp	Ps
CG1746	CG1746	444	204	√	√	√	√	√	√	√	√	√	√	√	√	√	-	-	-
CG2099	RpL35A	342	177	√	√	√	√	√	√	√	√	√	√	√	-	√	√	√	-
CG2746	RpL19	612	260	√	√	√	√	√	√	√	√	√	√	√	-	-	√	√	-
CG3186	eIF-5A	486	195	√	√	√	√	√	√	√	√	√	√	√	-	√	√	-	-
CG3446	CG3446	432	339	√	√	√	√	√	√	√	√	√	√	√	-	√	-	√	-
CG3661	RpL23	423	148	√	√	√	√	√	√	√	√	√	√	√	-	√	√	-	√
CG3997	RpL39	156	60	√	√	√	√	√	√	√	√	√	√	√	√	-	√	√	-
CG4097	Pros26	471	262	√	√	√	√	√	√	√	√	√	-	√	√	-	-	-	√
CG4169	CG4169	771	540	√	√	√	√	√	√	√	√	√	√	√	-	√	-	√	-
CG4800	Tctp	531	270	√	√	√	√	√	√	√	√	√	√	√	√	√	√	-	√
CG6770	CG6770	195	101	√	√	√	√	√	√	√	√	√	√	√	√	√	-	√	-
CG6779	RpS3	708	343	√	√	√	√	√	√	√	√	√	√	√	√	-	√	√	√
CG6803	Mf	318	180	√	√	√	√	√	√	√	√	√	√	√	-	√	-	√	√
CG7178	wupA	597	229	√	√	√	√	√	√	√	√	√	-	√	-	-	-	√	√
CG7424	RpL36A	309	119	√	√	√	√	√	√	√	√	√	√	√	√	-	√	√	-
CG7434	RpL22	378	197	√	√	√	√	√	√	√	√	√	√	√	√	-	√	-	-
CG7939	RpL32	405	186	√	√	√	√	√	√	√	√	√	√	√	-	-	√	√	-
CG8332	RpS15	456	192	√	√	√	√	√	√	√	√	√	√	-	-	√	√	√	-
CG8415	RpS23	429	157	√	√	√	√	√	√	√	√	√	√	-	√	√	-	√	√
CG8857	RpS11	471	206	√	√	√	√	√	√	√	√	√	√	√	-	-	√	-	√
CG8900	RpS18	498	186	√	√	√	√	√	√	√	√	√	√	-	-	√	√	-	√
CG11271	RpS12	429	230	√	√	√	√	√	√	√	√	√	√	√	-	-	-	√	√
CG11981	Prosβ3	618	327	√	√	√	√	√	√	√	√	√	-	-	√	√	-	-	√
CG15442	RpL27A	438	198	√	√	√	√	√	√	√	√	√	√	√	√	-	√	√	√
Total no. genes for each taxon				24	24	24	24	24	24	24	24	24	21	20	11	13	14	15	12
Percent missing data				5.8	1.5	1.6	1.3	1.9	4.0	1.1	1.0	10.7	20.9	22.3	66.2	51.0	52.1	49.2	51.7

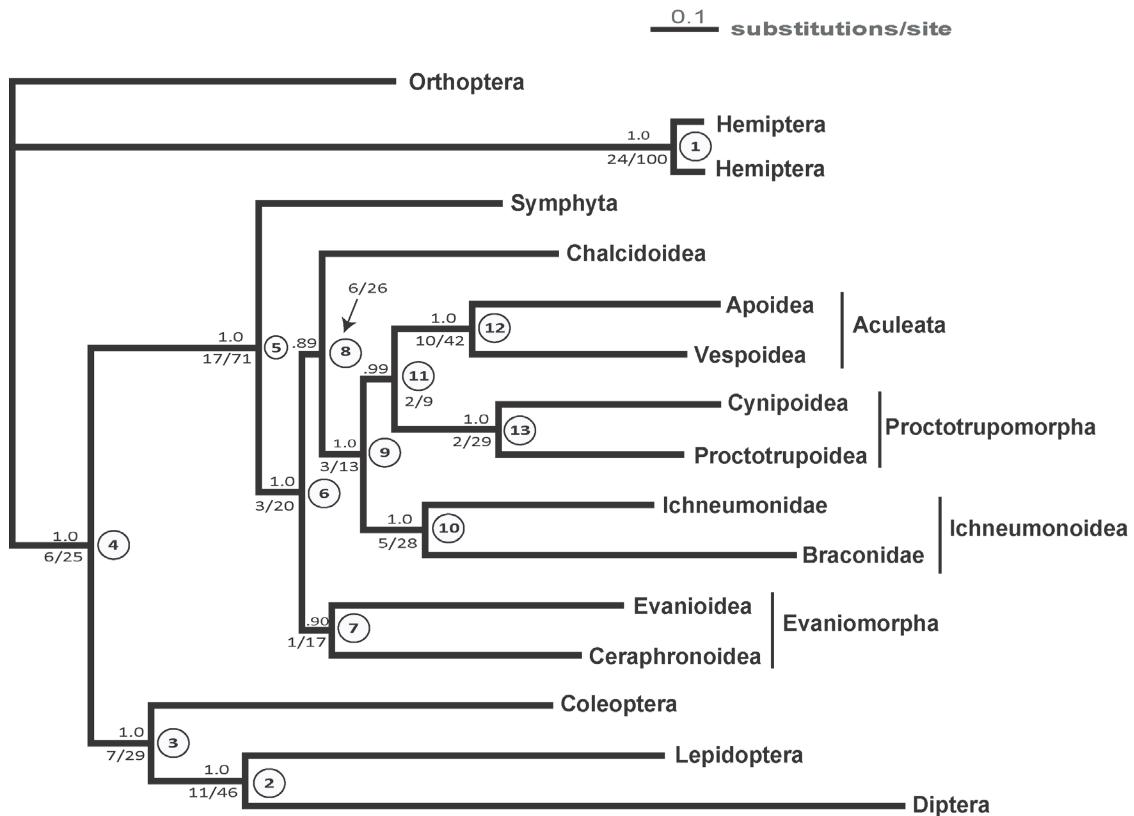


Figure 2.1. Bayesian phylogram inferred from the concatenated dataset of 24 genes. (50 million generations, burnin = 32 million generations). Posterior probabilities are listed above the node. The number of genes that recovered a clade are listed below the node, with the percentage of genes recovering the clade (out of total possible genes that could recover the clade) indicated after the forward slash. Nodes 8 and 11 have these figures above and to the right of the node, respectively. The circled numbers to the right of a node represent labels for ease of discussion and can be crossed reference with the node labels in Table 2.3.

Table 2.3 lists which genes recovered the clades depicted in Figure 2.1. The node numbers in Table 2.3 correspond to the circled node labels depicted in Figure 2.1. The highest supported node in terms of the percent of genes possible for clade recovery was unsurprisingly between the two most closely related taxa, the two hemipterans. The next highest supported node was the hymenopteran clade (node 5), with 17 out of the 24 genes indicating monophyly. While this clade has never been in doubt morphologically, the high level of support for this and other clades revealed the phylogenetic potential of these loci. Node 11, which represents a

Table 2.3. List of which genes supported the clades recovered in Figure 2.1. Refer to Figure 2.1 for node numbers. A checkmark indicates that node was recovered in the individual gene analysis; whereas a blank cell indicates the node was not recovered. A grey cell indicates that node could not be recovered due to missing taxa. A. Total number of genes supporting clade; B. Total number of genes possible for clade recovery; C. Percent of genes supporting clade.

	Node number												
gene	1	2	3	4	5	6	7	8	9	10	11	12	13
CG1746	√									√		√	
CG2099	√	√			√							√	
CG2746	√	√	√		√			√					√
CG3186	√		√	√	√								
CG3446	√	√	√					√					
CG3661	√	√											
CG3997	√												
CG4097	√				√							√	
CG4169	√				√								
CG4800	√									√		√	
CG6770	√				√								
CG6779	√	√		√	√								
CG6803	√	√	√	√	√	√				√		√	
CG7178	√				√			√	√		√	√	
CG7424	√				√			√		√			
CG7434	√									√			
CG7939	√		√		√			√					
CG8332	√		√	√	√	√		√	√			√	
CG8415	√				√	√			√				
CG8857	√	√	√	√	√							√	
CG8900	√	√											
CG11271	√	√			√							√	
CG11981	√	√			√		√						
CG15442	√	√		√	√						√	√	√
A	24	11	7	6	17	3	1	6	3	5	2	10	2
B	24	24	24	24	24	15	6	23	23	18	23	24	7
C	100	46	29	25	71	20	17	26	13	28	8.7	42	29

sister relationship between the aculeates and Cynipoidea + Proctotrupoidea, had the lowest percent of possible genes supporting the clade (Table 2.3). The validity of this clade is highly questionable, given that it has not been recovered by previous molecular, morphological, or

combined analyses (Castro & Dowton, 2006; Dowton & Austin, 1994; Dowton & Austin, 2001; Ronquist *et al.*, 1999). Although the nodal support was fairly high (0.99), it is well known that multi-gene datasets can produce very high support values, particularly posterior probabilities, for erroneous clades due to biases in the estimate (Simmons et al., 2004) or systematic error (Degnan & Rosenberg, 2006; Delsuc, 2006).

There are a number of reasons why unrelated taxa can group together on a phylogenetic tree, including: (1) long-branch attraction (Bergsten, 2005; Felsenstein, 1978); (2) violations of the model of evolution used to infer the tree, such as unequal base composition among taxa or rate heterogeneity (Gaut & Lewis, 1995; Huelsenbeck & Hillis, 1993); (3) evolutionary processes that do not conform to tree-like evolution, such as lateral gene transfer and hybridization (Beiko & Hamilton, 2006); and (4) inappropriate outgroup selection (Nixon & Carpenter, 1993; Tarríoa *et al.*, 2000). While this is not an exhaustive list, it comprises some of the most common reasons for inaccuracy in phylogenetic reconstruction. It is also important to note that these are not mutually exclusive problems. For example, long-branch attraction is often a corollary of rate heterogeneity, which violates the assumption of some evolutionary models (but see Whelan (2008)).

Given the disparity in branch lengths among the outgroup taxa, it is possible that outgroup rooting affected the result. To test for the effect of outgroup selection, three different analyses were performed. First, the orthopteran was excluded and the analysis was rooted on *A. pisum* (Hemiptera). Second, all outgroups were excluded except for the orthopteran, thereby excluding potential effects from the long branches of the panorpoid orders and the hemipterans. Finally, all outgroups were excluded except the coleopteran, potentially reducing the divergence time between the ingroup and outgroup.

When the analysis was rooted on a hemipteran and the orthopteran was excluded, some resolution was lost in the ingroup (Figure 2.2A), particularly among the relatively lower supported nodes depicted in Figure 2.1. However, the ingroup relationships were completely compatible with the phylogeny in Figure 2.1 and the aculeates are still recovered as sister to Proctotrupoidea + Cynipoidea, albeit with very low support. The Ichneumonoidea remained monophyletic but were recovered in a polytomy with the remaining apocritan lineages. Oddly, the Panorpoid orders were recovered as sister to the Hymenoptera. The second analysis, which excluded all outgroups except the orthopteran, recovered a similar tree (Figure 2.2B), however, the evanioid and ceraphronoid were recovered in a basal grade, rather than as sister taxa.

Additionally, many of the internal nodes between major apocritan lineages were weakly supported. Results were also similar when the coleopteran was the only included outgroup (Figure 2.2C). All of these analyses suggest there was very little support for branching order among Evanioidea, Ceraphronoidea, Chalcidoidea, and the remaining apocritans. In contrast, there was consistent, though weak, support for the sister relationship between Aculeata and Proctotrupoidea + Cynipoidea.

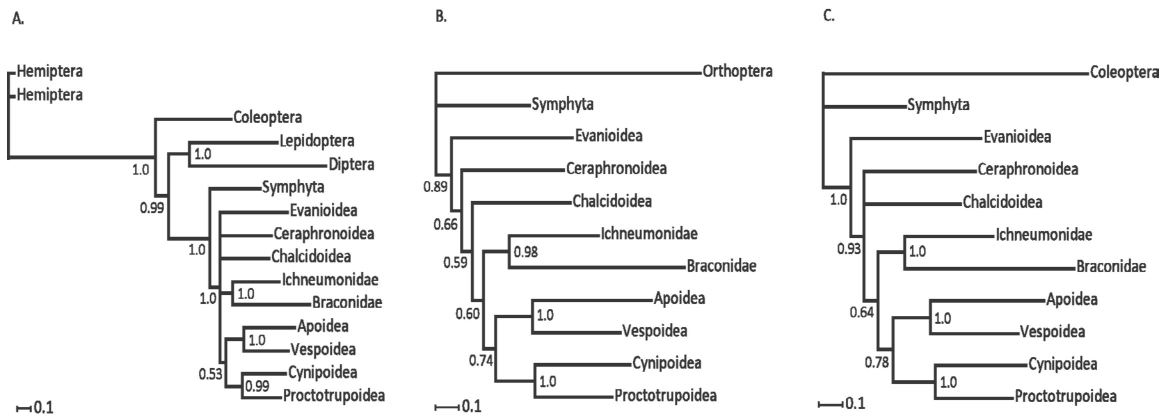


Figure 2.2. A-C. Bayesian phylograms inferred from the concatenated 24-gene dataset under varying subsets of outgroup taxa with posterior probabilities. Posterior probabilities are listed below the node or to the right of the node for terminal clades. A. Outgroup Hemiptera (A. pisum), Orthoptera excluded (120 million generations, burnin = 75 million generations). B. Outgroup Orthoptera, all other outgroups excluded (30 million generations, burnin = 12 million generations). C. Outgroup Coleoptera, all other outgroups excluded (150 million generations, burnin = 95 million generations).

Parsimony and maximum likelihood analyses were performed to test if phylogenetic method affected the results. A parsimony analysis of the concatenated dataset recovered one most parsimonious tree (Figure 2.3A). Additionally, a maximum likelihood analysis was performed and the resulting phylogeny is depicted in Figure 2.3B. Neither inference method recovered the same tree as the phylogeny inferred under a Bayesian framework, but there are similarities. All outgroup relationships were the same across all inference methods and the Hymenoptera were monophyletic. While the likelihood analysis recovered a monophyletic Aculeata and Apocrita, the parsimony analysis did not. Neither analysis placed the Chalcidoidea in a clade with Cynipoidea and Proctotrupoidea, contrary to Rasnitsyn's (1988) concept of

Proctotrupomorpha. Regardless of method, Ichneumonoidea, Aculeata, and Cynipoidea + Proctotrupoidea were recovered together in a clade, although the branching order was altered across the three methods. Most notably, the parsimony and likelihood analyses did not recover a monophyletic Ichneumonoidea, with the braconid placed in variable relationships with Aculeata. The monophyly of the Ichneumonoidea has never been controversial, thus, it is likely that either the braconid was misplaced or the ichneumonid was misplaced in both the parsimony and likelihood analyses. Given the extreme A-T bias for both Apoidea and Braconidae in the third position relative to the other taxa (Table A1, Appendix A), it is most likely that the braconid was misplaced.

It has been suggested (Akashi et al., 2007; Lockhart et al., 1994) that model based inference methods (using complex models) can handle some biased datasets better than parsimony (but see Conant and Lewis (2001)). This may account for the apparent non-monophyly of Aculeata or Apocrita under a parsimony criterion. However, this does not account for the non-monophyly of Ichneumonoidea under a likelihood framework. Since the analyses under Bayesian inference recovered all expected relationships, it was used for all further analyses.

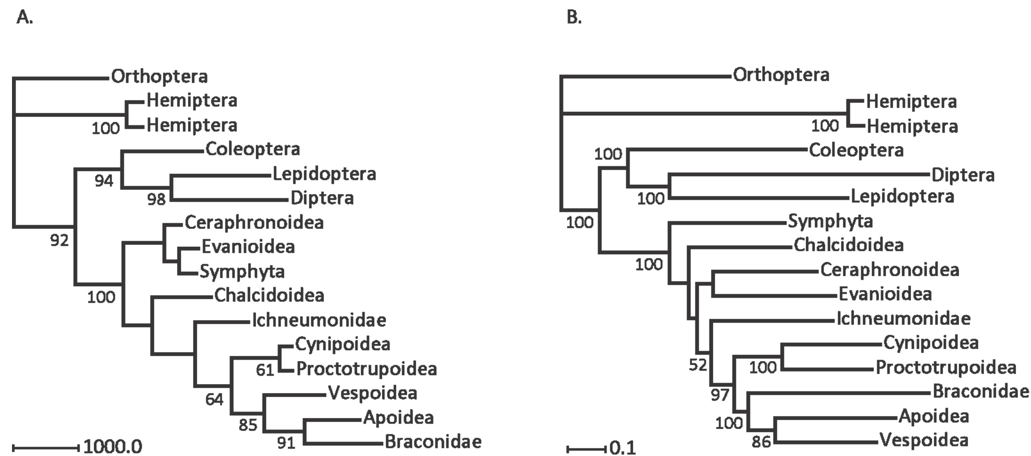


Figure 2.3. A-B. Parsimony and maximum likelihood analyses of the concatenated 24-gene dataset. A. Single most parsimonious tree (Length=21655, Consistency index = 0.48, Retention index = 0.38). Bootstrap values are depicted below the node. B. Phylogeny inferred under a maximum likelihood criterion using a GTRGAMMA model. Bootstrap values are listed above or below the node.

2.4.2 Nucleotide Composition Bias

The grouping of two unrelated taxa due to convergence and multiple substitutions is common with molecular data, particularly nucleotide data, as there are a limited number of character states (i.e. four states) (Andersson & Swofford, 2004). This is especially problematic in the third codon position due to the degenerate nature of the genetic code. To test if a nucleotide composition bias affected the analysis, chi-square tests for base composition homogeneity were performed (Table A2, Appendix A). Examining all of the individual gene alignments demonstrated that 22 out of 24 genes failed the test for base composition homogeneity ($p < 0.05$). Interestingly, when the dipteran was excluded from the test, only 10 of the 24 genes failed the homogeneity test with all data included (data not shown).

The concatenated dataset also demonstrated a lack of stationarity (Table A2, Appendix A). Each gene and the concatenated dataset were tested for nucleotide composition homogeneity for each codon position and with only the third position excluded. The null hypothesis of homogeneity was accepted for all genes with the third position excluded ($p<0.05$), but not for the concatenated dataset (Table A2, Appendix A), indicating systematic error. Only one gene (CG4169) failed the test for the first codon position (Table A2, Appendix A) and homogeneity was indicated for all genes and for the fully aligned dataset for the second position (data not shown).

There are several methods that can be used to adjust for base composition heterogeneity. A simple solution would be to exclude third codon positions in all genes that violate the assumption of homogeneity. This method does not account for synonymous changes in other positions (the 3 six-fold degenerate amino acids in the genetic code). Another method implemented with other problematic datasets (Blanquart & Lartillot, 2006; Phillips *et al.*, 2004; Regier *et al.*, 2008), called RY-coding, changes purines to the IUPAC ambiguity symbol 'R' and pyrimidines to 'Y,' thereby eliminating all transitions. Bases can be changed in all sites, just third position sites, or in third position sites and those first position sites that code for Leucine (the only 6-fold degenerate codon with synonymous transitions). Although characters are lost with both RY-coding and third position exclusion, the hope is that homoplasy is reduced, thus increasing the relative amount of historical phylogenetic signal in the data. Another solution might be to analyze the data as amino acids instead of as nucleotides, a method that has proven fairly successful using datasets with a large number genes (usually 50 or more) (Dunn et al., 2008; Philippe et al., 2005; Robbertse et al., 2006; Rokas et al., 2005; Rokas et al., 2003).

To reduce any potential effects from nucleotide composition biases on the phylogenetic inference, five different analyses were performed: (1) all third positions excluded (3-out); (2) all positions indicating heterogeneity for each individual gene excluded in a concatenated analysis (Mix-P); (3) RY-coding for third positions (RY-3); (4) RY-coding for all positions (RY-all); and (5) as amino acids (protein). To see if the outgroup had an effect when the third position was removed, the Mix-P dataset was analyzed with all outgroups excluded except the coleopteran (Mix-P-C-OG). Additionally, each codon position in the dataset was analyzed separately to test for conflicting signal across these partitions. Phylogenies for all the above analyses are presented in Figure 2.4A-H.

Phylogenetic inference of the 3 out and Mix-P datasets recovered identical topologies that included a monophyletic Holometabola, Hymenoptera, Apocrita, Aculeata, and Ichneumonoidea (Figure 2.4A), similar to the concatenated analysis with all data included (Figure 2.1). Additionally, the aculeates were again recovered sister to Proctotrupoidea + Cynipoidea, with high support. However this clade was the sister to all remaining apocritans. Additionally, the Ichneumonoidea were recovered as sister to Ceraphronoidea. Between the 3-out and Mix-P analyses, the 3-out phylogeny had lower support values, suggesting that the inclusion of the third position for the 2 genes with demonstrated homogeneity added conflicting signal.

The removal of third position characters dramatically altered the phylogenetic inference. The hypothesis in Figure 2.1 (with all data included) included the clade Ichneumonoidea (Aculeata (Proctotrupoidea + Cynipoidea)). Alternatively, the 3-out and Mix-P topologies inferred a completely different evolutionary history for the Ichneumonoidea and an earlier divergence for the clade Aculeata (Proctotrupoidea + Cynipoidea).

The phylogeny inferred from the RY- 3 dataset was somewhat similar to the Mix-P and 3-out topologies (Figure 2.4B). However, the relative positions of the Ceraphronoidea and Chalcidoidea were reversed. Additionally, there was a basal apocritan polytomy. Again, all of the expected relationships were recovered with high support, including a monophyletic Hymenoptera, Apocrita, Aculeata, and Ichneumonoidea. Most of the remaining nodes, which are of greatest interest, were poorly supported.

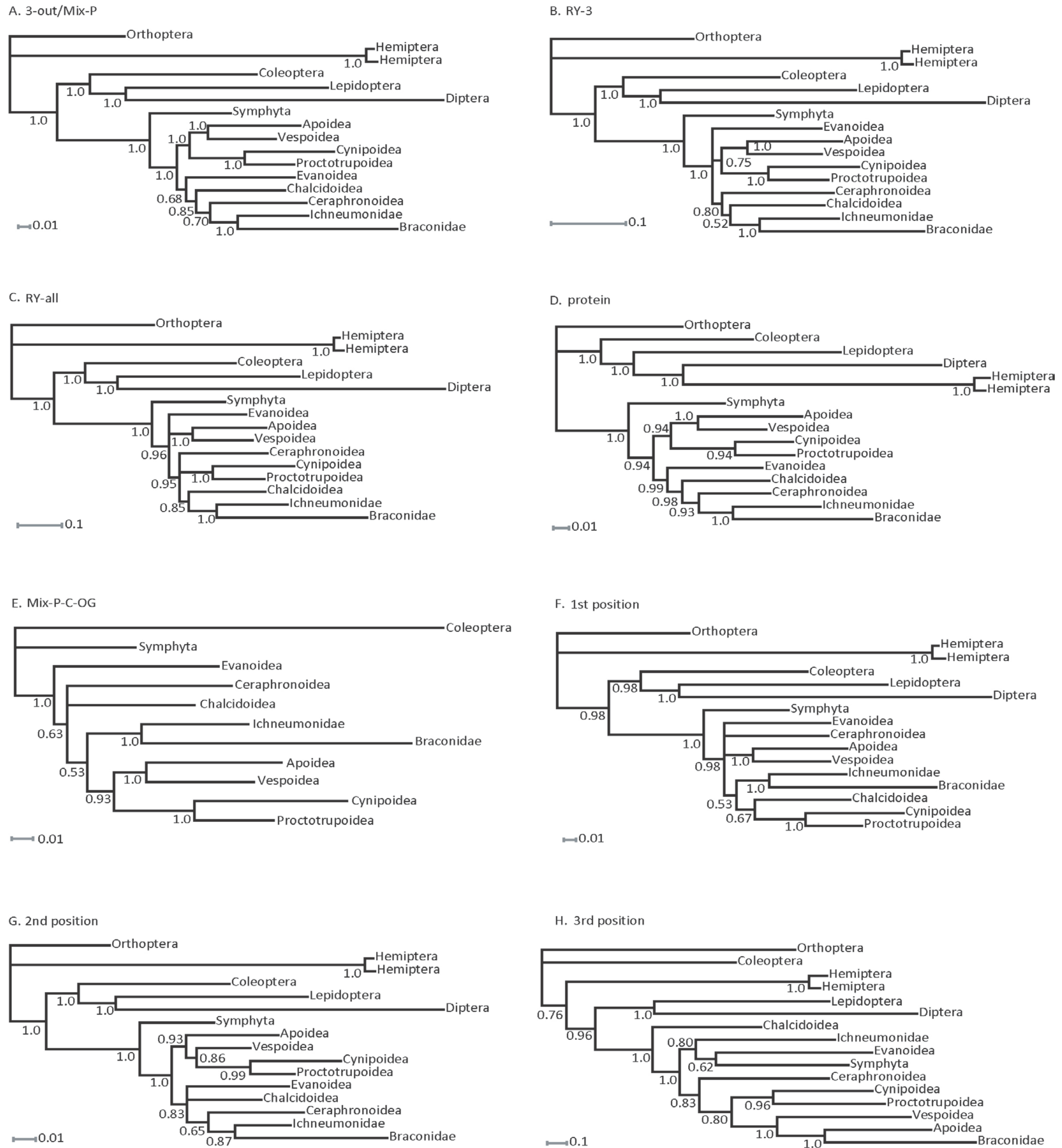


Figure 2.4. A-H. Bayesian phylograms inferred from various data partitioning schemes of the concatenated 24-gene dataset. See Results for additional details. Posterior probabilities are indicated near the relevant node. A. 3-out topology depicted (55 million generations, burnin = 28 million generations), and is the same topology inferred under the Mix-P dataset (40 million generations, burnin = 21 million generations). B. RY-3 (160 million generations, burnin = 120 million generations, runs may not have converged). C. RY-all (30 million generations, burnin = 10 million generations). D. Protein (2 million generations, burnin = 750 thous. gen.) E. Mix-P-C-OG (40 million generations, burnin = 24 million generations) F. 1st position (10 million generations, burnin = 4 million generations) G. 2nd position (10 million generations, burnin = 4 million generations). H. 3rd position (10 million generations, burnin = 4 million generations).

The topology based on the RY-all dataset did not recover the Aculeata as sister to Proctotrupoidea + Cynipoidea (Figure 2.4C). Rather, the aculeates were recovered in a basal polytomy with Evanioidea. Additionally, Proctotrupoidea + Cynipoidea were recovered in an unresolved clade with Ceraphronoidea and Chalcidoidea + Ichneumonoidea. RY-coding of the whole dataset could be advantageous in the unlikely situation where all transitional sites were saturated, thereby obscuring the historical signal. However, RY-coding an entire dataset also leads to increased homoplasy, as non-synonymous transitions in the first and second positions are converted to synapomorphies.

The topology inferred from the protein dataset was identical to the Mix-P and 3-out dataset for ingroup relationships (Figure 2.4D). However, the outgroup relationships were altered, as the Holometabola were recovered as paraphyletic. The two longest branches on the tree were grouped together, indicating that long-branch attraction may be responsible for the erroneous outgroup relationships. Given that the concatenated dataset failed the test for base composition heterogeneity in the first position, it is interesting that the protein dataset recovered the same ingroup relationships, suggesting that the systematic error in the first position played a small role in the outcome. Due to the strange outgroup relationships in the protein analysis and the long branch lengths among outgroup taxa in the Mix-P and RY-3 topologies, the Mix-P dataset was re-analyzed with all outgroups excluded except the coleopteran. Interestingly, the Ichneumonoidea were recovered as sister to the aculeate/ proctotrupoid + cynipoid clade, similar to the full data analysis. The topology was identical to the tree obtained with the same outgroup rooting, but with all data included (Figure 2.2C). This suggests that the variable positions of the Ceraphronoidea, Evanioidea, Chalcidoidea, and, Ichneumonoidea were sensitive to outgroup selection when the third position was excluded.

However, when all of the codon positions were analyzed separately, three very different topologies emerged. The topology inferred from the first codon position (Figure 2.4F) recovered the traditional Proctotrupomorpha (sensu Rasnitsyn (1988)), albeit with relatively weak support. Additionally, the Proctotrupomorpha were recovered as sister to the Ichneumonoidea, again with limited support. This Ichneumonoidea + Proctotrupomorpha clade was also recovered in a basal polytomy with the remaining apocritans, although the aculeates were recovered as monophyletic. All other relationships were as expected, including a monophyletic Holometabola, Hymenoptera, and Apocrita. Interestingly, this topology is most in line with

morphological data (Rasnitsyn, 1980b; Rasnitsyn, 1988) and evidence from other molecular analyses (Castro & Dowton, 2006). Furthermore, this very different topology suggests that the phylogenetic signal in the first position conflicted with the overall signal when both the first and second positions were included (3-out dataset, Figure 2.4A).

The inferred topology from the second codon position dataset was very similar to the topology inferred from the Mix-P, 3-out, and protein datasets (Figure 2.4G). Here, two separate lineages were recovered again, one containing Aculeata and Proctotrupoidea + Cynipoidea, the other containing all remaining apocritans in the dataset. The monophyletic Ichneumonoidea was again recovered as sister to Ceraphronoidea. Interestingly, the signal in the second codon position did not recover a monophyletic Aculeata. With the exception of the Proctotrupoidea + Cynipoidea clade, all apocritan clades were relatively weakly supported (posterior probabilities < 0.95).

Analysis of the third position dataset recovered an obviously erroneous topology (Figure 2.4H). Apocrita, Aculeata, and Ichneumonoidea are all recovered as paraphyletic. Interestingly, some of these paraphyletic clades were recovered with high support, demonstrating the power of saturation in the third position to mislead phylogenetic reconstruction. Examining the pattern of nucleotide substitution by codon position reveals a strong transitional bias in the third position, a moderate bias in the first position, and a weak transversion bias in the second position (Table A3-6, Appendix A). When the first and second positions were analyzed together, the transition bias remained fairly weak, suggesting that the second position compensated for the first position bias. These biases are further revealed across lineages using the disparity index (Kumar & Gadagkar, 2001b) to test for homogeneity in substitution patterns across different lineages. For each codon position, the null hypothesis of substitution pattern homogeneity was rejected ($p<0.05$) for 89% of the 45 pairwise comparisons for the ingroup for third position, 38% for the first position, and 11% for the second position (Tables A7-9, Appendix A). Given the transition-transversion biases and substitution pattern heterogeneity across much of the dataset, it is likely that the model of evolution used to infer the phylogenies was violated, particularly for the first and third codon positions. Clearly systematic bias has affected the phylogenetic inference, but to what extent cannot be readily determined.

2.4.3 Individual Gene Analyses

Individual gene trees displayed a serious lack of concordance with the concatenated analysis and with each other. Figure 2.5 depicts the phylogenies inferred for each gene. Although 3 of the gene trees were compatible with the ingroup relationships recovered in the concatenated analysis depicted in Figure 2.1 (CG6803, CG7178, and CG7939), two of these trees have very little resolution. Additionally, of the three compatible trees, only the topology inferred from gene CG6803 was entirely compatible with both ingroup and outgroup relationships (Figure 2.5).

Given that almost all of the genes violated the assumption of base composition homogeneity in the third position, each gene was reanalyzed with the third position removed if it failed the homogeneity test (Figure 2.6). Additionally, gene CG4169 was analyzed as a protein since the first codon position also failed the test. This mixed inclusion of sites across the different genes and the analysis of gene CG4169 as a protein was the exact mixture of partitions and genes run in the concatenated dataset called Mix-P, discussed earlier. Comparing the individual gene trees in Figure 2.6 to the topology inferred from the Mix-P dataset (Figure 2.4A), only 2 of the gene trees were compatible with the ingroup relationships. One of these gene trees (CG3661) recovered a polytomy that included the hymenopteran taxa and outgroups. The other gene tree (CG15442) was only resolved for Cynipoidea + Proctotrupoidea within the ingroup. Obviously, the base compositional biases in the third position were not enough to account for the vast majority of discordance among gene trees.

2.4.4. Visualizing Conflict and Compatibility With Filtered Supernetworks

Even though only 2 genes supported the exact relationship depicting aculeates as sister to Cynipoidea + Proctotrupoidea (Figure 2.1), the use of filtered supernetworks (Huson & Bryant, 2006) demonstrated that this relationship was compatible with several more of the individual gene trees. Filtered supernetworks have been successfully used to visualize the most common relationships given a set of taxonomically overlapping gene trees (Whitfield et al., 2008), an especially useful tool when there is a high degree of conflict among the input trees. Figure 2.7 A-D illustrates six filtered supernetworks, which include only those splits contained in (or compatible with) a set minimum number of gene trees.

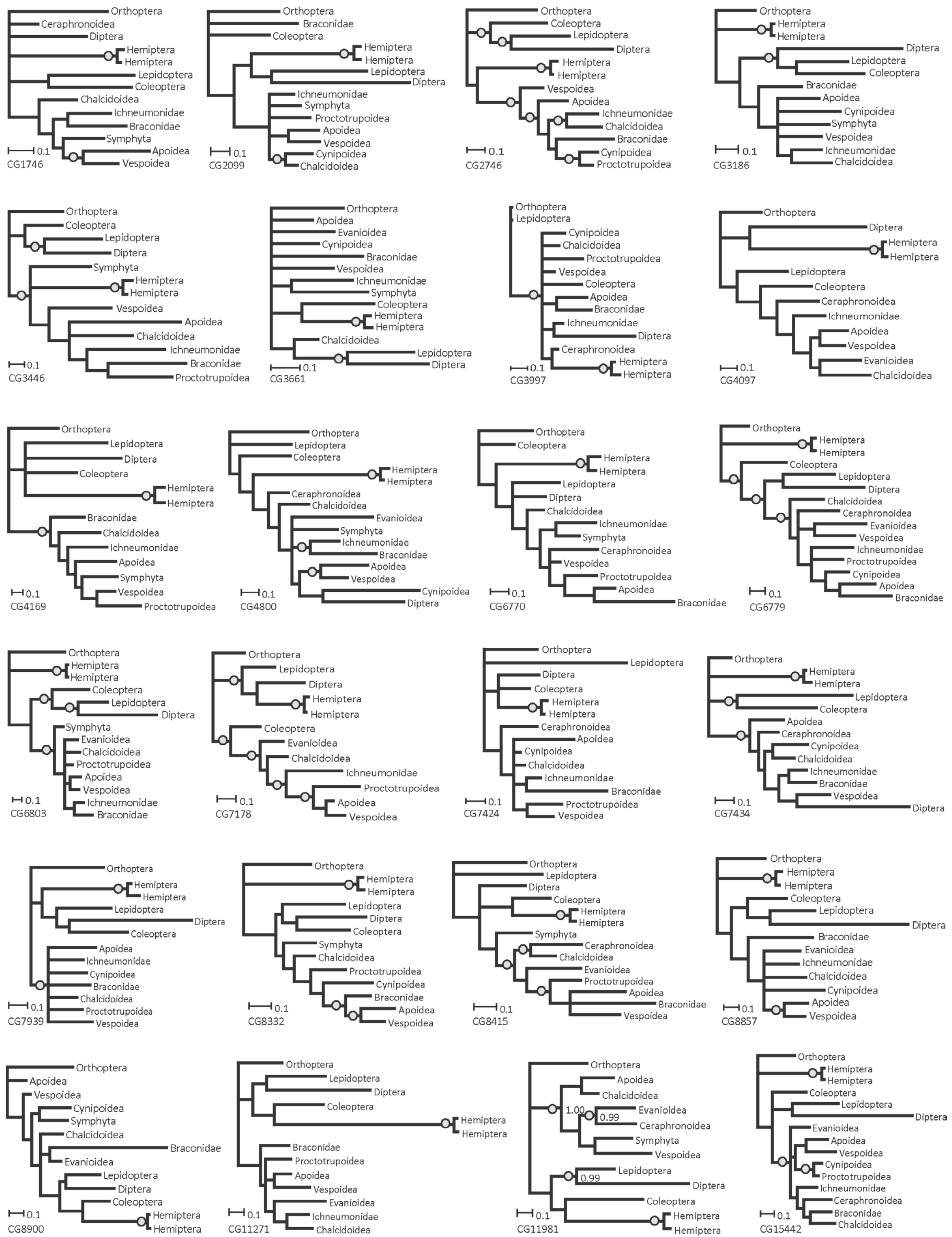


Figure 2.5. Bayesian phylograms inferred from individual gene datasets with all data included. For brevity, posterior probabilities ≥ 0.95 are indicated with a circle on the node.

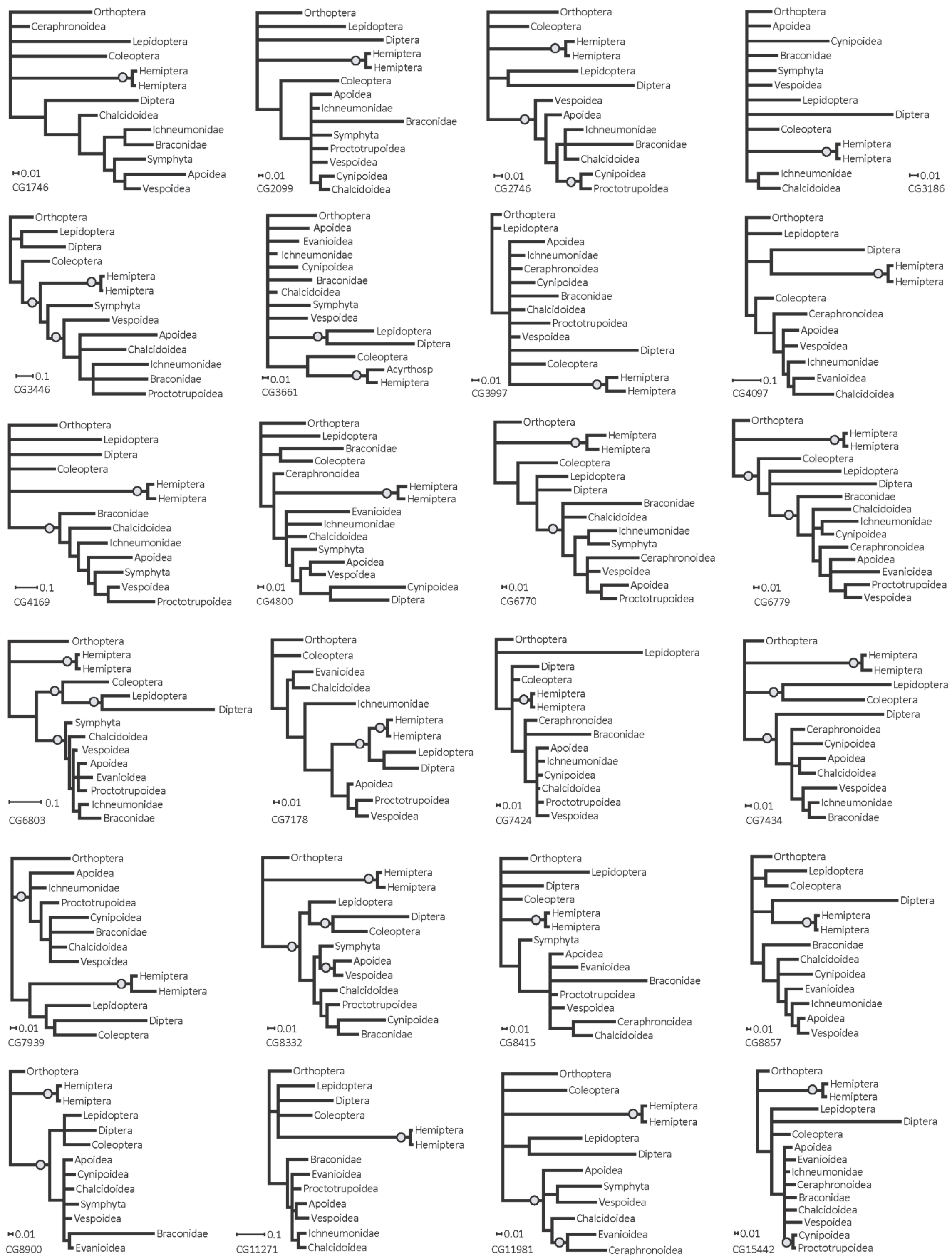


Figure 2.6. Bayesian phylograms inferred from individual gene datasets with the third position removed from genes demonstrating third position heterogeneity. Two gene trees include all codon positions (CG3997 and CG7178). One gene tree was analyzed with amino acid data (CG 4169). Posterior probabilities ≥ 0.95 are indicated with a circle. This data partitioning scheme is referred to as Mix-P (see Results for additional details).

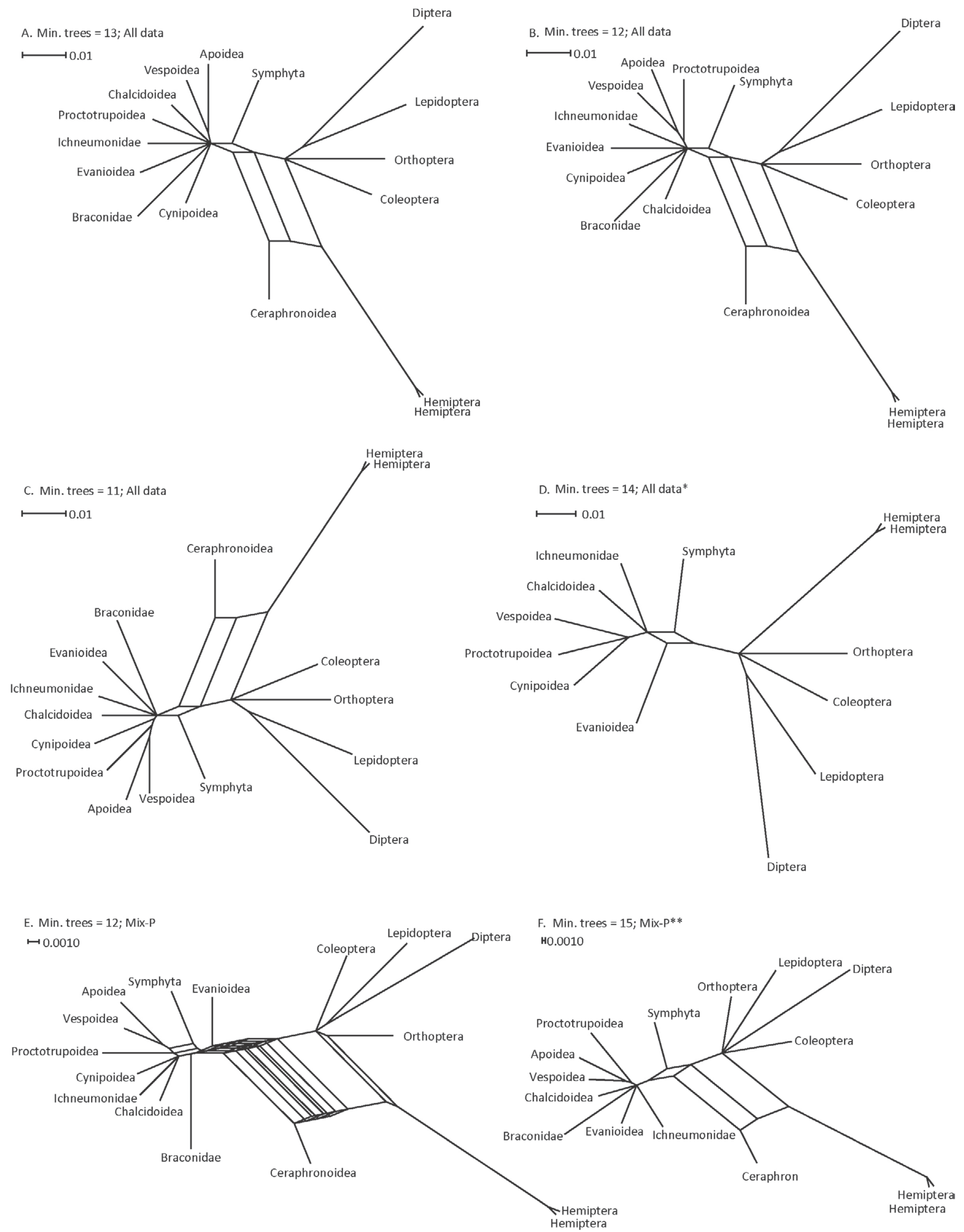


Figure 2.7. Filtered supernetworks of the 24-gene dataset. A-D. Calculated from 24 gene trees (cf. Figure 2.5). E-F. Calculated from all gene trees under the Mix-P partitioning scheme (cf. Figure 2.6). Lines that do not conform to a tree-like structure (i.e. parallel lines) represent incompatible splits. * Apoidea, Braconidae, and Ceraphronoidea excluded. ** Cynipoidea excluded.

As shown in Figure 2.7A, the split representing the aculeates was contained in more than 50% of the gene trees (mintrees=13). Most apocritan lineages were separated from Symphyta, but the position of Ceraphronoidea was reticulated with respect to outgroups and the ingroup. Clearly, the ceraphronoid demonstrated the most conflicting phylogenetic positions across the individual gene datasets with all data included, accounting for its volatile placement in the variety of analyses performed on the concatenated datasets discussed previously.

When the filter was set to 12 minimum trees, the proctotrupoid was recovered in a split with the aculeates (Figure 2.7B). When the filter was set to 11 minimum trees, the cynipoid was separated from the polytomy containing most apocritans and placed in a split with the aculeates + Proctotrupoidea. Thus, even though only 2 of the genes from the individual gene analyses recovered the aculeates with the proctotrupoid, the cynipoid, or both, there were several genes that were compatible with this relationship. Moreover, there were 5 genes that indicated a sister relationship between Apoidea and Braconidae (Figure 2.5) and 5 genes that recovered the accepted sister relationship between Ichneumonidae and Braconidae. When the third position was removed from genes with heterogeneous base composition, not one of the genes recovered the Braconidae + Apoidea relationship (Figure 2.6). Both of these taxa had similar A-T composition across a number of genes, and their recovery together in some individual gene trees was likely due to the convergent evolution of these nucleotides at the third position. The split containing Vespoidea, Proctotrupoidea, and Ceraphronoidea was recovered when the filter was set to 14 minimum trees (Figure 2.7D), provided Apoidea, Braconidae, and Ceraphronidae were excluded.

Although a 12-minimum tree filtered super-network with all data included revealed a close affinity between Proctotrupoidea and Aculeata (Figure 2.7B), this relationship was less clear when the third position was removed from genes with third position base composition heterogeneity (Figure 2.7E). Therefore, it may be third position synonymous changes were driving this relationship. However, the mixed model and protein datasets also recovered a sister relationship between aculeates and Cynipoidea + Proctotrupoidea, albeit with less support. One of the confounding issues was the volatile placement of Cynipoidea across the individual gene datasets when the third position was removed, which likely lowered the support for the Aculeata/Proctotrupoidea + Cynipoidea relationship. Indeed, when Cynipoidea was excluded,

the split containing the aculeates and Proctotrupoidea was recovered under a 15-minimum tree filtered super-network (Figure 2.7F). Clearly, volatile taxa were obscuring some of the historical signal, a problem likely over-exaggerated with low taxonomic sampling.

2.4.5. Visualizing Phylogenetic Signal Using Treeness Triangles

Treeness triangles allow for the visualization of competing signal in a given dataset by plotting the relative signal that corresponds to the internal (I) and external (E) branches of an estimated tree, and the residual (R) signal that does not correspond to any branches of a tree (White et al., 2007). These three data points correspond to the three apices (I, E, R) of the treeness triangle, and are determined by the set of splits calculated from the distance matrix in comparison to the estimated tree (see White et al., 2007 for a detailed explanation). Data points in the triangle that fall close to the internal-external (I-E) axis have the least conflicting or residual (R) signal. Points close to the external (E) apex demonstrate signal that corresponds mostly to terminal branches (at maximum a star tree) and therefore, are likely to be highly uninformative. Alternatively, data points closer to the internal apex (I) depict signal that corresponds to the internal branching order in the estimated tree. Thus, the most tree-like genes will have data points that lie close to the I-E axis and will be the most informative as the data point approaches the internal apex.

LogDet distances (Tamura & Kumar, 2002), which correct for multiple substitutions, were calculated for the first and second position for each individual gene dataset, and used to generate the data points on the treeness triangle (Figure 2.8). Three additional distance matrices were computed for the first and second positions separately for all genes concatenated together, and for the first and second positions of all genes concatenated together (Figure 2.8). Overall, 81 percent of the genes for both first and second position sites fell along the external-residual axis, demonstrating the high level of conflicting phylogenetic signal in the dataset. Almost all first position sites for all genes fell along the external-residual (E-R) axis, suggesting a large degree of conflicting and uninformative signal. Alternatively, 9 of the 24 (37.5%) genes for second position sites fell in the external-internal (E-I) axis, suggesting less conflicting signal and more characters supporting the estimated tree. Interestingly, all data points fell in the upper part of the triangle, suggesting more characters support terminal edges, likely a feature common to relatively conserved regulatory genes. When all of the first position sites were

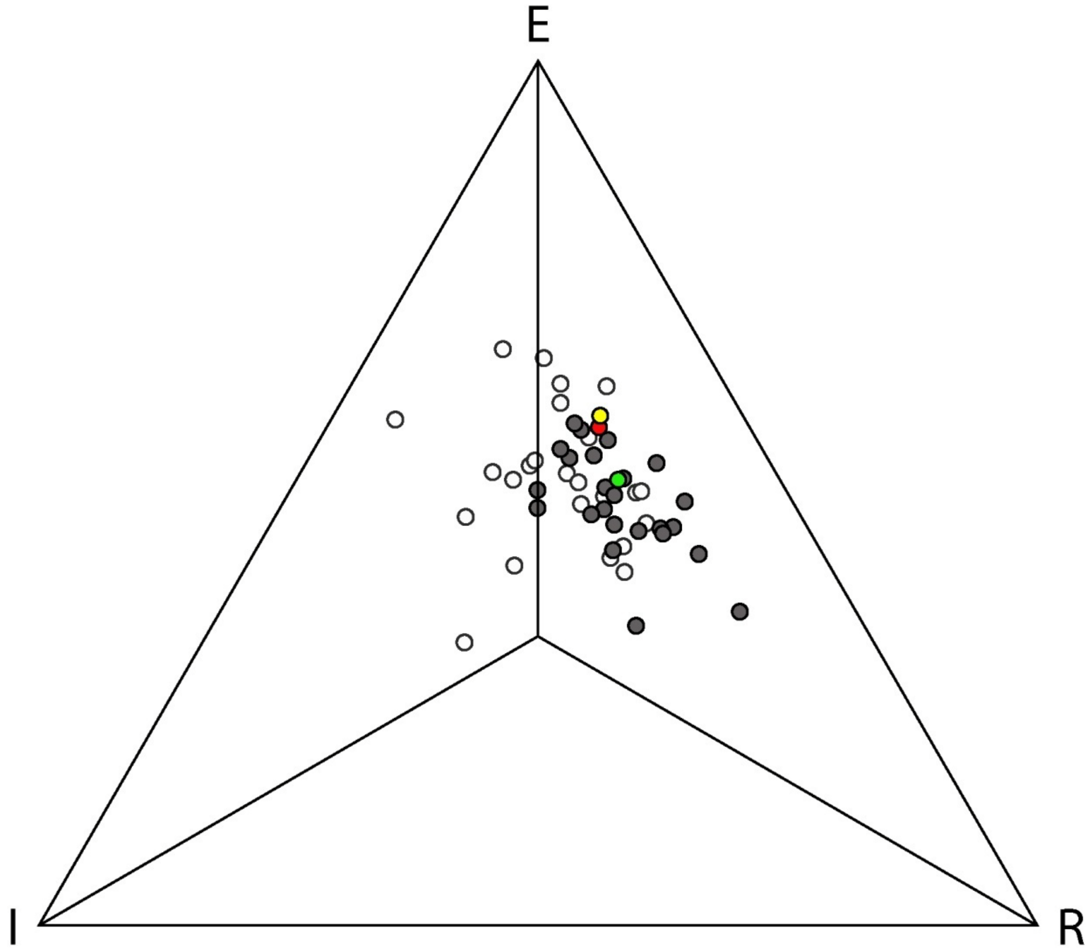


Figure 2. 8. Treeness triangle with data points calculated using the closest tree algorithm based on LogDet distance matrices of the individual genes and concatenated data for the 1st and 2nd codon positions. E = external edges, I = internal edges, R = residual signal. Data points near the External apex (E) represent unresolved trees (at maximum a start tree). Data points near the Internal (I) apex represent phylogenetic signal that corresponds to the internal branches of the tree. Data points near the Residual (R) apex represent phylogenetic signal that does not correspond to any branches on the tree calculated using the closets tree algorithm. Thus, genes with tree-like phylogenetic signal should have data points that fall along the I-E axis; although the closer the data point is the external (E) apex, the less informative the signal. The diagram is split into 3 sections for easier visualization of the data points relative to the axes of the triangle, but does not represent a 3-D image. Note that most data points fall within the E-R axis, suggesting a high level of conflict and a lack of informative signal. Grey fill = 1st position, individual genes; White fill = 2nd position, individual genes; Yellow fill = 1st position, concatenated dataset; Green fill = 2nd position, concatenated dataset; Red fill = 1st + 2nd positions, concatenated dataset.

concatenated together, the data point had a higher external signal and lower residual signal than expected from the distribution of the individual genes, indicating less conflict among the pooled information for first position sites. Alternatively, the data point for the concatenated second position sites had more residual signal than would be expected from the distribution of the individual genes, suggesting that the tree-like signal in the individual genes are conflicting with each other.

2.5. Discussion

Although the 24-gene, 17-taxon dataset likely suffered from both systematic and sampling error, the phylogenetic potential of these loci was revealed by the consistent recovery of well corroborated evolutionary relationships. Under Bayesian inference, all analyses of the concatenated nucleotides recovered a monophyletic Holometabola (with a sister relationship between the Panorpoid orders), Hymenoptera, Apocrita, Aculeata, Ichneumonoidea, and a sister relationship between the two most closely related putative proctotrupomorphs (Cynipoidea + Proctotrupoidea). Clearly, the method of inference and the model of evolution employed were appropriate enough to recover these relationships, regardless of missing data, heterogeneity in the pattern and rate of substitution, conflicting signal across data partitions and gene trees, and low taxonomic sampling. Thus, ESTs have great potential for resolving higher level Hymenopteran relationships, which will likely become even more apparent with greater taxonomic and genetic sampling. Even with only 24 genes, ESTs also have significant resolving power at the ordinal level within Holometabola.

It is difficult to speculate to what extent sampling and systematic error had on the unknown relationships within Hymenoptera. However, it cannot be ruled out, particularly since several relationships of interest demonstrated inconsistency across data partitions. For example, the phylogenetic position of Chalcidoidea, Ceraphronoidea, Evanioidea, and Ichneumonoidea varied across several different analyses. Since there was only one exemplar for each major lineage, it is possible that sampling error played a role in the lack of congruence among individual gene datasets. For a given gene, the pattern and rate of substitution of a given taxon may not have been characteristic for the group it represents. Thus, some gene trees may have recovered historical relationships, some may have had insufficient signal, and others may have recovered false relationships due to long-branch attraction artifacts and biases in the pattern of substitution. Increased taxonomic sampling has been the most common and

effective remedy for both phylogenetic conflict (Dunn et al., 2008; Hedtke et al., 2006)and long-branch attraction (Bergsten, 2005), and is the obvious next step for future empirical studies using ESTs for Hymenopteran relationships.

Missing data may have been another source of sampling error. Although other researchers have noted that missing data can have small effects on phylogenetic inference in large-scale phylogenomics (Philippe, 2004), these datasets have utilized more than 100 genes. Thus, if 50% data is missing for a given taxon, there are still 50 genes available to provide enough signals to potentially swamp out systematic biases. However, in this dataset, 5 taxa were missing over 49% of the characters (Table 2.2). In a smaller 24-gene dataset, high levels of missing data means there are limited number of genes available for inference. If even a handful of these genes contain signal that is not compatible with the true species tree, then the systematic biases inherent in some genes may not be overwhelmed by true historical signal. This is potentially the case in the dataset analyzed here as Ceraphronoidea and Evanioidea had high levels of missing data (66.2 and 51.7 percent, respectively) and seemed to be sensitive to the exclusion of various sites in concatenated datasets. However, this was not the case for Symphyta, Proctotrupoidea, and Cynipoidea (with 51.0, 49.2, and 52.1 percent missing data, respectively), whose phylogenetic positions were not altered in the concatenated analyses, regardless of the inclusion of characters. Chalcidoidea had only 1.1 percent missing data, although its volatile placement may have been affected by missing data in other taxa. However, the placement of Chalcidoidea was extremely volatile across the individual gene trees, where missing data was not an issue.

Clearly this dataset had widespread topological conflict among gene trees. Incongruity between individual gene trees and the species phylogeny is not a new or uncommon phenomenon (Philippe et al., 2005; Rokas et al., 2003). Inferring the right evolutionary tree can even be difficult using simulated data where the assumptions of the model are met and the true species phylogeny is known (Nei, 1996; Penny et al., 2001; White et al., 2007). The higher nodal support in several of the concatenated analyses might have been an artifact of systematic error. It is possible that multiple gene trees converged on an erroneous estimate and nodal support increased as the number of genes increased. As Degnan and Rosenberg (2006) point out, increased genetic sampling will not necessarily lead to an improved estimate of the species

phylogeny. However, this type of systematic error can only be tested with increased taxonomic sampling.

Base composition heterogeneity was another potential source of systematic error. Even though most of the individual gene datasets demonstrated homogeneity in base composition when the third position was excluded, the concatenated dataset violated the null hypothesis of homogeneity when the third position was excluded. This was likely a result of systematic error. Interestingly, when the two hemipterans and the dipteran were excluded from the test, both the first position and first and second positions together demonstrated homogeneity (Table A2, Appendix A). Thus, the majority of systematic error was accumulated from the base composition biases of these outgroup taxa. When the analysis was re-analyzed with these taxa removed (Figure 2.2B –C and 4E), the recovered relationships were highly congruent with the relationships recovered when all data and all taxa were included (Figure 2.1). Only the position of the ceraphronoid was slightly altered, demonstrating its sensitivity to the inclusion of various outgroup taxa. Thus, the dataset seemed to suffer more from compositional biases among outgroup taxa than ingroup taxa, particularly when the third position was removed.

Examinations of the phylogenetic signal across the different codon positions revealed the high degree of conflict present in the dataset. The three codon positions analyzed separately exposed three very different competing signals in the dataset (Figure 2.4F-H). Additionally, the tests for substitution pattern homogeneity indicated relatively strong transition biases in the first and third positions (Table A3 and A5, Appendix A). However, the bias in the first position was relatively weakened with the addition of second position sites (Table A6, Appendix A). Visualization of the conflict using the treeness triangle demonstrated just how much residual signal exists, not only in individual gene datasets, but also the concatenated datasets. Even if the true species tree was recovered in one of the concatenated analyses, surely the conflicting signal severely decreased the nodal support for several clades, thereby decreasing confidence in the inference. In the future, increased genetic and taxonomic sampling should lead to greater convergence among gene trees and therefore, less relative conflict across genes and data partitions.

2.5.1 Phylogenetic Implications

Previously the Ichneumonoidea has been proposed as the sister group to the Aculeata in both morphological (Brothers, 1975; Oeser, 1961; Rasnitsyn, 1988) and molecular studies (Dowton & Austin, 1994), albeit with limited evidence. In this study, the Ichneumonoidea were never recovered as sister to the aculeates in any of the concatenated analyses. While 5 of the individual gene trees (Figure 2.5) suggest some relationship between Braconidae and Aculeata, these relationships disappear in 4 of the 5 gene trees when the third position is removed (Figure 2.6). Thus, it is likely that the convergent A-T bias present in the third position of both Braconidae and Apoidea generated false homologies and an erroneous relationship when all data were included.

More recently, the Evaniomorpha have been suggested as the sister group to Aculeata, particularly Stephanidae, Trigonalyidae, and Megalyridae (Castro & Dowton, 2006; Dowton & Austin, 2001). Although these taxa were not included in this study, Aculeata was not recovered with Evanioidea (Aulacidae) or Ceraphronoidea in any of the concatenated analyses or individual gene trees. Since the two analyses that recovered the Aculeata + Evaniomorpha relationship were based largely on mitochondrial genes, it is possible that compositional biases may account for their results (see Castro *et al.*, 2002; Dowton & Austin, 1997).

In this study, the Cynipoidea + Proctotrupoidea were recovered as sister to the aculeates in all concatenated analyses, except the analysis of the first codon position alone (Figure 2.4F). This relationship was stable with the exclusion and inclusion of various outgroups, with all data included, and with corrections for base composition heterogeneity. Five gene trees suggested this relationship with all data included (Figure 2.5) and 3 gene trees when the third position was removed (Figure 2.6). In addition, there were several more gene trees compatible with this relationship (Figures 5-7).

Rasnitsyn (1980b) originally placed the Proctotrupomorpha as sister to the Ichneumonoidea, with these two groups sharing a close affiliation to Aculeata, based on the presence of articulating propodeal condyles (not present in Chalcidoidea). These three lineages were recovered together in several individual gene trees (Figures 5-6) and some concatenated analyses (Figures 1, 2B-C, and 4E). Rasnitsyn (1988) later revised his hypothesis, suggesting that the similarity between Proctotrupomorpha and Ichneumonoidea was a result of parallel evolution. Given that there at least some evidence in this study to suggest these three lineages

shared a common ancestor, these characters might need to be re-examined. The Ichneumonoidea were not recovered as sister to the Aculeata (Proctotrupoidea + Cynipoidea) clade in most of the analyses that corrected for base composition heterogeneity (Figure 2.4A-D). However, this clade was recovered when the analysis was rooted on the coleopteran and all other outgroups were excluded (Figure 2.4E). Thus, this relationship was sensitive to the inclusion or exclusion of characters and outgroups. It is possible that the differential rate of evolution in several outgroups (most notably in the dipteran and two hemipterans), as evidenced by the extreme branch lengths across concatenated analyses and gene trees (Figures 4A-C, 5-6), may have affected the ingroup topology. However, it is still unclear which lineage is sister to the Ichneumonoidea.

The placement of Evanioidea and Ceraphronoidea cannot be determined on the available evidence. However, across a vast majority of the individual gene trees, both Ceraphronoidea and Evanioidea demonstrated an earlier divergence from the remaining apocritans (Figures 5-6). The high levels of missing data in both of these taxa likely contributed to their highly volatile placements across the different analyses. Thus, resolving their true phylogenetic position will require further genetic and taxonomic sampling.

There is strong evidence suggesting that Chalcidoidea does not belong within the Proctotrupomorpha. This has a number of implications, including that Diaprioidea (sensu Sharkey, 2007) may also not belong in the Proctotrupomorpha, as it has been found in other molecular analyses to be the sister group to Chalcidoidea with strong support. Although missing data may have played a role in the variable placement of the ceraphronoid and evanioid, the chalcidoid was represented in every gene dataset. Interestingly, the chalcidoid demonstrated extreme volatility in individual gene trees, with almost every sister combination with other ingroup taxa recovered. The chalcidoid did not display any obvious compositional biases or extreme branch lengths that may account for its variable placement. If the Chalcidoidea are truly closely related to the other proctotrupomorphs, it is possible that non-phylogenetic evolutionary events, such as lateral gene transfer, may be affecting the placement of the chalcidoid. However, further testing is required before such a hypothesis can be invoked. Recently, Bezier et al. (2009) suggested that braconids have experienced lateral gene transfer in their association with symbiotic viruses used to manipulate host immune systems. To what

extent this prevails across the parasitic lineages within Hymenoptera is unknown, but may need to be considered for future empirical studies of Hymenopteran phylogenetics.

Based on the outcome of all of the analyses performed herein, the phylogeny depicted in Figure 2.4E represents the most supported relationships. The development of greater flexibility of the metasoma through a modified propodeal attachment with articulating condyles, as originally suggested by Rasnitsyn (1980b), may be a potential synapomorphy for this Ichneumonoidea as sister to Aculeata (Proctotrupoidea + Cynipoidea). However, the tree depicted in Figure 2.2A is a more conservative estimate, given the widespread conflict in the dataset and low taxonomic sampling.

2.6. Conclusions

From this study, it is evident that ESTs have huge potential to resolve higher level hymenopteran relationships. Even though holometabolan relationships were not the focus of this study, given the accurate resolution across the included orders, it is also clear that ESTs will be very useful for resolving long contested ordinal relationships. ESTs allow for greater taxonomic sampling beyond model organisms from genome projects. Indeed, the EST database is growing every day with transcripts from a variety of insects, and will allow for further testing of long debated phylogenetic relationships that have important evolutionary implications. Additionally, next-generation sequencing technologies that have recently become available will make large scale phylogenomic projects with high levels of taxonomic sampling a practical pursuit (e.g. Dunn *et al.*, 2008). The dataset analyzed here suffered from gene tree incongruity and large compositional and substitution pattern biases, which perhaps were too severe to overcome the low taxonomic sampling. However, recovery of all of the expected relationships across most of the analyses, not only points to the potential of these loci for phylogenetics, it also suggests that some of the recovered relationships, perhaps controversial, are indeed correct. The weight of the evidence here points to a sister relationship between Aculeata and Proctotrupoidea + Cynipoidea, contrary to previously proposed hypotheses. Additionally, there is evidence for the antiquity for both the evanioid and ceraphronoid lineages. Most of the evidence suggests that Chalcidoidea may not be contained within Proctotrupomorpha.

Copyright © Barbara J. Sharanowski 2009

CHAPTER 3: MOLECULAR SYSTEMATICS OF THE BRACONIDAE (HYMENOPTERA: ICHNEUMONOIDEA)

3.1. Introduction

Reconstructing the phylogenetic history of Braconidae has long been of interest to biologists in many fields. Aside from their valuable use in biological control, Braconidae provide an excellent system for studies on biodiversity and conservation, as well as evolutionary studies on the development of parasitism, host-parasite co-phylogenesis, morphological convergence, and the pattern of development of polydna viruses. Braconidae are one of the most diverse families of Hymenoptera with over 15,000 described species (Yu et al., 2001), and an estimated 25,000 species yet to be described (Marsh & Carlson, 1979). To date, however, the phylogeny of Braconidae is controversial and remains unresolved, especially at higher taxonomic levels.

Previous studies on braconid phylogenetics have been hampered by low taxonomic sampling, insufficient molecular sampling, or inconsistent morphological character coding, often based on preconceived notions of subfamilial membership (Wharton et al., 1992). Useful morphological characters are particularly difficult to ascertain in parasitoids, due to the high level of convergence among phenotypes adapted for a particular host group (Shaw & Huddleston, 1991). One of the greatest challenges to braconid systematics has been the inability to resolve higher level relationships among subfamilies, thus hampering ancestral state reconstructions and the testing of evolutionary theories (Zaldivar-Riverón et al., 2006).

Regardless of this lack of resolution, almost all researchers have assumed that ectoparasitism was the ancestral condition of all braconids. This belief is partly maintained by the lack of a robust phylogeny for Hymenoptera, with no solid conclusion on the sister group to Ichneumonoidea (Sharkey, 2007). Without an understanding of the placement of the Ichneumonoidea within Hymenoptera, it is difficult to speculate what life history strategy the ancestral ichneumonoid might have utilized. In an insightful review of the evolution and classification of Braconidae, Tobias (1967) stated that “transition from ectoparasitism to endoparasitism is one of the principle trends of biological evolution within the family (p. 387).” While this hypothesis, through an original transition from idiobiosis to koinobiois, has been well-

documented in the cyclostome subfamily Rogadinae (Shaw, 1983), it has not been conclusive for a majority of Braconidae (i.e. non-cyclostomes) (Wharton, 1993).

Understanding evolutionary transitions within Braconidae requires a robust phylogeny. Thus, the purpose of this paper is threefold: (i) to present a robust phylogeny of Braconidae based on multiple molecular markers, some new to braconid research, (ii) to examine higher level relationships, particularly among the helconoid complex of subfamilies, (iii) to test the ancestral mode of parasitism (endo- or ectoparasitism) of the Braconidae using ancestral state reconstruction.

3.2. Taxonomic Background

Members of Braconidae have typically been separated into two informal groups based primarily on mouthpart morphology and biology: the non-cyclostomes, with a flat or convex clypeus and flat labrum, and the cyclostomes, with a depressed clypeus and concave labrum (Quicke & van Achterberg, 1990b; Wharton, 1993). All members of the non-cyclostome lineage are koinobiont endoparasitoids, whereas members of the cyclostome lineage are primarily idiobiont ectoparasitoids, but demonstrate a wider range of biologies (Quicke & van Achterberg, 1990b; Tobias, 1967; van Achterberg, 1984). While most previous analyses suggest that these two groups form natural lineages, the membership of each group and relationships between the two groups have fluctuated across different analyses based on both morphological (Quicke & van Achterberg, 1990b; van Achterberg, 1984) and molecular datasets (Belshaw *et al.*, 1998; Belshaw & Quicke, 2002; Dowton *et al.*, 1998), and in combined analyses (Dowton et al., 2002; Pitz et al., 2007; Shi et al., 2005; Zaldivar-Riverón et al., 2006).

There have been three competing hypotheses on the evolution of Braconidae (Figure 3.1). First, the endoparasitic non-cyclostomes have been proposed as a derived lineage from cyclostome ancestors (Čapek, 1970; Quicke & van Achterberg, 1990b) (Figure 3.1A). Second, the non-cyclostomes have been proposed to form an independent lineage, sister to the cyclostomes, both having evolved from an unknown ancestor presumably ectoparasitic on concealed xylophagous coleopteran larvae (Belshaw et al., 1998; Gauld, 1988; Pitz et al., 2007; Shi et al., 2005; Tobias, 1967; van Achterberg, 1984; Wharton et al., 1992) (Figure 3.1B). Third, the non-cyclostomes have been proposed as a basal grade leading to the cyclostomes, suggesting an endoparasitic ancestor for Braconidae (Dowton et al., 1998) (Figure 3.1C). Most recent phylogenetic analyses suggest that the cyclostome lineage is sister to the endoparasitic

non-cyclostomes (Figure 3.1B). To date, however, the evidence for this hypothesis, or any of the competing ideas, has not been conclusive.

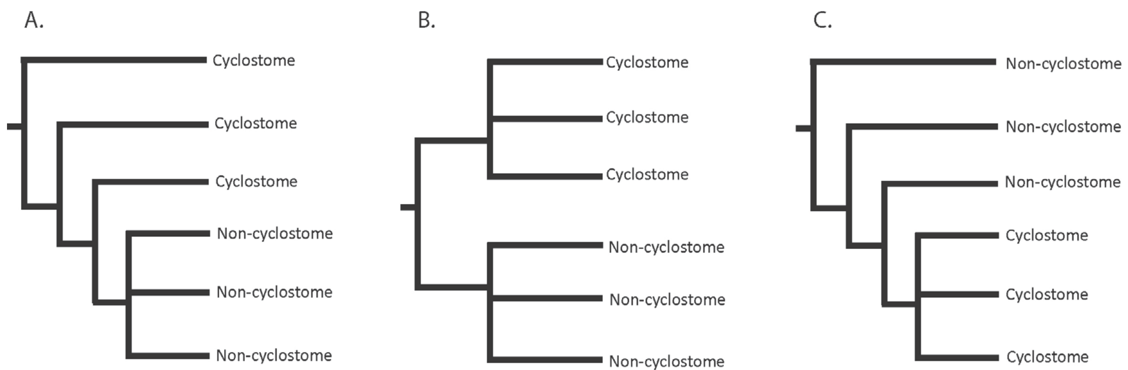


Figure 3.1. Alternative hypotheses on the relationships among the major lineages within Braconidae. A. Demonstrating a derived non-cyclostome lineage. B. Demonstrating two independent lineages: cyclostomes and non-cyclostomes. C. Demonstrating a derived cyclostome lineage.

In an effort to delineate relationships, the non-cyclostome lineage has previously been divided into two main complexes: the helconoid and microgastroid complexes (Wharton, 1993). Of these two lineages, only the microgastroid complex has been well supported in most molecular analyses (Banks & Whitfield, 2006; Belshaw *et al.*, 1998; Dowton & Austin, 1998; Dowton *et al.*, 1998; Murphy *et al.*, 2008; Whitfield, 1997), although the branching order among the representative subfamilies has fluctuated. Based on these analyses, the following subfamilies are recognized as part of the microgastroid complex: Microgastrinae, Cardiochilinae, Cheloninae, Adeliinae, Khoikhoiinae, Mendesellinae and Miracinae (Murphy et al., 2008). Additionally, the Ichneutinae have been suggested as the sister group to the microgastroid complex (Belshaw & Quicke, 2002; Dowton *et al.*, 2002; Quicke & van Achterberg, 1990b).

Wharton (1993) placed 14 subfamilies within the poorly understood helconoid complex, including: Amicrocentrinae, Agathidinae, Blacinae, Cenocoeliinae, Euphorinae, Helconinae, Homolobinae, Macrocentrinae, Meteorideinae, Meteorinae, Orgilinae, Sigalphinae, Trachypetinae, and Xiphozelinae. More recently, Belshaw and Quicke (2002) suggested that Euphorinae, Meteorinae, and the enigmatic Neoneurinae could be separated out into another lineage, referred to as the euphoroid complex. Additionally, they suggested that Cenocoeliinae may be affiliated with the euphoroid complex, as opposed to being closely related to Helconinae, as has been suggested in the past (Muesebeck & Walkley, 1951; Tobias, 1967). The helconoid complex has probably been the least understood lineage, partially due to the retention of primitive characters among several subfamilies contained within this putative complex (Tobias, 1967; van Achterberg, 1984). This ancestral morphology is most prominent among members of the Helconinae, thus, several authors have suggested that Helconinae is potentially one of the most ancestral lineages among the non-cyclostomes (Shaw & Huddleston, 1991; Tobias, 1967; van Achterberg, 1984). However, the branching order among the non-cyclostome subfamilies has been extremely unstable across different analyses, leaving no currently accepted picture of evolution.

The placement of Aphidiinae has also varied immensely across different analyses, being variably placed as sister to the cyclostomes (Dowton et al., 2002; van Achterberg, 1984), within the cyclostomes (Belshaw & Quicke, 2002; Dowton *et al.*, 1998), or within the non-cyclostomes (Čapek, 1970; Pitz et al., 2007; Shi et al., 2005). Recently, Zaldivar-Riverón et al. (2006) recovered an Aphidiinae + Mesostoinae clade as sister to the remaining cyclostomes with relatively high support, consistent with some previous analyses (Belshaw et al., 2000; Dowton et al., 2002).

Relationships among the cyclostome subfamilies have fluctuated across analyses, and the monophyly of several large subfamilial assemblages remains in doubt, including Doryctinae, Rogadinae, and Hormiinae. Not considering Aphidiinae + Mesostoinae, several analyses have recovered Rhyssalinae (including Histeromerinae) as sister to the remaining cyclostomes (Belshaw *et al.*, 1998; Dowton *et al.*, 2002; Pitz *et al.*, 2007; Quicke & van Achterberg, 1990b; Zaldivar-Riverón *et al.*, 2006). Doryctinae and Rogadinae have often been recovered as paraphyletic in molecular analyses (Dowton et al., 1998; Pitz et al., 2007; Zaldivar-Riverón et al., 2008), or if monophyletic with very little nodal support (Belshaw et al., 1998). Interestingly, phylogenetic inferences that have included morphological data have invariably recovered

Doryctinae as monophyletic (Dowton *et al.*, 2002; Quicke & van Achterberg, 1990b; Zaldivar-Riverón *et al.*, 2006), but not necessarily Rogadinae. Several analyses have also recovered a relatively well supported clade consisting of Braconinae, Gnamptodontinae, Exothecinae, Opiinae, and Alysiinae, which has been further confirmed with increased taxonomic sampling (Zaldivar-Riverón et al., 2006).

Figure 3.2 depicts a summary tree of what is currently known about relationships among subfamilies of Braconidae. The tree is based on previous molecular and morphological analyses (Belshaw *et al.*, 2000; Belshaw *et al.*, 1998; Belshaw *et al.*, 2003; Belshaw & Quicke, 2002; Dowton *et al.*, 1998; Dowton *et al.*, 2002; Pitz *et al.*, 2007; Quicke & van Achterberg, 1990b; Quicke & Belshaw, 1999; Wharton *et al.*, 1992; Zaldivar-Riverón *et al.*, 2006). Only relationships that have been well supported across multiple analyses are retained, whereas relationships that are in conflict across analyses are collapsed to polytomies. Subfamilies that have been variable recovered as polyphyletic are depicted with dashed lines, subfamilies that have never been analyzed in molecular datasets are depicted with vertical lines, and subfamilies that lack representations in this dataset are colored grey (see Figure 3.2 legend).

Thus, the evolutionary relationships among Braconidae have been highly unstable, regardless of whether morphological or molecular characters were utilized. One problem that plagues braconid scholarship is the continual use of morphological matrices coded at the level of subfamily, whereby the author's assumptions of subfamilial composition greatly influences the phylogenetic analysis if the subfamily is not a monophyletic assemblage (Wharton et al., 1992). Sampling error is another, and somewhat unavoidable, problem. Given the vast number of species within Braconidae it is difficult to have comprehensive taxonomic and character sampling. However, as information and evidence continues to accumulate, our understanding of evolution within Braconidae should become clearer. This study presents an examination of the relationships among braconid subfamilies, employing the largest taxonomic and genetic sampling of the family to date. Nearly 140 taxa and over 4kb of molecular data per taxon were used to infer braconid relationships. Additionally, nuclear protein-coding genes were utilized for the first time in braconid systematics, allowing for independent testing and corroboration of hypotheses on braconid evolution.

Figure 3.2. Tree summarizing current knowledge of braconid subfamilial relationships based on previous molecular and morphological analyses. Dashed lines indicate likely paraphyly. Vertical lines indicate subfamilies that have not previously been analyzed in molecular datasets for subfamilial relationships across Braconidae. Subfamily names colored grey are not represented in the current dataset. The placement of subfamilies with a question mark after the name are based on limited morphological or molecular data, and thus their placement represent s the current opinion in the field, rather than the results of phylogenetic testing.

3.3. Materials and Methods

3.3.1 Taxon sampling

Exemplars were obtained for 134 species of Braconidae and 5 species of Ichneumonidae that were employed as outgroups (Table 3.1). The number of recognized subfamilies within Braconidae is constantly in flux and differs depending on author. The subfamily classification employed here is an attempt to utilize the most current phylogenetic information available. Thus, the placement of the exemplars within subfamilies follows that of Belshaw et al. (1998) with modifications to the cyclostome subfamilies, based on the results of Zaldivar-Riverón et al. (2006) (Table 3.1). One exception is the placement of *Conobregma*, which is placed under its original designation within Betylobraconinae (van Achterberg, 1995), due to a lack of evidence suggesting an alternate placement. Employing this classification, and recognizing Maxfischeriinae, there are a total of 49 braconid subfamilies.

Thus, 40 subfamilies were represented in the dataset with at least one exemplar (Table 3.1). Subfamilies without representation include Telengainae, Lysiterminae, Ypistocerinae, Apozyginae, Vaepellinae, Dirrhopinae, Masoninae, Trachypetinae, and Adeliinae. The data matrix also includes newly sequenced subfamilies within a larger braconid phylogeny, including Amicrocentrinae, Ecnomiinae, Khoikhoiinae, Maxfischeriinae, Mendesellinae, Meteorideinae, and Microtypinae. Table 3.1 lists subfamilies under the putative complexes discussed previously.

There is an emphasis on Helconoid subfamilies, particularly Helconinae, due to the very contradictory and ambiguous placement in previous analyses (Belshaw *et al.*, 1998; Belshaw & Quicke, 2002; Dowton *et al.*, 2002; Pitz *et al.*, 2007). Additionally, all previous phylogenetic studies recovered a polyphyletic Helconinae with varying placement of its members at the base of the braconid phylogeny (Belshaw *et al.*, 1998; Belshaw & Quicke, 2002; Dowton *et al.*, 1998; Dowton *et al.*, 2002; Pitz *et al.*, 2007; Quicke & van Achterberg, 1990b; Shi *et al.*, 2005). Thus, Helconinae is a potentially important basal lineage and was heavily sampled in this dataset. Five outgroup taxa were selected from Ichneumonidae, well established to be the sister-group to Braconidae (Belshaw *et al.*, 1998; Dowton *et al.*, 2002; Sharkey & Wahl, 1992).

Table 3.1. Exemplars utilized in this study, including location of collection and the genes that were amplified for each taxon. Exemplars are divided by the putative containing lineages. An X indicates the gene region was amplified, whereas a dash indicates the gene was not amplified.

Exemplar	Internal Voucher Number	28S rDNA		18S	CAD CPSase*		ACC**	Country collected from
		D1D3	D3D5		54/405	apmod		
ICHNEUMONIDAE - Outgroups								
Odontocolon albotibiale (Bradley) (XORIDINAE)	DM054	x	x	x	x	x	x	USA
Baryceros texanus (Ashmead) (CRYPTINAE)	DM057	x	x	x	—	—	x	USA
Zagryphus nasutus (Cresson) (TRYPHONINAE)	DM059	x	x	x	—	x	—	USA
Pimpla sp. (PIMPLINAE)	DM094	x	x	x	—	x	x	USA
Dusona sp. Cameron (CAMPOPLEGINAE)	DM095	x	x	x	x	x	x	USA
BRACONIDAE - Putative Helconoid Complex								
HELCONINAE - Helconini								
Wroughtonia sp.1	BJS001	x	x	x	x	x	x	USA
Wroughtonia ferruginea (Brues)	BJS013	x	x	x	x	x	—	USA
Wroughtonia ligator (Say)	BJS017	x	x	x	x	x	x	USA
Wroughtonia sp.4	BJS022	x	x	x	x	x	x	USA
Eumacrocentrus americanus	BJS012	x	x	x	x	x	x	USA
Helcon texanus	BJS015	x	x	x	—	—	—	USA
Helcon tardator	BJS095	x	x	x	x	x	x	FRANCE
Helcon sp.3	BJS108	x	x	x	x	x	—	CHILE
Helcon sp.4	BJS110	x	x	x	x	x	x	CHILE
Helcon sp.5	BJS043	x	x	x	x	x	x	AUSTRALIA
Helcon sp.6	BJS045	x	x	x	x	x	—	AUSTRALIA
Helcon sp.7	BJS102	x	x	x	x	x	x	AUSTRALIA
Austrohelcon inornatus	BJS103	x	x	x	x	x	—	AUSTRALIA
Topaldios sp.	BJS040	x	x	x	x	—	x	CHILE
Helconini Gen. sp.1 unident.	BJS098	x	x	x	x	x	x	AUSTRALIA
Calohelcon sp.	BJS093	x	x	x	x	x	x	AUSTRALIA
Ussurohelcon nigricornis	BJS044	x	x	x	x	—	x	THAILAND
HELCONINAE - Diospilini								
Diospilini Gen. sp. unident.	BJS099	x	x	x	—	x	x	AUSTRALIA
Taphaeus sp.	BJS018	x	x	x	x	x	x	FRANCE
Diospilus sp.2	BJS020	x	x	x	x	x	x	FRANCE
Diospilus sp.3	JS059	x	x	x	—	—	x	COLUMBIA
Diospilus sp.4	JS093	x	x	x	x	x	x	PANAMA
Baeacis sp.1	JS091	x	x	x	x	x	x	PANAMA
Baeacis sp.2	BJS007	x	—	x	—	—	x	MADAGASCAR
Schaunislandia sp.1	BJS014	x	x	x	—	—	x	MADAGASCAR
Schaunislandia sp.2	BJS046	x	x	x	—	—	—	AUSTRALIA
Vadum sp.1	BJS087	x	x	x	x	x	x	USA
Vadum sp.2	BJS048	x	x	x	x	x	x	MEXICO
HELCONINAE - Brachistini								
Eubazus sp.1	BJS003	x	x	x	x	—	x	COLUMBIA
Eubazus (*Calyptus*) sp.2	BJS029	x	x	x	x	x	x	COLUMBIA
Eubazus (*Aliolus*) sp.3	BJS011	x	x	x	x	x	x	JAPAN
Eubazus (*Allodorus*) sp.4	BJS024	x	x	x	x	x	x	USA
Eubazus (*Brachistes*) sp.5	BJS026	x	x	x	x	x	x	COSTA RICA
Eubazus (*Brachistes*) sp.6	BJS034	x	—	x	x	x	—	FRANCE
Eubazus (*Aliolus*) sp.7	BJS010	x	x	x	x	x	x	USA
Eubazus (*Aliolus*) sp.8	BJS019	x	x	x	x	x	x	USA
Eubazus (*Aliolus*) sp.9	BJS037	x	x	x	x	x	—	COSTA RICA
Schizoprymnus sp.1	BJS008	x	x	x	x	x	x	SOUTH AFRICA
Schizoprymnus sp.2	BJS021	x	x	x	x	x	x	CHINA
Schizoprymnus sp.3	BJS023	x	x	x	—	x	x	USA
Nealiolus sp.	BJS025	x	x	x	x	x	x	USA
Triaspis sp.1	BJS027	x	x	—	—	x	x	SOUTH AFRICA
Triaspis sp.2	BJS036	x	—	x	x	x	x	USA

Table 3.1 cont'd.

BRACONIDAE - Putative Helconoid Complex cont'd.								
HELCONINAE - Brulleiini								
Flavihelcon distanti (Turner)	BJS085	x	x	x	x	x	x	MALAWI
Brulleia sp.	BJS113	x	x	x	x	x	x	THAILAND
ACAMPSOHELCONINAE								
Urosigalphus sp.1	BJS030	x	x	x	x	x	—	USA
Urosigalphus sp.2	BJS086	x	x	x	—	x	x	MEXICO
Urosigalphus sp.3	DM084	x	x	x	x	x	x	USA
BLACINAE								
Grypokeros sp.1	BJS112	x	x	x	x	x	x	CHILE
Grypokeros sp.2	JS214	x	x	x	—	x	—	CHILE
Apoblacus sp.	JS211	x	x	x	x	x	—	CHILE
Blacus sp.1	DM011	x	x	x	x	x	x	USA
Blacus sp.2	JS102	x	x	x	x	x	—	PANAMA
METEORIDEINAE								
Meteoridea sp.1	DM087	x	x	x	x	x	x	THAILAND
Meteoridea sp.2	JS228	x	x	x	x	x	—	THAILAND
AGATHIDINAE								
Cremnops montrealensis (Morrison)	JS031	x	x	x	x	x	x	USA
Bassus annulipes (Cresson)	JS046	x	x	x	x	x	x	USA
Earinus limitaris (Say)	JS106	x	x	x	x	x	x	CANADA
SIGALPHINAE								
Minanga serrata Cameron	JS209	x	x	x	x	—	x	SOUTH AFRICA
XIPHOZELINAE								
Xiphozele sp.	ZOO-35	x	x	x	x	x	x	THAILAND
MACROCENTRINAE								
Macrocentrus sp.	DM089	x	x	x	x	x	x	USA
Hymenochaonia sp.	JS008	x	x	x	x	x	x	USA
HOMOLOBINAE								
Homolobus sp.	JS027	x	x	x	x	—	x	USA
CHARMONTINAE								
Charmon cruentatus Haliday	JS012	x	x	x	x	x	x	USA
ORGILINAE								
Stantonia sp.	JS017	x	x	x	x	x	x	USA
Orgilus sp.	JS147	x	x	x	x	x	x	SOUTH AFRICA
AMICROCENTRINAE								
Amicrocentrum concolor (Szépligeti)	JS276	x	x	x	x	x	x	MALAWI
MICROTYPINAE								
Microtypus wesmaelii Ratzeberg	JS261	x	x	x	x	x	x	ENGLAND
CENOCOELIINAE								
Capitonius chontalensis (Cameron)	KP011	x	x	x	x	x	x	COSTA RICA
MAXFISCHERIINAE								
Maxfischeria sp.1	BJS114	x	x	x	x	—	x	AUSTRALIA
Maxfischeria sp.2	BJS115	x	x	x	x	x	x	AUSTRALIA
Maxfischeria sp.3	BJS116	x	x	x	x	x	x	AUSTRALIA
Maxfischeria tricolor Papp	BJS117	x	x	x	—	—	x	AUSTRALIA
Maxfischeria sp. 4	BJS089	x	x	x	x	—	x	AUSTRALIA
BRACONIDAE - Putative Euphoroid Complex								
EUPHORINAE								
Euphorinae Gen. sp. unident.	BJS035	x	x	x	x	x	x	USA
Planitorus sp.	BJS101	x	x	x	x	x	—	AUSTRALIA
Mannokeraia sp.1	BJS100	x	x	x	x	x	x	AUSTRALIA
Mannokeraia sp.2	BJS104	x	x	x	x	x	x	AUSTRALIA
Mannokeraia sp.3	BJS105	x	x	x	x	x	x	AUSTRALIA
Leiophron sp.	JS068	x	x	x	x	x	x	COLUMBIA
Perilitus sp.	JS124	x	x	x	x	—	—	MADAGASCAR

Table 3.1 cont'd.

Taxon	Code							Locality
BRACONIDAE - Putative Euphoroid Complex cont'd.								
METEORINAE								
Meteorinae Gen sp. unident.	BJS111	x	x	x	x	x	x	CHILE
Meteorus sp.1	BJS107	x	x	x	x	x	x	THAILAND
Meteorus sp.2	JS010	x	x	x	x	—	x	USA
ECNOMIINAE								
Ecnomios sp.	JS001	x	x	x	x	x	x	MADAGASCAR
NEONEURIINAE								
Kollasmosoma sp.	JS220	x	x	x	—	—	—	SPAIN
BRACONIDAE - Putative Microgastroid Complex								
MENDESELLINAE								
Espilogaster sp.	JS252	x	x	x	x	x	x	GUYANA
CHELONINAE								
Phanerotoma sp.	DM072	x	x	x	—	x	x	COLUMBIA
CARDIOCHILINAE								
Cardiochiles sp.	JS034	x	x	x	x	x	x	COLUMBIA
KHOIKHOIINAE								
Khoikhoia sp.	JS165	x	x	x	—	x	x	SOUTH AFRICA
MICROGASTRINAE								
Snellius sp.	JS078	x	x	x	x	x	x	COLUMBIA
Micropilitis sp.	DM037	x	x	x	x	—	x	
Fornicia sp.	JS222	—	—	x	x	x	x	THAILAND
MIRACINAE								
Miracinae Gen. sp. unident.	JS272	x	x	x	x	x	x	
ICHNEUTINAE								
Ichneutes sp.	DM090	x	x	x	x	x	x	USA
Proterops nigripennis Wesmael	JS003	x	x	x	x	x	x	USA
Muesonia straminea Sharkey & Wharton	JS042	x	x	x	x	x	x	COLUMBIA
BRACONIDAE - Putative Cyclostome Lineage								
APHIDIINAE								
Ephedrus sp.	JS207	x	x	x	x	—	x	FRANCE
Pseudopraon sp.	JS208	x	x	x	x	x	x	FRANCE
MESOSTOINAE								
Andesipolis sp.	JS225	x	x	x	x	x	x	CHILE
Aspilodemon sp.	JS007	x	x	x	x	—	—	COLUMBIA
Hydrangeocola sp.	JS054	x	x	x	x	x	x	COLUMBIA
ROGADINAE								
Aleiodes sp.	DM070	x	x	x	x	x	x	COLUMBIA
Clinocentrus sp.1	JS058	x	x	x	x	—	x	COLUMBIA
Clinocentrus sp.2	ZOO-8	x	x	x	x	x	x	USA
Macrostomion sp.	JS079	x	x	x	x	x	x	COLUMBIA
Cystomastax sp.	JS069	x	x	x	x	x	x	COLUMBIA
Polystenidea sp.	JS024	x	x	x	x	x	x	USA
DORYCTINAE								
Doryctes anatolikus Marsh	DM086	x	x	x	x	x	x	USA
Doryctes sp.	ZOO12	x	x	x	x	x	x	USA
Leluthia sp.	ZOO18	x	x	x	x	x	x	COLUMBIA
Liobracon sp.	ZOO20	x	x	x	x	x	x	KENYA
Heterospilus sp.1	ZOO11	x	x	x	x	—	—	USA
Heterospilus sp.2	DM103	x	x	x	x	—	—	USA
Notiospathius sp.	DM071	x	x	x	x	x	x	COLUMBIA
RHYSIPOLINAE								
Rhysipolis sp.1	DM081	x	x	x	x	x	x	USA
Rhysipolis sp.2	JS243	x	x	x	x	x	x	HUNGARY

Table 3.1 cont'd.

BRACONIDAE - Putative Cyclostomes cont'd.								
BRACONINAE								
Bracon sp.	DM073	x	x	x	x	x	x	COLUMBIA
Cyanopterus sp.	ZOO23	x	x	x	—	—	x	USA
Vipio texanus (Cresson)	JS005	x	x	x	—	—	—	USA
Hemibracon sp.	JS086	x	x	x	—	—	x	PANAMA
EXOTHECINAE								
Colastes sp.	JS081	x	x	x	x	x	x	PANAMA
Shawiana sp.	JS195	x	x	x	x	—	x	USA
OPIINAE								
Opius sp.	JS025	x	x	x	x	—	x	COLUMBIA
ALYSIINAE								
Hoplitalysia slossonae Ashmead	JS029	x	x	x	x	x	x	USA
HORMIINAE								
Hormius sp.	JS094	x	x	x	—	x	x	PANAMA
RHYSSALINAE								
Histeromerus sp.	JS202	x	x	x	—	x	x	USA
Oncophanes sp.	JS023	x	x	x	x	x	x	USA
Acrisis sp.	DM100	x	x	x	—	x	x	USA
Dolopsidea sp.	JS223	x	x	x	—	—	x	USA
PAMBOLINAE								
Pambolus sp.	DM074	x	x	x	x	x	x	USA
Pseudorhysipolis sp.	JS082	x	x	x	x	—	x	COLUMBIA
GNAMPTODONTINAE								
Pseudognamptodon sp.	JS020	x	x	x	—	x	x	USA
BETYLOBRACONINAE								
Conobregma sp.	ZOO27	x	x	x	x	x	x	DOMINICAN

*CAD, carbamoyl-phosphate sythetase-asparate transcarbamoylase-dihydroorotase, CPSase, carbamoylphosphate synthetase
**ACC, acetyl-coenzyme A carboxylase

3.3.2 DNA protocols

Genomic DNA was extracted from ethanol-preserved or previously mounted specimens following Qiagen protocols in conjunction with the DNeasy™ Tissue Kit (Qiagen, Valencia, CA). Voucher specimens were deposited in the University of Kentucky Insect Collection (Table 3.1). Several previous studies (Belshaw *et al.*, 1998; Mardulyn & Whitfield, 1999; Pitz *et al.*, 2007; Shi *et al.*, 2005; Zaldivar-Riverón *et al.*, 2006) have utilized 28S and/or 18S rDNA for inferring braconid relationships. However, 28S seems to have the greatest utility for tribal and generic relationships, while the highly conserved 18S has limited utility for inferring relationships between subfamilies. Nuclear protein-coding genes have not yet been utilized for inferring relationships among braconid subfamilies. However, the recent development of primers for protein-coding genes in other insects offers a possible new source of genetic information that

may be useful for relationships among Braconidae. The phylogenetic utility of CAD (carbamoyl-phosphate sythetase-asparate transcarbamoylase-dihydroorotase, occasionally called rudimentary) has been demonstrated for several insects, including flies (Moulton & Wiegmann, 2004), green lacewings (Winterton & de Freitas, 2006), and more recently for hymenopterans, including pteromalids (Desjardins et al., 2007) and megachilids (Praz et al., 2008). Acetyl-coenzyme A carboxylase, or ACC, has been suggested as a useful marker for Lepidoptera, although has never been explicitly tested (Regier, 2007; Regier et al., 2008). Here, these four genes were targeted for amplification, including: 28S rDNA (expansion regions D1-D5, sequenced in 2 fragments); 18S rDNA, two non-contiguous segments of the CPSase (carbamoylphosphate synthetase) region of CAD and one region of ACC. All primer pairs and associated references are listed in Table 3.2.

Table 3.2. Primer pairs used to amplify gene regions employed in this study and associated references.

Gene (region)	Primer Name	Primer Sequence (5' to 3')	Primer Length	Source
28S rDNA	28SD1F	ACC CGC TGA ATT TAA GCA TAT	21-mer	Harry et al. (1996)
(D1-D3)	28SD2R	TTG GTC CGT GTT TCA AGA CGG G	22-mer	Campbell et al. (1993)
	28SD1shortF	GUG GUA AAC UCC AUC UAA G	19-mer	current paper
	28SD2shortR	ACA TGT TAG ACT CCT TGG TC	20-mer	current paper
28S rDNA	28SD3F	GAC CCG TCT TGA AAC ACG GA	20-mer	Nunn et al. (1996)
(D3-D5)	28SD5R	CCC ACA GCG CCA GTT CTG CTT ACC	24-mer	Schulmeister (2003)
18S rDNA	18S-H17F	AAA TTA CCC ACT CCC GGC A	19-mer	Ocampo & Hawks (2006)
	18S-H35R	TGG TGA GGT TTC CCG TGT T	19-mer	Ocampo & Hawks (2006)
CAD	54F	GTN GTN TTY CAR ACN GGN ATG GT	23-mer	Moulton & Wiegmann (2004)
(CPSase)	405R	GCN GTR TGY TCN GGR TGR AAY TG	23-mer	Moulton & Wiegmann (2004)
CAD	apCADfor1	GGW TAT CCC GTD ATG GCB MGW GC	23-mer	Danforth et al. (2004)
(CPSase)	apCADrev1mod	GCC ATY RCY TCB CCY ACR CTY TTC AT	26-mer	Danforth et al. (2004)
ACC	1F	GTN TGG GCN GGN TGG GG	17-mer	Regier (2007)
	2F	TGG GCN GGN TGG GGN CAY GC	20-mer	Regier (2007)
	4R	GCY TCY TCD ATD ATY TTY TG	20-mer	Regier (2007)

The CPSase small chain of CAD (54F/405R) was amplified using primers developed by Mouton and Wiegmann (2004). Cycling conditions were slightly modified from the published protocols and included an initial denaturation at 94°C for 4 min, 3 cycles of 94°C for 30 sec, 59°C for 30 sec , and 72°C for 90 sec, followed by 5 cycles of 94°C for 30 sec, 57.5°C for 30 sec , and

72°C for 90 sec, then 28 cycles of 94°C for 30 sec, 54°C for 30 sec , and 72°C for 90 sec, with a final extension for 3 min at 72°C . A region of the CPSase large chain of CAD (primer pair apCADfor1/apCADrev1mod) was amplified using the primers and protocols developed by Danforth et al. (2004). An approximately 500 bp region of ACC was amplified using the primers of Regier (2007). A touchdown protocol was used to amplify ACC, which included an initial denaturation at 95°C for 4 min, followed by 29 cycles of a 30 sec denaturation at 95°C, a 30 sec annealing step starting at 60°C and decreasing 0.5°C every cycle, and an elongation step at 72°C for 40 sec. This touchdown protocol was followed by 8 cycles of 95°C for 30 sec, 45°C for 30 sec, and 72°C for 40 sec, with a final elongation step for 7 min at 72°C. Both regions (D1-D2 and D3-D5) of 28S rDNA were amplified with an initial denaturation of 3 min at 94°C, followed by 35 cycles of 94°C for 30 sec, 52°C for 30 sec, and 72°C for 70 sec, and a final elongation for 7 min at 72°C. The 18S rDNA fragment was amplified using the same protocol for 28S rDNA except the denaturation and annealing steps were lengthened to 45 sec each.

All PCR reactions were performed on a Bio-Rad PTC-0200 DNA Engine thermal cycler, using 1-2 µL of DNA extract, 2.5 µL of 1 X Standard *Taq* Buffer (New England Biolabs) (10 mM Tris-HCl, 50 mM KCl, 1.5 mM $MgCl_2$), 2.5 µL dNTP (10 µM), 2.0 µL of MgSO4 (50 µM), 1.0 µL of each primer (10 µM), 0.2 µL of *Taq* DNA polymerase (New England Biolabs), and purified water to a final volume of 25 µL. Both product purification and sequencing were performed at the Advanced Genetic Technologies Center, University of Kentucky using Agencourt CleanSEQ magnetic beads and an Applied Biosystems 3730xl DNA Analyzer, respectively. Contigs were assembled and edited using Contig Express (Vector NTI Advance10™Invitrogen™).

3.3.3. Multiple Sequence Alignment

Multiple sequence alignment is critically important to phylogenetic analysis, as it establishes statements of homology among molecular characters (Kjer, 1995). Ribosomal DNA has particularly difficult regions to align, that vary based on automated alignment program parameter settings (Sharkey et al., 2006; Thompson et al., 1994; Wheeler, 1995). Thus, the rDNA genes were aligned based on a secondary structure model for Ichneumonoidea developed by Yoder and Gillespie (2004) and Gillespie et al. (2005). Small regions of alignment ambiguity (RAAs), regions of slipped-strand compensation (RSCs), and regions of expansion and contraction (RECs) were aligned by eye (Yoder & Gillespie, 2004) (see Gillespie, 2004 for details on the determination of ambiguous regions). However, recently, Yoder and Gillespie added models for the large ambiguous regions (designated NHR1 and NHR2), and these models were

utilized to guide the alignment in these larger regions of ambiguity. Often, regions of ambiguous alignment are excluded from phylogenetic analysis as the homology statements might not be valid (Kjer, 1995). However, a significant amount of phylogenetic information is lost with the exclusion of these often variable, but informative regions. Thus, the 28SrDNA gene regions, which contain the largest number of ambiguous sites, were analyzed both with and without the defined regions of ambiguity.

For the protein-coding genes, alignment was performed using MUSCLE (Edgar, 2004) on the European Bioinformatics Institute (EBI) server and hand corrected in BioEdit (Hall, 1999) for reading frame accuracy. Protein-coding alignments were run through GBlocks v. 0.91b (Castresana, 2000) under default settings, to remove regions of ambiguous alignment. This treatment effectively removed all introns and uninformative indels.

3.3.4 Phylogenetic Analyses

Bayesian analyses were performed using MrBayes v3.1.2 (Huelsenbeck & Ronquist, 2001). MrModeltest v2.3 (Nylander, 2004; Posada & Crandall, 1998) was used with Paup* (Swofford, 2000) to test for the best evolutionary model for each gene region. For all protein-coding genes, the general time reversible model of evolution with a parameter for invariant sites and rate heterogeneity modeled under a gamma distribution (GTR+I+G) was determined as the best-fitting model using hierarchal likelihood ratio tests as implemented in MrModeltest. Stem regions subject to compensatory base changes of the rDNA genes were analyzed using the doublet model (Schöniger & von Haeseler, 1994), whereas loops and ambiguous regions were analyzed under the GTR+I+G model. All Bayesian analyses were performed with two independent searches, four chains and default priors. The number of generations required to reach stationarity varied across the different analyses and are reported in the respective figure legends. Stationarity was determined by the convergence statistics and log likelihood plots in MrBayes. After discarding trees for the burnin-phase, trees and branch lengths were summarized from the two independent searches with a majority rule consensus method. All protein-coding genes were partitioned by codon position, with model parameters unlinked across all partitions. Ribosomal DNA genes were partitioned into stem, loop, and ambiguous regions with model parameters unlinked. Tests for base composition homogeneity were performed in Paup* 4.0b10 (Swofford, 2000) using the PaupUp graphical interface (Calendini &

Martin, 2005). Base composition of different gene partitions were calculated in MEGA 4.0 (Tamura et al., 2007).

3.3.5 Ancestral State Reconstructions

Taxa were coded for endo- or ectoparasitism using the available biological information known for the closest taxonomic level. Often, information was available for one species of a given genus, and this information was used to code for that genus, even if a different species in the same genus with unknown biology was utilized. If the biology for a given genus was unknown, the most common mode of parasitism for the containing subfamily was used to infer the biology. One taxon, *Hydrangeocola* sp., has been reared from galls (Brèthes, 1927) and was originally coded as phytophagous. However, the characters state was re-coded as missing to retain a binary character necessary for some calculations. Ancestral state reconstructions were performed using maximum likelihood in Mesquite v. 2.5 (Maddison & Maddison, 2008). Two models were explicitly tested, including the Markov k-state 1 parameter model and the Asymmetrical Markov k-state 2 parameter model. The former model assumes an equal rate of transition between character states, whereas the latter model allows for unequal transition rates. The best-fitting model was determined using the asymmetry likelihood ratio test, which compares the likelihood of the two models on a tree for a given character. Reconstructions were performed using the topology and branch lengths with the highest likelihood from the posterior distribution of trees from the Bayesian analysis of the concatenated dataset.

3.4. Results

3.4.1 Gene statistics

Of the 139 total taxa examined in this study, 60% were amplified for all 6 gene regions, 85% for at least 5 regions, and 96% for at least 4 regions (Table 3.1). Thus, missing data were minimal. The final concatenated dataset had an aligned length of 4337 base pairs of which 43% of the sites were parsimony informative (Table 3.3). Among the individual gene datasets, the CPSase small chain of CAD (54-405) had the greatest number of informative sites. With the third position removed, nearly 50% of the sites were parsimony informative. Generally, the other two protein-coding gene regions (ACC and CAD-apmod) were more conserved. Both regions lost a significant portion of informative sites when the third position was removed (Table 3.3). Of the rDNA genes, the D1-D3 expansion region of 28S had the greatest percentage of parsimony informative sites. Most of the variability was between the D2-D3 regions, which has been the

most widely used amplicon for braconid systematics (Belshaw *et al.*, 1998; Belshaw & Quicke, 2002; Zaldivar-Riverón *et al.*, 2006).

Table 3.3. Gene regions utilized for individual and concatenated analyses, with the number of included taxa, aligned length, and percentage of parsimony informative sites. * For protein coding genes, the aligned length is reported after treatment with GBlocks; for rDNA genes, the aligned length includes regions of ambiguity.

Gene region	No. of included taxa	Aligned length*	Percent parsimony informative sites
28S (D1-D3)	138	1070	0.49
28S (D3-D5)	135	608	0.29
18S	138	754	0.15
CAD (54-405)	114	726	0.65
CAD (54-405) 3-out	114	484	0.48
CAD (apmod)	109	642	0.47
CAD (apmod 3-out)	109	428	0.22
acc	123	537	0.50
ACC 3-out	123	358	0.28
All genes	139	4337	0.43
All genes 3-out	139	3702	0.34

Across the 6 gene regions sampled here, there was a slight A-T composition bias in the concatenated dataset (A=28.2%, G=25.3%, C=20.0%, T=26.6%). When nucleotide composition was examined across the individual gene partitions, the stem regions of all rDNA gene fragments exhibited slight G-C biases (Figure 3.3A). Alternatively the loop regions, which are not inhibited by compensatory base changes, demonstrated distinct A-T biases. These biases were further exaggerated in ambiguous regions (Figure 3.3A). The stem regions of all of the rDNA fragments had relatively high proportions of both guanine and thymine, higher than any other combination of paired nucleotides in the 18S and 28S (D1-D3) rDNA fragments. This finding is unsurprising given that stem regions commonly have both Watson-Crick base pairings as well as guanine-thymine (G-T) interactions (Zuker, 1989). All of the protein-coding gene regions exhibited slight A-T biases when all positions were considered (Figure 3.3A). However, this bias was largely removed when the third position was excluded (Figure 3.3B).

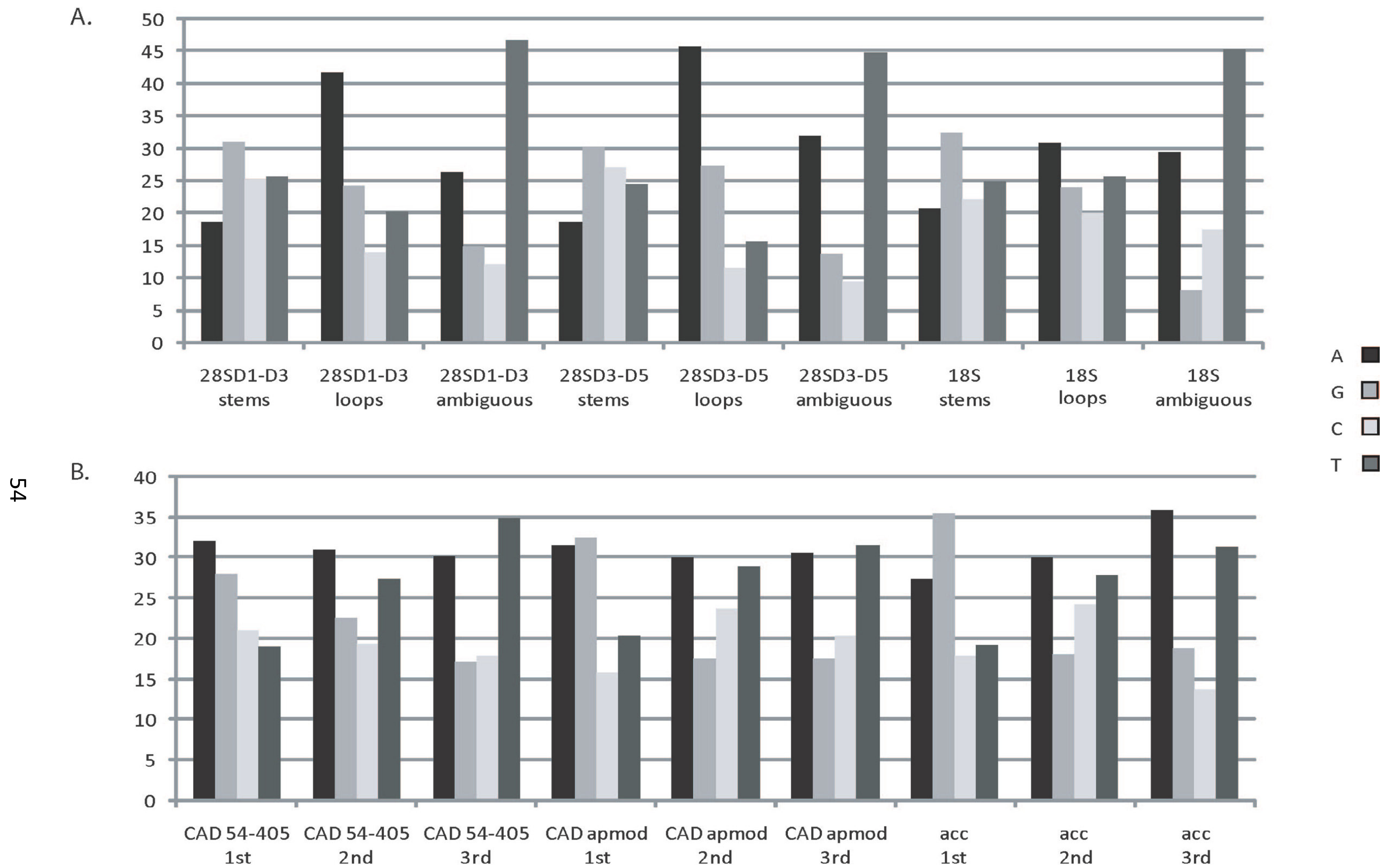


Figure 3.3. Nucleotide composition for all gene partitions. A. Ribosomal rDNA B. Protein-coding genes.

All protein-coding genes demonstrated non-stationarity in base composition when all data were included (Table B1, Appendix B). However, the null hypothesis of base composition homogeneity was accepted when the third position was removed, suggesting saturation in third position sites for all protein-coding genes. Although all sites of the D3-D5 region of 28S Rdna exhibited homogeneity, the stem regions of the D1-D3 fragment did not (Table B1, Appendix B). This lack of stationarity in the stem regions affected the outcome of the test for homogeneity for all sites in this amplicon, as both the loops and ambiguous regions of 28S (D1-D3) rDNA demonstrated homogeneity, but all sites did not (Table B1, Appendix B). Although the base composition bias may have affected the results of the phylogenetic inference of the 28S (D1-D3) rDNA fragment, the utilization of the doublet model may account for some of the compositional biases. Given that G-T interactions are common in rDNA, some taxa may have exhibited higher G-T content among paired sites, while others retained higher A-T content at these same sites, potentially causing the test for homogeneity to fail. Interestingly, the regions of ambiguity for all rDNA genes demonstrated base composition stationarity (Table B1, Appendix B). Thus, the higher A-T composition in these regions was relatively consistent across all taxa.

3.4.2. Individual gene analyses

3.4.2.1 Ribosomal DNA

Bayesian inference of the 28S (D1-D3) rDNA fragment recovered two distinct lineages: one containing the cyclostomes *sensu lato*[1] and one containing the non-cyclostomes (Figure 3.4). The cyclostome lineage, in the broad sense, also exhibited two main clades. The first clade consisted of Mesostoinae, Maxfischeriinae, and Aphidiinae (minus *Pseudopraon*), further referred to as the aphidioid complex. This clade was recovered as sister to the remaining cyclostomes. These relationships are consistent with the findings of Zaldivar-Riverón (2006) and Dowton et al. (2002), who both recovered Mesostoinae + Aphidiinae as sister to the remaining cyclostomes (Maxfischeriinae was not included in either analysis). The relationships among the remaining cyclostomes are poorly resolved, although a relatively well supported clade consisting of Opiinae, Alysiinae, Exothecinae, and Gnamptodontinae was recovered. The

---

[1] Although members of Mesostoinae possess the cyclostome condition, members of Aphidiinae and Maxfischeriinae do not (although occasionally the Aphidiinae have been considered secondarily cyclostome). These three subfamilies were recovered in a clade sister to the remaining cyclostomes. Thus, the entire lineage expands the strict definition of the cyclostomes (hence, *sensu lato* (s.l.). The remaining cyclostome families are further referred to as cyclostomes *sensu stricto* (s.s.).

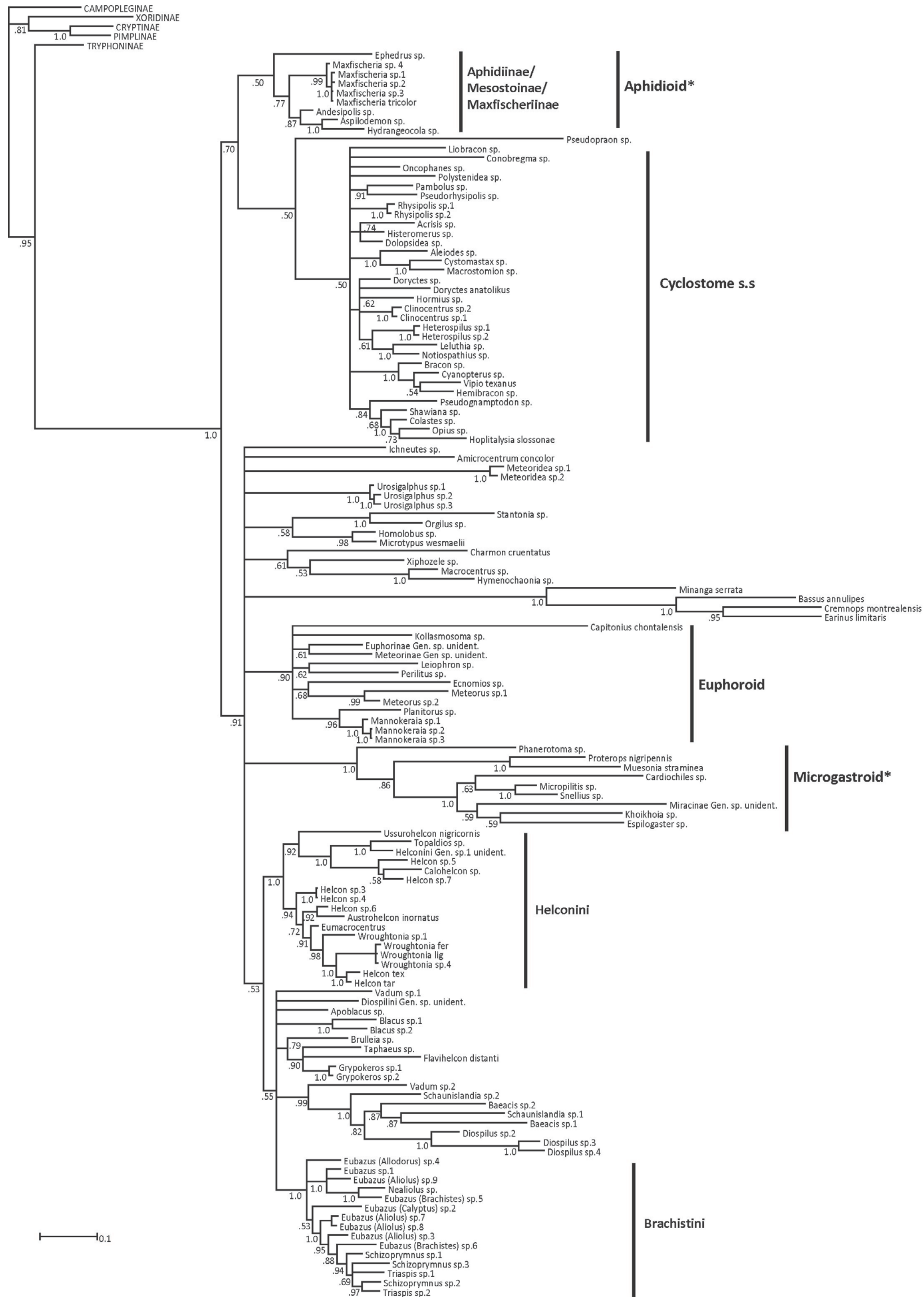


Figure 3.4. Inferred topology from the Bayesian analysis of 28S (D1-D3) rDNA. Posterior probabilities are listed below the node (12 million generations; burnin = 6.5 million generations). Asterisk indicates non-monophyly.

branching order among non-cyclostome subfamilies was also poorly resolved, particularly among helconoid subfamilies. However, there was strong support for a microgastroid lineage, including a paraphyletic Ichneutinae (*Ichneutes* is not recovered with the other two ichneutine genera, *Muesonia* and *Proterops*). Additionally, the euphoroid complex (including Euphorinae, Cenocoeliinae, Ecnomiinae, Meteorinae, Neoneurinae, and the enigmatic genera *Planitorus* and *Mannokeraia*) was recovered as monophyletic with high support. Within a clade containing Helconinae and Blacinae, Helconini and Brachistini were both recovered as monophyletic with strong support. However, the relationships among Blacinae and the other helconine tribes, Diospilini and Brulleiini, were poorly resolved.

The phylogeny recovered from the 28S (D3-D5) dataset had very little resolution, with almost all taxa contained within a large polytomy (Figure 3.5). In contrast to the other 28S dataset, Bayesian inference of this fragment recovered a basal Helconini, albeit without convincing support. The aphidioid clade (minus *Pseudopraon*) was again recovered, but with no relationship to the remaining cyclostomes. Sigalphinae + Agathidinae were recovered in a sister relationship with Acampsohelconinae, but with very little support. Almost all taxa were recovered in paraphyletic assemblages.

When the two fragments of 28S were analyzed together, there was greater resolution in the recovered phylogeny than when either gene region was analyzed alone (Figure B1, Appendix B). Again the cyclostomes and non-cyclostomes were recovered as two independent lineages. Generally, there was more resolution among cyclostome subfamilies in the combined dataset. Additionally, the aphidioid clade containing Maxfischeriinae, Mesostoinae, and Aphidiinae were recovered as sister to the remaining cyclostomes with very high support. While the two aphidiine taxa were not recovered as monophyletic when either gene fragment was analyzed alone, they were recovered together when the datasets were combined. Among the non-cyclostomes, the microgastroid complex of subfamilies was again recovered, but this time was monophyletic, including all representatives of the Ichneutinae, and with the ichneutines as sister to the remaining microgastroids. These relationships are consistent with previously published analyses (Dowton *et al.*, 2002; Murphy *et al.*, 2008). Similar to the 28S (D1-D3) rDNA analysis, the euphoroid complex was recovered as monophyletic, although the branching order among the subfamilial representatives was not well resolved. Consistent with both individual analyses, the combined dataset did not recover a monophyletic helconoid complex.

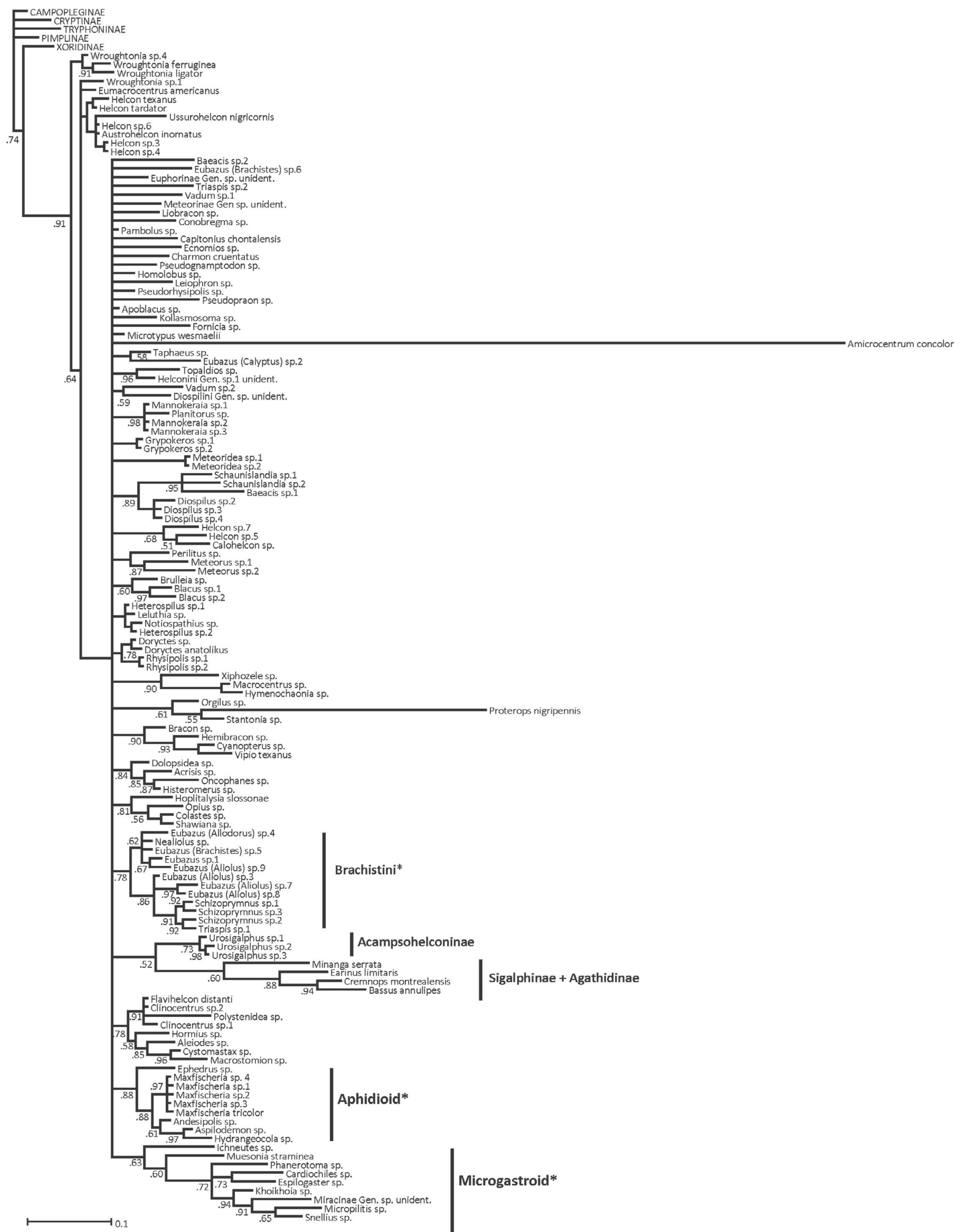


Figure 3.5. Inferred topology from the Bayesian analysis of 28S (D3-D5) rDNA. Posterior probabilities are listed below the node (23 million generations; burnin = 17.5 million generations). Asterisk indicates non-monophyly.

Both fragments of 28S rDNA were also analyzed with the ambiguous regions deleted (Figure B2, Appendix B). The recovered phylogeny using Bayesian inference is very similar to the tree recovered when all data were included (cf. Figure B1, Appendix B). Generally resolution was lost when the ambiguous regions were excluded and the nodal support for recovered clades was diminished. Occasionally, the branching order was altered, which was more notable among terminal relationships than more internal relationships. Relationships that were altered tended to be poorly supported in both analyses. Thus, inclusion of ambiguous regions primarily enhanced support for clades recovered when these regions were excluded. However, the poorly supported relationships that were altered when the ambiguous regions were included may have had a compounding effect in concatenated analyses.

The 18S rDNA gene region employed here was fairly conserved across the Braconidae, causing a lack of resolution in the inferred phylogeny (Figure 3.6). Relationships among the cyclostome subfamilies were the least resolved, with almost all taxa contained within the large basal polytomy. Among the non-cyclostomes, the Helconini and Brachistini (Helconinae) were both recovered as monophyletic. Other groups with a smaller number of representatives that were recovered as monophyletic include: Agathidinae, Microgastrinae, Macrocentrinae, Meteorideinae, and Orgilinae. Additionally, there was some evidence for the microgastroid complex *sensu stricto* (i.e. not including Ichneutinae), but it was not recovered as monophyletic.

3.4.2.2 Protein-coding genes

Bayesian inference of CAD (54-405) resulted in a well resolved topology with many highly supported nodes (Figure 3.7). The recovered phylogeny depicted the euphoroid complex as a basal braconid lineage, although the support was not overly convincing. Within the euphoroid complex, Cenocoeliinae was recovered as sister to the remaining subfamilies. The other braconid subfamilies were recovered in a polytomy with 3 main lineages indicated. The first lineage included the microgastroid complex as sister to Sigalphinae + Agathidinae. Although the microgastroid complex did not include the erratic taxon *Ichneutes*, the remaining microgastroid subfamilies were recovered as monophyletic. The second lineage emerging from the polytomy included the cyclostomes s.s and the aphidioid complex, which were both monophyletic. However, the erroneous placement of *Ichneutes* obscured the sister relationships between the cyclostomes s.s. and the aphidioid complex. Within the cyclostome

Figure 3.6. Inferred topology from the Bayesian analysis of 18S rDNA. Posterior probabilities are listed below the node (8 million generations; burnin = 3.5 million generations). Asterisk indicates non-monophyly.

Figure 3.7. Inferred topology from the Bayesian analysis of CAD (54-405). Posterior probabilities are listed below the node (6 million generations; burnin = 3.5 million generations). Asterisk indicates non-monophyly.

s.s. lineage, most relationships were poorly resolved, with several paraphyletic subfamilies (e.g. Rogadinae, Doryctinae). The third lineage within the polytomy included the helconoid complex. Acampsohelconinae was recovered as sister to all other subfamilies within the clade. Additionally, two main lineages were recovered within the complex. One clade included the subfamilies Amicrocentrinae, Orgilinae, Homolobinae, Microtypinae, Charmontinae, Xiphozelinae, and Macrocentrinae. These subfamilies are similar morphologically and biologically and have variably been placed together in different classification schemes of the Braconidae (van Achterberg, 1984). For ease of discussion this clade is hereafter referred to as the macrocentroid complex. The other lineage included Helconinae and Blacinae, neither of which was recovered as monophyletic. Once again, Helconini and Brachistini were recovered as monophyletic groups with very strong nodal support. Meteorideinae was recovered within the large polytomy; however, with only two representatives and an unresolved branching order, very little can be said about the relationship of Meteorideinae to other braconid subfamilies.

Given that the CAD (54-405) gene fragment failed the test for base composition homogeneity in the third position, it was excluded in a subsequent analysis. When the third position was excluded, the recovered topology was similar to the relationships recovered with all data included (Figure B3, Appendix B). However, the monophyly of the cyclostomes s.s and Helconini was eroded with the exclusion of the third position. Additionally, Ichneutinae (minus *Ichneutes*) was recovered in a polytomy with the microgastroid complex and Sigalphinae + Agathidinae. The branching order between the main lineages was altered slightly, as the cyclostomes were recovered as sister to the helconoid complex, but with very low nodal support. Generally, all of the basal nodes depicting the relationships among the major braconid lineages were poorly supported.

The relationships recovered under Bayesian inference of the CPSase large chain of CAD (CAD-apmod) differed significantly from the relationships recovered using the CPSase small chain (CAD 54-405). Generally, there was a lack of basal resolution among braconid lineages and nodes that were resolved tended to be poorly supported (Figure 3.8). In contrast to the other region of CAD, the ichneutine *Proterops* was recovered as sister to all remaining braconid taxa, albeit with relatively weak support. The branch length of *Proterops* was relatively long,

Figure 3.8. Inferred topology from the Bayesian analysis of CAD (apmod). Posterior probabilities are listed below the node (10 million generations; burnin = 2.5 million generations). Asterisk indicates non-monophyly.

possibly indicating long-branch attraction to the outgroup taxa. Discounting the position of *Proterops*, some members of the aphidioid complex were recovered as the ancestral lineage of the Braconidae, again with limited support. This group included Maxfischeriinae and Mesostoinae, but not Aphidiinae. The only aphidiine (*Pseudopraon*) included in this analysis fell within the cyclostomes s.s. Most of the remaining subfamilies were recovered in a large polytomy. However, Helconini and Brachistini were recovered as monophyletic, the latter with strong support. Oddly, the cyclostomes were recovered as a derived lineage within a basal grade consisting of some macrocentroid, microgastroid, euphoroid, and helconoid taxa. Other strongly supported clades among the non-cyclostomes included: Agathidinae + Meteorideinae, Maxfischeriinae + Mesostoinae, a paraphyletic euphoroid complex, and a paraphyletic microgastroid complex. The cyclostomes, although paraphyletic with respect to the placement of the alysiine *Hoplitalysia slossonae*, were recovered with strong support and included the aphidiine, *Pseudopraon* sp. Within the cyclostomes, the rogadines were paraphyletic with most taxa falling in a basal polytomy with respect to the other cyclostome taxa. However, there was strong support for a clade consisting of all other cyclostome taxa. Additionally, Rhysipolinae was recovered as sister to the remaining cyclostomes with strong support.

When the third position was excluded from the analysis, all of the major braconid lineages were recovered in a large polytomy (Figure B4, Appendix B), suggesting that the third position was driving the branching order recovered in Figure 3.8. Agathidinae, Meteorideinae, Maxfischeriinae, Acampsohelconinae, Microgastrinae, and Brachistini were all recovered as monophyletic. Additionally, with the exception of the alysiine (*Hoplitalysia* sp.) which had a clearly erroneous placement, the cyclostomes were recovered as a distinct clade. Additionally, Euphorinae (if Meteorinae is included) was recovered as monophyletic.

The relationships recovered under Bayesian inference of ACC were very conservative, with most lineages recovered in a large polytomy. The aphidioid complex was once again recovered as monophyletic, although there was no relationship to the remaining cyclostomes. Among members of Helconinae, Brachistini was recovered as a strongly supported monophyletic clade derived from a basal grade of both diospiline and blacine taxa, along with an erroneously placed euphorine (*Mannokeraia* sp.3). Acampsohelconinae was recovered as sister to the strongly supported monophyletic Helconini. Euphorinae was also recovered as a distinct clade, but was paraphyletic with respect to the misplaced *Mannokeraia*. Although the

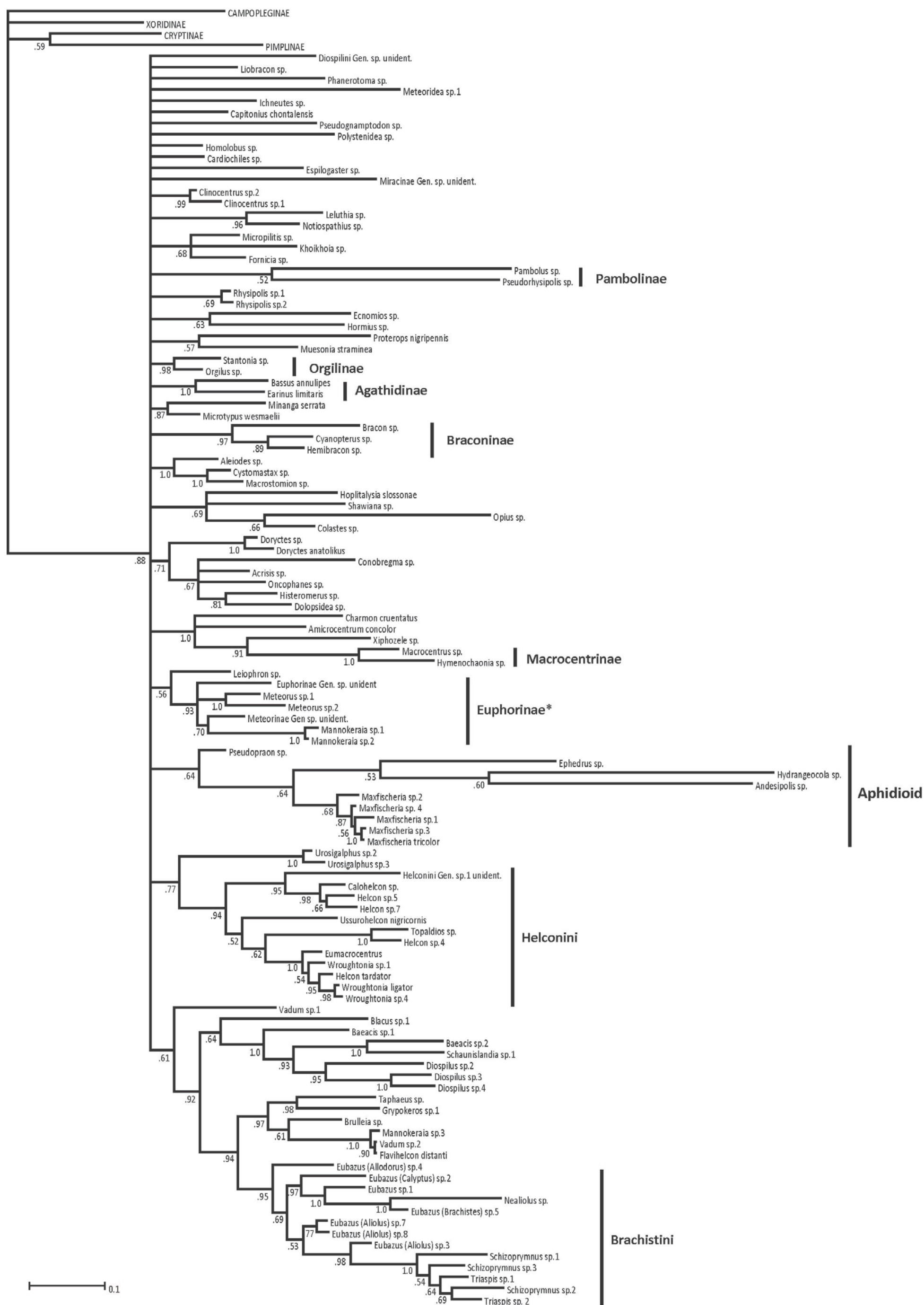


Figure 3.9. Inferred topology from the Bayesian analysis of ACC. Posterior probabilities are listed below the node (20 million generations; burnin = 5 million generations). Asterisk indicates non-monophyly.

branching order among the subfamilies was poorly resolved, a number of additional subfamilies were recovered as monophyletic, including: Macrocentrinae, Braconinae, Agathidinae, Orgilinae, and Pambolinae. When the third position was removed, most of the resolution recovered when all data was included was eroded (Figure B5, Appendix B). However, with the exception of *Hormius*, the cyclostomes were recovered as a distinct clade, that included a monophyletic aphidioid complex.

3.4.3 Concatenated analysis

Bayesian inference of the concatenated dataset recovered a well resolved and strongly supported topology (Figure 3.10). Given the extensive resolution of the recovered phylogeny, nodal support was represented by the thickness and shade of supporting branches (see, Figure 3.10 legend). Similar to the 28S rDNA analysis (Figure B1, Appendix B), the non-cyclostomes were recovered as sister to the cyclostomes. Both lineages were strongly supported, with posterior probabilities of 0.96 and 1.0, respectively.

Within the cyclostomes, the aphidioid complex was once again robustly recovered as sister to the remaining cyclostomes. Additionally, all subfamilies within the aphidioid complex were recovered as monophyletic, although Mesostoinae was weakly supported. Among the remaining cyclostomes with multiple representatives, Pambolinae, Braconinae, Rhyssalinae, and Rhysipolinae were all recovered as monophyletic with robust support. Additionally, there was strong evidence suggesting Opiinae, Alysiinae, Exothecinae, and Gnamptodontinae were closely related. Given that Doryctinae was never recovered as monophyletic across the individual gene trees, it was unsurprising that this subfamily was recovered as paraphyletic in the concatenated analysis. Rogadinae (including *Conobregma*) was recovered as a distinct clade, but did not include the volatile taxon *Polystenidea*.

Among the non-cyclostome subfamilies, Meteorideinae was recovered as the sister group to the remaining non-cyclostomes, albeit with weak support. The unusually long branch length of the meteorideine clade is suggestive of long-branch attraction. Apart from Meteorideinae, there were two distinct clades recovered within the non-cyclostome lineage. The first clade, which was moderately supported (posterior probability =0.77), contained the euphoroid and microgastroid complexes and the subfamilies Agathidinae + Sigalphinae. The microgastroid complex was recovered as monophyletic, including all Ichneutinae. However,

Figure 3.10. Inferred topology from the Bayesian analysis of all genes concatenated. The third position was excluded from protein-coding genes. Thick black branches indicate posterior probabilities ≥ 0.95. Thin black branches indicate posterior probabilities between .90 and .98. Thin grey branches indicate posterior probabilities < 0.90. Posterior probabilities are listed below the node (40 million generations; burnin = 28 million generations). Asterisk indicates non-monophyly.

Ichneutinae was recovered as paraphyletic, but sister to all remaining microgastroid subfamilies. Among the remaining microgastroid subfamilies, the branching order was variably supported, but there was strong evidence indicating Cheloninae as the sister lineage to the other subfamilies. Sigalphinae + Agathidinae was robustly recovered as the sister group to the microgastroid complex, suggesting these subfamilies do not belong within the helconoid complex. The euphoroid complex was robustly recovered with Cenocoeliinae as sister to all remaining euphoroid subfamilies. Neither Euphorinae nor Meteorinae were recovered as monophyletic.

The second major non-cyclostome lineage contained all the remaining taxa traditionally placed in the helconoid complex. Acampsohelconinae was recovered as sister to the remaining taxa, but with relatively weak support. Similar to the CAD (54-405) gene tree, the macrocentroid clade containing the subfamilies Orgilinae, Homolobinae, Microtypinae, Charmontinae, Amicrocentrinae, Xiphozelinae, and Macrocentrinae was recovered with impressive support. There was also strong evidence indicating that the latter 4 subfamilies are closely related. Additionally, Orgilinae was recovered as sister to Homolobinae + Microtypinae with robust support. This entire lineage was weakly recovered as sister to a clade consisting of all taxa within the Helconinae and Blacinae. Similar to most individual gene trees, Brachistini and Helconini were robustly supported as monophyletic. However, the remaining helconine tribes, Diospilini and Brulleiini, were recovered as paraphyletic with respect to each other and the Blacinae.

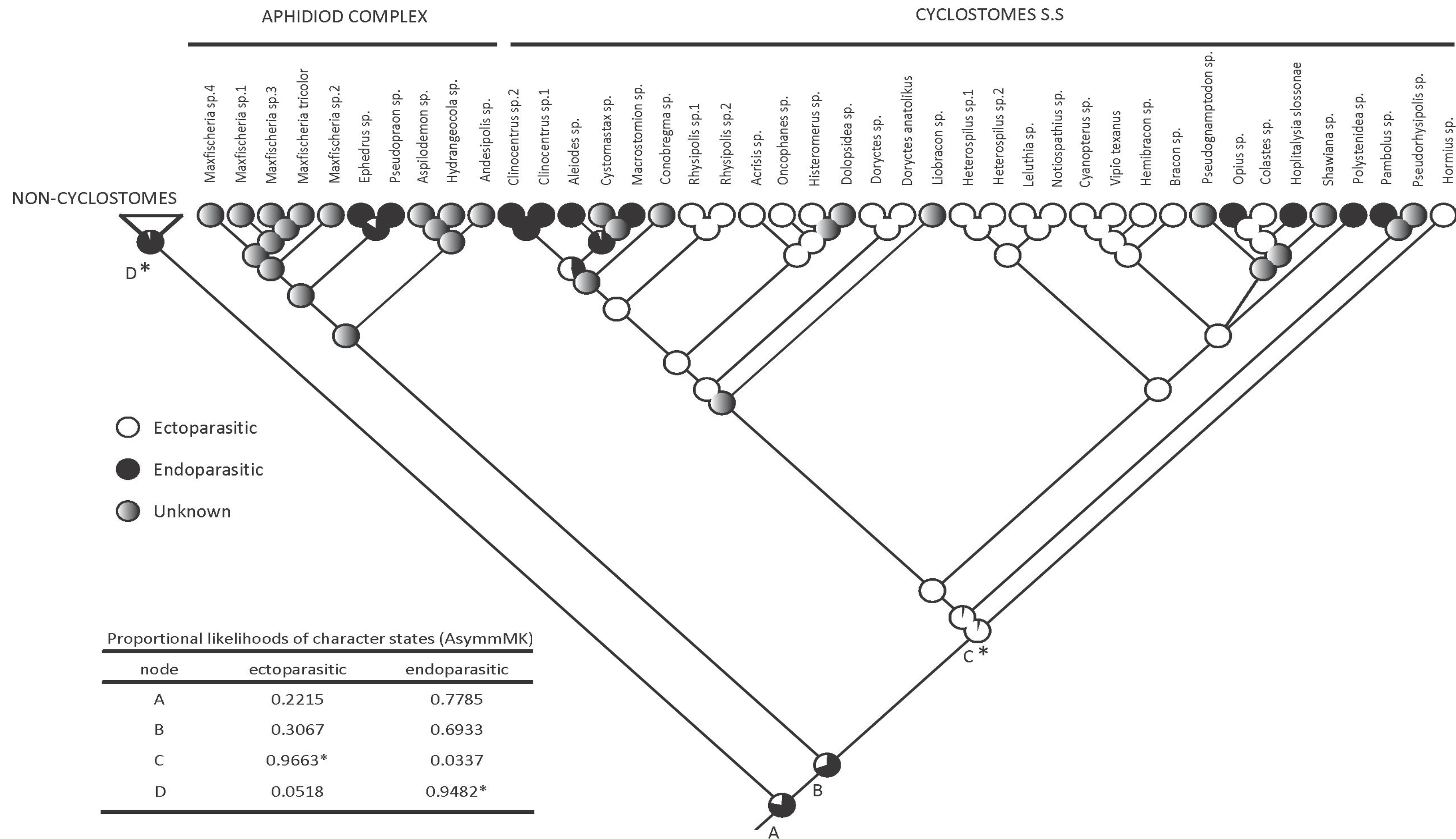
3.4.4 Ancestral State Reconstruction

Life history strategies were mapped onto the topology with the highest probability from the Bayesian inference of the concatenated dataset using maximum likelihood ancestral state reconstruction under the MK1 and AsymmMK models. Based on the likelihood ratio test, the AsymmMK model was significantly better than the MK1 model (χ^2 = 12.344, 1df, p=0.05). Thus, the AsymmMK model, which calculates differential rates of transition between two characters states, provided a better fit for the data given the topology. The AsymmMK model can only be utilized with binary characters thus, the phytophagous taxon *Hydrangeocola* was coded as unknown. Outgroups were removed and the tree was re-rooted between the two main braconid lineages, the cyclostomes s.l. and non-cyclostomes. Additionally, several non-cyclostomes were removed to balance the number of taxa on each side of the root point, to

prevent any bias created by the greater sampling of the endoparasitic non-cyclostomes. Since all known non-cyclostomes were coded as endoparasitic, this clade was collapsed in the resulting figure for easier visualization (Figure 3.11). Based on the recovered topology from the Bayesian inference, the ancestral condition for the two main braconid lineages was reconstructed as endoparasitic (Figure 3.11, node A). However, this reconstruction was not statistically significant. Of the selected nodes of interest labeled in Figure 3.11, only the ancestral nodes leading directly to the non-cyclostomes and the cyclostomes s.s were statistically significant for endoparasitism and ectoparasitism, respectively (Figure 3.11, nodes D and C). The ancestral node leading to the aphidioid complex could not be ascertained under the AsymmMK model due to missing data. Calculations for the rate of gains and losses was calculated under the model and demonstrated a higher rate of transition from ecto- to endoparasitism ($0 \rightarrow 1 = 3.1134$, $1 \rightarrow 0 = 0.1780$). This calculation makes intuitive sense given the topology, as there was only one possible transition from endo- to ectoparasitism, from node B to C in Figure 3.11.

If the ancestral braconid was indeed endoparasitic, then there was one independent transition to ectoparasitism (at node C) and at least four independent transitions to endoparasitism within the cyclostome lineage (five if the placement of the rogadine *Polystenidea* is correct). The transitions to endoparasitism in Opiinae (*Opius* sp.) and Alysiinae (*Hoplitalysia slossonae*) are independent in this reconstruction (Figure 3.11). Given the lack of statistical significance of the most ancestral node for the Braconidae (node A) and the node leading to the aphidioid + cyclostome s.s lineage, it is possible that the ancestor of the Braconidae was ectoparasitic. If this were true, then there would be at least six independent transitions to endoparasitism (seven considering the placement of *Polystenidea*).

Most of the internal branches within the cyclostome lineage were poorly supported in the concatenated analysis (Figure 3.10). Thus, the reconstruction was reanalyzed after collapsing the cyclostome s.s. lineage to a complete polytomy. Interestingly, the ancestral condition for the Braconidae became statistically significant as endoparasitic when the cyclostomes were collapsed to a polytomy (Figure 3.11, node A, proportional likelihood for endoparasitism = 0.9167). Thus, the recovered topology among the cyclostomes had a significant effect on the ancestral state reconstruction for the most ancestral braconid node.



Proportional likelihoods of character states (AsymmMK)

node	ectoparasitic	endoparasitic
A	0.2215	0.7785
B	0.3067	0.6933
C	0.9663*	0.0337
D	0.0518	0.9482*

Figure 3.11. Ancestral state reconstruction under the Asymmetrical Markov k-state 2 parameter model (AsymmMK) using maximum likelihood. Taxa were coded as ectoparasitic (0), endoparasitic (1), or missing (?). Outgroups were removed. The non-cyclostome lineage was collapsed for easier visualization, but was reduced to 42 taxa, the same number as the taxa in the aphidioid and cyclostome lineages combined. All included non-cyclostomes were coded as endoparasitic. The proportional likelihoods are listed in the table for the selected nodes of interest, labeled A-D. An asterisk indicates significance at $p < 0.05$. Calculations were not possible for nodes depicted as unknown.

To test how the outgroups's character state would affect the reconstruction, one analysis was run with outgroups included and coded as endoparasitic, and once with outgroups included and coded as ectoparasitic. When outgroups were coded as endoparasitic, the most ancestral code was again recovered as significantly endoparasitic (Figure 3.11, node A, proportional likelihood for endoparasitism = 0.9574). However, when outgroups were coded as ectoparasitic, the reconstruction of the ancestral most node was recovered as ectoparasitic, but not significantly (Figure 3.11, node A, proportional likelihood for ectoparasitism = 0.7154). While it makes intuitive sense that the character state of the outgroup might affect the analysis, it highlights the importance of understanding the evolutionary transitions among the Ichneumonidae and to determine the sister group to the Ichneumonoidea.

The biology for the aphidioid complex was primarily coded as unknown due to the lack of available host records for the included taxa. This ambiguity may have affected the outcome of the reconstruction of character states. Thus, different biologies were coded for the members of Maxfischeriinae to test the effect of the biology of this clade on the reconstructed ancestral state for Braconidae. Aphidiinae was coded as endoparasitic. Mesostoinae was coded as unknown, since the only known host record for the included taxa indicated phytophagy (*Hydrangeocola*). When, the members of Maxfischeriinae were coded as endoparasitic, the ancestral condition for the Braconidae was significantly endoparasitic (Figure 3.11, node A, proportional likelihood for endoparasitism= 0.9398). However, when these taxa were coded as ectoparasitic, the ancestral condition was equivocal (Figure 3.11, node A, proportional likelihood for ectoparasitism = 0.5530). Thus, the biology of the aphidioid clade can have a large impact on the reconstruction of the ancestral life history strategy for the Braconidae.

3.5. Discussion

3.5.1 Utility of protein-coding markers

The phylogenetic utility of CAD, or rudimentary, has been well-documented in other insects (Desjardins *et al.*, 2007; Moulton & Wiegmann, 2004; Praz *et al.*, 2008; Winterton & de Freitas, 2006). Of the two regions of the CAD utilized here, the small chain fragment of the CPSase region (CAD 54-405) is considerably more informative for resolving relationships among braconid subfamilies. Both regions demonstrate heterogeneity in base composition in the third position, potentially indicating saturation. However, the phylogeny inferred from the large chain fragment of the CPSase region of CAD (apmod) (Figure 3.8) is much less resolved than the

CAD (54-405) gene tree (Figure 3.7). Thus, the large chain of the CPSase region of CAD may have greater phylogenetic utility for higher level relationships than those analyzed here. For subfamilial relationships within the Braconidae, acetyl-coenzyme A carboxylase (ACC) seems to be too conserved to have any significant resolving power, particularly when the third position is removed. However, the slow rate of evolution and ease of amplification and alignment of this gene may be advantageous for higher level phylogenetic studies of the Hymenoptera and other insect orders.

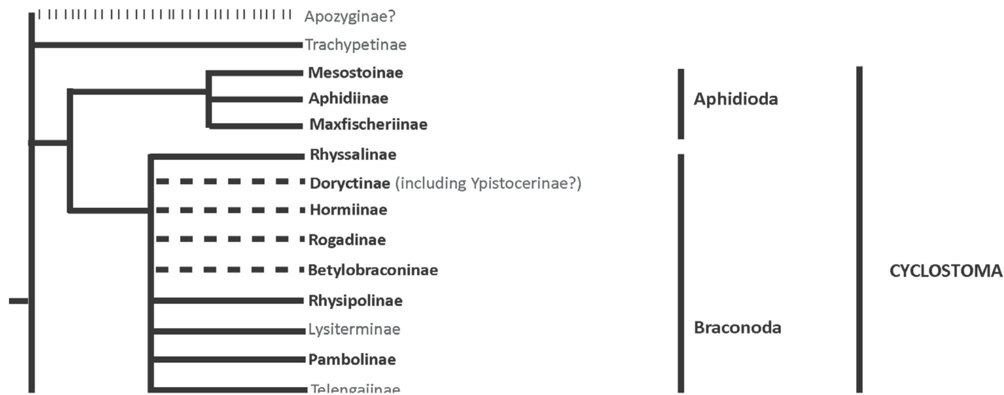
3.5.2 Phylogenetic implications

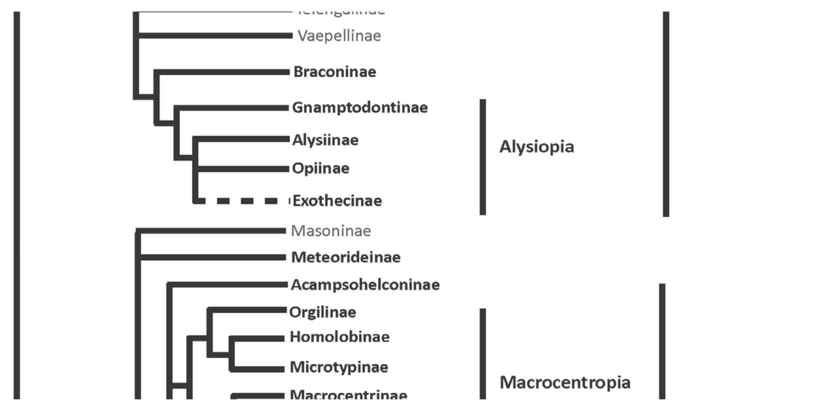
Three competing hypotheses have been proposed for the evolution of the major lineages of Braconidae (Figure 3.1). Among the genetic markers analyzed here, there is no evidence suggesting that the non-cyclostomes were derived from within the cyclostome lineage, contrary to findings based strictly on morphology (Quicke & van Achterberg, 1990b). There is some evidence in support of the viewpoint proposed by Dowton et al. (1998) that the cyclostomes represent a derived braconid lineage (cf. Figure 3.1C). The individual gene analyses of 28S (D3-D5) rDNA and both regions of CAD recover a derived cyclostome lineage, but with relatively weak support (Figures 2.5, 2.7, and 2.8). One gene recovers Helconinae as the ancestral braconid lineage (Figure 3.5) and another recovers the euphoroid complex as ancestral (Figure 3.7). Additionally, inference of the CAD (apmod) dataset recovers the cyclostomes as a derived lineage from within the euphoroid complex. However, when the third position is excluded from CAD (apmod), the branching order suggesting a derived cyclostome lineage dissolves (Figure B4, Appendix B). Although the euphoroid complex is weakly recovered as ancestral using CAD (54-405), support for this topology diminishes when the third position is removed (Figure B3, Appendix B). Based on the weight of the evidence herein, it is unlikely that the cyclostomes are a derived lineage from within the non-cyclostomes; rather, the cyclostomes most likely represent a sister lineage to the non-cyclostomes.

The individual gene analyses using 18S rDNA and ACC are uninformative as to the branching order among the major lineages within Braconidae (Figures 2.6 and 2.9, respectively). However, the analysis of 28S (D1-D3) rDNA, arguably the most informative gene fragment of all the rDNA markers, provides some evidence for the hypothesis that the cyclostomes and non-cyclostomes are sister lineages (Figure 3.4). Additionally, similar topologies are recovered when the 2 regions of 28S rDNA are analyzed together, regardless of the inclusion or exclusion of

regions of ambiguous alignment (Figures B1 and B2, Appendix B). The weight of the total evidence, depicted in the topology of the concatenated dataset (Figure2.10) also suggests that the cyclostomes and non-cyclostomes are sister lineages, as has been most commonly recovered in previous analyses (Belshaw *et al.*, 1998; Belshaw & Quicke, 2002; Pitz *et al.*, 2007; Shi *et al.*, 2005; Wharton *et al.*, 1992).

Based on the individual gene analyses and on the concatenated dataset, there is increased support for several previously proposed relationships and strong evidence for several new relationships among braconid subfamilies. These relationships are summarized below and are depicted in Figure 3.12 with newly proposed taxonomic names (Table 3.4). Currently, there is no division between family and subfamily that is formally recognized by the International Code of Zoological Nomenclature (ICZN). Within Braconidae, this has led to severe fluctuation in the number and arrangement of subfamilies (see Wharton, 2000 for a detailed discussion). Some authors have proposed names to reflect phylogenetic relationships among subfamilies, such as the complexes with the -oid suffix discussed previously. While these complexes attempt to portray relationships among subfamilies, the –oid suffix is confusing, as animal superfamily names use the ending –oidea. Additionally, the ending –oidae is typically used for the informal rank epifamily (above family). Therefore, the suffix for previously named and newly proposed complexes are changed to –oda (Greek, meaning 'like' or 'a resemblance' (Borror, 1960)). This change is similar enough to the previous suffix to minimize confusion with older literature and misunderstandings between other taxonomic rankings. To better reflect phylogenetic relationships, some zoologists (Lambert, 1990; Sereno, 1986) have made informal rankings between order and suborder using the following prefixes (in order): parv–; nan–; hypo–; and min–. Thus, these complexes might be referred to as hypofamilies, an unofficial ranking between family and subfamily. This rank designation allows for flexibility within the classification scheme, as there is at least one rank above (nanfamily) and below (minfamily) hypofamily. Based on abundant molecular and/or morphological evidence suggesting further relationships between subfamilies, the rank of minfamily is also utilized here. The suffix –opia (Greek, meaning 'vision') is utilized to denote minfamily. It should be stressed, however, that these taxonomic ranks are informal and are not recognized by the ICZN, but offer phylogenetic meaning and information content within the zoological code, as the senior generic name is used as the root, and an original suffix is used to denote the ranks.





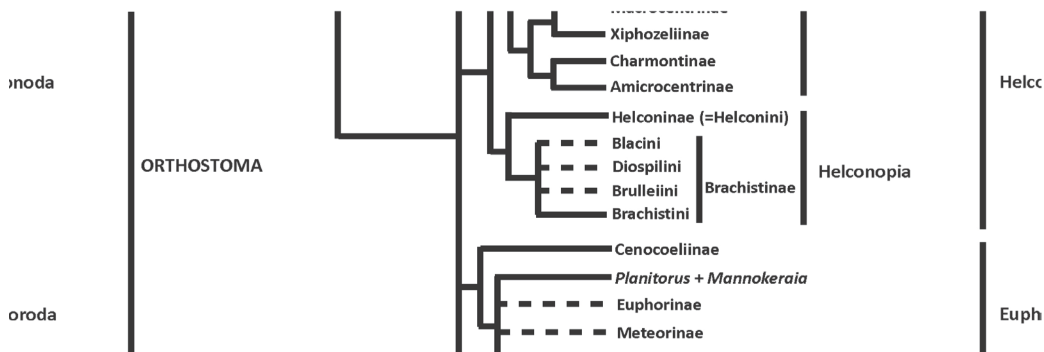


Figure 3.12. Summary tree of well-supported relationships among Braconidae with proposed taxonomic changes. See Table 3.4 and discussion for further information. Dashed lines indicate likely paraphyly. Vertical lines indicate subfamilies that have been analyzed in any molecular datasets. Subfamily names colored grey are not represented in the current dataset. The placement of subfamilies with a question mark after the name are based on limited morphological or molecular data, and thus their placement represent s the current opinion in the field, rather than the results of repeated phylogenetic testing.

Table 3.4. Proposed classification scheme for Braconidae.

FAMILY BRACONIDAE

Unplaced Subfamilies

Apozyginae

Trachypetinae

A. ORTHOSTOMA	**B. CYCLOSTOMA**
Unplaced Subfamilies	**1. HYPOFAMILY APHIDIODA**
Meteorideinae	Maxfischeriinae
Masoninae	Aphidiinae
1. HYPOFAMILY HELCONODA	Mesostoinae
Unplaced Subfamilies	**2. HYPOFAMILY BRACONODA**
Acampsohelconinae	Unplaced Subfamilies
a. Minfamily Helconopia	Betylobraconinae
Helconinae	Braconinae
Brachistinae	Doryctinae
(incl. Brachistini,	
Blacini,	Hormiinae
Brulleiini, Diospilini)	Lysiterminae
b. Minfamily Macrocentropia	Pambolinae
Amicrocentrinae	Rhyssalinae
Charmontinae	Rogadinae
Macrocentrinae	Telengainae
Xiphozeliinae	Vaepellinae
Orgilinae	Ypistocerinae
Homolobinae	**a. Minfamily Alysiopia**
Microtypinae	Alysiinae
2. HYPOFAMILY EUPHORODA	Exothecinae
Cenocoeliinae	Gnamptodontinae
Ecnomiinae	Opiinae
Euphorinae	
Meteorinae	
Neoneurinae	
3. HYPOFAMILY SIGALPHODA	
Agathidinae	
Sigalphinae	
4. HYPOFAMILY MICROGASTRODA	
Unplaced Subfamilies	
Dirrhopinae	
a. Minfamily Microgastropia	
Cardiochelinae	
Cheloninae(including	
Adeliinae)	
Khoikhoiinae	
Microgastrinae	
Miracinae	
b. Minfamily Ichneutopia	
Ichneutinae	

Beyond altering complexes to hypofamilies and minfamilies, the commonly used informal names cyclostome and non-cyclostome are also in need of revision. Here, the latter informal division is renamed to reflect the phylogenetic relationships recovered in this study and to minimize confusion with previous literature. Consistent with previous literature (van Achterberg, 1984), the name Cyclostoma (Greek, meaning 'round mouth' (Brown, 1956)) will refer to the cyclostomes s.l. (Figure 3.12). The name Orthostoma (Greek, meaning 'normal mouth' (Brown, 1956)) will refer to the non-cyclostomes. These names might be referred to as a rank above hypofamily (i.e. nanfamily); however, the nomenclature does not incorporate the senior generic names of group members and therefore have less meaning within the ICZN.

3.5.2.1 Orthostoma

Orthostoma, formerly referred to as the non-cyclostomes, includes the following hypofamilies: Microgastroda, Sigalphoda, Euphoroda, Helconoda, and Macrocentroda (Figure 3.12). Based on a lack of evidence to suggest otherwise, the subfamilies Meteorideinae and Masoninae (the latter not analyzed in this study) are not placed within any minfamily, but are included within Orthostoma. Previous molecular studies have demonstrated members of Trachypetinae to be sister to all remaining braconids (Belshaw *et al.*, 2000; Belshaw *et al.*, 1998; Belshaw & Quicke, 2002; Dowton *et al.*, 2002; Pitz *et al.*, 2007). However, long-branch attraction has been suspected to contribute to this placement (Belshaw & Quicke, 2002). It is possible upon future sequencing efforts that Trachypetinae may also be included within Orthostoma, but currently are designated as unplaced (Table3.4).

3.5.2.1.1 Microgastroda

Microgastroda (previously entitled the microgastroid complex) has been the subject of numerous studies due to the extensive utility of its members as biological control agents and as model group to understand the evolution of polydna viruses (Banks & Whitfield, 2006; Murphy *et al.*, 2008; Whitfield, 1997). The monophyly of Microgastroda has been well supported in numerous molecular analyses (Banks & Whitfield, 2006; Belshaw *et al.*, 1998; Dowton & Austin, 1998; Dowton *et al.*, 1998; Murphy *et al.*, 2008; Whitfield, 1997), although the branching order of the included subfamilies has varied. Here, the monophyly of Microgastroda, including Ichneutinae, is robustly demonstrated across numerous gene trees and in the concatenated analysis (Figure 3.10). However, similar to other molecular analyses (Belshaw *et al.*, 2000; Belshaw & Quicke, 2002; Pitz *et al.*, 2007), a monophyletic Ichneutinae is never recovered,

partially due to the volatile placement of *Ichneutes* across the individual gene trees. Thus, two minfamilies are created within Microgastroda, Ichneutopia and Microgastropia (Table 3.4), with the former paraphyletic. The subfamily Dirrhopinae, which has never been analyzed with molecular data is left as unplaced within Microgastroda.

Most of the internal branches within Microgastroda are not strongly supported (Figure 3.10), a common issue with phylogenetic studies of microgastrode subfamilies (for a detailed discussion, see Murphy et al., 2008). However, consistent with previous studies (Banks & Whitfield, 2006; Belshaw *et al.*, 1998; Dowton *et al.*, 1998; Dowton *et al.*, 2002; Murphy *et al.*, 2008), Cheloninae is robustly recovered as a basal lineage, sister to a clade consisting of Mendesellinae, Khoikhoiinae, Miracinae, Cardiochilinae, and Microgastrinae.

3.5.2.1.2 Sigalphoda

The robustly recovered sister relationship between Sigalphinae and Agathidinae, called the sigalphoid complex by Belshaw and Quicke (2002), confirms the findings of several recent analyses (Belshaw *et al.*, 1998; Belshaw & Quicke, 2002; Dowton *et al.*, 2002; Pitz *et al.*, 2007; Shi *et al.*, 2005). However, the sister group to this complex has never been robustly recovered and somewhat debated (for a detailed discussion, see Sharkey, 1992). From the concatenated analysis, Sigalphoda is recovered as sister to Microgastroda, with relatively strong support (Figure 3.10). However, this relationship is only recovered in one of the six individual gene analyses (Figure 3.7). There is some morphological evidence to suggest a close affinity between Sigalphoda and Ichneutinae. Sharkey and Wharton (1994) hypothesized that the Agathidinae + Sigalphinae (including *Pselephanus*) were sister to Ichneutinae. However, Ichneutinae has more recently been placed as sister to Microgastropia (Belshaw et al., 1998; Dowton et al., 2002; Murphy et al., 2008; Pitz et al., 2007; Shi et al., 2005), and this relationship was also recovered here with variable support (Figure 3.10).

As an alternative hypothesis, Meteorideinae is recovered as the sister group to Sigalphoda in the CAD (apmod) gene tree (Figure 3.8), a relationship that has stronger support when the third position is removed (Figure B4, Appendix B). A close relationship between Sigalphoda and Meteorideinae has been recovered in some morphological analyses (Belshaw *et al.*, 2003; Quicke & van Achterberg, 1990b; Wharton *et al.*, 1992) and molecular analyses (Belshaw & Quicke, 2002), but not consistently. Thus, the sister group relationship between Sigalphoda and Microgastroda remains tentative.

3.5.2.1.3 Euphoroda

Euphoroda , formerly called the Euphoroid complex by Belshaw and Quicke(Belshaw & Quicke, 2002), is robustly supported in the concatenated and several individual gene analyses. Based on the analyses performed herein, the following subfamilies are contained within Euphoroda: Cenocoeliinae, Ecnomiinae, Neoneurinae, Euphorinae, and Meteorinae (Figure 3.12 and Table 3.4). Although there is only one member of Cenocoeliinae included in the dataset, it was convincingly recovered as the sister group to all remaining subfamilies of Euphoroda (Figure 3.10). The relationships among the remaining subfamilies are less clear, in part due to limited taxonomic sampling, particularly for rare subfamilies such as the Neoneurinae and Ecnomiinae.

Although the unusual genera *Planitorus* and *Mannokeraia* have previously been considered as part of the Betylobraconinae and Masoninae, respectively (van Achterberg, 1995), a recent analysis suggested that *Mannokeraia* was more closely related to the Euphorinae (Belshaw & Quicke, 2002) The genus *Planitorus* is also suspected of having a close relationship to Euphorinae (Quicke, *personal communication*). This finding is confirmed here, as both taxa are consistently recovered together within Euphoroda (Figure 3.10). The limits of Euphorinae and Meteorinae are not clear from these analyses, and certainly will require greater taxonomic sampling to resolve the question of monophyly of these two subfamilies.

3.5.2.1.4 Helconoda

The 10 remaining subfamilies within Orthostoma are recovered as monophyletic with strong support, including: Acampsohelconinae, Blacinae, Helconinae, Amicrocentrinae, Charmontinae, Homolobinae, Macrocentrinae, Microtypinae, Orgilinae, and Xiphozelinae (Figure 3.10). There is also strong evidence for at least two separate clades within this larger group, one containing Helconinae and Blacinae, and the other containing the latter 7 subfamilies. These are divided into minfamilies and newly termed Helconopia and Macrocentropia, respectively (Figure 3.1.2 and Table 3.4).

Based on the individual and concatenated analyses, the phylogenetic placement of Acampsohelconinae remains unclear. There is some evidence supporting *Urosigalphus* (the only member of Acampsohelconinae analyzed) as separate from Helconinae or Blacinae, as proposed by van Achterberg (van Achterberg, 2002). However, the placement of Acampsohelconinae as a basal lineage within Helconoda is weakly supported in the concatenated analysis (Figure 3.10). Additionally, the placement of Acampsohelconinae is contradictory or unresolved across the

individual gene trees, recovered as sister to Meteorideinae (Figure B1 and B2, Appendix B), Helconini (Figure 3.9), Sigalphinae + Agathidinae (Figure 3.5), or the traditional helconoid complex (Figure 3.6). The relatively long branch of the acampsohelconine clade and low taxonomic sampling may be affecting the placement of these taxa (Figure 3.10). Thus, the phylogenetic position of Acampsohelconinae is left as unplaced within Helconoda (Table 3.4), until further evidence can be ascertained.

As discussed previously, Meteorideinae is also recovered as a separate lineage from Helconoda, as the sister group to all other members of Orthostoma. However, the placement of this subfamily is unresolved in 5 out of the 6 individual gene analyses. Inference of the CAD (apmod) dataset recovers Meteorideinae as sister to Agathidinae with relatively strong support (Figure 3.8 and Figure B3, Appendix B), but this is not supported in the concatenated analysis (Figure 3.10).

3.5.2.1.4.1 Helconopia

Two subfamilies (Helconinae and Blacinae, as previously defined) are placed within the newly formed minfamily Helconopia. The limits of these two subfamilies have never been well-defined, with genera from each group variably included within the two different subfamilies (Martin, 1956; van Achterberg, 1988), and with Blacinae often considered as a tribe of Helconinae (Sharkey, 1993; van Achterberg, 1975). In this study, members of Blacinae are consistently recovered in clades with members of the helconine tribes Diospilini and Brulleiini. Additionally, none of these 3 groups are recovered as monophyletic in any of the individual gene trees or the concatenated analysis. Thus, the current definitions of Blacinae, Diospilini and Brulleiini need to be tested with further phylogenetic analyses and deeper taxonomic sampling.

There is strong evidence in the concatenated analysis and across most of the individual gene trees for a monophyletic Helconini. Thus, Helconini should be elevated to the rank of subfamily. Brachistini is also consistently and robustly recovered as monophyletic. However, Brachistini is typically recovered as a derived lineage from both diospiline and blacine members. To rectify the classification of the remaining three tribes of Helconinae (Diospilini, Brulleiini, and Brachistini) and the Blacinae, it is proposed that Brachistini also be elevated to subfamily status and contain the following four tribes: Brachistini, Diospilini, Blacini, and Brulleiini (Table 3.4). With the current understanding of these groups, and based on this study, only Brachistini is monophyletic. Determination of the limits of the remaining three tribes will require further

phylogenetic testing with a greater sampling of diospiline, brulleiine, and blacine taxa. Thus, the newly formed Helconopia contains two subfamilies: Helconinae (formerly Helconini) and Brachistinae, all members of which are parasitoids of Coleoptera.

3.5.2.1.4.2. Macrocentropia

Closely related to Helconopia is a separate lineage of parasitoids that utilize lepidopteran hosts. This lineage, newly termed Macrocentropia, is strongly supported in the concatenated analysis (Figure 3.10) and across several gene trees. Macrocentropia includes the following subfamilies: Orgilinae, Homolobinae, Microtypinae, Macrocentrinae, Charmontinae, Amicrocentrinae, and Xiphozelinae (Figure 3.12 and Table 3.4). The latter 4 subfamilies are recovered in a strongly supported clade in the concatenated analysis (Figure 3.10). Additionally, 4 of the 6 individual gene analyses demonstrate some support for this lineage. Amicrocentrinae is robustly recovered as the sister group to Charmontinae, and Xiphozelinae is recovered as the sister group to the Macrocentrinae. Charmontinae, which has variably been placed within Homolobinae (Van Achterberg, 1979), Macrocentrinae(Čapek, 1970), Orgilinae (Čapek, 1973) or as its own subfamily (Quicke & van Achterberg, 1990b), is never recovered as sister to Homolobinae or Orgilinae. Rather, there is strong evidence placing Charmontinae closer to Macrocentrinae and related subfamilies, as has been suggested by van Achterberg and Quicke (1992) based on ovipositor morphology and Čapek (1970) based on larval cephalic structures. For simplicity, it may be prudent in the future to demote these four subfamilies (Amicrocentrinae, Charmontinae, Macrocentrinae, and Xiphozelinae) to tribes contained within Macrocentrinae, if future morphological evidence warrants this classification.

The relationship Orgilinae (Homolobinae + Microtypinae) is also robustly supported across a number of individual gene analyses and the concatenated analysis (Figure 3.10). This sister relationship has also been proposed by a number of authors based on larval and adult morphology and biology (Čapek, 1970; van Achterberg, 1984; van Achterberg, 1992). Again, future morphological studies may demonstrate that these subfamilies could be demoted to tribes within the subfamily Orgilinae.

3.5.2.2 Cyclostoma

Cyclostoma, previously referred to as cyclostomes s.l., includes two sister hypofamilies: Aphidioda and Braconoda (Figure 3.12 and Table 3.4). Aphidioda is robustly recovered as monophyletic and sister to Braconoda (previously referred to as cyclostomes s.s.) (Figure 3.10),

confirming the findings of other recent analyses (Belshaw et al., 2000; Dowton et al., 2002; Zaldivar-Riverón et al., 2006). The phylogenetic placement of Apozyginae (not analyzed here) remains uncertain, although the retention of the second recurrent vein suggests a basal phylogenetic position within Braconidae (Sharkey & Wahl, 1992). However, future studies may indicate that Apozyginae belongs within Cyclostoma, as members possess the hypoclypeal depression and share many similarities with members of Doryctinae (Sharkey, 1993). Thus, Apozyginae is left as unplaced within Braconidae (Table 3.4).

3.5.2.2 .1 Aphidioda

Based on the analyses preformed herein, Aphidioda is established to include Aphidiinae, Mesostoinae, and Maxfischeriinae. This lineage is recovered in 5 of the 6 individual gene datasets and in the concatenated analysis (Figure 3.10). The monotypic genus, *Maxfischeria*, can be firmly elevated to the rank of subfamily and excluded as a tribe of Helconinae, where it was originally placed (Papp, 1994). The relationships within Aphidioda are less clear. Mesostoinae is recovered as the sister group to Maxfischeriinae in the 5 gene trees that recovered Aphidioda. However, Aphidiinae is recovered as the sister group to Maxfischeriinae in the analysis that included both fragments of 28S rDNA (Figures B1 and B2, Appendix B), and in the concatenated analysis (Figure 3.10). The aphidiine, *Pseudopraon* sp., demonstrates volatility in its placement across the individual gene trees, and may have contributed to the differential branching order recovered across the individual and concatenated analyses.

3.5.2.2.2 Braconoda

Braconoda (cyclostomes s.s.) is recovered as monophyletic in only 2 of the 6 individual gene trees (Figures 2.4 and 2.7). However, this clade is strongly supported in the concatenated analysis (Figure 3.10) and in the combined 28S rDNA datasets (Figure B1 and B2, Appendix B). These results confirm the findings of several previous analyses (Belshaw et al., 2000; Dowton et al., 2002; Zaldivar-Riverón et al., 2006) and further suggest that Braconoda forms a natural group. Nonetheless, most of the relationships among the subfamilies of Braconoda lack convincing nodal support. These results are likely due to the limited taxon sampling among the cyclostome subfamilies (only 32 of the 135 braconid taxa were members of Braconoda), particularly for the protein coding genes (Table 3.1).

Exceptions include a clade consisting of Gnamptodontinae, Exothecinae, Opiinae, and Alysiinae, which is robustly recovered in the concatenated analysis (Figure 3.10) and in analyses

of 28S rDNA (Figures 2.4, and B1 and B2, Appendix B). This confirms the findings of several previous analyses (Belshaw *et al.*, 2000; Belshaw *et al.*, 1998; Dowton *et al.*, 2002; Zaldivar-Riverón *et al.*, 2006). This clade is designated the rank of minfamily and is termed Alysiopia (Table 3.4). Contrary to previous studies (Quicke, 1993; Whitfield, 1992b), there is additional evidence suggesting a close relationship between Alysiinae, Opiinae, and Exothecinae to the exclusion of Gnamptodontinae, which are recovered together in several individual gene analyses (Figures 2.4, 2.5, and 2.9, and Figures B1-B3, and B5, Appendix B) and the concatenated analysis (Figure 3.10). However, the branching order among these three subfamilies varies across the different analyses, with Exothecinae often recovered as paraphyletic. Future morphological studies may warrant demoting these three subfamilies as tribes within Alysiinae, and possibly reclassifying some members of Exothecinae. However, the current composition of Alysiopia includes Gnamptodontinae, Alysiinae, Exothecinae, and Opiinae (Table 3.4), with Gnamptodontinae as the likely sister group to the latter three subfamilies (Figure 3.12).

Based on this research and previous studies (Belshaw *et al.*, 2000; Belshaw *et al.*, 1998; Dowton *et al.*, 2002; Zaldivar-Riverón *et al.*, 2006), Braconinae is the likely sister group to Alysiopia. However, Braconinae is not included within Alysiopia as there are several morphological features uniting Alysiopia to the exclusion of Braconinae (see Quicke, 1993 and Whitfield, 1992b for a detailed discussion of these features). Rather, Braconinae might be elevated to minfamily and included with Telengaiinae and Vaepellinae, as has been suggested by some previous studies (Quicke & van Achterberg, 1990a; Tobias, 1988; Wharton, 2000; Wharton *et al.*, 1992; Zaldivar-Riverón *et al.*, 2006). However, here it is left unplaced within Braconoda until further evidence suggests otherwise.

All other subfamilies within Braconoda are left as unplaced (Table 3.4) due to a lack of evidence across the multiple genes utilized here and previous studies suggesting consistent phylogenetic relationships. A recent study of cyclostome relationships based on 28S rDNA, morphological data, and comprehensive taxonomic sampling recovered Rhyssalinae as the ancestral lineage of Braconoda (Zaldivar-Riverón et al., 2006). Additionally, this basal placement of Rhyssalinae was indicated using 16S rDNA (Dowton et al., 1998) and was robustly recovered using a combination of 16S and 28S rDNA gene fragments (Belshaw et al., 2000). Similarly, the combined 28S rDNA analyses performed here also recover Rhyssalinae as the ancestral braconode lineage (Figures B1 and B2, Appendix B). However, this relationship is not

recovered in any other gene trees or in the concatenated analysis. Rather, Rhyssalinae is recovered as a derived group from doryctine ancestors (Figure 3.10), but without convincing support. Thus, the placement of this subfamily needs to be confirmed with additional genetic and morphological evidence.

Rogadinae, excluding *Polystenidea* sp., is recovered with strong support. These results are somewhat consistent with previous analyses (Zaldivar-Riverón et al., 2006; Zaldivar-Riverón et al., 2008), that have found weak support for a monophyletic Rogadinae and variable placement of the Stiropiini (which includes *Polystenidea*, the only representative of the tribe in this analysis).

Monophyly of Doryctinae has rarely been recovered in molecular analyses, and this study is no exception. Several morphological synapomorphies have been identified for Doryctinae (Belokoblylskij *et al.*, 2004; Quicke & van Achterberg, 1990b) and thus, the inclusion of morphological data into phylogenetic analyses has typically recovered this group as monophyletic (Dowton et al., 2002; Zaldivar-Riverón et al., 2006). An in-depth phylogenetic examination of Doryctinae, using both morphological and molecular data remains a fertile area of research.

3.5.3. Evolution of life history among the Braconidae

The notion of ectoparasitism as the ancestral ground plan within the Braconidae has been a long-standing assumption in braconid scholarship (Shaw & Huddleston, 1991), but has never been critically tested. The assumption of ectoparasitism leading to endoparasitism is somewhat intuitive in that endoparasitoids, particularly koinobionts, had to evolve a variety of complex mechanisms to exploit their host's immune system and prevent encapsulation of the egg. Additionally, Dollo's law has dominated ideas on the evolution of traits for the last century, suggesting that once a complex trait has been lost, it is unlikely to be regained (Dollo, 1893). Thus, evolutionary reversals to elaborate traits are considered highly improbable (but see Marshall et al., 1994). Evolutionary transitions in Hymenoptera, presumably from an endophytic lifestyle to an ectoparasitic lifestyle, have been well argued by a number of authors (Gauld, 1988; Gauld & Bolton, 1988; Handlirsch, 1908; Königsmann, 1976; Rasnitsyn, 1980a; for a detailed review, see Whitfield *et al.*, 2003). However, the sister group to the Ichneumonoidea remains unknown, and thus, an ectoparasitic ground plan cannot be an automatic assumption for the Ichneumonoidea.

The ancestral state reconstructions of Braconidae in this study raise the possibility that the ancestral condition was endoparasitic. At minimum, it suggests that researchers of Braconidae cannot assume ectoparasitic origins without rigorous testing of that hypothesis. Only when outgroups are included and coded as ectoparasitic is an ectoparasitic ancestral condition favored, although the result was highly ambiguous. Alternatively, an endoparasitic outgroup resulted in a significantly endoparasitic reconstruction for the ancestral node of Braconidae. Thus, the determination of the sister-group to the Ichneumonoidea is a necessary first step to rigorously test theories on the biological transitions of parasitism within Braconidae. Unfortunately, the phylogeny and evolutionary transitions among the Hymenoptera have been largely contradictory, and the sister group to the Ichneumonoidea remains in question (Sharkey, 2007; Whitfield et al., 2003).

One of the salient results of this study is the importance of understanding the phylogeny relationships among Cyclostoma. Endoparasitism is reconstructed significantly when the branching order of Braconoda (cyclostomes s.s) is not considered (i.e. collapsed to a polytomy). Additionally, the ancestral condition is significantly endoparasitic if Maxfischeriinae are coded as endoparasitic (currently their biology is unknown). Thus, a robust phylogeny of these lineages will be required to fully understand braconid evolution and rigorously test long-standing notions on the evolution of modes of parasitism. A recent study with comprehensive taxonomic sampling presented a fairly robust phylogeny of the cyclostome subfamilies (Zaldivar-Riverón et al., 2006) . However, greater genetic sampling is needed to confirm these results, given the incongruity in cyclostome relationships, particularly within Braconoda, across the different gene trees analyzed here. Given that Aphidioda is a distinct lineage sister to Braconoda, understanding the biology and phylogeny of the members of Aphidioda will also be critical for future ancestral state reconstructions.

3.6. Conclusions

Examining evolutionary transitions through phylogeny has become a vital component of evolutionary biology. However, robust phylogenies based on multiple lines of evidence are necessary to understand patterns of evolutionary change through time. The Braconidae provide an excellent system to study evolutionary processes, such as transitional patterns of host utilization, the evolution of host finding mechanisms, phenotypic convergence, and the evolution of parasitic life strategies. These are basic theoretical research avenues but have

fundamental applications to several biological fields, particularly for parasitoid mediated biological control or genetic manipulation of the symbiotic organisms that parasitoids use to manipulate host immune systems. In this study a robust phylogeny of Braconidae was generated using several molecular markers. Several higher level relationships were recovered with significant support across multiple genes, providing independent lines of evidence to support the phylogenetic hypotheses.

Among braconid subfamilies, there was strong evidence supporting an independent clade, termed Aphidioda, as separate and distinct from the remaining members of Cyclostoma. Maxfischeriinae was recovered firmly within Aphidioda, and can be definitively removed from Helconinae. There was also abundant evidence confirming Euphoroda as distinct from other subfamilies traditionally placed within the helconoid complex. A close relationship between the Microgastroda and Sigalphoda was recovered. Helconoda was recovered with two distinct lineages, one endoparasitic on coleopteran larvae (Helconopia) and the other on lepidopteran larvae (Macrocentropia). Helconinae was never recovered as monophyletic with respect to Blacinae. Thus, the rank of subfamily for Blacinae needs to be re-examined, as representative taxa were consistently recovered with members of the helconine tribes Diospilini and Brulleiini. Members of Helconini were robustly recovered as monophyletic and distinct from the other helconine tribes, and were elevated to subfamily status. Brachistini was also elevated to subfamily with four recognized tribes, Brachistini, Brulleiini, Diospilini, and Blacini, the latter three of which are paraphyletic. This study focused primarily on the non-cyclostome subfamilies, newly termed Orthostoma, but the monophyly of the Cyclostoma (including a sister relationship between Aphidioda and Braconoda) was strongly supported. However, relationships within Braconoda were poorly supported, primarily due to the weaker taxonomic sampling of exemplars from this group. Several taxonomic changes were proposed based on the robust results of this study as well as consistent results with several previous studies (summarized in Table 3.4 and depicted in Figure 3.12).

Based on the robustly recovered phylogeny, the hypothesis of an ectoparasitic ancestor for Braconidae was tested using ancestral state reconstruction. Ectoparasitism was significantly recovered as the ancestral condition for the Braconoda (cyclostomes s.s.). The analyses suggested an endoparasitic ancestor for Braconidae, Cyclostoma, and Orthostoma, but reconstructions were only statistically significant for the latter taxon. These results confirm the

suspicions espoused by Wharton (Wharton, 1993) and alluded to by others (Shaw & Huddleston, 1991; Tobias, 1967), that the Orthostoma had a separate evolutionary history with respect to Cyclostoma, and may not have originated from an ectoparasitic ancestor. Although the reconstruction of the ancestral node of Braconidae was not statistically significant, if some of the taxa within Aphidioda (e.g. Maxfischeriinae) are discovered to be endoparasitic, then an endoparasitic ancestor for the Braconidae is highly probable. This study highlights the need to intensively investigate the natural history of braconid parasitoids to further our understanding of the evolutionary pathways within the system. Additionally, the sister group to the Ichneumonoidea is still desperately needed for a full understanding of braconid evolution.

Copyright © Barbara J. Sharanowski 2009

CHAPTER 4: OVERALL SUMMARY AND FUTURE DIRECTIONS

From Chapter 2, it is evident that ESTs have potential to resolve relatively deep divergences among insect lineages, not only within Hymenoptera, but likely across the entirety of Holometabola. However, the scale of phylogenomic datasets for studies of any given lineage will need to be larger, encompassing a greater sample of the genetic and taxonomic diversity. Increasing the number of exemplars in a dataset can have a counter-effect on systematic error and vastly improve phylogenetic inference (Hedtke et al., 2006). Additionally, increased genetic sampling will help to increase the number of transcripts common to all taxa and thereby decrease the amount of missing data.

As seen in the study outlined in Chapter 2, gene tree phylogenies do not necessarily match species tree phylogenies. Although non-phylogenetic events can be a cause of the incongruity between gene trees and species trees, more commonly it is due to biases in the characteristics of the data that are not accurately accounted for in the model employed (Sullivan & Joyce, 2005; Whelan *et al.*, 2001). However, new models are constantly being developed (Galtier & Gouy, 1998; Tuffley & Steel, 1998; Wang *et al.*, 2007; Whelan, 2008), and although most are still too computationally intensive to be useful for most practical datasets, this is likely to change rapidly in the near future. Additionally, increased genetic sampling with limited missing data will allow for the use of newer methods for analyzing discordance among gene trees (Ané et al., 2007; Larget, 2006).

The advent of new sequencing technology (e.g. pyrosequencing) is making large-scale phylogenomics economically feasible for most laboratories (Hudson, 2008). Thus, an obvious next-step for investigation on higher-level Hymenopteran relationships is to sequence a larger number of taxa for a greater number of transcripts using next-generation sequencing technology. Additionally, as a two-pronged approach, primers can be developed from the EST alignments, for amplification of rare taxa that have already been extracted for genomic DNA, thus further increasing the breadth of exemplars that can be examined. This kind of future study can hopefully confirm or dispute the results of this study, leading to a better understanding of hymenopteran evolution.

A robust phylogeny of Hymenoptera impacts evolutionary studies on social behavior, symbiotic associations, and morphological convergence. Additionally, understanding higher-level hymenopteran relationships has an impact on the evolutionary transitions of lower level

relationships, such as the study outlined in Chapter 3. Clearly the ancestral reconstruction performed in Chapter 3 was limited by the lack of certainty of the sister group to the Ichneumonoidea. The character state of the outgroups had an immense effect on the recovered pattern of transition between different modes of parasitism. These reconstructions were also limited by the lack of knowledge on the biology of many lineages within Braconidae. Thus, it is strongly recommended that future research entail natural history investigations within Braconidae and Ichneumonoidea as a whole, thus, leading to a better understanding of the evolutionary processes that have occurred. However, from Chapter 3, a clearer picture of the phylogeny of the Braconidae emerged, particularly for early branching lineages. This robust phylogeny will allow for the reclassification of some taxa into monophyletic groups and hopefully provide a scaffold for testing additional evolutionary theories as more biological information becomes available.

Copyright © Barbara J. Sharanowski 2009

Appendix A

Table A1. Nucleotide composition for all taxa and codon positions.

	All positions					1st position				2nd position				3rd position			
Taxon (short name)	T	C	A	G	Total	T-1	C-1	A-1	G-1	T-2	C-2	A-2	G-2	T-3	C-3	A-3	G-3
Orthoptera (Lm)	24.1	19.4	31.4	25.1	10281	16.3	19.4	32.7	31.5	26.2	19.6	35.3	18.9	29.7	19.1	26.1	25
Hemiptera (Ap)	24.9	19.5	32.2	23.4	10715	18.9	16.2	33.8	31.2	26.7	18.9	36	18.5	29.2	23.3	26.8	20.7
Hemiptera (Mp)	24.8	19.5	32	23.7	10484	18.8	16.4	33.7	31.1	26.6	18.8	36	18.7	29.2	23.4	26.2	21.2
Lepidoptera (Bm)	22.2	23.2	29.5	25.1	10749	15.9	19.7	32.2	32.1	26.6	20.4	35	18	24.2	29.5	21.2	25.2
Diptera (Dm)	18.7	28.9	24.2	28.2	10740	14.7	22.9	30.2	32.2	26.3	19.9	34.9	18.9	15.1	43.9	7.4	33.6
Coleoptera (Tc)	22	22.9	28.8	26.2	10779	17.5	18.3	32.3	31.9	26.2	19.3	35.2	19.3	22.3	31.2	19	27.5
Apoidea (Am)	28.3	15.9	34.6	21.2	10803	18.1	18	33.1	30.8	26.4	19.4	35.5	18.7	40.5	10.3	35.2	14.1
Aulacidae (Ps)	22	22.8	28.5	26.7	5273	15.8	19.9	31.1	33.2	26.7	18.1	35.5	19.7	23.4	30.4	19	27.2
Ichneumonidae (Cs)	21.9	23.5	29.8	24.9	8484	15.8	20.3	32.5	31.4	25.5	19.8	35.6	19.1	24.3	30.3	21.2	24.2
Ceraphronoidea (Ce)	23.3	22.2	29.6	24.9	3688	18	18.2	33.6	30.1	27.4	18.1	34.9	19.6	24.4	30.2	20.3	25.1
Cynipoidea (Fe)	24.4	19.4	33.4	22.9	5231	16.6	18.6	35.2	29.5	27.2	16.5	36.4	19.9	29.3	23.1	28.5	19.1
Braconidae (Lt)	28.2	17.4	34.4	19.9	8638	17.4	18.4	32.7	31.5	26.8	19.8	34.6	18.8	40.5	14.1	36	9.4
Chalcidoidea (Nv)	23.1	22.2	29.6	25.1	10797	16.1	19.2	33.5	31.2	26.5	19.1	35.7	18.8	26.6	28.3	19.8	25.3
Symphyta (Ns)	24.4	23	27.6	25	5350	16.7	20.3	29.9	33.2	27.6	20.8	33.1	18.5	28.8	28	19.8	23.4
Proctotrupoidea (Pp)	23.3	20.2	32.8	23.7	5551	15.7	20.8	34.1	29.4	24.9	18.4	36.8	19.8	29.3	21.5	27.4	21.8
Vespoidea (Si)	24	20.7	30.6	24.7	9749	16.6	19.8	33	30.7	26.8	19.2	35	19	28.6	23.1	23.9	24.4
Average	23.7	21.2	30.6	24.4	8160	16.8	19.1	32.7	31.4	26.5	19.3	35.3	18.9	27.9	25.4	23.7	23

Table A2. Chi-square tests for base composition heterogeneity

Gene#	Characters included	χ^2	*df*	p-value	homogenous/ heterogeneous
CG1746	all	59.907	36	0.007	heterogeneous
CG1746	3 out	19.451	36	0.989	homogenous
CG2099	all	89.092	39	0.000	heterogeneous
CG2099	3 out	25.940	39	0.946	homogenous
CG2746	all	135.036	36	0.000	heterogeneous
CG2746	3 out	9.588	36	1.000	homogenous
CG3186	all	72.691	36	0.000	heterogeneous
CG3186	3 out	7.244	36	1.000	homogenous
CG3446	all	139.691	39	0.000	heterogeneous
CG3446	3 out	23.720	39	0.974	homogenous
CG3661	all	55.109	39	0.045	heterogeneous
CG3661	3 out	4.509	39	1.000	homogenous
CG3997	all	23.832	39	0.973	homogenous
CG3997	3 out	6.817	39	1.000	homogenous
CG4097	all	97.402	33	0.000	heterogeneous
CG4097	3 out	9.873	33	1.000	homogenous
CG4169	all	258.462	36	0.000	heterogeneous
CG4169	3 out	48.490	36	0.080	homogenous
CG4169	1st position	51.834	36	0.042	heterogeneous
CG4800	all	117.486	42	0.000	heterogeneous
CG4800	3 out	10.393	42	1.000	homogenous
CG6770	all	53.762	39	0.058	homogenous
CG6770	3 out	14.037	39	1.000	homogenous
CG6779	all	153.717	42	0.000	heterogeneous
CG6779	3 out	11.021	42	1.000	homogenous
CG6803	all	73.658	39	0.001	heterogeneous
CG6803	3 out	16.873	39	0.999	homogenous
CG7178	all	34.596	33	0.392	homogenous
CG7178	3 out	13.422	33	0.999	homogenous
CG7424	all	60.482	39	0.015	heterogeneous
CG7424	3 out	4.315	39	1.000	homogenous
CG7434	all	89.089	36	0.000	heterogeneous
CG7434	3 out	8.638	36	1.000	homogenous
CG7939	all	89.363	36	0.000	heterogeneous
CG7939	3 out	10.145	36	1.000	homogenous
CG8332	all	88.114	36	0.000	heterogeneous
CG8332	3 out	7.324	36	1.000	homogenous
CG8415	all	82.135	39	0.000	heterogeneous
CG8415	3 out	4.192	39	1.000	homogenous
CG8857	all	73.297	36	0.000	heterogeneous
CG8857	3 out	8.795	36	1.000	homogenous
CG8900	all	81.985	36	0.000	heterogeneous
CG8900	3 out	10.530	36	1.000	homogenous
CG11271	all	101.298	36	0.000	heterogeneous
CG11271	3 out	20.697	36	0.981	homogenous
CG11981	all	187.413	33	0.000	heterogeneous
CG11981	3 out	17.328	33	0.989	homogenous
CG15442	all	80.035	42	0.000	heterogeneous
CG15442	3 out	7.763	42	1.000	homogenous
concatenated	all	1530.186	45	0.000	heterogeneous
concatenated	3-out	105.997	45	0.000	heterogeneous
concatenated	1st position	142.696	45	0.000	heterogeneous
concatenated	2nd position	33.335	45	0.900	homogenous
concatenated	RY-3	212.749	45	0.000	heterogeneous
concatenated	RY-all	71.883	45	0.007	heterogeneous
concatenated	3-out*	47.401	36	0.0968	homogenous

*With the dipteran and both hemipterans excluded

Tables A3-A6 depicts the pattern of nucleotide substitution for various data partitions (Tamura et al. 2004). Each entry shows the probability of substitution from one base (row) to another base (column) instantaneously. Only entries within a row should be compared. Rates of transitional substitutions are shown in bold and transversional substitutions in italics. Rate variation among sites was modeled with a gamma distribution, with an estimated shape parameter (see note below each table). The overall transition/transversion bias (R) were calculated using the following formula [A*G*k_1 + T*C*k_2]/ [(A+G)*(T+C)]. Differences in the composition bias among sequences were considered in evolutionary comparisons. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (pairwise deletion option). All calculations were conducted in MEGA 4 (Tamura et al. 2007).

Table A3. Maximum composite likelihood estimate of the pattern of nucleotide substitution for the first codon position.

	A	**T**	**C**	**G**
A	-	*4.08*	*4.74*	**5.94**
T	*8.03*	-	**20.9**	*7.6*
C	*8.03*	**18.01**	-	*7.6*
G	**6.27**	*4.08*	*4.74*	-

NOTE: The nucleotide frequencies were A = 0.328, T = 0.167, C = 0.194, G = 0.311. The transition/transversion rate ratios were k_1 = 0.781 (purines) and k_2 = 4.413 (pyrimidines). The overall transition/transversion bias was R = 0.945. The gamma distribution shape parameter was 0.5. There were a total of 3639 positions in the final dataset.

Table A4. Maximum composite likelihood estimate of the pattern of nucleotide substitution for the second codon position.

	A	T	C	G
A	-	*7.78*	*5.51*	**8.22**
T	*10.37*	-	**7.47**	*5.59*
C	*10.37*	**10.56**	-	*5.59*
G	**15.24**	*7.78*	*5.51*	-

NOTE: The nucleotide frequencies were A = 0.355, T = 0.266, C = 0.188, G = 0.191. The transition/transversion rate ratios were k_1 = 1.469 (purines) and k_2 = 1.357 (pyrimidines). The overall transition/transversion bias was R = 0.493. The gamma distribution shape parameter was 0.5. There were a total of 3639 positions in the final dataset.

Table A5. Maximum Composite Likelihood Estimate of the Pattern of Nucleotide Substitution Third Position.

	A	T	C	G
A	-	*4.32*	*4.01*	**14.48**
T	*3.65*	-	**19.07**	*3.57*
C	*3.65*	**20.53**	-	*3.57*
G	**14.81**	*4.32*	*4.01*	-

NOTE: The nucleotide frequencies were A = 0.235, T = 0.278, C = 0.258, G = 0.23. The transition/transversion rate ratios were k_1 = 4.053 (purines) and k_2 = 4.757 (pyrimidines). The overall transition/transversion bias was R = 2.254. The gamma distribution shape parameter was set to 100,000,000, but was estimated as infinity in Paup* (Swofford, 2000). There were a total of 3639 positions in the final dataset.

Table A6. Maximum Composite Likelihood Estimate of the Pattern of Nucleotide Substitution for First and Second Positions Together.

	A	T	C	G
A	-	*5.96*	*5.26*	**7.39**
T	*9.4*	-	**12.89**	*6.91*
C	*9.4*	**14.61**	-	*6.91*
G	**10.05**	*5.96*	*5.26*	-

NOTE: The nucleotide frequencies were A = 0.341, T = 0.216, C = 0.191, G = 0.251. The transition/transversion rate ratios were k_1 = 1.069 (purines) and k_2 = 2.451 (pyrimidines). The overall transition/transversion bias was R = 0.655. The gamma distribution shape parameter was 0.5. There were a total of 7278 positions in the final dataset.

Tables 8-10 depict the results of the test for the homogeneity of substitution patterns between sequences. This test is similar to the chi-square test for base composition homogeneity, but is a more powerful test (Kumar & Gadagkar, 2001b). In addition, the test calculates pairwise comparisons for all taxa, and thus, is more informative. The test calculates the probability of rejecting the null hypothesis that sequences have evolved with the same pattern of substitution, as judged from the extent of differences in base composition biases between sequences (Kumar & Gadagkar, 2001b). A Monte Carlo test (1000 replicates was used to estimate the p-values (significance < 0.05). All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). All analyses were conducted in MEGA 4 (Tamura et al., 2007).

Table A7. Test of the homogeneity of substitution patterns between sequences for the first position for all 24 genes together. Grey colored cells are significant for the pairwise comparison.

	Lm	Ap	Mp	Bm	Dm	Tc	Am	Ps	Cs	Ce	Fe	Lt	Nv	Pp	Si
Orthoptera (Lm)															
Hemiptera (Ap)	0.00														
Hemiptera (Mp)	0.00	1.00													
Lepidoptera (Bm)	1.00	0.00	0.00												
Diptera (Dm)	0.00	0.00	0.00	0.00											
Coleoptera (Tc)	0.05	0.00	0.00	0.01	0.00										
Apoidea (Am)	0.01	0.03	0.02	0.00	0.00	0.09									
Evanioidea (Ps)	1.00	0.00	0.00	0.36	0.01	0.01	0.00								
Ichneumonidae (Cs)	0.07	0.00	0.00	1.00	0.00	0.00	0.00	1.00							
Ceraphronoidea (Ce)	1.00	0.01	0.00	0.25	0.00	1.00	0.07	0.05	0.10						
Cynipoidea (Fe)	0.23	0.05	0.06	0.01	0.00	1.00	1.00	0.02	0.01	1.00					
Braconidae (Lt)	0.31	0.01	0.02	0.00	0.00	1.00	1.00	0.03	0.00	1.00	0.31				
Chalcidoidea (Nv)	0.12	0.00	0.00	0.08	0.00	0.00	0.00	0.12	0.03	1.00	0.03	0.00			
Symphyta (Ns)	1.00	0.00	0.00	1.00	0.32	0.03	0.00	0.37	1.00	1.00	0.00	0.00	0.10		
Proctotrupoidea (Pp)	0.30	0.02	0.04	0.32	0.00	0.34	1.00	0.09	0.00	0.18	1.00	0.32	0.04	0.07	
Vespoidea (Si)	0.39	0.00	0.00	0.22	0.00	0.03	0.00	0.22	0.07	0.16	0.26	0.02	1.00	0.08	1.00

Table A8. Test of the homogeneity of substitution patterns between sequences for the second position for all 24 genes together. Grey colored cells are significant for the pairwise comparison.

	Lm	Ap	Mp	Bm	Dm	Tc	Am	Ps	Cs	Ce	Fe	Lt	Nv	Pp	Si
Orthoptera (Lm)															
Hemiptera (Ap)	0.17														
Hemiptera (Mp)	0.11	1.00													
Lepidoptera (Bm)	0.10	0.02	0.05												
Diptera (Dm)	1.00	0.07	0.29	0.13											
Coleoptera (Tc)	1.00	0.15	0.34	0.00	1.00										
Apoidea (Am)	0.28	1.00	1.00	0.07	1.00	1.00									
Evanioidea (Ps)	0.22	0.11	0.12	0.02	1.00	1.00	0.08								
Ichneumonidae (Cs)	1.00	0.08	0.32	0.12	1.00	1.00	1.00	0.11							
Ceraphronoidea (Ce)	0.30	1.00	1.00	0.21	1.00	0.32	0.33	1.00	1.00						
Cynipoidea (Fe)	1.00	1.00	1.00	0.05	1.00	1.00	0.01	1.00	0.10	0.02					
Braconidae (Lt)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.08				
Chalcidoidea (Nv)	0.07	1.00	1.00	0.01	0.19	0.30	1.00	0.22	0.03	1.00	0.02	0.17			
Symphyta (Ns)	0.38	1.00	1.00	0.10	1.00	1.00	1.00	1.00	1.00	1.00	0.01	1.00	1.00		
Proctotrupoidea (Pp)	1.00	0.39	1.00	0.15	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
Vespoidea (Si)	0.20	0.25	1.00	0.04	1.00	1.00	1.00	0.22	1.00	1.00	0.05	1.00	1.00	1.00	1.00

Table A9. Test of the homogeneity of substitution patterns between sequences for the third position for all 24 genes together. Grey colored cells are significant for the pairwise comparison.

	Lm	Ap	Mp	Bm	Dm	Tc	Am	Ps	Cs	Ce	Fe	Lt	Nv	Pp	Si
Orthoptera (Lm)															
Hemiptera (Ap)	0.00														
Hemiptera (Mp)	0.00	1.00													
Lepidoptera (Bm)	0.00	0.00	0.00												
Diptera (Dm)	0.00	0.00	0.00	0.00											
Coleoptera (Tc)	0.00	0.00	0.00	0.00	0.00										
Apoidea (Am)	0.00	0.00	0.00	0.00	0.00	0.00									
Evanioidea (Ps)	0.00	0.00	0.00	0.14	0.00	1.00	0.00								
Ichneumonidae (Cs)	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00							
Ceraphronoidea (Ce)	0.00	0.00	0.00	0.33	0.00	0.01	0.00	0.03	0.17						
Cynipoidea (Fe)	0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00					
Braconidae (Lt)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00				
Chalcidoidea (Nv)	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.03	0.39	0.00	0.00			
Symphyta (Ns)	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.09	0.02	1.00	0.00	0.00	1.00		
Proctotrupoidea (Pp)	0.00	0.10	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
Vespoidea (Si)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01

Copyright © Barbara J. Sharanowski 2009

Appendix B.

Table B1. Chi-square tests for base composition homogeneity for the gene partitions utilized in the analyses in Chapter 3.

Gene region	Positions included	χ^2	*df*	P-value	Base composition
CAD (54-405)	all	1150.516	339	0.000	heterogeneous
CAD (54-405)	1st and 2nd	157.536	339	1.000	homogeneous
CAD (apmod)	all	655.924	324	0.000	heterogeneous
CAD (apmod)	1st and 2nd	47.096	324	1.000	homogeneous
ACC	all	617.967	366	0.000	heterogeneous
ACC	1st and 2nd	53.865	366	1.000	homogeneous
28S (D1D3)	all	670.915	411	0.000	heterogeneous
28S (D1D3)	stems	614.679	411	0.000	heterogeneous
28S (D1D3)	loops	64.872	411	1.000	homogeneous
28S (D1D3)	ambiguous	394.453	411	0.748	homogeneous
28S (D3D5)	all	172.007	402	1.000	homogeneous
28S (D3D5)	stems	100.978	402	1.000	homogeneous
28S (D3D5)	loops	37.200	402	1.000	homogeneous
28S (D3D5)	ambiguous	386.059	402	0.708	homogeneous
18S	all	40.815	411	1.000	homogeneous
18S	stems	28.421	411	1.000	homogeneous
18S	loops	12.470	411	1.000	homogeneous
18S	ambiguous	165.703	411	1.000	homogeneous

Figure B1. Inferred topology from the Bayesian analysis of both 28S gene fragments together with regions of ambiguity included. Posterior probabilities are listed below the node (30 million generations; burnin = 20 million generations.). Asterisk indicates non-monophyly.

Figure B2. Inferred topology from the Bayesian analysis of both 28S gene fragments together with regions of ambiguity excluded. Posterior probabilities are listed below the node (30 million generations; burnin = 20 million generations). Asterisk indicates non-monophyly.

Figure B3. **Inferred topology from the Bayesian analysis of CAD (54-405) with the 3rd position excluded.** Posterior probabilities are listed below the node (4 million generations; burnin = 1.5 million generations). Asterisk indicates non-monophyly.

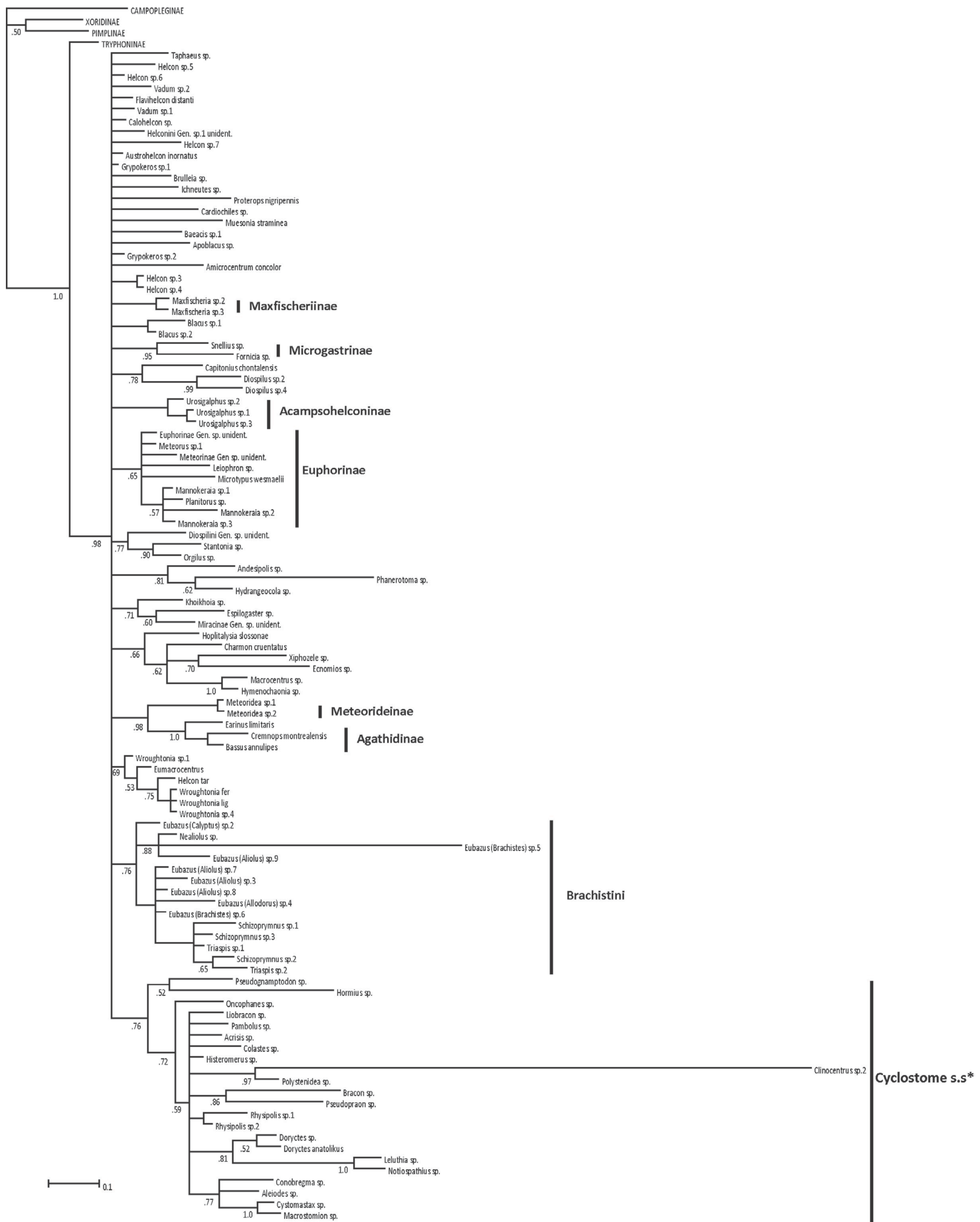


Figure B4. Inferred topology from the Bayesian analysis of CAD (apmod) with the 3rd position excluded. Posterior probabilities are listed below the node (4 million generations; burnin = 1.5 million generations). Asterisk indicates non-monophyly.

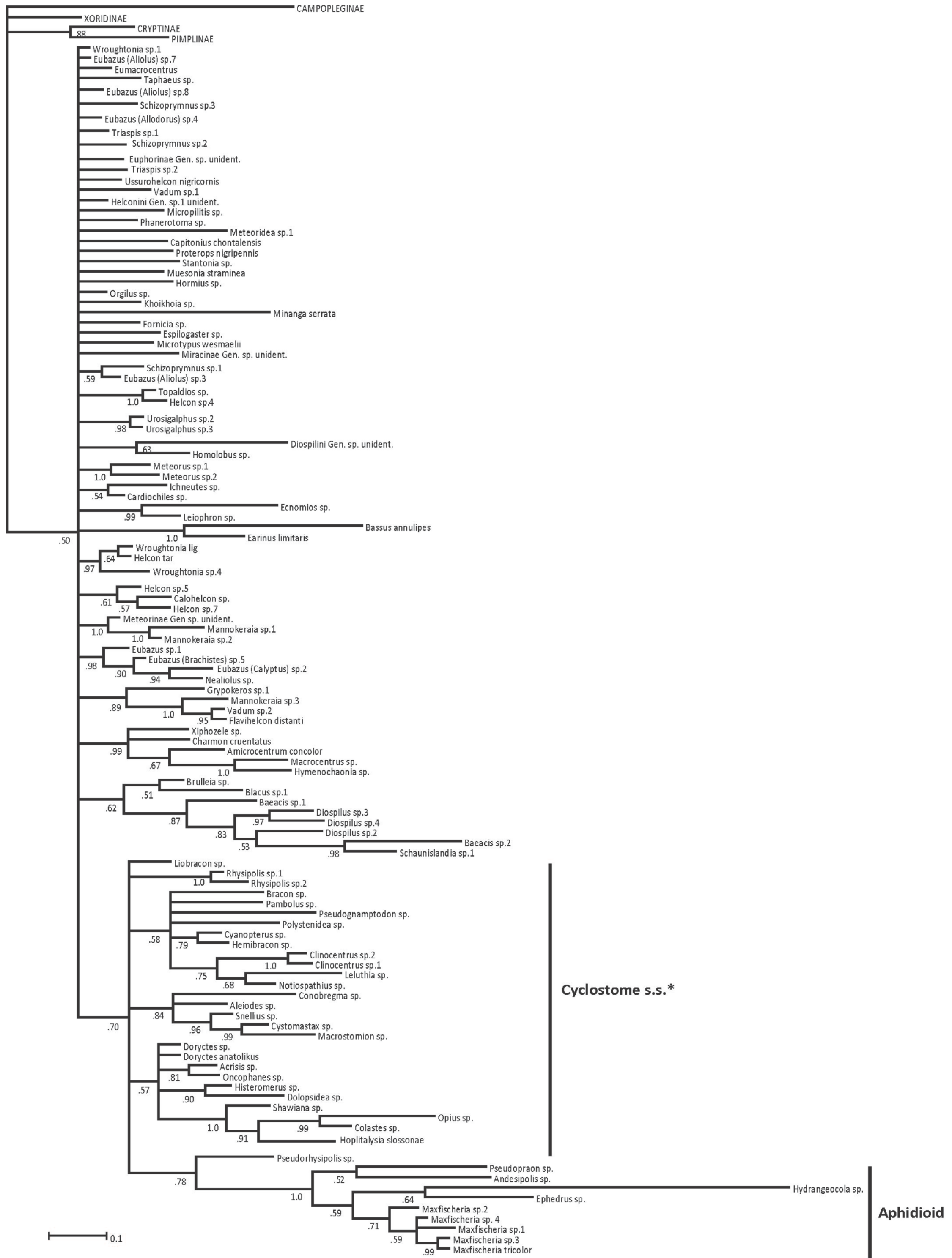


Figure B5. Inferred topology from the Bayesian analysis of ACC with the 3rd position excluded. Posterior probabilities are listed below the node (4 million generations; burnin = 1.5 million generations). Asterisk indicates non-monophyly.

Copyright © Barbara J. Sharanowski 2009

References

Akashi, H., Goel, P. & John, A. (2007). Ancestral Inference and the study of codon bias evolution: Implications for molecular evolutionary analyses of the *Drosophila melanogaster* subgroup. *Public Library of Science One* **2,** e1065.

Alexander, R. D. (1974). The evolution of social behavior. *Annual Review of Ecology and Systematics* **5,** 325-383.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215,** 403-410.

Andersson, F. E.Swofford, D. L. (2004). Should we be worried about long-branch attraction in real data sets? Investigations using metazoan 18S rDNA. *Molecular Phylogenetics and Evolution* **33,** 440-451.

Ané, C., Larget, B., Baum, D. A., Smith, S. D. & Rokas, A. (2007). Bayesian estimation of concordance among gene trees. *Molecular Biology and Evolution* **24,** 412-426.

Austin, A.Dowton, M. (2000). *Hymenoptera: Evolution, Biodiversity, and Biological Control*. CSIRO, Canberra.

Banks, J. C.Whitfield, J. B. (2006). Dissecting the ancient rapid radiation of microgastrine wasp genera using additional nuclear genes. *Molecular Phylogenetics and Evolution* **41,** 690–703.

Barnes, W. M. (1994). PCR amplification of up to 35-kb DNA with high fidelity and high yield from λ bacteriophage templates. *Proceedings of the National Academy of Sciences USA* **91,** 2216-2220.

Baurain, D., Brinkmann, H. & Philippe, H. (2007). Lack of resolution in the animal phylogeny: Closely spaced cladogenesis or undetected systematic errors? *Molecular Biology and Evolution* **24,** 6-9.

Beijing Genomics Institute (200). Silkworm Knowledge Base: SilkDB, Vol. 2008.

Beiko, R.Hamilton, N. (2006). Phylogenetic identification of lateral genetic transfer events. *BMC Evolutionary Biology* **6,** 15.

Belokoblylskij, S. A., Zaldivar-Riverón, A. & Quicke, D. L. J. (2004). Phylogeny of the genera of the parasitic wasps subfamily Doryctinae (Hymenoptera: Braconidae) based on morphological evidence. *Zoological Journal of the Linnean Society* **142,** 369-404.

Belshaw, R., Dowton, M., Quicke, D. & Austin, A. (2000). Estimating ancestral geographical distributions: A gondwanan origin for aphid parasitoids? *Proceedings of the Royal Society of London B* **267,** 491-496.

Belshaw, R., Fitton, M., Herniou, E., Gimeno, C. & Quicke, D. (1998). A phylogenetic reconstruction of the Ichneumonoidea (Hymenoptera) based on the D2 variable region of 28S ribosomal RNA. *Systematic Entomology* **23,** 109-123.

Belshaw, R., Grafen, A. & Quicke, D. (2003). Inferring life history from ovipositor morphology in parasitoid wasps using phylogenetic regression and discriminant analysis. *Zoological Journal of the Linnean Society* **139,** 213-228.

Belshaw, R.Quicke, D. (2002). Robustness of ancestral state estimates: evolution of life history strategy in Ichneumonoid parasitoids. *Systematic Biology* **51,** 450-477.

Bergsten, J. (2005). A review of long-branch attraction. *Cladistics* **21,** 163-193.

Bezier, A., Annaheim, M., Herbiniere, J., Wetterwald, C., Gyapay, G., Bernard-Samain, S*., et al.* (2009). Polydnaviruses of braconid wasps derive from an ancestral nudivirus. *Science* **323,** 926-930.

Blanquart, S.Lartillot, N. (2006). A Bayesian compound stochastic process for modeling nonstationary and nonhomogeneous sequence evolution. *Molecular Biology and Evolution* **23,** 2058-2071.

Borror, D. J. (1960). *Dictionary of Word Roots and Combining Forms*. N-P Publications, Palo Alto, California.

Brèthes (1927). Nouveaux Hyménoptères parasites du Chili. *Revista Chilena de Historia Natural* **31,** 194-200.

Brothers, D. L. (1975). Phylogeny and classification of the aculeate Hymenoptera, with special reference to Mutillidae. *University of Kansas Scientific Bulletin* **50,** 483-648.

Brown, R. W. (1956). *Composition of Scientific Words: A Manual of Methods and a Lexicon of Materials for the Practice of Logotechnics*. Smithsonian Institution Press, Washington.

Bull, J. J. (1981). Coevolution of haplo-diploidy and sex determination in the Hymenoptera. *Evolution* **35,** 568-580.

Calendini, F.Martin, J.-F. (2005). PaupUP v1.0.3.1: A free graphical frontend for Paup* Dos software.

Čapek, M. (1970). A new classification of the Braconidae (Hymenoptera) based on the cephalic structures of the final instar larva and biological evidence. *The Canadian Entomologist* **102,** 846-875.

Čapek, M. (1973). Key to the final instar larvae of the Braconidae (Hymenoptera). *Acta Instituti Forestalis Zvolenensis***,** 259-268.

Carpenter, J. M.Wheeler, W. C. (1999). Towards simultaneous analysis of molecular and morphological data in Hymenoptera. *Zoologica Scripta* **28,** 251-260.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* **17,** 540-552.

Castro, L.Dowton, M. (2006). Molecular analyses of the Apocrita (Insecta : Hymenoptera) suggest that the Chalcidoidea are sister to the diaprioid complex. *Invertebrate Systematics* **20,** 603-614.

Castro, L.Dowton, M. (2007). Mitochondrial genomes in the Hymenoptera and their utility as phylogenetic markers. *Systematic Entomology* **32,** 60-69.

Castro, L. R., Austin, A. D. & Dowton, M. (2002). Contrasting rates of mitochondrial molecular evolution in parasitic Diptera and Hymenoptera. *Molecular Biology and Evolution* **19,** 1100-1113.

Chenchik, A., Zhu, Y. Y., Diatchenko, L., Li, R., Hill, J. & Siebert, P. D. (1998). Generation and use of high-quality cDNA from small amounts of total RNA by SMART PCR. Gene Cloning and Analysis by RT-PCR., pp. 305-319. BioTechniques Books, Westborough.

Chomczynski, P.Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry* **162,** 156-159.

CIPRES Collaborative Group (2005-2008). CIPRES: Cyberinfrastructure for Phylogenetic Research. http://www.phylo.org/. [Accessed: November 2008], Vol. 2008. San Diego Supercomputing Center.

Conant, G. C.Lewis, P. O. (2001). Effects of nucleotide composition bias on the success of the parsimony criterion in phylogenetic inference. *Molecular Biology and Evolution* **18,** 1024-1033.

Danforth, B. N., Fang, J., Sipes, S., Brady, S. G. & Almeida, E. (2004). Phylogeny and molecular systematics of bees (Hymenoptera: Apoidea). http://www.entomology.cornell.edu/BeePhylogeny/. [Accessed: January 2007]. Cornell University, Ithaca, NY

Darwin, C. R. (1872). *The origin of species by means of natural selection, or the preservation of flavoured races in the struggle for life. 6th edition.* John Murray, London.
Dávalos, L. M.Perkins, S. L. (2008). Saturation and base composition bias explain phylogenomic conflict in Plasmodium. *Genomics* **91,** 433-442.
Deans, A. R., Gillespie, J. J. & Yoder, M. J. (2006). An evaluation of ensign wasp classification (Hymenoptera: Evaniidae) based on molecular data and insights from ribosomal RNA secondary structure. *Systematic Entomology* **31,** 517-528.
Degnan, J. H.Rosenberg, N. A. (2006). Discordance of species trees with their most likely gene trees. *Public Library of Science Genetics* **2,** e68.
Delsuc, F. (2006). Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* **439,** 965-968.
Desjardins, C. A., Regier, J. C. & Mitter, C. (2007). Phylogeny of Pteromalid parasitic wasps (Hymenoptera: Pteromalidae): Initial evidence from four protein-coding nuclear genes. *Molecular Phylogenetics and Evolution* **45,** 454-469.
Dollo, L. (1893). Les lois de l'évolution. *Bull. Soc. Belge Geol. Paleont. Hydrol* **7,** 164-166.
Dowton, M.Austin, A. (1994). Molecular phylogeny of the insect order Hymenoptera: Apocritan relationships. *Proceedings of the National Academy of Sciences USA* **91,** 9911-9915.
Dowton, M.Austin, A. (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, A. D. (1997). 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, A. D. (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, A. D. & Antolin, M. F. (1998). Evolutionary Relationships among the Braconidae (Hymenoptera: Ichneumonoidea) inferred from partial 16S rDNA gene sequences. *Insect Molecular Biology* **7,** 129-150.
Dowton, M., Austin, A. D., Dillon, N. & Bartowsky, E. (1997). Molecular phylogeny of the apocritan wasps: the Proctotrupomorpha and Evaniomorpha. *Systematic Entomology* **22,** 245-255.
Dowton, M., Belshaw, R., Austin, A. D. & Quicke, D. (2002). Simultaneous molecular and morphological analysis of Braconid relationships (Insecta: Hymenoptera: Braconidae) indicates independent mt-tRNA gene inversions within a single wasp family. *Journal of Molecular Evolution* **54,** 210-226.
Dunn, C. W., Hejnol, A., Matus, D. Q., Pang, K., Browne, W. E., Smith, S. A.*, et al.* (2008). Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* **452,** 745-749.
Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5,** 113.
Elsik, C. G., Worley, K. C., Zhang, L., Milshina, N. V., Jiang, H., Reese, J. T.*, et al.* (2006). Community annotation: Procedures, protocols, and supporting tools. *Genome Research* **16,** 1329-1333.
Enright, A. J., Van Dongen, S. & Ouzounis, C. A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research* **30,** 1575-1584.
Felsenstein, J. (1978). Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* **27,** 401-410.

Fontaine, C., Dajoz, I., Meriguet, J. & Loreau, M. (2006). Functional diversity of plant-pollinator interaction webs enhances the persistence of plant communities. *Public Library of Science Biology* **4,** e1.

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.

Gauld, I. (1988). Evolutionary patterns of host utilization by ichneumoid parasitoids (Hymenoptera: Ichneumonidae and Braconidae). *Biological Journal of the Linnean Society* **35,** 351-377.

Gauld, I.Bolton, B. (1988). *The Hymenoptera*. British Museum of Natural History, Oxford University Press, London.

Gaut, B.Lewis, P. (1995). Success of maximum likelihood in the four taxon case. *Molecular Biology and Evolution* **12,** 152-162.

Gillespie, J. J. (2004). Characterizing regions of ambiguous alignment caused by the expansion and contraction of hairpin-stem loops in ribosomal RNA molecules. *Molecular Phylogenetics and Evolution* **33,** 936-943.

Gillespie, J. J., Munro, J. B., Heraty, J. M., Yoder, M. J., Owen, A. K. & Carmichael, A. E. (2005). Secondary structural model of the 28S rRNA expansion segments D2 and D3 for chalcidoid wasps (Hymenoptera: Chalcidoidea). *Molecular Biology and Evolution* **22,** 1593-1608.

Goodstadt, L.Ponting, C. P. (2006). Phylogenetic reconstruction of orthology, paralogy, and conserved synteny for dog and human. *Public Library of Science Computational Biology* **2,** e133.

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

Handlirsch, A. (1908). *Die fossilen Insekten und die Phylogenie der rezenten Formen: Ein Handbuch für Paläontologen und Zoologen*. W. Engelmann.

Hedtke, S., Townsend, T. & Hillis, D. (2006). Resolution of phylogenetic conflict in large data sets by increased taxon sampling *Systematic Biology* **55,** 522-529.

Hendy, M. D. (1991). A combinatorial description of the closest tree algorithm for finding evolutionary trees. *Discrete Mathematics* **96,** 51-58.

Hölldobler, B.Wilson, E. O. (1990). *The Ants*. Bel knap Press of Harvard University Press, Cambridge, Massachusetts.

Hudson, M. E. (2008). Sequencing breakthroughs for genomic ecology and evolutionary biology. *Molecular Ecology Resources* **8,** 3-17.

Huelsenbeck, J. P.Hillis, D. M. (1993). Success of phylogenetic methods in the four taxon case. *Systematic Biology* **42,** 247-264.

Huelsenbeck, J. P.Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* **17,** 754-755.

Hughes, J., Longhorn, S., Papadopoulou, A., Theodorides, K., de Riva, A., Mejia-Chang, M.*, et al.* (2006). Dense Taxonomic EST sampling and its applications for molecular systematics of the Coleoptera (Beetles). *Molecular Biology and Evolution* **23,** 268-278.

Huson, D. H.Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23,** 254-267.

Huson, D. H., Dezulian, T., Kloepper, T. & Steel, M. (2004). Phylogenetic super-networks from partial trees. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **1,** 151-158.

Huson, D. H., Richter, D. C., Rausch, C., Dezulian, T., Franz, M. & Rupp, R. (2007). Dendroscope- An interactive viewer for large phylogentic trees. *BMC Bioinformatics* **8,** 460.

Janzen, D. (1966). Coevolution of mutualism between ants and acacias in Central America. *Evolution* **20,** 249-275.

Kiester, R. A., Lande, R. & Schemske, D. W. (1984). Models of coevolution and speciation in plants and their pollinators. *The American Naturalist* **124,** 220-243.

Kjer, K. M. (1995). Use of rRNA secondary structure in phylogenetic studies to identify homologous positions: an example of alignment and data presentation from the frogs. *Molecular Phylogenetics and Evolution* **4,** 314-330.

Königsmann, E. (1976). Das phylogenetische system der Hymenoptera. Tiel 1: Einführung, Grundplanmerkmale, Schwestergruppe und Fossilfunde. *Deutsche Entomologische Zeitung* **23,** 253-279.

Königsmann, E. (1978a). Das phylogenetische system der Hymenoptera. Tiel 3: Terebrantes (Unterordnung Apocrita). *Deutsche Entomologische Zeitung* **25,** 1-55.

Königsmann, E. (1978b). Das phylogenetische system der Hymenoptera. Tiel 4: Aculeata (Unterordnung Apocrita). *Deutsche Entomologische Zeitung* **25,** 365-435.

Kumar, S.Gadagkar, S. R. (2001a). Disparity index: A simple statistic to measure and test the homogeneity of substitution patterns between molecular sequences. *Genetics* **158,** 1321-1327.

Kumar, S.Gadagkar, S. R. (2001b). Disparity Index: A simple statistic to measure and test the homogeneity of substitution patterns between molecular sequences. *Genetics* **158,** 1321-1327.

Lambert, D. (1990). *Dinosaur Data Book*. Facts On File & British Museum (Natural History), Oxford.

Larget, B. (2006). Bayesian untangling of Concordance Knots (BUCKy), version 1.1. 1.1 ed. Department of Statistics, University of Wisconsin

LaSalle, J.Gauld, I. D. (1993). *Hymenoptera and Biodiversity*. C.A.B. International, Wallingford, Oxon, UK.

Lockhart, P. J., Steel, M. A., Hendy, M. D. & Penny, D. (1994). Recovering evolutionary trees under a more realistic model of sequence evolution. *Molecular Biology and Evolution* **11,** 605-612.

Maddison, W. P.Maddison, D. R. (2008). Mesquite: a modular system for evolutionary analysis. Version 2.5 http://mesquiteproject.org.

Mardulyn, P.Whitfield, J. B. (1999). Phylogenetic signal in the COI, 16S and 28S genes for inferring relationships among genera of Microgastrinae (Hymenoptera: Braconidae): evidence of a high diversification rate in this group of parasitoids *Molecular Phylogenetics and Evolution* **12,** 282-294.

Marsh, P. M.Carlson, R. W. (1979). Superfamily Ichneumonoidea. Catalog of Hymenoptera in America North of Mexico. Volume I, pp. 143-144. Smithsonian Institution Press, Washington, D.C.

Marshall, C. R., Raff, E. C. & Raff, R. A. (1994). Dollo's law and the death and resurrection of genes. *Proceedings of the National Academy of Sciences USA* **91,** 12283-12287.

Martin, J. C. (1956). *A Taxonomic Revision of the Triaspidine Braconid Wasps of Nearctic America (Hymenoptera)*. Canadian Department of Agriculture, Ottawa.

Moulton, J. K.Wiegmann, B. M. (2004). Evolution and phylogenetic utility of CAD (rudimentary) among Mesozoic-aged Eremoneuran Diptera (Insecta). *Molecular Phylogenetics and Evolution* **31,** 363-378.

Muesebeck, C. F. W.Walkley, L. M. (1951). Hymenoptera of America north of Mexico, synoptic catalog. *In* "Agriculture Monographs" (U. D. o. Agriculture, Ed.), pp. 90-184.
Murphy, N., Banks, J., Whitfield, J. B. & Austin, A. (2008). Phylogeny of the parasitic microgastroid subfamilies (Hymenoptera: Braconidae) based on sequence data from seven genes, with an improved time estimate of the origin of the lineage. *Molecular Phylogenetics and Evolution* **47,** 378-395.
Mushegian, A. R., Garey, J. R., Martin, J. & Liu, L. X. (1998). Large-scale taxonomic profiling of eukaryotic model organisms: A comparison of orthologous proteins encoded by the human, fly, nematode, and yeast genomes *Genome Research* **8,** 590-598.
Nei, M. (1996). Phylogenetic analysis in moleculary evolutionary genetics. *Annual Review of Genetics* **30,** 371-403.
Nixon, K.Carpenter, J. M. (1993). On outgroups. *Cladistics* **9,** 413-426.
Nylander, J. A. A. (2004). MrModeltest v2. *In* "*Program distributed by the author.* Evolutionary Biology Centre, Uppsala University."
Oeser, R. (1961). Vergleichend-morphologische Untersuchungen über den Ovipositor der Hymenopteren. *Mitteilungen des Zoologisches Museum Berlin* **37,** 1-119.
Papp, J. (1994). *Maxfischeria tricolor* gen.n. et sp.n. from Australia (Insecta: Hymenoptera: Braconidae). *Annalen Des Naturhistorischen Museums Wien* **96,** 143-147.
Penny, D., McComish, B. J., Charleston, M. A. & Hendy, M. D. (2001). Mathematical elegance with biochemical realism: the covarion model of molecular evolution. *Journal of Molecular Evolution* **53,** 711-723.
Philippe, H. (2004). Phylogenomics of eukaryotes: impact of missing data on large alignments. *Mol. Biol. Evol.* **21,** 1740-1752.
Philippe, H., Lartillot, N. & Brinkmann, H. (2005). Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. *Mol. Biol. Evol.* **22,** 1246-1253.
Phillips, M. J., Delsuc, F. & Penny, D. (2004). Genome-scale phylogeny and the detection of systematic biases. *Molecular Biology and Evolution* **21,** 1455-1458.
Pitz, K., Dowling, A. P. G., Sharanowski, B. J., Boring, C. A. B., Seltmann, K. C. & Sharkey, M. J. (2007). Phylogenetic relationships among the Braconidae (Hymenoptera: Ichneumonoidea) as proposed by Shi et al.: a reassessment. *Molecular Phylogenetics and Evolution* **43,** 338-343.
Posada, D.Crandall, K. A. (1998). Modeltest: testing the model of DNA substitution. *Bioinformatics* **14,** 817-818.
Praz, C. J., Muller, A., Danforth, B. N., Griswold, T. L., Widmer, A. & Dorn, S. (2008). Phylogeny and biogeography of the tribe Osmini (Hymenoptera: Megachilidae). *Molecular Phylogenetics and Evolution* **49,** 185-197.
Quicke, D.van Achterberg, C. (1990a). Phylogeny of the subfamilies of the family Braconidae (Hymenoptera: Ichnemonoidea). *Zoologische Verhandelingen Leiden* **258,** 1-180.
Quicke, D.van Achterberg, C. (1990b). Phylogeny of the subfamilies of the family Braconidae (Hymenoptera: Ichneumonoidea). *In* "Zoologische Verhandelingen Leiden", Vol. 258, pp. 1-180.
Quicke, D. L. J. (1993). The polyphyletic origin of endoparasitism in the cyclostome ineages of Braconidae (Hymenoptera): a reassessment. *Zoologische Mededelingen Leiden* **67,** 159-177.
Quicke, D. L. J.Belshaw, R. (1999). Incongruence Between Morphological Data Sets: An Example from the Evolution of Endoparasitism Among Parasitic Wasps (Hymenoptera: Braconidae). *Systematic Biology* **48,** 436-454.

Rasnitsyn, A. P. (1980a). The origin and evolution of Hymenoptera. *Transactions of the Paleontological Institute of the Academy of Sciences of the USSR* **174,** 1-192.
Rasnitsyn, A. P. (1980b). The origin and evolution of Hymenoptera [In Russian]. *Transactions of the Paleontological Institute of the Academy of Sciences of the USSR* **174,** 1-192.
Rasnitsyn, A. P. (1988). An outline of evolution of hymenopterous insects (order Vespida). *Oriental Insects* **22,** 115-145.
Regier, J. C. (2007). Protocols, concepts, and reagents for preparing DNA sequence templates v. 9/19/07. www.umbi.umd.edu/users/jcrlab/PCR_primers.pdf. [Accessed: January, 2008], Vol. 2008.
Regier, J. C., Shultz, J. W., Ganley, A. R. D., Hussey, A., Shi, D., Ball, B.*, et al.* (2008). Resolving arthropod phylogeny: Exploring phylogenetic signal within 41 kb of protein-coding nuclear gene sequence. *Systematic Biology* **57,** 920 - 938.
Remm, M., Storm, C. E. V. & Sonnhammer, E. L. L. (2001). Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *Journal of Molecular Biology* **314,** 1041-1052.
Robbertse, B., Reeves, J. B., Schoch, C. L. & Spatafora, J. W. (2006). A phylogenomic analysis of the Ascomycota. *Fungal Genetics and Biology* **43,** 715-725.
Rokas, A., Kruger, D. & Carroll, S. B. (2005). Animal evolution and the molecular signature of radiations compressed in time. *Science* **310,** 1933-1938.
Rokas, A., Williams, B. L., King, N. & Carroll, S. B. (2003). Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425,** 798-804.
Ronquist, F.Huelsenbeck, J. P. (2003). MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19,** 1572-1574.
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.
Savard, J., Tautz, D., Richards, S., Weinstock, G. M., Gibbs, R. A., Werren, J. H.*, et al.* (2006). Phylogenomic analysis reveals bees and wasps (Hymenoptera) at the base of the radiation of Holometabolous insects. *Genome Research* **16,** 1334-1338.
Schöniger, M.von Haeseler, A. (1994). A stochastic model and the evolution of autocorrelated DNA sequences. *Molecular Phylogenetics and Evolution* **3,** 240-247.
Schulmeister, S., Wheeler, W. C. & Carpenter, J. M. (2002). Simultaneous analysis of the basal lineages of Hymenoptera (Insecta) using sensitivity analysis. *Cladistics* **18,** 455-484.
Sereno, P. C. (1986). Phylogeny of the bird-hipped dinosaurs. *National Geographic Research* **2,** 234-256.
Sharkey, M. J. (1992). Cladistics and tribal classification of the Agathidinae (Hymenoptera: Braconidae). *Journal of Natural History* **26,** 425-447.
Sharkey, M. J. (1993). Family Braconidae. Hymenoptera of the World: an Identification Guide to Families, pp. 362-395. Research Branch, Agriculture Canada and Canada Communications Group, Ottawa, Canada.
Sharkey, M. J. (2007). Phylogeny and classification of Hymenoptera. *Zootaxa* **1668,** 521-548.
Sharkey, M. J., Laurenne, N., Sharanowski, B. J., Quicke, D. & Murray, D. (2006). Revision of the Agathidinae (Hymenoptera: Braconidae) with comparisons of static and dynamic alignments. *Cladistics* **22,** 1-22.
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.

Sharkey, M. J.Wahl, D. B. (1992). Cladistics of the Ichneumonoidea (Hymenoptera). *Journal of Hymenoptera Research* **1,** 15-24.

Sharkey, M. J.Wharton, R. A. (1994). A revision of the genera of the world Ichneutinae (Hymenoptera: Braconidae). *Journal of Natural History* **28,** 873-912.

Shaw, M. R. (1983). On evolution of endoparasitism: the biology of some genera of Rogadinae (Braconidae). *Contributions to the American Entomological Institute* **20,** 307-328.

Shaw, M. R.Huddleston, T. (1991). Classification and biology of Braconid wasps (Hymenoptera: Braconidae). *Handbooks for the Identification of British Insects* **7,** 1-126.

Shi, M., Chen, X. X. & van Achterberg, C. (2005). Phylogenetic relationships among the Braconidae (Hymenoptera: Ichneumonoidea) inferred from partial 16S rDNA, 28S rDNA D2, 18S rDNA gene sequences and morphological characters. *Molecular Phylogenetics and Evolution* **37,** 104-116.

Simmons, M. P., Pickett, K. M. & Miya, M. (2004). How meaningful are Bayesian support values? *Molecular Biology and Evolution* **21,** 188-199.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22,** 2688-2690.

Stamatakis, A., Hoover, P. & Rougemont, J. (2008). A rapid bootstrap algorithm for the RAxML web-servers. *Systematic Biology* **75,** 758-771.

Sullivan, J., Holsinger, K. E. & Simon, C. (1995). Among-site rate variation and phylogenetic analysis of 12S rRNA in Sigmodontine rodents. *Molecular Biology and Evolution* **12,** 988-1001.

Sullivan, J.Joyce, P. (2005). Model Selection in Phylogenetics. *Annual Review of Ecology, Evolution, and Systematics* **36,** 445-466.

Swofford, D. L. (2000). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. . Sinauer Associates, Sunderland, Mass.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24,** 1596-1599.

Tamura, K.Kumar, S. (2002). Evolutionary distance estimation under heterogeneous substitution pattern among lineages. *Molecular Biology and Evolution* **19,** 1727-1736.

Tamura, K., Nei, M. & Kumar, S. (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences (USA)* **101,** 11030-11035.

Tarríoa, R., Rodríguez-Trellesb, F. & Ayalaa, F. J. (2000). Tree rooting with outgroups when they differ in their nucleotide composition from the ingroup: The *Drosophila saltans* and *willistoni* groups, a case study *Molecular Phylogenetics and Evolution* **16,** 344-349.

The FlyBase Consortium (2008). FlyBase: A Database of *Drosophila* Genes & Genomes. http://flybase.org/. [Accessed: June 2008]. FB2008_05 Dmel Release 5.8 ed, Vol. 2008.

The Honeybee Genome Sequencing Consortium (2008). BeeBase: Hymenoptera Genome Database. http://www.beebase.org/ [Accessed: June 2008]. Amel_pre_release2_OGS_cds ed.

Theodorides, K., De Riva, A., GÃ³mez-Zurita, J., Foster, P. G. & Vogler, A. P. (2002). Comparison of EST libraries from seven beetle species: towards a framework for phylogenomics of the Coleoptera. *Insect Molecular Biology* **11,** 467-475.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22,** 4673-4680.

Tobias, V. I. (1967). A review of the classification, phylogeny and evolution of the family Braconidae (Hymenoptera). *Entomol. Obozr. 46 (3): 645-669. (In Russian; translation in Entomological Review of Washington. 46: 387-399)*.

Tobias, V. I. (1988). A new subfamily of braconids (Hymenoptera, Braconidae) from Australia. . *Entomol. Obozr.* **67,** 644-648.

Toth, A. L., Varala, K., Newman, T. C., Miguez, F. E., Hutchison, S. K., Willoughby, D. A.*, et al.* (2007). Wasp gene expression supports an evolutionary link between maternal behavior and eusociality. *Science* **318,** 441-444.

Trivers, R. L.Hare, H. (1976). Haploidploidy and the evolution of the social insect. *Science* **191,** 249-263.

Tuffley, C.Steel, M. (1998). Modeling the covarion hypothesis of nucleotide substitution. *Mathematical Biosciences* **147,** 63-91.

Tweedie, S., Ashburner, M., Falls, K., Leyland, P., McQuilton, P., Marygold, S.*, et al.* (2009). Flybase: enhancing Drosophila Gene Ontology annotations. *Nucleic Acids Research* **37,** D555-D559; doi: 10.1093/nar/gkn788.

van Achterberg, C. (1975). A revision of the tribus Blacini (Hymenoptera: Braconidae: Helconinae). *Tijdschrift Voor Entomologie* **118,** 159-322.

Van Achterberg, C. (1979). A revision of the subfamily Zelinae Auct. (Hymenoptera: Braconidae). *Tijdschrift voor Entomologie* **122,** 241-479.

van Achterberg, C. (1984). Essay on the phylogeny of Braconidae (Hymenoptera: Ichneumonoidea). *Entomologisk Tidskrift* **105,** 41-58.

van Achterberg, C. (1988). Revision of the subfamily Blacinae Foerster (Hymenoptera: Braconidae). *Zoologische Verhandelingen Leiden* **249,** 1-324.

van Achterberg, C. (1992). Revision of the genera of the subfamily Microtypinae (Hymenoptera: Braconidae). *Zoologische Mededelingen* **66,** 369-380.

van Achterberg, C. (1995). Generic revision of the subfamily Betylobraconinae (Hymenoptera: Braconidae) and other groups with modified fore tarsus. *Zoologische Verhandelingen Leiden* **298, 31,** 1-242.

van Achterberg, C. (2002). Revision of the genus *Canalicephalus* Gibson and the recognition of the Acampsohelconinae (Hymenoptera: Braconidae) as extant. *Zoologische Mededelingen* **76,** 247-270.

van Achterberg, C.Quicke, D. (1992). Phylogeny of the subfamilies of the family Braconidae: A reassessment assessed. *Cladistics* **8,** 237-264.

Vilhelmsen, L. (2001). Phylogeny and classification of the extant basal lineages of the Hymenoptera (Insecta). *Zoological Journal of the Linnean Society* **131,** 393-442.

Vilhelmsen, L. (2003). Phylogeny and classification of the Orrusidae (Insecta: Hymenoptera), a basal parasitic wasp taxon. *Zoological Journal of the Linnean Society* **139,** 337-418.

Wang, H. C., Spencer, M., Susko, E. & Roger, A. J. (2007). Testing for covarion-like evolution in protein sequences. *Molecular Biology and Evolution* **24,** 294-305.

Wang, J., Xia, Q., He, X., Dai, M., Ruan, J., Chen, J.*, et al.* (2005). SilkDB: a knowledgebase for silkworm biology and genomics. *Nucleic Acids Research* **33,** D399-402.

Wertheim, B., Kraaijeveld, A., Schuster, E., Blanc, E., Hopkins, M., Pletcher, S.*, et al.* (2005). Genome-wide gene expression in response to parasitoid attack in Drosophila. *Genome Biology* **6,** R94.

Wharton, R. (2000). Can Braconid Calssification Be Reconstructed to Facilitate Portrayal of Relationships? Hymenoptera: Evolution, Biodiversity, and Biological Control, pp. 143-153. CSIRO, Canberra.

Wharton, R. A. (1993). Bionomics of the Braconidae. *Annual Review of Entomology* **38,** 121-143.

Wharton, R. A., Shaw, M. R., Sharkey, M. J., Wahl, D. B., Woolley, J. B., Whitfield, J. B., *et al.* (1992). Phylogeny of the subfamilies of the family Braconidae (Hymenoptera: Ichneumonoidea): A reassessment. *Cladistics* **8,** 199-235.

Wheeler, W. C. (1995). Sequence alignment, parameter sensitivity, and the phylogenetic analysis of molecular data. *Systematic Biology* **44,** 321-331.

Wheeler, W. C., Whiting, M., Wheeler, Q. D. & Carpenter, J. M. (2001). The phylogeny of the extant hexapod orders. *Cladistics* **17,** 113-169.

Whelan, S. (2008). Spatial and temporal heterogeneity in nucleotide sequence evolution. *Molecular Biology and Evolution* **25,** 1683-1694.

Whelan, S., Liò, P. & Goldman, N. (2001). Molecular phylogenetics: state-of-the-art methods for looking into the past. *Trends in Genetics* **17,** 262-272.

White, W. T., Hills, S. F., Gaddam, R., Holland, B. R. & Penny, D. (2007). Treeness triangles: Visualizing the loss of phylogenetic signal. *Molecular Biology and Evolution* **24,** 2029-2039.

Whitfield, J. B. (1992a). Phylogeny of the non-aculeate Apocrita and the evolution of parasitism in the Hymenoptera. *Journal of Hymenoptera Research* **1,** 3-14.

Whitfield, J. B. (1992b). The polyphyletic origin of endoparasitism in the cyclostome ineages of Braconidae (Hymenoptera). *Systematic Entomology* **17,** 273-286.

Whitfield, J. B. (1997). Molecular and morphological data suggest a common origin for the polydnaviruses among braconid wasps. *Naturwissenschaften***,** 502-507.

Whitfield, J. B. (1998). Phylogeny and evolution of host-parasitoid interactions in Hymenoptera. *Annual Review of Entomology* **43,** 129-151.

Whitfield, J. B., Baker, M. R. & Littlewood (2003). Phylogenetic Insights into the Evolution of Parasitism in Hymenoptera. Advances in Parasitology, pp. 69-100. Academic Press.

Whitfield, J. B., Cameron, S. A., Huson, D. H. & Steel, M. A. (2008). Filtered z-closure supernetworks for extracting and visualizing recurrent signal from incongruent gene trees. *Systematic Biology* **57,** 939 - 947.

Whiting, M. F. (2002). Phylogeny of the holometabolous insect orders: Molecular evidence. *Zoologica Scripta* **31,** 3-15.

Wiebes, J. T. (1979). Co-evolution of figs and their insect pollinators. *Annual Review of Ecology and Systematics* **10,** 1-12.

Wilfert, L., Gadau, J. & Schmid-Hempel, P. (2007). Variation in genomic recombination rates among animal taxa and the case of social insects. *Heredity* **98,** 189-197.

Winterton, S.de Freitas, S. (2006). Molecular phylogeny of the green lacewings (Neuroptera: Chrysopidae). *Australian Journal of Entomology* **45,** 235-243.

Yoder, M.Gillespie, J. (2004). jRNA. Exploring insect phylogeny using RNA secondary structure. http://hymenoptera.tamu.edu/rna.

Yu, D. S., Horstmann, K. & van Achterberg, C. (2001). Taxapad 2004: Scientific Names for Information Management. *In* "Biological and Taxonomical Information: Ichneumonoidea 2004, CD". Taxapad, Vancouver.

Zaldivar-Riverón, A., Mori, M. & Quicke, D. (2006). Systematics of the cyclostome subfamilies of braconid parasitic wasps (Hymenoptera: Ichneumonoidea): A simultaneous molecular and morphological Bayesian approach. *Molecular Phylogenetics and Evolution* **38,** 130-145.

Zaldivar-Riverón, A., Shaw, M. R., Sáez, A. G., Mori, M., Belokoblylskij, S. A., Shaw, S. R., *et al.* (2008). Evolution of the parasitic wasp subfamily Rogadinae (Braconidae): phylogeny and evolution of lepidopteran host ranges and mummy characteristics. *BMC Evolutionary Biology* **8,** 329.

Zuker, M. (1989). Methods in Enzymology, pp. 262-288. Academic Press, San Diego, California.
Zwickl, D. J.Hillis, D. M. (2002). Increased taxon sampling greatly reduces phylogenetic error. *Systematic Biology* **51,** 588-598.

Copyright © Barbara J. Sharanowski 2009

VITA of Barbara J. Sharanowski

Date and Place of Birth:

- November 19, 1974, Stratford, Ontario, Canada

Education:

- *M.A. Forensic Entomology*, University of Saskatchewan; 2004
- *B.A. Anthropology and Archaeology*, University of Saskatchewan, 1999

-

Professional Employment:

- *Research Assistant*, Department of Entomology, University of Kentucky, 2004 – 2009
- *Lecturer*, ENT 684 Phylogenetic Systematics, University of Kentucky, 2008
- *Extension assistant,* Department of Entomology, University of Kentucky, 2006
- *Forensic Consultant*, (Freelance), Mississauga, Ontario, Carey, McCallum, Nimjee, and Smith Law Office, 2001 – 2002
- *Teaching Assistant*, University of Saskatchewan, Department of Anthropology and Archaeology, 1999 – 2000
- *Substitute Lecturer*, Introduction to Biology, LCC, Kentucky, 2006

Publications:

- **Sharanowski**, B.J., Walker, E.G., and Anderson, G.S. 2008. Insect Succession and Decomposition Patterns on Shaded and Sunlit Carrion in Saskatchewan in Three Different Seasons. *Forensic Science International*, 179: 219-240.
- **Sharanowski**, B.J. and M.J. Sharkey. 2007. A description of three new species of *Helconichia* Sharkey & Wharton (Hymenoptera: Braconidae: Ichneutinae) with a revised key to all species. *Zootaxa* 1502: 45-53.
- Pitz, K.M., A. P.G. Dowling, B.J. **Sharanowski**, C.A. Boring, K.C. Seltmann and M.J. Sharkey. 2007. Phylogenetic relationships among the Braconidae (Hymenoptera: Ichneumonoidea) as proposed by Shi et al.: a reassessment. *Molecular Phylogenetics and Evolution* 43(1): 338-343.
- Sharkey, M.J., N. Laurenne, B. **Sharanowski**, D. Quicke, and D. Murray. 2006. Revision of the Agathidinae (Hymenoptera: Braconidae) with comparisons of static and dynamic alignments. Cladistics 22: 1-22.
- **Sharanowski**, B.J. 2003. Initial Studies in Forensic Entomology in Saskatchewan: Decomposition and Insect Succession on Pig Carrion in the Prairie Ecozone. Master's Thesis, University of Saskatchewan.

VITA of Barbara J. Sharanowski cont'd.

Scholastic and Professional Honors:

- *Program Enhancement Funds*, Section Symposia, Entomological Society of America, 2008
- *National Science Foundation Travel Grant*, 2008
- *Kentucky Opportunity Fellowship***,** 2007
- *President's Prize*, Oral Presentation, Entomological Society of America
- *University of Kentucky Women's Club Fellowship*, 2006
- *Departmental Scholarship*, Dept. Anthropology and Archaeology, University of Saskatchewan, 2001
- *Nasser Travel Bursary*, University of Saskatchewan, 1999