

OUTGROUP MISLEADING EFFECT AND NEW ARTHROPOD PHYLOGENY

Omar Rota Stabelli*, David S Horner^ and Maximilian J. Telford*.

*Department of Biology, University College of London, UK; ^Dip. di Scienze Biomolecolari e Biotecnologie, Università degli studi di Milano; contact: o.stabelli@ucl.ac.uk



UNCERTAIN AFFINITIES IN ARTHROPODS. There are at least two main questions regarding arthropod phylogeny still under strong debate. First, are the hexapods a monophyletic group or, for instance, has the six-legged body plan evolved independently at least two times? Second, is the Mandibulata a monophyletic clade or should we rather believe in the Paradoxopoda hypothesis (chelicerates plus myriapods) recently proposed based on molecular analyses? Mitochondrial and nuclear based phylogenies give support for conflicting hypothesis and morphological data seem to contradict both sources of molecular evidence. Interestingly, as far as hexapods monophyly is concerned, conflict is even present between different mitochondrial based phylogenetic reconstructions. In order to analyze the relationships among basal arthropods and to understand the conflict among different mitochondrial base reconstructions we have investigated the impact of compositional heterogeneity in these sequences and the impact of outgroup choice on the inference of phylogenetic trees.

Case 1: ARE HEXAPODA PARAPHYLETIC?

Phylogenetic analysis of concatenated mitochondrial proteins (Nardi et al. 2003) has suggested the paraphyly of hexapods (Insecta, Collembola, Diplura, Protura). In these analyses, collembolans constitute the sister group of crustaceans and insects (Figure 1 B). By way of response, Delsuc et al. (2003), analysed the corresponding nucleotide dataset and recovered the hexapods as a monophyletic group (Figure 1 A). More recent mtDNA based phylogenies (Cook et al. 2005, Hassanin et al. 2005) confirmed the paraphyletic hypothesis, but the criticisms of Delsuc et al. have strongly reduced the confidence of mtDNA in resolving hexapods affinities.

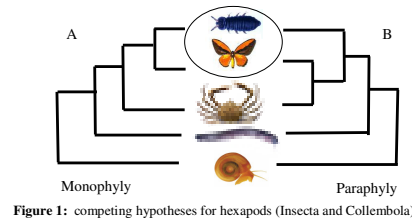


Figure 1: competing hypotheses for hexapods (Insecta and Collembola)

OUTGROUP MISLEADING EFFECT

A suitable outgroup will simply identify the position of the root of a phylogenetic tree, it will not affect the topology of the ingroup (Holland et al. 2003). We analysed the 25 taxa dataset of Delsuc et al. in the presence and absence of outgroup sequences, under the same conditions used by the authors. In the absence of outgroup sequences (molluscs and annelids) crustaceans and insects are recovered as a strongly supported monophyletic group and Collembola emerge basal to this group (in accordance with other mtDNA based phylogenies).

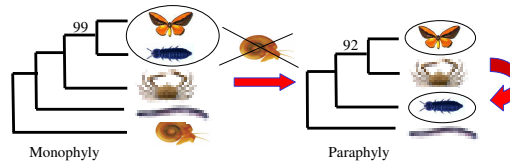


Figure 2: Absence of outgroup leads to changes in the tree topology and recover paraphyletic hexapods.

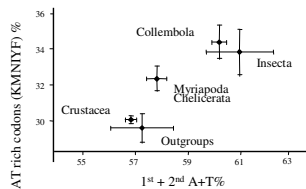


Figure 3: AT content plotted against AT rich amino acid

COMPOSITIONAL BIAS LINKS OUTGROUPS AND CRUSTACEA

In order to investigate the influence of outgroup sequences on the ingroup topology we inspected the A+T mean percentage of the 1st + 2nd codon position for each principal lineage in the tree (Delsuc dataset) against amino acid frequencies expressed in terms of codons with A or T at both positions 1 and 2 (Fig 3). The mean nucleotide and amino acid frequencies of the outgroups and crustaceans are similar (moderate A+T) and insect and collembolan sequences both have a high A+T percentage. Shared nucleotide compositional biases lead to increased homoplasy at unconstrained sites and create erroneous signal (Foster and Hickey 1997, 1999) that may cause crustaceans to be pulled to the root of the tree and insects and collembolans to be recovered as monophyletic.

HETEROGENEOUS MODEL OF EVOLUTION CONFIRMS COMPOSITIONAL EFFECT

In order to investigate whether the perturbation of ingroup topology in the presence of outgroup sequences is linked to compositional heterogeneity in the data, we evaluated the support for specific nodes in the tree when the underlying model of sequence composition is allowed to vary over the tree (program P4, Foster 2004). We investigated Bayesian posterior support for monophyletic hexapods (with monophyletic insects as control) as the number of composition categories was increased in the Delsuc dataset. Table 1 illustrates that while under a homogeneous composition model (1 comp. vector) posterior support for monophyletic hexapods was high, as the number of composition categories used was increased, support for this grouping decreased. Moreover in the absence of the outgroup, support for hexapod paraphyly remains high. When the model fits the data (pt statistics more than 0.05 corresponding to at least 4 compositional vectors) hexapod monophyly is scarcely supported (bayesian support of 0.55) suggesting that the heterogeneous model overcame the misleading effect of outgroup.

N. Comp. vectors	With outgroup		Without outgroup		pt
	Hexapoda monophyly	Insecta monophyly	Hexapoda monophyly	Hexapoda paraphyly	
1	0.97	0.99	0.99	0.00	
2	0.85	0.97	1.00	0.00	
4	0.72	0.99	1.00	0.09	
8	0.55	1.00	1.00	0.55	0.55

Table 1: support for hexapod monophyly decrease under a heterogeneous models of evolution.

Case 2: IS MANDIBULATA A MONOPHYLETIC CLADE?

From a morphological point of view (Scholtz and Edgecombe 2006, Harzsch et al. 2005) Myriapoda, Crustacea and Hexapoda share a common head structure that lead to the definition of the Mandibulata clade (Figure 4 B). By the way many mitochondrial and nuclear based phylogenies suggested that Myriapoda share more affinities with Chelicerata forming a group called either Paradoxopoda (Mallat and Giribet 2005) or Myriochelata (Pisani et al. 2004) (Figure 4 A). It was shown that when using mitochondrial genes the choice of outgroups (Cameron et al. 2005) may play a key role in the correct inference of arthropods phylogeny. We analysed a selection of available metazoan outgroups and the partially sequenced Onychophora and newly sequenced Priapulida. For each major metazoan group that has an invertebrate mtDNA code, we chose no more than two species on the basis of a reduced genetic ML distance to the basal arthropod *Limulus polyphemus* and "arthropod-like" compositional properties (expressed in terms of GC content and GC skew values)

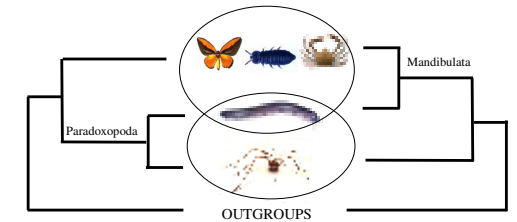


Figure 4: Competing hypotheses for arthropods phylogeny. Lower circle shows Paradoxopoda (myriapods + chelicerates) while upper one show Mandibulata (hexapods + crustaceans + myriapods)

PRIAPULID IS GENETICALLY AND COMPOSITIONALLY THE BEST OUTGROUP

Results indicate that some outgroups, in particular from lophotrochozoans, have a GC content (data not shown) and strand specific bias (Figure 5 black dots) very different from the one of arthropods, suggesting that the use of such outgroups may interfere with the stationarity of the evolutionary model and be responsible for misleading phylogenetic signals (Hassanin et al. 2005 Gibson et al. 2005). On the other hand we were able to detect Priapulida sequence as the closer to arthropods in term of genetic distance (data not shown) and compositional similarity to arthropods, expressed both in terms of GC content and GC skew.

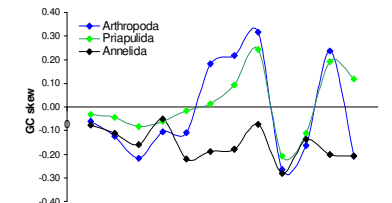
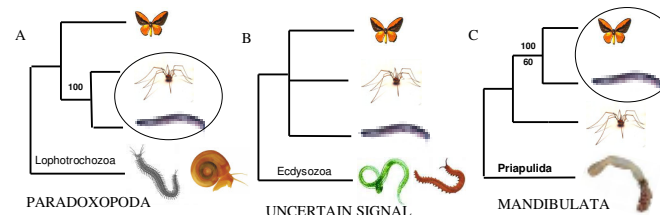


Figure 5: G-C skew values calculated on 1st and 2nd codon position for each mitochondrial gene plotted in accordance with arthropod ancestral gene order. GC value is an index that describes strand related bias. A positive value of the skew indicate for example that a sequence is enriched in G.

THE CLOSER THE OUTGROUP, THE MORE MANDIBULATA IS SUPPORTED

Inference of phylogeny shows that distant or compositionally dissimilar outgroups, such as the lophotrochozoans (molluscs and annelids), support the Paradoxopoda hypothesis (Figure 6 A). Moreover the use of more closely related - even if compositionally dissimilar - outgroups, in particular from the ecdysozoans, determine a significant decay in Paradoxopoda support and some evidence for the Mandibulata clade (Figure 6 B). In the presence of the less distant and more compositionally similar outgroup, Priapulida, phylogenetic reconstruction recovers with high confidence Mandibulata as a monophyletic group, in conflict with previous molecular results and in accordance with the morphological point of view (Figure 6 C).



REFERENCES:

Nardi F. et al. *Science* 299: 1887-9 (2003);
 Foster P. *Syst Biol.* 53, 485 (2004);
 Delsuc F. et al. *Science* 301, 1482 (2003);
 Saccone C. et al. *Gene*, 328 195 (1999);
 Holland B. et al. *Syst. Biol.* 52 (2): 229 (2003);
 Foster P, Hickey D., *J. Mol. Evol.* 44 (3):282-8 (1997);
 Scholz G *Dev Gen Evol.* 216:395 (2006);
 Harzsch S and Edgecombe. *Dev Gen Evol* 215 53 (2005);
 Cameron SL. et al. *Clastic* 20 534 (2004);
 Mallat J and Giribet G 2005 *Mol Phy Evol* 40 772 (2006);
 Pisani D et al. *BMC Biol* 19 2:1 (2004);
 Gibson A et al. *MBE* 22 251 (2004);
 Hassanin A et al. *Syst Biol* 54 277 (2005).

CONCLUSIONS: We have shown two cases in which outgroup affects the topology of the reconstructed ingroups phylogeny. In both cases the misleading effect is correlated with high genetic divergence and compositional properties of outgroups. We suggest that both the support for Hexapod monophyly and Paradoxopoda clade from the analysis of concatenated mitochondrial coding regions is likely to be an artefact of shared compositional bias and inappropriate outgroup choice rather than a true phylogenetic signal.