Molecular Phylogenetics and Evolution 95 (2016) 171–182

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Phylogenomic analyses resolve an ancient trichotomy at the base of Ischyropsalidoidea (Arachnida, Opiliones) despite high levels of gene tree conflict and unequal minority resolution frequencies ☆

Casey H. Richart [a,b,*], Cheryl Y. Hayashi [b], Marshal Hedin [a]

[a] Department of Biology, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182, USA
[b] Department of Biology, University of California, Riverside, CA 92521, USA

ARTICLE INFO

Article history:
Received 13 March 2015
Revised 16 September 2015
Accepted 13 November 2015
Available online 9 December 2015

Keywords:
Multispecies coalescence
Concatenation
Concordance factors
Unequal minority resolution frequencies
Incomplete lineage sorting
Gene trees

ABSTRACT

Phylogenetic resolution of ancient rapid radiations has remained problematic despite major advances in statistical approaches and DNA sequencing technologies. Here we report on a combined phylogenetic approach utilizing transcriptome data in conjunction with Sanger sequence data to investigate a tandem of ancient divergences in the harvestmen superfamily Ischyropsalidoidea (Arachnida, Opiliones, Dyspnoi). We rely on Sanger sequences to resolve nodes within and between closely related genera, and use RNA-seq data from a subset of taxa to resolve a short and ancient internal branch. We use several analytical approaches to explore this succession of ancient diversification events, including concatenated and coalescent-based analyses and maximum likelihood gene trees for each locus. We evaluate the robustness of phylogenetic inferences using a randomized locus sub-sampling approach, and find congruence across these methods despite considerable incongruence across gene trees. Incongruent gene trees are not recovered in frequencies expected from a simple multispecies coalescent model, and we reject incomplete lineage sorting as the sole contributor to gene tree conflict. Using these approaches we attain robust support for higher-level phylogenetic relationships within Ischyropsalidoidea.

1. Introduction

Harvestmen (Opiliones) are among the most species-rich arachnid orders (Harvey, 2002), and have an ancient diversification history (Hedin et al., 2012; Sharma and Giribet, 2014). Within Opiliones there are 46 recognized families, approximately 1500 genera, and more than 6500 described species (Machado et al., 2007; Kury et al., 2014). Phylogenomics of higher-level relationships within Opiliones (Hedin et al., 2012) has found strong support for four primary clades (suborders Cyphophthalmi, Laniatores, Dyspnoi, and Eupnoi), with recent amendments based on novel fossil discoveries (Giribet and Sharma, 2015). Here we build upon the research of Hedin et al. (2012) to analyze relationships within the Dyspnoi superfamily Ischyropsalidoidea. The Ischyropsalidoidea are confined to the northern hemisphere, with the age of the root estimated to be as recent as 137 MYA (Schönhofer et al., 2013) or as old as 240–360 MYA (Sharma and Giribet, 2014). Currently, 85 species are classified into seven genera: *Ischyropsalis*, *Sabacon*, *Taracus*, *Ceratolasma*, *Acuclavella*, *Hesperonemastoma*, and *Crosbycus* (Kury, 2013). The superfamily is defined on the basis of genitalic characters (Martens, 1976), palpal morphology (Martens et al., 1981), and by having metapeltidial sensory cones (Shear, 1986; though see Shultz, 1998). Each genus is morphologically distinct (Fig. 1), and there has been little controversy regarding their respective monophyly. In fact, monogeneric families have been proposed for a number of genera (e.g., Dresco, 1970; Martens, 1976; Shear, 1986; Schönhofer, 2013), but this taxonomic solution was criticized by Gruber (1978). Conversely, family level hypotheses within Ischyropsalidoidea have been problematic, with two of the last three non-monogeneric familial hypotheses (Sabaconidae of Giribet et al., 2010, Taracidae of Schönhofer, 2013) failing to identify diagnostic morphological synapomorphies.

Abbreviations: AA, amino acid; AT3, AT content at the 3rd base-pair position; BCA, Bayesian concordance analysis; BIC, Bayesian information criterion; BLAST, Basic Local Alignment Search Tool; BSV, bootstrap value; CF, concordance factor; ESS, effective sample size; ILS, incomplete lineage sorting; LBA, long-branch attraction; ML, maximum likelihood; MYA, millions of years ago; NGS, next generation sequencing; NNI, nearest neighbor interchange; OTU, operational taxonomic unit; PP, posterior probability; SPR, subtree-pruning-regrafting; UMRFs, unequal minority resolution frequencies.

☆ This paper was edited by the Associate Editor Francesco Frati.

* Corresponding author at: Department of Biology, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182, USA.

E-mail address: pileated@gmail.com (C.H. Richart).

http://dx.doi.org/10.1016/j.ympev.2015.11.010
1055-7903/© 2015 Elsevier Inc. All rights reserved.

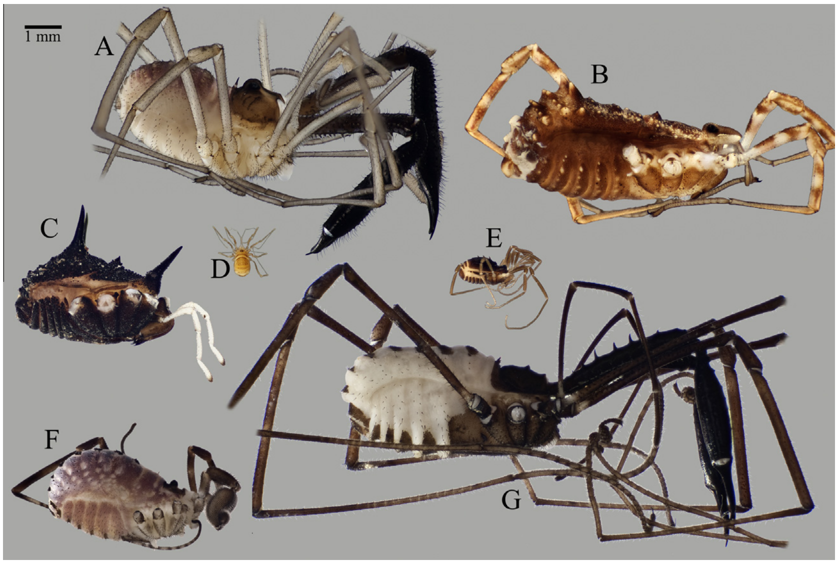

Fig. 1. Ischyropsalidoidea. Generic representatives from the superfamily Ischyropsalidoidea. A. *Taracus gertschi* (851092), B. *Ceratolasma tricantha* (850889), C. *Acuclavella makah* (829726), D. *Crosbycus dasycnemus* (851086), E. *Hesperonemastoma* sp. (851085), F. *Sabacon* sp. (851091), and G. *Ischyropsalis h. hellwigi* (851090). Full sized high-resolution images can be seen at MorphBank using the specimen identification numbers listed behind each name. Specimens were imaged using a Visionary Digital BK Plus system (http://www.visionarydigital.com) with composite images combined using Zerene Stacker 1.04 (http://www.zerenesystems.com), and edited with Adobe Photoshop CS6.

Systematics has moved into an era where phylogenetic hypotheses are being resolved at an unprecedented rate. For example, well-studied systems, such as mammalian interordinal relationships, now contain few controversial nodes (e.g., Murphy et al., 2001; Meredith et al., 2011). This development is due in part to the arrival of next-generation sequencing (NGS) technologies and continued advance in statistical phylogenetics. With the ability to generate matrices containing hundreds of loci (e.g. Hedin et al., 2012; Faircloth et al., 2012), NGS data have both supported pre-existing hypotheses, and recovered novel taxonomic hypotheses that are robustly supported. Furthermore, NGS technologies have allowed for the identification of rapid, ancient radiations (e.g., McCormack et al., 2013; Teeling and Hedges, 2013). These radiations are notoriously hard to resolve (e.g. Faircloth et al., 2012; Springer and Gatesy, 2014), with short internal branches that are in part characterized by high levels of gene tree conflict. Gene trees can conflict with a species tree for numerous reasons (Maddison, 1997; Maddison and Knowles, 2006; Degnan and Rosenberg, 2009) including undetected paralogy, recombination, hybridization, saturation, and long-branch attraction (LBA). A major source of gene tree conflict inherent to rapid diversifications is coalescent stochasticity – the random sorting of ancestral polymorphisms across successive speciation events (Kingman, 1982; Degnan and Rosenberg, 2009). This phenomenon is known as incomplete lineage sorting (ILS), and its occurrence is expected to increase as a function of shorter internal branch lengths and larger ancestral population sizes (Maddison, 1997). ILS can occur to such an extent that the most likely gene tree is incongruent with the species tree, a situation that has been defined as an "anomaly zone" (Degnan and Rosenberg, 2006).

A problem inherent to molecular phylogenetic reconstruction of ancient and rapid successive diversification events is that short internal branches do not provide enough time for slowly evolving loci to accumulate informative substitutions, whereas more rapidly evolving loci accumulate homoplastic substitutions along descending long branches (Regier et al., 2008). Such internal branches can have so little phylogenetic signal that even small amounts of non-phylogenetic signal can yield support for an incorrect phylogeny (Huelsenbeck and Hillis, 1993; Swofford et al., 2001; Philippe et al., 2011), and this can occur to such an extent that it is positively misleading (Huelsenbeck and Hillis, 1993; Bull et al., 1993). Therefore, although the selective use of slowly evolving coding regions (nucleotides or amino acids) has been a successful approach for reconstructing the backbone of numerous higher-level phylogenies (e.g., Iwabe et al., 1989; Hedin et al., 2012; Zhang et al., 2012; Lang et al., 2013; Raymann et al., 2014), these loci are not expected to contain enough informative characters to resolve the branching order of ancient internal nodes across short branches, and simply adding more data does not guarantee that analyses will resolve the correct topology (Swofford et al., 2001; Philippe et al., 2011).

Longer loci have been shown to improve phylogenetic inference in situations where long terminal branches relative to short internal branches cause "zones" of inconsistent estimation (Swofford et al., 2001). Accordingly, independent loci have traditionally been concatenated into a supermatrix, with the assumption that this will allow for the emergence of hidden support, or the increased support for a clade relative to the sum of support for the clade when data partitions are analyzed separately (Gatesy et al., 1999). The theoretical argument against the supermatrix approach is that recombination and coalescent stochasticity result in genes having different evolutionary histories, and that concatenating these loci into a supermatrix (which in effect treats all data as a single locus) violates the assumption of recombination (Kubatko and Degnan, 2007). Simulation studies under these conditions (short internal branches with high levels of gene tree conflict) have supported this contention, showing that concatenation can result in support for incorrect topologies (Seo, 2008), with support increasing as more loci are added (Kubatko and Degnan, 2007). To address these concerns, methods of phylogenetic inference have been developed under multispecies coalescent models that co-estimate gene trees, divergence times, population sizes, and a species tree from multiple unlinked loci (e.g., BEST, Liu and Pearl, 2007; *BEAST, Heled and Drummond, 2010). Simulation studies have shown that fully-parametric multispecies coalescent methods can be highly accurate even with high levels of gene tree incongruence (Liu and Edwards, 2009), and outperform supermatrix methods (Heled and Drummond, 2010). Currently, the major