Available online at www.sciencedirect.com

ScienceDirect

Molecular Phylogenetics and Evolution 43 (2007) 583–595

MOLECULAR
PHYLOGENETICS
AND
EVOLUTION

www.elsevier.com/locate/ympev

The effect of model choice on phylogenetic inference using mitochondrial sequence data: Lessons from the scorpions

Martin Jones [a,*], Benjamin Gantenbein [b], Victor Fet [c], Mark Blaxter [a]

[a] *Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JT, UK*
[b] *AO Research Institute, Clavadelerstrasse 8, Davos Platz CH-7270, Switzerland*
[c] *Department of Biological Sciences, Marshall University, Huntington, WV 25755-2510, USA*

Received 25 April 2006; revised 14 November 2006; accepted 14 November 2006
Available online 29 November 2006

Abstract

Chelicerates are a diverse group of arthropods, with around 65,000 described species occupying a wide range of habitats. Many phylogenies describing the relationships between the various chelicerate orders have been proposed. While some relationships are widely accepted, others remain contentious. To increase the taxonomic sampling of species available for phylogenetic study based on mitochondrial genomes we produced the nearly complete sequence of the mitochondrial genome of the scorpion *Mesobuthus gibbosus*. Mitochondrial gene order in *M. gibbosus* largely mirrors that in *Limulus polyphemus* but tRNA secondary structures are truncated. A recent analysis argued that independent reversal of mitochondrial genome strand-bias in several groups of arthropods, including spiders and scorpions, could compromise phylogenetic reconstruction and proposed an evolutionary model that excludes mutational events caused by strand-bias (Neutral Transitions Excluded, NTE). An arthropod dataset of six mitochondrial genes, when analyzed under NTE, yields strong support for scorpions as sister taxon to the rest of Chelicerata. We investigated the robustness of this result by exploring the effect of adding additional chelicerate genes and taxa and comparing the phylogenies obtained under different models. We find evidence that (1) placement of scorpions arising at the base of the Chelicerata is an artifact of model mis-specification and scorpions are strongly supported as basal arachnids and (2) an expanded chelicerate dataset finds support for several proposed interordinal relationships (ticks plus mites [Acari] and spiders plus whip spiders plus whip scorpions [Araneae + Pedipalpi]). Mitochondrial sequence data are subject to systematic bias that is positively misleading for evolutionary inference and thus extreme methodological care must be taken when using them to infer phylogenies.

Keywords: Phylongeny; Scorpiones; Chelicerata; Mitochondrion; Strand-bias; NTE

1. Introduction

1.1. Chelicerates

The chelicerates (subphylum Chelicerata) are a diverse group of arthropods including many organisms of medical (*Ixodes scapularis*, a Lyme disease vector), economic (*Rhipicephalus* [formerly *Boophilus*, see Barker and Murrell, 2004] *microplus*, a cattle tick) and evolutionary (Xiphosura, horseshoe crabs) interest. Sea spiders (class: Pycnogonida) are sometimes considered chelicerates but several alternative hypotheses have not been ruled out (Dunlop and Arango, 2005). The phylogenetics of chelicerates has been studied at many levels, from the species (Dobson and Barker, 1999) to the phylum (Mallatt et al., 2004). Major problems are the relationships between the various chelicerate orders, for which many solutions have been proposed but on which a consensus is yet to be reached (Weygoldt, 1998). In particular, the phylogenetic position of the Scorpiones, a key order in the arachnid phylogeny, is highly disputed (Weygoldt, 1998; Dunlop and Braddy, 2001; Dunlop and Webster, 1999; Giribet et al., 2002; Wheeler and Hayashi, 1998). Scorpions

---

[*] Corresponding author. Fax: +44 131 650 7489.
E-mail address: martin.jones@ed.ac.uk (M. Jones).

1055-7903/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2006.11.017

have been variously proposed to be basal (Hypothesis 1, Fig. 1a) or derived (Hypothesis 3, Fig. 1c) arachnids (reviewed in Wheeler and Hayashi, 1998). Although scorpions are a key taxon for the understanding of chelicerate evolution, current sequence datasets are limited in extent. In particular, scorpion mitochondrial genomes have not been fully exploited (Gantenbein and Largiadèr, 2003). Fourteen complete mitochondrial genomes from Acari (ticks and mites), three from spiders, and one from a New World scorpion (*Centruroides limpidus*: Buthidae) have been sequenced. Here, we describe the mitochondrial genome of the East Mediterranean scorpion *Mesobuthus gibbosus* (Buthidae) and use it to further explore chelicerate relationships.

1.2. Multigene datasets

Hypotheses regarding the relationships between chelicerate orders have been largely based on morphology (Weygoldt and Paulus, 1979; Giribet et al., 2002; Wheeler and Hayashi, 1998; Regier and Shultz, 2001). Nuclear ribosomal RNA genes have been used for wider arthropod and chelicerate phylogeny (Mallatt and Giribet, 2006; Wheeler and Hayashi, 1998). Here, we examine the utility of mitochondrial sequence data for addressing questions of chelicerate phylogeny. Recent data suggest that large datasets, comprising many genes, can resolve problematic phylogenies with a high degree of confidence (Philippe et al., 2005a; Rokas et al., 2003). By combining the information from multiple gene sequences, clades can be recovered that are not recovered under analysis of any of the individual genes ('hidden support', described in de Queiroz and Gatesy, 2006). Different genes with different evolutionary rates may give strong phylogenetic signals at different depths in a phylogenetic tree (Giribet, 2002). Thus, by including multiple genes in a phylogenetic study, one could obtain a tree that any single gene would be unable to resolve, as advocated earlier by the total evidence proponents (e.g., Kluge, 1989; Nixon and Carpenter, 1996). The use of multiple genes for phylogenetics comes with its own set of difficulties, the most significant of which are computational complexity, and the need for evolutionary models that describe the variation between genes (Pupko et al., 2002). As the number of genes (and hence the number of characters) included in a multiple sequence alignment grows, so does the time required to evaluate the likelihood or parsimony score of a corresponding phylogenetic tree and hence the time required to execute tree search algorithms. The choice of evolutionary model, always a critical issue in phylogenetic reconstruction (Lemmon and Moriarty, 2004), is particularly important where multiple genes are involved. If the genes evolve under different evolutionary constraints, a single model of DNA evolution may not accurately describe the history of all characters in the alignment, and separate models and parameters may have to be assigned to each gene. Additionally, if some gene sequences are unavailable for some taxa, the alignment may have an appreciable proportion of missing data which may adversely affect the robustness of the tree (Wiens, 2003).

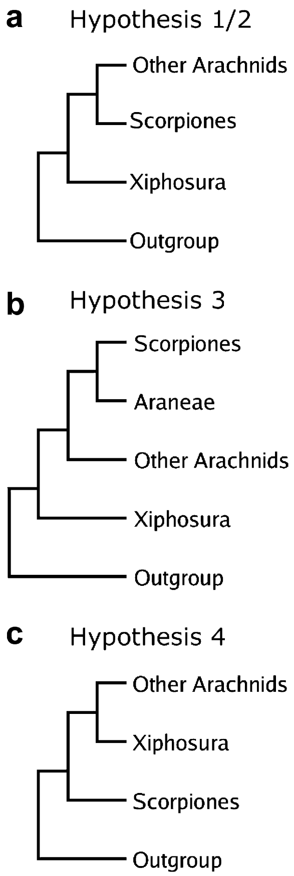

Fig. 1. Summary trees showing various phylogenetic hypotheses regarding the placement of Scorpiones. Scorpiones are shown (a) as sister taxon to other arachnids (b) as sister taxon to other chelicerates (c) as derived arachnids forming a sister taxon to Araneae.

1.3. Strand-bias

In a single gene phylogenetic study involving few characters, the accuracy of the tree is limited by the amount of phylogenetic signal present. In contrast, in a multigene study, the accuracy of the tree is more likely to be limited by systematic errors. Phenomena such as differing substitution rates between lineages (Brinkmann et al., 2005; Felsenstein, 1978), differing patterns of rates between lineages (Gadagkar and Kumar, 2005; Philippe et al., 2005b) and compositional bias (Galtier and Gouy, 1995) have all been found to lead to tree reconstruction artifacts. Model choice has been shown to be a key factor in overcoming difficulties associated with analysis of biased sequences (Lemmon and Moriarty, 2004; Posada and Buckley, 2004; Sullivan and Swofford, 2001). Additionally, differing patterns of evolution between genes can cause problems in phylogenetic reconstruction under models that fail to take inter-gene differences into account (Nylander et al., 2004).

All sequences in non-recombining animal mitochondrial genomes are tightly linked and, in some respects, behave as a single gene. It might therefore be supposed that mitochondrial gene sets would be affected by biases to a similar