Molecular Phylogenetics and Evolution 108 (2017) 22–33

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Towards a higher-level Ensifera phylogeny inferred from mitogenome sequences

Zhijun Zhou [a,*], Ling Zhao [b], Nian Liu [c], Huifang Guo [a], Bei Guan [a], Juanxia Di [a], Fuming Shi [a,*]

[a] Key Laboratory of Invertebrate Systematics and Application of Hebei Province, College of Life Sciences, Hebei University, Baoding, Hebei 071002, China
[b] College of Life Sciences and Biotechnology, Mianyang Normal University, Mianyang, Sichuan 621000, China
[c] College of Life Sciences, Shaanxi Normal University, Xi'an, Shaanxi 710062, China

ARTICLE INFO

Article history:
Received 4 October 2016
Revised 24 January 2017
Accepted 31 January 2017
Available online 8 February 2017

Keywords:
Ensifera
Gene rearrangement
Mitogenome
Next generation sequencing
Phylogeny

ABSTRACT

Although mitogenomes are useful tools for inferring evolutionary history, only a few representative ones can be used for most Ensifera lineages. Thirty-two ensiferan mitogenomes were determined using ABI Sanger sequencing and standard primer walking of 2–3 overlapping Long-PCR fragments, or Illumina® HiSeq2000 for "shotgun" sequenced long-PCR-amplified mitochondrial or total genomic DNA. Six patterns of gene arrangements, including the novel *trnR-trnS^{AGN}-trnA-trnN-trnG-nad3* in *Lipotactes tripyrga* (Lipotactinae), were identified from 59 ensiferan mitogenomes. The results suggest that *trnM-trnI-trnQ* and *trnA-trnR-trnE-trnS^{AGN}-trnN-trnF* rearrangements might be a shared derived character in Pseudophyllinae and Gryllidae, respectively. We found base composition biases in our dataset, which potentially complicate the inference of higher-level ensiferan phylogeny. Site-heterogeneous Bayesian inference (BI) and site-homogeneous maximum likelihood (ML) analyses recovered all ensiferan superfamilies as monophyletic. The site-homogeneous BI analysis failed to recover the monophyly of Stenopelmatoidea. As Schizodactyloidea was only represented by *Comicus campestris*, its monophyly could not be tested. In the Triassic/Jurassic boundary, Ensifera diverged into grylloid and non-grylloid clades. All analyses confirmed Grylloidea and Gryllotalpoidea as sister groups. Site-heterogeneous BI analysis found Schizodactyloidea as the most basal lineage and sister to the clade formed by Grylloidea and Gryllotalpoidea, but the site-homogeneous analyses placed it basally to the non-grylloid clade and recovered a sister relationship between Tettigonioidea and (Hagloidea, Rhaphidophoroidea, Stenopelmatoidea), although this clade had a low support. The site-heterogeneous BI analysis found Tettigonioidea and Hagloidea were sister groups (posterior probability (PP) = 0.99), Stenopelmatoidea was sister to (Tettigonioidea, Hagloidea) (PP > 0.91), and Rhaphidophoroidea was basal to the non-grylloid clade. At a lower level, all analyses divided Tettigonioidea into Phaneropteridae and Tettigoniidae.

1. Introduction

Ensifera is the oldest group of orthopterans and includes familiar singing insects, such as crickets, katydids, wetas, and their kin. It has over 12,000 species in approximately 2000 genera. Ensifera is interest to many biologists because several members of this groups communicate by sounds (Gwynne, 1995). However, due to a long history of conflicting classification schemes based on different interpretations of morphological characters, a fully resolved phylogeny is still lacking for Ensifera, and no stable classification has been proposed for this group to date. Taxonomists have divided Ensifera into infraorders or superfamilies, and the current classification of extant Ensifera in the Orthoptera species file v.5.0/5.0 (Eades et al., 2016) includes two extant infraorders plus two extant superfamilies, namely Rhaphidophoroidea Walker, 1869 and Schizodactyloidea Blanchard, 1845. The infraorder Gryllidea contains the superfamilies Grylloidea Laicharting, 1781 and Gryllotalpoidea Leach, 1815, and the infraorder Tettigoniidea contains three superfamilies Hagloidea Handlirsch, 1906, Stenopelmatoidea Burmeister, 1838, and Tettigonioidea Krauss, 1902 (Eades et al., 2016). The Ensifera phylogeny remains largely unresolved, its main queries include: (1) the placement of Schizodactyloidea within a non-grylloid clade based on cladistic analyses (Desutter-Grandcolas, 2003), or within a Gryllidea clade, based on the reappraisal of morphological and molecular evidence (Gwynne, 1995; Heads and Leuzinger, 2011). Based on mitogenome analysis, Song

* Corresponding authors.
E-mail addresses: zhijunzhou@163.com (Z. Zhou), shif_m@126.com (F. Shi).

http://dx.doi.org/10.1016/j.ympev.2017.01.014
1055-7903/

et al. (2015) found that Schizodactyloidea was the sister group of Grylloidea and it was placed basally to the non-grylloid clade in the total evidence analysis; (2) the single extant family within Hagloidea, i.e., Prophalangopsidae, was proposed ancestral to two large superfamilies: Tettigonioidea and Stenopelmatoidea (Gorochov, 2003). Recent molecular phylogenies have proposed various hypotheses for the sister group of katydids, including Stenopelmatidae (Fenn et al., 2008), Raphidophoridae (Mugleston et al., 2013), and (Raphidophoridae, (Stenopelmatidae, Hagloidea)) (Song et al., 2015); and (3) the widely accepted placement of katydids as a single family, Tettigoniidae (Eades et al., 2016). Recently, Phaneropterinae has been raised to the family Phaneropteridae, which includes four plant-feeding subfamilies: Phaneropterinae, Pseudophyllinae, Mecopodinae, and Phyllophorinae (Heller et al., 2014). However, most orthopterists do not support the two-family classification scheme (Song et al., 2015).

Previous studies have demonstrated that the mitogenome can be extensively applied to infer metazoan phylogeny at both deep and shallow taxonomic levels (Cameron, 2014; Miya and Nishida, 2000; Shao and Barker, 2006; Song et al., 2015; Wu et al., 2014). Mitogenome-based phylogenies are likely to be better resolved than that based on nuclear genes and provide valuable information for coalescent-based methods of species-tree estimation (Meikejohn et al., 2014). Ensiferan phylogenetic studies have increasingly relied on molecular markers over morphological characters (Yang et al., 2015), following the widespread trend of using insect mitogenomes as a source of sequence data for investigating phylogenetic relationships (Cameron, 2014). The gene content of insect mitogenome is usually highly conserved, including 13 protein coding genes (PCGs), 22 transfer RNAs (tRNAs), two ribosomal RNAs (rRNAs), and at least one large non-coding region, including a control region (CR) (Boore, 1999). To date, few molecular phylogenies have been reconstructed for Ensifera, compared to other insect orders such as Caelifera. Several publications have inferred the mitogenome phylogeny of Orthoptera based on newly sequenced mitogenomes and available mitogenomes, but these studies often had a relatively small ensiferan taxon sampling, which was not enough to assess previous classification schemes (Fenn et al., 2008; Zhou et al., 2010, 2014). Recently, a large-scale phylogenetic study of Orthoptera covered 36 of the 40 families representing all 15 currently recognized superfamilies, using complete mitogenomes and four nuclear loci data, but only including 19 ensiferan mitogenomes (Song et al., 2015).

Initially, mitogenomes were obtained by Sanger sequencing after mitochondrial isolation and DNA extraction (Flook et al., 1995) followed by conventional PCR. The former was effective for large organisms but not for small insects, and the latter was beset by the nuclear copies of mitogenome fragments, i.e., nuclear mitochondrial insertions (numts). The long-PCR based approach was used for mitogenomic sequencing at the end of the last century (Miya and Nishida, 1999; Sorenson et al., 1999), and it was regarded as the best method for small insects' mitogenomic sequencing in the long term. Although Sanger sequencing has been used to obtain high-quality mitogenomic sequences, the time-consuming primer design, and the cost of recovering large numbers of mitogenome sequences remained challenging (Wu et al., 2014). Recently, high-throughput "shotgun" sequenced Long-PCR-amplified mtDNA was utilized for sequencing mitogenome and high throughput sequencing platforms are presently used for generating mitogenomes without using PCR (Cameron 2014; Zhou et al., 2013). This has led to a rapid increase in the number of sequenced mitogenomes (Pons et al., 2014) and next-generation sequencing (NGS) technology has made acquisition of entire mitogenomes both practical and economically viable. To date (Jan. 2016), only about 27 complete or nearly complete Ensifera mitogenomes have been reported or registered in GenBank (https://www.ncbi.nlm.nih.gov/genbank/). Although the majority of these ensiferan mitogenomes were sequenced through the conventional PCR-based approach (Fenn et al., 2008; Yang et al., 2016; Zhou et al., 2014), NGS-based mitogenomes are currently replacing them. Published mitogenomes have the potential to advance the understanding of ensiferan classification, evolution, and genetics.

The recommendation to increase sampling in any phylogenetic study may not seem important; however, it is especially important for Ensifera, provided it is performed in a comprehensive manner (Legendre et al., 2010). In the present study, 26 full and six partial ensiferan mitogenomes were newly sequenced by both classic and NGS techniques to resolve the phylogenetic relationships within the suborder Ensifera and to establish a time frame for its diversification. The remaining mitogenomes used in the present study were previously sequenced and published by our team (Zhou et al., 2007, 2008, 2009, 2010, 2011, 2013, 2014) or obtained from GenBank (Suppl. Table S1). The ensiferan mitogenomes obtained here (comprising four superfamilies, four families, and 12 subfamilies) were added to those publically available, totalizing seven superfamilies, providing data to better resolve ensiferan phylogeny. Thus, the present study comprised a phylogenetic evaluation of Ensifera based on their mitogenomes, significantly expanding the data set from previous studies and contributing to a stable Ensifera phylogenetic framework that will facilitate complete investigation of the origins of sound communication. The aims of the present study were: (a) to estimate Ensifera phylogeny using mitogenome data; (b) to test the monophyly of major katydid subfamilies in China, and infer their subfamilial relationships; and (c) to further compare the phylogenetic relationships among the groups established in prior studies.

2. Materials and methods

2.1. Taxon sampling and DNA extraction

Specimens used in this study are listed in Supplementary Table S1. After collection, individuals were immediately preserved in 100% ethanol in the field, and then transferred to −20 °C at Hebei University. Total DNA was extracted from the leg muscle tissue of a single adult specimen using the TIANamp Genomic DNA Kit according to the manufacturer's instructions (Tiangen Biotech, Beijing, China).

2.2. Mitogenomes sequencing, assembly, and annotation

The mitogenomes of three Meconematinae species, namely *Decma fissa*, *Pseudokuzicus pieli*, and *Pseudocosmetura anjiensis*, were sequenced using the standard primer walking of 2–3 overlapping Long-PCR fragments protocol (Yang et al., 2012; Zhou et al., 2007). Remaining ensiferan mitogenomes were determined using Illumina® HiSeq2000 (Illumina®, San Diego, CA, USA) for "shotgun" sequencing Long-PCR-amplified mtDNA or total genomic DNA. Two or three Long-PCR amplicons (not less than 2 μg each) spanning the mitogenome were pooled in equimolar ratios. Shotgun DNA libraries with a mean fragment size of 250 bp were prepared using the Illumina® TruSeq DNA Sample Prep v2 kit according to the manufacturer's instructions. DNA libraries were pooled and paired-end sequenced (PE: 2 × 150 bp) on the Illumina® HiSeq2000 MiSeq platform using ~0.5% of the machine run capacity. The Illumina® HiSeq2000 technology was also used for mitogenome sequencing without the need for PCR, at BGI-Shenzhen (China). Based on a total genomic DNA library (insert size = 250 bp), prepared using the method mentioned for MiSeq, a 2.5-Gb mitogenome, corresponding to ~1/16th lane, was paired-end sequenced (PE: 2 × 150 bp).