

1-1-2012

## The first mitochondrial genome for the wasp superfamily Platygastroidea: the egg parasitoid *Trissolcus basalis*

Meng Mao

*University of Wollongong*, mm663@uowmail.edu.au

Alejandro Valerio

*Ohio State University*

Andrew D. Austin

*University of Adelaide*

Mark Dowton

*University of Wollongong*, mdowton@uow.edu.au

Norman F. Johnson

*Ohio State University*

Follow this and additional works at: <https://ro.uow.edu.au/scipapers>



Part of the [Life Sciences Commons](#), [Physical Sciences and Mathematics Commons](#), and the [Social and Behavioral Sciences Commons](#)

---

### Recommended Citation

Mao, Meng; Valerio, Alejandro; Austin, Andrew D.; Dowton, Mark; and Johnson, Norman F.: The first mitochondrial genome for the wasp superfamily Platygastroidea: the egg parasitoid *Trissolcus basalis* 2012, 194-204.

<https://ro.uow.edu.au/scipapers/4735>

---

# The first mitochondrial genome for the wasp superfamily Platygastroidea: the egg parasitoid *Trissolcus basalis*

## Abstract

The nearly complete mitochondrial (mt) genome of an egg parasitoid, *Trissolcus basalis* (Wollaston), was sequenced using both 454 and Illumina next-generation sequencing technologies. A portion of the noncoding region remained unsequenced, possibly owing to the presence of repeats. The sequenced portion of the genome is 15 768 bp and has a high A+T content (84.2%), as is typical for hymenopteran mt genomes. A total of 36 of the 37 genes normally present in animal mt genomes were located. The one exception was *trnR*; a truncated version of this gene is present between *trnS1* and *nd5*, but it is unclear whether this gene fragment could code for the entire *trnR* gene. The mt gene arrangement of *T. basalis* is different from other Proctotrupomorpha mt genomes, with a number of *trn* genes in different positions. However, no shared derived gene rearrangements were identified in the present study. Bayesian analyses of mt genomes from 29 hymenopteran taxa and seven other orders of holometabolous insects support some uncontroversial evolutionary relationships, but indicate that much higher levels of taxonomic sampling are necessary for the resolution of family and superfamily relationships.

## Keywords

parasitoid, egg, platygastroidea, trissolcus, superfamily, wasp, genome, mitochondrial, basalis, first

## Disciplines

Life Sciences | Physical Sciences and Mathematics | Social and Behavioral Sciences

## Publication Details

Mao, M., Valerio, A., Austin, A. D., Dowton, M. & Johnson, N. F. (2012). The first mitochondrial genome for the wasp superfamily Platygastroidea: the egg parasitoid *Trissolcus basalis*. *Genome*, 55 (3), 194-204.

# **The first mitochondrial genome for the wasp superfamily**

## **Platygastroidea: the egg parasitoid *Trissolcus basalis***

Meng Mao<sup>1</sup>, Alejandro Valerio<sup>2</sup>, Andrew D. Austin<sup>3</sup>, Mark Dowton<sup>1</sup>, Norman F. Johnson<sup>2</sup>

<sup>1</sup>Centre for Medical Bioscience, School of Biological Sciences, Wollongong University, Wollongong, NSW, 2522, Australia

<sup>2</sup>Department of Evolution, Ecology and Organismal Biology, The Ohio State University, 1315 Kinnear Road, Columbus, OH 43212, USA

<sup>3</sup>Center for Evolutionary Biology and Biodiversity, School of Earth and Environmental Sciences, The University of Adelaide, S.A. 5005, Australia;

Keywords: mitochondrial, Hymenoptera, Proctotrupomorpha, *Trissolcus*, gene rearrangement, phylogenetic relationship

Corresponding author: Meng Mao, Centre for Medical Bioscience, School of Biological Sciences, University of Wollongong, Northfields Avenue, Wollongong, NSW, 2522, Australia. Telephone: 612-42-213709; Fax: 612-42-214135; email: [mm663@uow.edu.au](mailto:mm663@uow.edu.au)

## Abstract

The nearly complete mitochondrial (mt) genome of an egg parasitoid, *Trissolcus basalis* (Wollaston), was sequenced. This genome is 15768 bp in size and has a high A+T content (84.2%). In the protein-coding genes, the A+T content is as high as 92.3% in the third codon position. Four of the protein-coding genes have incomplete termination codons: TA or T. All transfer RNA (*trn*) genes have the typical cloverleaf structure except for *trnS<sub>2</sub>*. The mt gene arrangement of *T. basalis* is different from other Proctotrupomorpha mt genomes. Bayesian analyses of mt genomes from 29 hymenopteran taxa and 7 other orders of holometabolous insects support some uncontroversial evolutionary relationships, but indicate that much higher levels of taxonomic sampling are necessary for the resolution of family and superfamily relationships.

## Introduction

Mitochondrial (mt) genomes have been widely used for the reconstruction of phylogenetic relationships at different taxonomic levels (Miya et al. 2003; Cameron et al. 2007; Kim et al. 2009; Li et al. 2011). In insects, mtDNA is typically a double-stranded circular molecule of about 16 kb in size. It contains 13 protein-coding genes, 2 ribosomal RNA (*rrn*) genes and 22 transfer RNA (*trn*) genes (Boore 1999). Additionally, there is a major non-coding region, known as the AT-rich region or control region, which plays a role in initiation of transcription and replication (Zhang and Hewitt 1997).

The Hymenoptera is one of the most species-rich and biologically diverse insect orders (Lasalle and Gauld 1993). Although much progress has been made to investigate the phylogeny of the Hymenoptera (Dowton and Austin 1994; Ronquist et al. 1999; Dowton and Austin 2001; Castro and Dowton 2006; Heraty et al. 2011; Sharkey et al. 2011), there are also many relationships remaining unresolved. For example, the monophyly of the Proctotrupomorpha (*sensu* Rasnitsyn 1988) is well supported by recent studies, but the relationships among the proctotrupomorph families are still controversial (Dowton and Austin 2001; Castro and Dowton 2006; Rasnitsyn and Zhang 2010; Heraty et al. 2011). Several analyses have demonstrated that the whole mt genome is a useful tool to resolve hymenopteran evolutionary relationships (Castro and Dowton 2007; Cameron et al. 2008; Dowton et al. 2009a; Wei et al. 2010a). To date, 18 complete mt genomes

and 15 nearly complete mt genomes have been successfully sequenced in the Hymenoptera, with 18 taxa added in the last 3 years. This large amount of data holds promise to improve phylogenetic analysis. Notwithstanding, there remain shortcomings with this dataset, with only 11 superfamilies represented. For approximately half of the hymenopteran superfamilies, there is no mt genome data available.

In this paper, we provide the mt genomic sequence of *Trissolcus basalis* (Wollaston) from the superfamily Platygastroidea, the third largest of the parasitic superfamilies after the Ichneumonoidea and Chalcidoidea (Austin et al. 2005). No mitochondrial genome sequences for the Platygastroidea have previously been published. The characteristics of this mt genome, with respect to genome composition, nucleotide content and codon usage, were compared with other hymenopteran mt genomes. We conducted phylogenetic analyses of mt genomes from 29 hymenopteran taxa and 7 other orders of the holometabolous insects. Bayesian analyses were performed using nucleotide sequences (including and excluding the third codon position). The performance of two alignment methods (Clustal W and Muscle) was also compared. Another goal of this research was to further explore the evolution of mt gene rearrangement. The hymenopteran mt genome has been reported to rearrange more frequently than those from other insect orders (Croizer and Crozier 1993; Dowton and Austin 1999; Dowton et al. 2009b). Here we compared the genome structures of four Proctotrupomorpha taxa.

## **Material and Methods**

### **DNA extraction, PCR amplification and sequencing**

**454 sequencing.** Genomic DNA was extracted from whole bodies of 25 adult male specimens of *T. basalis* from a laboratory culture maintained by Dr. Ferdinando Bin at the Università di Perugia. The extract was sent to the DNA Sequencing Facility at the University of Pennsylvania Perelman School of Medicine. Five full plates were run on a Roche/454 GS FLX sequencer using Titanium chemistry. This resulted in 5,080,113 reads of a total of 1,535,920,544 bases. These reads were assembled using gsAssembler 2.5.3 with a minimum overlap of 30 bases and the complex genome assembly option.

**Illumina sequencing.** To address the issue of possible homopolymer errors in 454 sequencing, genomic DNA was extracted from an additional five female specimens from the same culture. The extract was processed using the Nextera DNA sample preparation kit from Epicentre Biotechnologies following the manufacturer's instructions. The sample was submitted to the Nucleic Acid Shared Resource of the College of Medicine at The Ohio State University. A single lane of 51 base reads was run on an Illumina Genome Analyzer IIx resulting in 29,780,645 reads and a total of 1,518,812,895 bases. These reads were assembled using VCAKE1.5 into repetitive and non-repetitive contigs, and short contigs (< 70 base) were then removed. The remaining contigs were then assembled together with the final 454 assembly using gsAssembler 2.5.3. Neson 0.42 was used to correct errors in the hybrid assembly.

Contigs of mitochondrial origin from the entire assembly were identified by a BLAST search using known hymenopteran mitochondrial proteins. Two contigs emerged from the search, with lengths of 8,067 bases and 7,719 bases.

### **Sequence analyses**

The two contigs were imported into Bioedit version 7.0.5 (Hall 1999). Visual inspection revealed that the same dinucleotide repeat [(TA)<sub>n</sub>] was present at one end of each of the contigs. When the contigs were oriented in the ancestral orientation (i.e. with protein-coding genes oriented on the same strand and direction as the ancestral pancrustacean), the shorter contig had a (TA)<sub>13</sub> repeat at the 3' end, while the longer contig had a (TA)<sub>9</sub> repeat at the 5' end. We interpret this as indicating that a dinucleotide repeat that varies in length is situated between these two contigs. Joining these, the ancestral orientation of all protein-coding and *rrn* genes is maintained. The dinucleotide repeat is situated in the intergenic space, between the *trnH* and *nd4* genes.

19 *trn* genes were identified by tRNA-Scan SE version 1.21 (lowelab.ucsc.edu/tRNAscan-SE/) (Lowe and Eddy 1997), specifying mitochondrial/chloroplast DNA as the source and choosing the invertebrate mitochondrial genetic code for tRNA isotype prediction. The cove cutoff score was set to 5 in order to avoid missing some *trn* genes with lower cutoff scores. Two *trn* genes (*trnN*

and *trnS<sub>I</sub>*) failed to be detected by tRNA-Scan SE, but were identified by visually comparing unassigned regions of the genome with previously determined tRNA counterparts from the Hymenoptera. Thirteen protein-coding genes were identified using open reading frame (ORF) finder ([ncbi.nlm.nih.gov/gorf/orfig.cgi](http://ncbi.nlm.nih.gov/gorf/orfig.cgi)), specifying the invertebrate mitochondrial genetic code. The initiation and termination codons of some genes were identified according to the boundaries of *trn* genes and comparison with other insect mt sequences. The precise boundaries of *rrn* genes are difficult to define. They were assumed to be bounded by the neighboring *trn* genes (Beckenbach 2011; Beckenbach and Stewart 2009; Downton et al. 2009a). However, the boundary between *rrnL* and *rrnS* was difficult to define in *T. basalis*, as the *trnV* that is usually between the two *rrn* genes is not present here. Instead, we determined this boundary by aligning the *rrn* regions from *T. basalis* with other closely related insect *rrnL* and *rrnS* genes.

### Sequence alignment

Mitogenomic sequences from 29 hymenopteran taxa and 7 taxa from other holometabolous insects were obtained from GenBank (Table 1). Nucleotide sequences for each of the 13 protein-coding genes and the 2 *rrn* genes were imported into separate files using MEGA (version 5.05, Tamura et al. 2011) and aligned using Clustal W (Thompson et al. 1994) or Muscle (Edgar and Robert 2004) as implemented within MEGA 5. For the protein-coding genes (excluding the stop codons), the nucleotide sequences were translated into the amino acid sequences using the invertebrate mitochondrial genetic code, and the amino acid sequences aligned using Clustal W or Muscle. MEGA 5 then creates a nucleotide alignment using the amino acid sequences as a guide. The alignment parameters were the default settings for all genes. The Clustal W alignment parameters were as specified in Cameron et al. (2008). The MUSCLE alignment parameters for the rRNA genes were gap open penalty = -400, gap extend penalty = 0; and for the protein-coding genes were gap open penalty = -2.9, gap extend penalty = 0. For both types of gene, the maximum memory allocated was 2047 MB, with a maximum of 8 iterations. The clustering method used for all iterations was UPGMB, with the length of the minimum diagonal set to 24. Following alignment, individual genes were concatenated prior to phylogenetic analysis.

### Phylogenetic analysis

Previous phylogenetic studies using hymenopteran mt genome sequences indicated that nucleotide data analyses were superior to amino acid data analyses, and that Bayesian analyses were superior to parsimony analyses (Castro and Dowton 2007; Dowton et al. 2009a). For this reason, we employed Bayesian analyses in this study. The analyses were conducted using MrBayes (version 3.1.2; Ronquist and Huelsenbeck 2003) at the freely available Bioportal server ([www.biportal.uio.no](http://www.biportal.uio.no)). The two datasets (one aligned using Clustal, the other aligned using Muscle) were divided into four partitions: the first, second and third codon positions and *rrn* genes. The GTR+G+I model (nst=6, rate=invgamma) was chosen as the best-fit model for all of the partitions (MrModelTest 2.3, Nylander et al. 2004). Two analyses (all codon positions included or the third codon positions excluded) were performed for the two alignment datasets. The Markov chain Monte Carlo (MCMC) process was set so that four chains (three heated and one cold) ran simultaneously. For each analysis, we conducted four independent runs for 1,000,000 generations, with trees being sampled every 100 generations. Burn-in was discarded according to the plot of generation against the likelihood scores and the sump command in MrBayes. The percentage of trees recovering a particular clade was used as a measure of that clade's posterior probability (Huelsenbeck and Ronquist 2001).

## **Result and Discussion**

### **Genome size and gene content**

The almost complete mt genome sequence of *T. basalis* is 15,768 bp in size (Fig. 1). We were not able to obtain the sequence of part of the non-coding region. Although the sequencing strategy should have sequenced the non-coding region just as frequently as the coding regions, the inability to identify fragments that could be added to our mitochondrial contigs is possibly due to the presence of repeats. We did attempt to amplify the noncoding region by conventional PCR, but were unable to obtain amplicons, as has been reported for a number of hymenopteran mt genomes (Castro and Dowton 2005; Castro et al. 2006; Oliveira et al. 2008). We were able to identify 36 of the 37 genes usually found in animal mt genomes (Table 2); 13 protein coding genes, 21 of the 22 *trn* genes, 2 *rrn* genes and part of the AT-rich region. Several short non-coding regions were identified. The longest of these is 60 bp long, and is located between the *trnS<sub>I</sub>* and *nd5* genes.



There are 5 genes that overlap: 2 between protein-coding genes (*a8* and *a6*, and *nd4* and *nd4L*) and three between *trn* genes (*trnD* and *trnK*, *trnN* and *trnE*, and *trnT* and *trnP*).

### **Nucleotide content**

As has been widely published in other insect mt genomes (Simon et al. 1994), the nucleotide content of *T. basalis* is heavily biased towards adenine and thymine (Table 3). The entire sequence has 43.7% A, 9.8% C, 40.5% T and 6.0% G. The total A+T content is 84.2%, which is typical of other hymenopteran mt genomes. For example, it is almost as high as that of *Apis mellifera* [84.9% (Crozier and Crozier, 1993)] but higher than that of *Perga condei* [78% (Castro and Dowton 2005)] and *Vanhornia eucnemidarum* [80.1% (Castro et al. 2006)].

For the mt protein-coding genes of *T. basalis*, the A+T content is a little lower than that of the entire mt genome sequence. However, the first and second codon positions have a lower A+T content (78.9% and 75.4%, respectively), while the third codon position shows a much higher A+T content (92.3%). Conversely, both the *rrn* and *trn* genes have a higher A+T content (87.9% and 88.7%, respectively) when compared with the average for the genome. The nucleotide content bias is inferred to result from mutational pressure (Asakawa et al. 1991; Foster et al. 1997).

### **Codon usage and amino acid composition**

The codon usage in the *T. basalis* mt genome is highly skewed towards codons that are high in AT-content (Table 4). Codons such as CTG (Leu) and CGC (Arg), encoded by both C- and G-rich codons, are rarely used. Conversely, the AT-rich codons ATT (Ile), TTA (Leu), TTT (Phe) and ATA (Met) are the four most frequently used codons (465, 460, 407 and 341 times, respectively). Within a particular synonymous codon family, AT-rich codons are predominantly used. For example, leucine can be coded by six alternative codons, but TTA is used much more frequently than any other. The relative synonymous codon usage (RSCU: Nei and Kumar 2000) for TTA is 5.23. The RSCU measures the relative proportion that a codon is used within a particular synonymous codon family – the maximum RSCU value for the leucine codon family is 6. For each codon family, the sum of the A- and T-rich codons is close to the maximum possible RSCU value. Similar codon usage bias has been observed in other hymenopteran mt genomes (Castro et

al. 2006; Cha et al. 2007; Wei et al. 2009).

### **Translation initiation and termination codons**

Thirteen protein-coding genes were identified using ORF finder. Table 2 shows the position of these genes in the mt genome of *T. basalis*, their lengths, and translation initiation and termination codons. Conventional ATA, ATT or ATG initiation codons are assigned to all of the thirteen protein-coding genes. Nine genes (*a8*, *cob*, *cox1*, *cox2*, *nd1*, *nd3*, *nd4* and *nd4L* and *nd5*) are predicted to use the complete termination codons TAA or TAG. Four other genes have incomplete termination codons: TA for *a6* and *nd6*, and T for *nd2* and *cox3*. Incomplete termination codons have been commonly reported for other insect species (Clary and Wolstenholme 1985; Crozier and Crozier 1993). It has been proposed that the complete termination codon (TAA) is created by polyadenylation after transcription (Ojala et al. 1981).

### **Transfer RNA genes**

Twenty-one *trn* genes (59-73 bp in size) were identified in the *T. basalis* mt genome. Fourteen *trn* genes are encoded on the J-strand (the strand encoding the maJority of genes, as defined by Simon et al 1994), with 7 encoded on the N-strand (the miNority strand). Their predicted secondary structures are typical cloverleaf structures, except for *trnS<sub>I</sub>* (Fig. 2). The D-stem in the dihydrouridine (DHU) arm is absent in the *trnS<sub>I</sub>* gene, which has been reported in other hymenopteran species (Castro et al. 2006; Wei et al. 2010b). In addition, mismatched base pairs appear in the DHU stem and acceptor stem, which is common in the insect *trn* genes (Shao and Barker 2003; Zhou et al. 2009). The anticodons are identical to their counterparts in most other published hymenopteran mt genomes.

### **Ribosomal RNA genes**

The small and large ribosomal RNA genes (*rrnS* and *rrnL*) are located on the N-strand. Unlike other hymenopteran mt genomes, they are directly adjacent to each other, without the *trnV* gene between them. *rrnS* is 752 bp in size with an A+T content of 86.2%. *rrnL* is 1270 bp in size with an A+T content of 88.8%. These sizes are similar when compared with other Hymenoptera. Based on the range of hymenopteran taxa included in Table 1, *rrnS* ranges from 600 to 811 nucleotides,

while *rrnL* ranges from 868 to 1541 nucleotides. The A+T content of *rrnS* is similar when compared with other hymenopteran taxa (Table 1), in which the A+T content ranges from 74.1% to 91.2%. The A+T content of *rrnL* is at the higher end of the range; it is only lower than that of *Cotesia* (89.72%) and *Macrocentrus* (89.55%).

### Non-coding regions

The longest non-coding region (partial AT-rich region) of *T. basalis* is 974 bp in size, which is located between the *rrnS* and *trnC* genes (Fig. 1). The A+T content is 89.8%, which is not markedly higher compared with the remainder of the genome. There are several other short non-coding regions ranging in size from 2 bp to 60 bp, which are located throughout the genome. Interestingly, the 60 bp non-coding region has the anticodon of *trnR* (the *trn* gene that could not be found) and the length is similar to the *trn* genes, but it could not be folded into a clover-leaf structure unless overlapping more than 30 bp with *nd5*. If this section is ascribed to the *trnR* gene, the slightly shorter *nd5* gene lacks a stop codon (either complete or incomplete). Recently, some truncated *trn* genes were reported in the mitochondrial genomes of velvet worms, with *trn* editing converting the primary transcripts to conventional *trn* structures (Segovia et al. 2011). It is conceivable that the missing *trnR* gene is present in this non-coding region, and is completed by a similar mechanism.

### Genome organization

The mitochondrial genome organization of *T. basalis* is shown in Figs 1 and 3. All of the protein-coding and *rrn* genes have identical positions and transcriptional directions when compared with the organization of the ancestral pancrustacean (Crease 1999; Cook 2005) (Fig. 3), which (as far as is known) is the same organization as the ancestral hymenopteran (Dowton et al. 2009b). However, the positions of a number of *trn* genes differ from the ancestral organization. For example, the ancestral positions of *trnC*, *trnY* and *trnW* are between the *nd2* and *coxI* genes, while in *T. basalis*, *trnC* and *trnY* are between *nd2* and the non-sequenced region (presumably the non-coding region). Similarly, *trnA*, *trnD*, *trnI*, *trnM*, *trnQ*, *trnR* and *trnS<sub>I</sub>* are not in the ancestral positions.

We then compared the organization of the *T. basalis* mt genome with three other hymenopteran taxa (all from the Proctotrupomorpha), to investigate whether any of the *trn* gene rearrangements were shared with close allies. The Proctotrupomorpha have been consistently recovered as a monophyletic group in molecular phylogenetic studies (Dowton and Austin 2001; Castro and Dowton 2006; Heraty et al. 2011). Shared gene rearrangements are considered reliable indicators of shared ancestry (Boore et al. 1995), and have been used to infer phylogeny in the Hymenoptera (Dowton 1999), although examples of homoplastic gene rearrangements have been reported (Flook et al. 1995; Dowton and Austin 1999). The mt genome organizations of these three additional taxa from the Proctotrupomorpha are shown in Fig. 3 to facilitate comparisons. All of the protein-coding and *rrn* genes are positioned identically in *T. basalis*, *Vanhornia* and the available mitogenomic sequences of *Ceratosolen*, while *Nasonia* has an inversion of six protein-coding genes and a translocation of one protein-coding gene (Oliveira et al. 2008). There are only two junctions (*cox2-cox3* and *nd4-rrnL*) that are present in all four taxa. The pattern that emerges from this comparison is that the positions of *trn* genes among the Proctotrupomorpha is extremely variable, with only two *trn* genes remaining in their ancestral locations in all four taxa (*trnT* and *trnP*). *trnH* and *trnS<sub>2</sub>* are present in their ancestral locations in three Proctotrupomorpha taxa, with their positions unknown in *Ceratosolen* and *Nasonia*, respectively. All remaining *trn* genes have rearranged in one or more taxa. Interestingly, the *trnA* gene is in the same derived position (next to the *rrnS* gene) in both *Ceratosolen* and *Nasonia*. These are the only two included representatives from the Chalcidoidea – this gene rearrangement may thus be a synapomorphy for a subset of the Chalcidoidea.

### Phylogenetic analysis and considerations

Phylogenetic analyses were performed using 36 complete or partial mt genomes, with 29 of these taxa from the Hymenoptera. This is more than double the number of hymenopteran taxa compared with previous studies (14 taxa: Dowton et al., 2009). The most comprehensive analyses of holometabolan relationships place the Hymenoptera at the base of the Holometabola (Savard et al. 2006; Wiegmann et al. 2009), which make it difficult to choose an outgroup. For this reason, we included a range of homometabolous taxa in the analysis. MrBayes allows only a single taxon to be the outgroup, so we chose *Tribolium* (Coleoptera) as the outgroup, but any of the other

holometabolous insects would have been appropriate. Two Bayesian analyses (all codon positions included or the third codon positions excluded) were performed for each of the alignment datasets (Clustal and Muscle).

The intraordinal relationships within the Hymenoptera were broadly congruent across the four analyses, although there were some differences in the topologies. The major difference occurred between analyses that included all codon positions and those that excluded the third codon positions.

In the 29 hymenopteran taxa, there are 10 taxa from the Aculeata, which contains the superfamilies Apoidea, Chrysidoidea and Vespoidea. The analyses with the third codon positions excluded recovered the Aculeata as a monophyletic group, while the analyses with all codon positions recovered a controversial clade, with *Evania* (a member of Evaniomorpha) disrupting the Aculeata. This contrasts with previous studies (Dowton et al. 2009a), in which the Aculeata were recovered as a clade in both analyses (i.e. including or excluding third codon positions). This suggests that the exclusion of third codon positions improves phylogenetic accuracy when analyzing hymenopteran relationships using the whole mt genome, which is consistent with previous analyses (Castro and Dowton 2007; Dowton et al. 2009a; Cameron et al. 2009). By contrast, exclusion of third codon positions has not had a positive effect on phylogenetic reconstruction in some insect orders (Diptera: Cameron et al. 2007; Orthoptera: Fenn et al. 2008). Dataset alignment strategies may also affect the recovered topology (Cameron et al. 2004; Cameron et al. 2009). Comparison of the topologies recovered from the datasets aligned by Clustal and Muscle (both with third codon positions excluded) indicated that the topology produced from the Muscle alignment was more congruent with expected relationships than the Clustal alignment. Analysis of both datasets recovered a range of monophyletic superfamilies, such as Apoidea, Chalcidoidea and Ichneumonoidea, but the relationships within the superfamilies varied between the two strategies. For example, within the Braconidae, the Muscle analysis recovered relationships entirely congruent with the most recent analysis of Sharanowski et al. (2011); the monophyly of the noncyclostomes (*Cotesia* + *Phanerotoma* + *Macrocentrus* + *Meteorus*), the monophyly of the cyclostomes (*Spathius* + *Diachasmimorpha*), the monophyly of

the microgastroids (*Cotesia* + *Phanerotoma*), and a sister relationship between the cyclostomes and the Aphidiinae (*Aphidius*). By contrast, the Clustal topology failed to support these expected relationships. This suggests that the phylogenetic relationships at lower taxonomic level are more sensitive to the alignment strategy.

For the reasons outlined above, we present the tree of the dataset aligned by Muscle, excluding the third codon positions (Fig. 4). Relationships among the Symphyta are recovered as expected – there are only three symphytans included, and they form a paraphyletic grade at the base of the Hymenoptera; Tenthredinoidea + (Cephoidea + (Orussoidea + Apocrita). As expected, a monophyletic Aculeata is recovered, as is a monophyletic Apoidea. *Evania* + Aculeata form a clade, consistent with the resolution of the Aculeata inside the Evaniomorpha in more taxon-rich analyses (Castro and Dowton 2006; Heraty et al. 2011; Sharkey et al. 2011). However, support for a range of relationships within the Aculeata is low, with posterior probabilities below 0.95. Terminal branch lengths are very long compared with internode lengths. Further, compared with previous analyses, internal relationships within the Aculeata are less congruent with expectation, despite the increase in taxonomic sampling. For example, the Chrysidoidea are not recovered as monophyletic; although Heraty et al. (2011) also failed to recover a monophyletic Chrysidoidea. Similarly, the Vespoidea are polyphyletic, but this is in agreement with recent, broader analyses of hymenopteran relationships (Heraty et al., 2011; Sharkey et al., 2011). Alternatively, Pilgrim et al. (2008) recovered a paraphyletic Vespoidea. Together, these observations suggest that the uncontroversial relationships recovered by Dowton et al. (2009) may have been due to the narrow taxon sampling – for example, the two vespoids are closely related (both from the Vespidae), biasing the analysis toward a finding of monophyly. It suggests that a higher taxonomic sampling is necessary to reliably recover relationships within the aculeates.

With respect to the Proctotrupomorpha, four taxa from three superfamilies (Proctotrupoidea, Platygastroidea and Chalcidoidea) are included in our analysis. The four taxa failed to form a clade as *Vanhornia* (Proctotrupoidea) was misplaced, while the Chalcidoidea and Platygastroidea were recovered as a grade above the Ichneumonoidea. Within the Ichneumonoidea, the well-established sister group relationship between the Ichneumonidae and Braconidae are well

supported (Rasnitsyn 1988; Dowton et al. 1997). In addition, the two groups (cyclostomes and non-cyclostomes) of Braconidae were well recovered.

## Conclusions

We obtained the nearly complete mt genome of *T. basalis*, which is the first record of a mt genome of the superfamily Platygastroidea. Comparison of the genome organization of *T. basalis* with other Proctotrupomorpha taxa identifies a number of gene rearrangements, which might provide informative characters for the resolution of relationships among the Proctotrupomorpha. Phylogenetic analyses using the entire mt genome sequences indicate that, despite the large number of characters provided by the mitochondrial genome, increased taxonomic sampling is likely necessary to reliably recover evolutionary relationships at the family and superfamily levels in the Hymenoptera.

## Acknowledgments

Thanks to H. Klompen (Ohio State University), R.W. Hall (Ohio State University), P. Yan (Ohio State University) and T. Ganguly (University of Pennsylvania) for assistance with the laboratory work. This material is based upon work supported in part by the National Science Foundation under grant No. DEB-0614764 to N.F. Johnson and A.D. Austin. Additional financial support for sequencing was provided by the College of Arts & Sciences of The Ohio State University and the Dorothy and Josef Knull Funds.

## References

- Asakawa, S., Kumazawa, Y., Araki, T., Himeno, H., Miura, K., and Watanabe, K. 1991. Strand-specific nucleotide composition bias in echinoderm and vertebrate mitochondrial genomes . J. Mol. Evol. **32**: 511-520.
- Austin, A.D., Johnson, N.F., and Dowton, M. 2005. Systematics, evolution, and biology of scelionid and platygastriid wasps. Annu. Rev. Entomol. **50**: 553-582.
- Beckenbach, A.T. 2011. Mitochondrial genome sequences of representatives of three families of scorpionflies (Order Mecoptera) and evolution in a major duplication of coding sequence. Genome **54**: 368-376.

- Beckenbach, A.T., and Stewart, J.B. 2009. Insect mitochondrial genomics 3: the complete mitochondrial genome sequences of representatives from two neuropteroid orders: a dobsonfly (order Megaloptera) and a giant lacewing and an owlfly (order Neuroptera). *Genome* **52**: 31-38.
- Boore, J.L., Collins, T.M., Stanton, D., Daehler, L.L., and Brown, W.M. 1995. Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements. *Nature* **376**: 163-165.
- Boore, J. L. 1999. Animal mitochondrial genomes. *Nucleic Acids Res.* **27**: 1767-1780.
- Cameron, S.L., Miller, K.B., D'Haese, C.A., Whiting, M.F., and Barker, S.C. 2004. Mitochondrial genome data alone are not enough to unambiguously resolve the relationships of Entognatha, Insecta and Crustacea sensu lato (Arthropoda). *Cladistics* **20**: 534-557.
- Cameron, S.L., Lambkin, C.L., Barker, S.C., and Whiting, M.F. 2007. A mitochondrial genome phylogeny of Diptera: whole genome sequence data accurately resolve relationships over broad timescales with high precision. *Syst. Entomol.* **32**: 40-59.
- Cameron, S.L., Dowton, M., Castro, L.R., Ruberu, K., Whiting, M.F., Austin, A.D., Diement, K., and Stevens, J. 2008. Mitochondrial genome organization and phylogeny of two vespid wasps. *Genome* **51**: 800-808.
- Cameron, S.L., Sullivan, J., Song, H.J., Miller, K.B., and Whiting, M.F. 2009. A mitochondrial genome phylogeny of the Neuropterida (lace-wings, alderflies and snakeflies) and their relationship to the other holometabolous insect orders. *Zool. Scr.* **38**: 575-590.
- Castro, L. R., and Dowton, M. 2005. The position of the Hymenoptera within the Holometabola as inferred from the mitochondrial genome of *Perga condei* (Hymenoptera: Symphyta: Pergidae). *Mol. Phylogenet. Evol.* **34**: 469-479.
- Castro, L.R., and Dowton, M. 2006. Molecular analyses of the Apocrita (Insecta : Hymenoptera) suggest that the Chalcidoidea are sister to the diaprioid complex. *Invertebr. Syst.* **20**: 603-614.
- Castro, L.R., and Dowton, M. 2007. Mitochondrial genomes in the Hymenoptera and their utility as phylogenetic markers. *Syst. Entomol.* **32**: 60-69.
- Castro, L.R., Ruberu, K., and Dowton, M. 2006. Mitochondrial genomes of *Vanhornia eucnemidarum* (Apocrita: Vanhorniidae) and *Primeuchroeus* spp. (Aculeata: Chrysididae): evidence of rearranged mitochondrial genomes within the Apocrita (Insecta: Hymenoptera). *Genome* **49**: 752-766.
- Cha, S.Y., Yoon, H.J., Lee, E.M., Yoon, M.H., Hwang, J.S., Jin, B.R., Han, Y.S., and Kim, I. 2007. The



- complete nucleotide sequence and gene organization of the mitochondrial genome of the bumblebee, *Bombus ignitus* (Hymenoptera: Apidae). *Gene* **392**: 206-220.
- Clary, D.C., and Wolstenholme, D.R. 1985. The Mitochondrial DNA Molecule of *Drosophila yakuba*: Nucleotide Sequence, Gene Organization, and Genetic Code. *J. Mol. Evol.* **22**: 252-271.
- Cook, C.E. 2005. The complete mitochondrial genome of the stomatopod crustacean *Squilla mantis*. *BMC Genomics* **6**: 105.
- Crease, T.J. 1999. The complete sequence of the mitochondrial genome of *Daphnia pulex* (Cladocera: Crustacea). *Gene* **233**: 89-99.
- Crozier, R.H., and Crozier, Y.C. 1993. The Mitochondrial Genome of the Honeybee *Apis mellifera*: Complete Sequence and Genome Organization. *Genetics* **133**: 97-117.
- Dowton, M. 1999. Relationships among the cyclostome braconid (Hymenoptera: Braconidae) subfamilies inferred from a mitochondrial tRNA gene rearrangement. *Mol. Phylogenet. Evol.* **11**: 283-287.
- Dowton, M., Cameron, S. L., Austin, A. D., and Whiting, M.F. 2009a. Phylogenetic approaches for the analysis of mitochondrial genome sequence data in the Hymenoptera – A lineage with both rapidly and slowly evolving mitochondrial genomes. *J. Mol. Evol.* **52**: 512-519.
- Dowton, M., Cameron, S.L., Dowavic, J.I., Austin, A.D., and Whiting, M.F. 2009b. Characterization of 67 mitochondrial tRNA gene rearrangements in the Hymenoptera suggests that mitochondrial tRNA gene position is selectively neutral. *Mol. Biol. Evol.* **26**: 1607-1617.
- Dowton, M., and Austin, A.D. 1994. Molecular Phylogeny of the Insect Order Hymenoptera: Apocritan Relationships. *Proc. Natl. Acad. Sci. USA* **91**: 9911-9915.
- Dowton M., and Austin, A.D. 1999. Evolutionary dynamics of a mitochondrial rearrangement " hot spot" in the Hymenoptera. *Mol. Biol. Evol.* **16**: 298-309.
- Dowton, M., and Austin, A.D. 2001. Simultaneous analysis of 16S, 28S, COI and morphology in the Hymenoptera: Apocrita – evolutionary transitions among parasitic wasps. *Biol. J. Linn. Soc.* **74**: 87-111.
- Dowton, M., Austin, A.D., Dillon, N., and Bartowsky, E. 1997. Molecular phylogeny of the apocritan wasps: the Proctotrupomorpha and Evanioromorpha. *Syst. Entomol.* **22**: 245-255.
- Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**: 1792-1797.

- Fenn, J.D., Song, H., Cameron, S.L., and Whiting, M.F. 2008. A preliminary mitochondrial genome phylogeny of Orthoptera (Insecta) and approaches to maximizing phylogenetic signal found within mitochondrial genome data. *Mol. Phylogenet. Evol.* **49**: 59-68.
- Flook, P., Rowell, H., and Gellissen, G. 1995. Homoplastic rearrangements of insect mitochondrial tRNA genes *Naturwissenschaften* **82**: 336-337.
- Foster, P.G., Jermini, L.S., and Hickey, D.A. 1997. Nucleotide composition bias affects amino acid content in proteins coded by animal mitochondria. *J. Mol. Evol.* **44**: 282-288.
- Friedrich, M., and Muqim, N. 2003. Sequence and phylogenetic analysis of the complete mitochondrial genome of the flour beetle *Tribolium castaneum*. *Mol. Phylogenet. Evol.* **26**: 502-512.
- Gotzek, D., Clarke, J., and Shoemaker, D. 2010. Mitochondrial genome evolution in fire ants (Hymenoptera: Formicidae). *BMC Evol. Biol.* **10**: 300.
- Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**: 95-98.
- Hasegawa, E., Kobayashi, K., Yagi, N., and Tsuji, K. 2011. Complete mitochondrial genomes of normal and cheater morphs in the parthenogenetic ant *Pristomyrmex punctatus* (Hymenoptera: Formicidae). *Myrmecological News* **15**: 85-90.
- Heraty, J., Ronquist, F., Carpenter, J.M., Hawks, D., Schulmeister, S., Dowling, A.P., Murray, D., Munro, J., Wheeler, W.C., Schiff, N., and Sharkey, M. 2011. Evolution of the hymenopteran megaradiation. *Mol. Phylogenet. Evol.* **60**: 73-88.
- Huelsenbeck, J.P., and Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**: 754-755.
- Kim, K.G., Hong, M.Y., Kim, M.J., Im, H.H., Kim, M.I., Bae, C.H., Seo, S.J., Lee, S.H., and Kim, I. 2009. Complete mitochondrial genome sequence of the yellow-spotted long-horned beetle *Psacotha hilaris* (Coleoptera: Cerambycidae) and phylogenetic analysis among coleopteran insects. *Mol. Cells* **27**: 429-441.
- LaSalle, J., and Gauld, I.D. 1993. Hymenoptera: their diversity and their impact on the diversity of other organisms. *In Hymenoptera and Biodiversity. Edit by J. LaSalle and Gauld I. D.* CAB International Wallingford, U.K. pp.1-26.
- Lee, E.-S., Shin, K.S., Kim, M.-S., Park, H., Cho, S., and Kim, C.-B. 2006. The mitochondrial genome of the smaller tea tortrix *Adoxophyes honmai* (Lepidoptera: Tortricidae). *Gene* **373**: 52-57.

- Li, H., Gao, J., Liu, H., Liu, H., Liang, A., Zhou, X., and Cai, W. 2011. The architecture and complete sequence of mitochondrial genome of an assassin bug *Agriosphodrus dohrni* (Hemiptera: Reduviidae). *Int. J. Biol. Sci.* **7**: 792-804.
- Lowe, T.M., and Eddy, S.R. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**: 955-964.
- Matsumoto, Y., Yanase, T., Tsuda, T., and Noda, H. 2009. Species-specific mitochondrial gene rearrangements in biting midges and vector species identification. *Med. Vet. Entomol.* **23**: 47-55.
- Miya, M., Takeshima, H., Endo, H., Ishiguro, N.B., Inoue, J.G., Mukai, T., Satoh, T.P., Yamaguchi, M., Kawaguchi, A., Mabuchi, K., Shirai, S.M., and Nishida, M. 2003. Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. *Mol. Phylogenet. Evol.* **26**: 121-138.
- Nei, M., and Kumar, S. 2000. *Molecular evolution and phylogenetics*. Oxford University Press, Oxford. pp. 1-333.
- Nylander, J.A.A., Ronquist, F., Huelsenbeck, J.P., and Nieves-Aldrey, J.L. 2004. Bayesian Phylogenetic Analysis of Combined Data. *Syst. Biol.* **53**: 47-67.
- Ojala, D., Montoya, J., and Attardi, G. 1981. tRNA punctuation model of RNA processing in human mitochondria. *Nature* **290**: 470-474.
- Oliveira, D.C.S.G., Raychoudhury, R., Lavrov, D.V., and Werren, J.H. 2008. Rapidly evolving mitochondrial genome and directional selection in mitochondrial genes in the parasitic wasp *nasonia* (Hymenoptera: Pteromalidae). *Mol. Biol. Evol.* **25**: 2167-2180.
- Pilgrim, E.M., Von Dohlen, C.D., and Pitts, J.P. 2008. Molecular phylogenetics of Vespoidea indicate paraphyly of the superfamily and novel relationships of its component families and subfamilies. *Zool. Scr.* **37**: 539-560.
- Rasnitsyn, A.P. 1988. An outline of the evolution of hymenopterous insects. *Orient. Insects* **22**: 115-145.
- Rasnitsyn, A.P., and Zhang, H.C. 2010. Early evolution of Apocrita (Insecta, Hymenoptera) as indicated by new findings in the middle Jurassic of Daohugou, northeast China. *Acta. Geol. Sin-Engl.* **84**: 834-873.
- Ronquist, F., and 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., and Lindgren, M. 1999. Phylogeny of the Hymenoptera: A cladistic reanalysis of Rasnitsyn's (1988) data. *Zool. Scr.* **28**: 13-50.
- Savard, J., Tautz, D., Richards, S., Weinstock, G.M., Gibbs, R.A., Werren, J.H., Tettelin, H., and Lercher, M.J. 2006. Phylogenomic analysis reveals bees and wasps (Hymenoptera) at the base of the radiation of Holometabolous insects. *Genome Res.* **16**: 1334-1338.
- Segovia, R., Pett, W., Trewick, S., and Lavrov, D.V. 2011. Extensive and evolutionarily persistent mitochondrial tRNA editing in velvet worms (phylum Onychophora). *Mol. Biol. Evol.* **28**: 2873-2881.
- Shao, R., and Barker, S.C. 2003. The highly rearranged mitochondrial genome of the plague thrips, *Thrips imaginis* (Insecta: Thysanoptera): convergence of two novel gene boundaries and an extraordinary arrangement of rRNA genes. *Mol. Biol. Evol.* **20**: 362-370.
- Sharanowski, B.J., Dowling, A.P.G., and Sharkey, M.J. 2011. Molecular phylogenetics of Braconidae (Hymenoptera: Ichneumonoidea), based on multiple nuclear genes, and implications for classification. *Syst. Entomol.* **36**: 549-572.
- Sharkey, M.J., Carpenter, J.M., Vilhelmsen, L., Heraty, J., Liljeblad, J., Dowling, A.P.G., Schulmeister, S., Murray, D., Deans, A.R., and Ronquist, F. 2011. Phylogenetic relationships among superfamilies of Hymenoptera. *Cladistics* **27**: 1-33.
- Silvestre, D., Dowton, M., and Arias, M.C. 2008. The mitochondrial genome of the stingless bee *Melipona bicolor* (Hymenoptera, Apidae, Meliponini): Sequence, gene organization and a unique tRNA translocation event conserved across the tribe Meliponini. *Genet. Mol. Biol.* **31**: 451-460.
- Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., and Flook, P. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Am.* **87**: 651-701.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* (submitted).
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673-4680.
- Wei, S.J., Shi, M., He, J. H., Sharkey, M., and Chen, X.X. 2009. The complete mitochondrial genome

- of *Diadegma semiclausum* (Hymenoptera: Ichneumonidae) indicates extensive independent evolutionary events. *Genome* **52**: 308-319.
- Wei, S.J., Shi, M., Sharkey, M.J., van Achterberg, C., and Chen, X.X. 2010a. Comparative mitogenomics of Braconidae (Insecta: Hymenoptera) and the phylogenetic utility of mitochondrial genomes with special reference to Holometabolous insects. *BMC Genomics* **11**: 371.
- Wei, S.J., Tang, P., Zheng, L.H., Shi, M., and Chen, X.X. 2010b. The complete mitochondrial genome of *Evania appendigaster* (Hymenoptera: Evaniidae) has low A+T content and a long intergenic spacer between atp8 and atp6. *Mol. Biol. Rep.* **37**: 1931-1942.
- Wiegmann, B.M., Trautwein, M.D., Kim, J.W., Cassel, B.K., Bertone, M.A., Winterton, S.L., and Yeates, D.K. 2009. Single-copy nuclear genes resolve the phylogeny of the holometabolous insects. *BMC Biol.* **7**: 34.
- Zhang, D.X., and Hewitt, G.M. 1997. Insect mitochondrial control region: A review of its structure, evolution and usefulness in evolutionary studies. *Biochem. Syst. Ecol.* **25**: 99-120.
- Zhou, Z., Huang, Y., Shi, F., and Ye, H. 2009. The complete mitochondrial genome of *Deracantha onos* (Orthoptera: Bradyporidae). *Mol. Biol. Rep.* **36**: 7-12.

## Figure Legends

Figure 1. Map of the circular mitochondrial genome of *T. basalis*. Gene abbreviations are as follow: *a6* and *a8* refer to ATPase subunit genes, *cob* refers to cytochrome b gene, *cox1*, *cox2* and *cox3* refer to cytochrome oxidase subunit genes, *nd1-nd6* refers to NADH dehydrogenase subunit genes, *rrnL* and *rrnS* refer to the large and small ribosomal RNA subunit genes. *trn* genes are indicated by single-letter amino acid codes, *L<sub>1</sub>*, *L<sub>2</sub>*, *S<sub>1</sub>* and *S<sub>2</sub>* denote *trnL<sup>UAG</sup>*, *trnL<sup>UAA</sup>*, *trnS<sup>UCU</sup>* and *trnS<sup>UGA</sup>*. Genes are transcribed in the anticlockwise direction except those indicated by underlining.

Figure 2. Inferred secondary structure of the 21 *trn* genes identified in the mt genome of *T. basalis*. Watson-Crick pairs are indicated by lines, wobble GU pairs are indicated by dots and other noncanonical pairs are indicated by asterisks.

Figure 3. Mitochondrial genome organizations of four Proctotrupomorpha taxa, compared with the ancestral pancrutacean mt genome organization (Crease 1999, Cook 2005). Gene transcription direction and abbreviations are as described in Fig. 1. Gene movements, relative to the ancestral organization, are indicated with arrows.

Figure 4. Bayesian analysis of mitogenomic sequences based on the Muscle alignment dataset, including 1<sup>st</sup> and 2<sup>nd</sup> codon positions from protein-coding genes, and the two *rrn* genes. *Tribolium* (Coleoptera) was specified as the outgroup. Posterior probabilities are shown at each node.



0.2

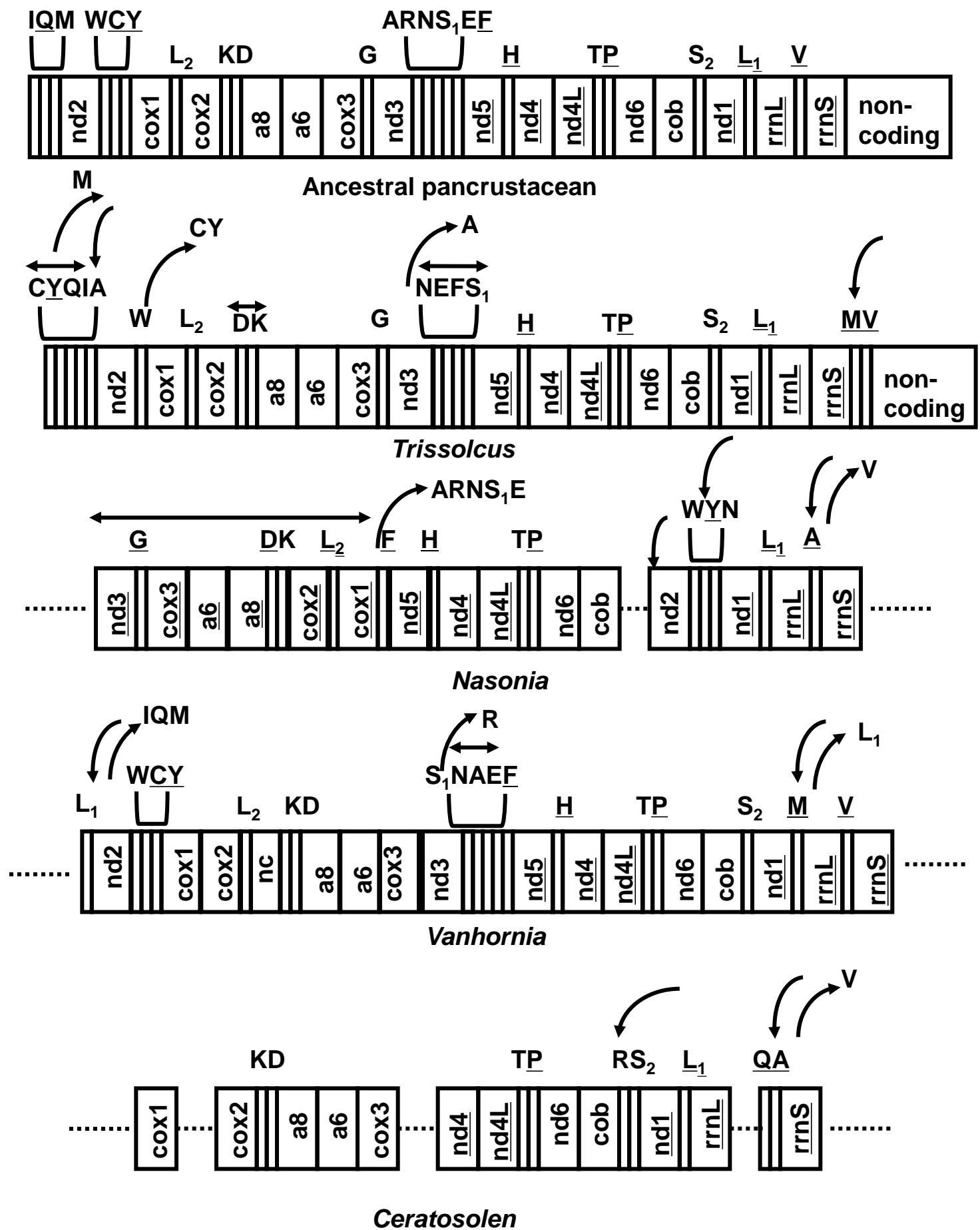




Table 1 List of taxa used in phylogenetic analyses

Order	Family	Genus	GenBank Accession	Referance
Coleoptera	Tenebrionidae	<i>Tribolium</i>	AJ312413	Friedrich and Muqim 2003
Diptera	Ceratopogonidae	<i>Culicoides</i>	AB361004	Matsumoto et al. 2009
Mecoptera	Bittacidae	<i>Bittacus</i>	HQ696578	Beckenbach unpublished
Megaloptera	Sialidae	<i>Sialis</i>	FJ859905	Cameron et al. 2009
Neuroptera	Ascalaphidae	<i>Ascaloptynx</i>	FJ171324	Beckenbach and Stewart 2009
Lepidoptera	Tortricidae	<i>Adoxophyes</i>	NC_008141	Shin 2006
Raphidioptera	Raphidiidae	<i>Mongoloraphidia</i>	FJ859902	Cameron et al. 2009
Hymenoptera				
<b>Suborder</b>				
Symphyta				
Cephoidea	Cephidae	<i>Cephus</i>	NC_012688	Dowton et al. 2009
Orussoidea	Orussidae	<i>Orussus</i>	NC_012689	Dowton et al. 2009
Tenthredinoidea	Pergidae	<i>Perga</i>	AY787816	Castro and Dowton 2005
Apocrita				
Apoidea	Apidae	<i>Apis</i>	NC_001566	Crozier and Crozier 1993
	Apidae	<i>Bombus</i>	DQ870926	Kim and Cha 2007
	Apidae	<i>Melipona</i>	NC_004529	Silvestre and Arias 2004
Chalcidoidea	Agaonidae	<i>Ceratosolen</i>	AH016005	Chen et al. unpublished
	Pteromalidae	<i>Nasonia</i>	NW_001815691	Oliveira et al. 2008
Chrysidoidea	Bethylidae	<i>Cephalonomia</i>	FJ823227	Wei et al. unpublished
	Chrysididae	<i>Primeuchroeus</i>	AH015389	Castro et al. 2006
Evanioidea	Evaniidae	<i>Evania</i>	NC_013238	Wei et al. 2010
Ichneumonoidea	Braconidae	<i>Aphidius</i>	GU097658	Wei et al. 2010
	Braconidae	<i>Cotesia</i>	NC_014272	Wei et al. 2010
	Braconidae	<i>Diachasmimorpha</i>	GU097655	Wei et al. 2010
	Braconidae	<i>Macrocentrus</i>	GU097656	Wei et al. 2010
	Braconidae	<i>Meteorus</i>	GU097657	Wei et al. 2010
	Braconidae	<i>Phanerotoma</i>	GU097654	Wei et al. 2010
	Braconidae	<i>Spathius</i>	NC_014278	Wei et al. 2010
	Ichneumonidae	<i>Diadegma</i>	NC_012708	Wei et al. 2009
	Ichneumonidae	<i>Enicospilus</i>	FJ478177	Dowton et al. 2009
	Ichneumonidae	<i>Venturia</i>	FJ478176	Dowton et al. 2009
Platygastroidea	Scelionidae	<i>Trissolcus</i>	present study	
Proctotrupoidea	Vanhorniidae	<i>Vanhornia</i>	NC_008323	Castro et al. 2006
Stephanoidea	Stephanidae	<i>Schlettererius</i>	FJ478175	Dowton et al. 2009
Vespoidea	Formicidae	<i>Pristomyrmex</i>	NC_015075	Yagi et al. 2011
	Formicidae	<i>Solenopsis</i>	NC_014669	Gotzek et al. 2010
	Mutillidae	<i>Radoszkowskius</i>	NC_014485	Wei et al. 2010
	Vespidae	<i>Abispa</i>	NC_011520	Cameron et al. 2008
	Vespidae	<i>Polistes</i>	EU024653	Cameron et al. 2008

Table 2 Locations and nucleotide (nt) sequence lengths of genes in the *Trissolcus* mt genome, the lengths of predicted amino acid (aa) sequences of protein-coding genes and their initiation and termination codons

Gene/region	nt position	Size		Codon	
		No. of nt	No. of aa	Initiation	Termination
<i>trnC</i>	2-62	61			
<i>trnY</i>	82-147	66			
<i>trnQ</i>	178-249	73			
<i>trnI</i>	253-318	66			
<i>trnA</i>	322-387	66			
<i>nd2</i>	447-1425	979	327	ATA	T
<i>trnW</i>	1426-1489	64			
<i>cox1</i>	1490-3025	1536	512	ATA	TAA
<i>trnAL<sub>2</sub></i>	3040-3109	70			
<i>cox2</i>	3110-3787	678	226	ATT	TAA
<i>trnD</i>	3790-3858	69			
<i>trnK</i>	3856-3927	72			
<i>a8</i>	3928-4095	168	56	ATT	TAA
<i>a6</i>	4079-4737	659	220	ATG	TA
<i>cox3</i>	4738-5530	793	265	ATG	T
<i>trnG</i>	5531-5598	68			
<i>nd3</i>	5605-5949	345	115	ATA	TAA
<i>trnN</i>	5951-6019	68			
<i>trnE</i>	6017-6080	64			
<i>trnF</i>	6084-6150	67			
<i>trnS<sub>1</sub></i>	6154-6212	59			
<i>nd5</i>	6275-7984	1710	570	ATT	TAA
<i>trnH</i>	7985-8049	65			
<i>nd4</i>	8080-9429	1350	450	ATG	TAA
<i>nd4L</i>	9423-9695	273	91	ATA	TAG
<i>trnT</i>	9709-9774	66			
<i>trnP</i>	9774-9837	64			
<i>nd6</i>	9839-10428	590	197	ATA	TA
<i>cob</i>	10429-11565	1137	379	ATG	TAA
<i>trnS<sub>2</sub></i>	11581-11650	70			
<i>nd1</i>	11655-12566	912	304	ATA	TAA
<i>trnL<sub>1</sub></i>	12573-12638	66			
<i>rrnL</i>	12639-13908	1270			
<i>rrnS</i>	13909-14660	752			
<i>trnM</i>	14661-14727	67			
<i>trnV</i>	14729-14793	65			

Table 3 Nucleotide composition (%) of the *Trissolcus* mt genome

Nucleotide		Length (bp)	A	C	T	G	A+T
Entire sequence		15768	43.7	9.8	40.5	6.0	84.2
Protein-coding sequence		11130	36.5	9.1	45.7	8.6	82.2
J-strand		6885	38.5	11.4	42.8	7.4	81.3
N-strand		4245	33.4	5.5	50.5	10.6	83.9
Codon position	1st	3710	40.9	8.7	38.0	12.4	78.9
	2nd	3710	23.6	14.0	51.8	10.6	75.4
	3rd	3710	45.0	4.8	47.3	2.9	92.3
Ribosomal RNA gene sequences		2022	42.7	4.0	45.1	8.1	87.9
Transfer RNA gene sequences		1396	46.4	4.7	42.3	6.5	88.7
J-strand		1002	46.6	5.3	42.1	5.9	88.7
N-strand		394	45.9	3.3	42.6	7.9	88.5
AT-rich region		974	42.8	5.1	47.0	5.0	89.8

Table 4 Codon usage for protein-coding genes of the *Trissolcus* mt genome

codon	aa	nc	RSCU	codon	aa	nc	RSCU	codon	aa	nc	RSCU	codon	aa	nc	RSCU
TTT	Phe	407	1.92	TCT	Ser	92	2.45	TAT	Tyr	161	1.69	TGT	Cys	25	1.85
TTC	Phe	18	0.08	TCC	Ser	9	0.24	TAC	Tyr	30	0.31	TGC	Cys	2	0.15
TTA	Leu	460	5.23	TCA	Ser	102	2.71	TAA	*	12	1.85	TGA	Trp	79	1.93
TTG	Leu	15	0.17	TCG	Ser	4	0.11	TAG	*	1	0.15	TGG	Trp	3	0.07
CTT	Leu	28	0.32	CCT	Pro	41	1.45	CAT	His	50	1.64	CGT	Arg	8	0.84
CTC	Leu	1	0.01	CCC	Pro	17	0.6	CAC	His	11	0.36	CGC	Arg	1	0.11
CTA	Leu	23	0.26	CCA	Pro	55	1.95	CAA	Gln	54	1.93	CGA	Arg	26	2.74
CTG	Leu	1	0.01	CCG	Pro	0	0.0	CAG	Gln	2	0.07	CGG	Arg	3	0.32
ATT	Ile	465	1.88	ACT	Thr	48	1.52	AAT	Asn	252	1.82	AGT	Ser	20	0.53
ATC	Ile	30	0.12	ACC	Thr	11	0.35	AAC	Asn	25	0.18	AGC	Ser	0	0.0
ATA	Met	341	1.92	ACA	Thr	66	2.1	AAA	Lys	155	1.9	AGA	Ser	70	1.86
ATG	Met	15	0.08	ACG	Thr	1	0.03	AAG	Lys	8	0.1	AGG	Ser	4	0.11
GTT	Val	55	1.93	GCT	Ala	34	1.92	GAT	Asp	46	1.74	GGT	Gly	17	0.45
GTC	Val	4	0.14	GCC	Ala	7	0.39	GAC	Asp	7	0.26	GGC	Gly	4	0.11
GTA	Val	48	1.68	GCA	Ala	27	1.52	GAA	Glu	61	1.74	GGA	Gly	99	2.61
GTG	Val	7	0.25	GCG	Ala	3	0.17	GAG	Glu	9	0.26	GGG	Gly	32	0.84

aa, amino acid; nc, number of codons; RSCU, relative synonymous codon usage; \*, stop codon.