Molecular Phylogenetics and Evolution xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Higher-level phylogeny of the Hymenoptera inferred from mitochondrial genomes

Meng Mao *, Tracey Gibson, Mark Dowton

Centre for Medical Bioscience, School of Biological Sciences, University of Wollongong, Wollongong, NSW 2522, Australia

ARTICLE INFO

Article history:
Received 16 July 2014
Revised 27 November 2014
Accepted 16 December 2014
Available online xxxx

Keywords:
Mitochondrial genome
Apocrita
Phylogeny
Gene rearrangement
Inverted duplicated A + T rich regions

ABSTRACT

Higher-level hymenopteran relationships remain unresolved in both morphological and molecular analyses. In this study, we present the most comprehensive analyses of hymenopteran relationships based on 48 mitochondrial (mt) genomes. One complete and two nearly complete mt genomes representing three hymenopteran superfamilies were newly sequenced. We assessed the influence of inclusion/exclusion of 3rd codon positions, alignment approaches, partition schemes and phylogenetic approaches on topology and nodal support within the Hymenoptera. The results showed that the topologies were sensitive to the variation of dataset and analytical approach. However, some robust and highly supported relationships were recovered: the Ichneumonomorpha was monophyletic; the Trigonalyoidea + Megalyroidea and the Diaprioidea + Chalcidoidea were consistently recovered; the Cynipoidea was generally recovered as the sister group to the Diaprioidea + Chalcidoidea. In addition, the monophyletic Aculeata and Proctotrupomorpha were recovered in some analyses. Several gene rearrangements were detected in each of the three newly sequenced mt genomes. Specifically, the *Ibalia leucospoides* mt genome harbors a large inversion of a gene block from *trnE* to $trnS_2$. Inverted, duplicated A + T rich regions were detected in the *Ibalia leucospoides* mt genome, which probably played an important role during the formation of the large gene block inversion via recombination.

1. Introduction

The Hymenoptera, comprising of sawflies, wasps, bees and ants, is one of the most speciose insect orders. They live in all terrestrial and some aquatic habitats (Huber, 2009). The Hymenoptera conventionally comprises two suborders, the Symphyta and the Apocrita, with 27 superfamilies (9 superfamilies in the Symphyta and 18 in the Apocrita) (Aguiar et al., 2013). Members of the Symphyta are more primitive, with phytophagous lifestyle. The Orussidae is the only parasitoid symphytan family (Sharkey, 2007). More than 90% of the described hymenopteran species belong to the Apocrita (Huber, 2009). As one of the most biologically diverse suborder of insects, the Apocrita exhibits a rich spectrum of lifestyles including parasitoidism, predation, omnivory, mycophagy and secondary reversals to phytophagy (Huber, 2009; New, 2012).

Robust phylogenetic relationships could provide a framework to understand the evolution of the various biologies exhibited by the Apocrita (Dowton and Austin, 2001). However, the higher level relationships within the Apocrita remain contentious (Sharkey, 2007). Based on morphological and fossil evidence, Rasnitsyn (1988) proposed a new infraorder system, which formed a framework for subsequent research of hymenopteran phylogeny. In this system, the Apocrita was divided into four infraorders, the Ichneumonomorpha, the Vespomorpha (Aculeata), the Proctotrupomorpha and the Evaniomorpha, with the Ichneumonomorpha sister to the Aculeata and the Proctotrupomorpha sister to the Evaniomorpha. The monophyly of the Ichneumonomorpha and Aculeata was consistently supported by subsequent morphological and molecular analyses (Castro and Dowton, 2006; Dowton and Austin, 1994, 2001; Dowton et al., 1997; Heraty et al., 2011; Klopfstein et al., 2013; Ronquist et al., 1999; Sharkey et al., 2011; Vilhelmsen et al., 2010). However, the sister relationship between the two infraorders was only recovered by some morphological and early molecular analyses (Dowton and Austin, 1994; Dowton et al., 1997; Rasnitsyn and Zhang, 2010; Vilhelmsen et al., 2010). Instead, most of the later molecular analyses supported the Aculeata nested within the paraphyletic Evaniomorpha and a sister relationship between the Ichneumonomorpha and the Proc-

* Corresponding author at: Centre for Medical Bioscience, School of Biological Sciences, University of Wollongong, Northfields Avenue, Wollongong, NSW 2522, Australia. Fax: +61 612 42 214135.
E-mail address: mm663@uowmail.edu.au (M. Mao).

http://dx.doi.org/10.1016/j.ympev.2014.12.009
1055-7903/© 2014 Published by Elsevier Inc.

totrupomorpha (Dowton and Austin, 2001; Heraty et al., 2011; Klopfstein et al., 2013; Sharkey et al., 2011).

The Proctotrupomorpha (including the Proctotrupoidea, Cynipoidea, Diaprioidea, Mymarommatoidea, Platygastroidea and Chalcidoidea) and Evaniomorpha (including the Ceraphronoidea, Evanioidea, Megalyroidea, Stephanoidea and Trigonalyoidea) were two novel concepts in Rasnitsyn's infraorder system. The clade Evaniomorpha has not been consistently recovered in both morphological and molecular analyses (Castro and Dowton, 2006; Gibson, 1999; Heraty et al., 2011; Klopfstein et al., 2013; Rasnitsyn and Zhang, 2010; Ronquist et al., 1999; Sharkey et al., 2011). For example, most recent analyses supported the Stephanoidea as the most basal apocritan lineage (Heraty et al., 2011; Peters et al., 2011; Sharkey et al., 2011; Vilhelmsen et al., 2010). Furthermore, recent molecular analyses recovered the Evaniomorpha as paraphyletic relative to the Aculeata (Castro and Dowton, 2006; Heraty et al., 2011; Klopfstein et al., 2013; Sharkey et al., 2011). Similarly, the monophyly of the Proctotrupomorpha was not supported by early morphological or molecular analyses. In morphological analyses, Ronquist et al. (1999) reanalyzed Rasnitsyn's (1988) data using numerical cladistic method and the trees showed that the Ceraphronoidea fell among the Proctotrupomorpha, close to the Chalcidoidea and Platygastroidea (Ronquist et al., 1999). An early molecular analysis conducted by Dowton et al. (1997) using 37 taxa based on 16S rRNA gene also failed to retrieve a monophyletic Proctotrupomorpha, as the Cynipoidea was placed as a relatively basal apocritan lineage. But the reason for this might be the limited taxon samplings (only two cynipoid representatives) (Dowton et al., 1997). Later, Dowton and Austin (2001) performed multiple analyses with a dramatically increased taxon dataset and more molecular markers (16S rRNA, COI and 28S rRNA). However, the Heloridae (Proctotrupoidea) was recovered outside the Proctotrupomorpha in most of these analyses. Interestingly, when the morphological characters were included, the Ceraphronoidea was placed within the Proctotrupomorpha (Dowton and Austin, 2001). Although later and more comprehensive molecular analyses consistently recovered the Proctotrupomorpha as monophyletic, the internal relationships remain elusive (Castro and Dowton, 2006; Heraty et al., 2011; Klopfstein et al., 2013; Sharkey et al., 2011). For example, Sharkey (2007) separated the Diaprioidea from the Proctotrupoidea as a new superfamily based on previous molecular evidence (Castro and Dowton, 2006; Dowton and Austin, 2001), but the monophyly of the Diaprioidea was not well supported by later molecular analyses (Heraty et al., 2011; Klopfstein et al., 2013).

The entire mitochondrial (mt) genome is a popular molecular marker used in phylogenetic studies mainly because of its maternal inheritance and higher rate of nucleotide substitution compared with the nuclear DNA (Moritz et al., 1987). In addition, some other mt genome characters, such as gene rearrangements, also provide useful phylogenetic information (Boore and Brown, 1998). However, poor representation of a broad range of lineages restricts the evolutionary utility of the mt genome. Among the insect orders, hymenopteran mt genomes are not well represented (Cameron, 2014). Therefore, the main purpose of the present study is to increase our understanding of the phylogeny of the Hymenoptera by extending the taxonomic range of available mt genome sequences. Here, we present three new mt genomes for representatives of three hymenopteran superfamilies: *Ibalia leucospoides* (Cynipoidea), *Monomachis antipodalis* (Diaprioidea) and *Pelecinus polyturator* (Proctotrupoidea). The mt genomes from the superfamilies Cynipoidea and Diaprioidea are sequenced for the first time. We conducted phylogenetic analyses using these new mt genomes together with the previously published mt genomes of the Hymenoptera. For the first time, we present analyses of mt genome data from representatives of every extant apocritan superfamily, with the exception of Mymarommatoidea. A range of analytical approaches were employed. In addition, we compared the organization of the newly sequenced mt genomes with other Proctotrupomorpha taxa and with the putative ancestral organization for the Hymenoptera, in order to identify any shared, derived gene rearrangements.

2. Materials and methods

2.1. DNA extraction

The collection details for each species sequenced for this study are listed in Table 1. All specimens were stored at 4 °C in the UOW wasp alcohol library before extraction. Genomic DNA was extracted from 100% ethanol preserved specimens using the 'salting out' protocol (Aljanabi and Martinez, 1997). The DNA was resuspended in 100 μl of fresh TE solution (1 mM Tris–HCl, 0.1 mM EDTA [pH 8]) and stored at 4 °C.

2.2. Mt genome amplification, sequencing, annotation and bioinformatic analysis

PCR amplifications and sequencing reactions were conducted as previously described (Mao et al., 2014a). Briefly, short fragments were amplified and sequenced using a range of universal insect mitochondrial primers (Simon et al., 1994, 2006) and primers that had been previously designed from consensus hymenopteran mt sequences. Using the sequence information obtained, taxon-specific primers were designed for each sample to amplify the remaining regions by long PCR. The primer walking method was employed to determine the complete, double stranded sequence for each long PCR product.

Raw sequences were assembled into contigs in ChromasPro Ver 1.33 (Technelysium Ltd., Tewantin, Australia). tRNA genes were identified using tRNA-scan SE 1.21 with a cove cutoff score of 5 (lowelab.ucsc.edu/tRNAscan-SE/) (Lowe and Eddy, 1997) and ARWEN 1.2 (http://130.235.46.10/ARWEN/) (Laslett and Canbäck, 2008). ORFinder (www.ncbi.nlm.nih.gov/gorf/gorf.html) was used to identify protein-coding genes, specifying the invertebrate mt genetic code. The start and stop codons of some genes were corrected according to the boundaries of tRNA genes and through alignment with other hymenopteran mt sequences. Ribosomal RNA genes were identified by sequence comparison with published hymenopteran mt ribosomal RNA sequences.

Nucleotide compositions were determined using MEGA5 (Tamura et al., 2011). The AT and GC skews were measured for the major (J) strand of each genome. The formulae used were AT-skew = (A − T)/(A + T) and GC-skew = (G − C)/(G + C) (Perna and Kocher, 1995).

2.3. Sequence alignment and phylogenetic analysis

Nucleotide sequences for each of the 13 protein-coding genes and the 2 rRNA genes were imported into separate files using MEGA5 (Tamura et al., 2011) and aligned using Muscle (Edgar, 2004) or MAFFT (Katoh et al., 2005). For the protein-coding genes (excluding the stop codons), an amino acid alignment was generated first for each gene in Muscle as implemented within MEGA5, or MAFFT at the freely available TranslatorX server (http://translatorx.co.uk/) (Abascal et al., 2010). A nucleotide alignment was then inferred from the amino acid alignment. The alignment parameters for all genes in Muscle were the default settings, which have been specified in a previous study (Mao et al., 2012). For the MAFFT alignment of the rRNA genes, we used the G-INS-i algorithm as implemented in the MAFFT web server (http://mafft.cbrc.jp/align-