Molecular Phylogenetics and Evolution 62 (2012) 27–34

Contents lists available at SciVerse ScienceDirect

Molecular Phylogenetics and Evolution

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

Resolving phylogenetic signal from noise when divergence is rapid: A new look at the old problem of echinoderm class relationships

Davide Pisani [a], Roberto Feuda [a], Kevin J. Peterson [b], Andrew B. Smith [c,*]

[a] The National University of Ireland, Maynooth, Co. Kildare, Ireland
[b] Department of Biological Sciences, Dartmouth College, N. College Street, Hanover, NH 03755, USA
[c] Department of Palaeontology, The Natural History Museum, Cromwell Road, London SW75BD, UK

ARTICLE INFO

Article history:
Received 1 March 2011
Revised 8 July 2011
Accepted 31 August 2011
Available online 14 September 2011

Keywords:
Molecular evolution
Long-branch attraction
Echinoderms
Phylogeny
MicroRNA
Molecular clocks

ABSTRACT

Resolving evolutionary relationships in groups that underwent fast radiation in deep time is a problem for molecular phylogeny, as the scant phylogenetic signal that characterises short internal branches is generally swamped by more recent substitutions. We implement an approach, that maps how the support for rival phylogenies changes when analysing subsets of sites with either faster and more heterogeneous rates or slower and more homogeneous rates, to address a long-standing problem in deuterostome phylogeny – the interrelationships of the eleutherozoan echinoderm classes. We show that miRNA genes are phylogenetically uninformative as to the relationships of asteroids, echinoids and ophiuroids, consistent with a rapid radiation of these groups as suggested by their fossil record. Using three nuclear rRNAs and seven nuclear housekeeping genes, we map the support for the three possible phylogenetic arrangements of asteroids, ophiuroids and echinoids when moving between subsets of the data with very similar or very different rates of evolution. Only one of the three possible topologies (asteroids (ophiuroids + echinoids)) strengthens when the most rate-homogeneous subset of data are analysed. The other two possible pairings become stronger in a less reliable data subset, which includes the fastest and thus homoplasy-rich data in our alignment. Thus, while superficial analysis of our concatenated alignment identifies asteroids and ophiuroids as sister taxa, more thorough analyses suggest that ophiuroids may be more closely related to echinoids. Divergence of these echinoderm groups, using a relaxed molecular clock, is estimated to have occurred within ~5 million years. Our results illustrate that the analytic approach of phylogenetic signal dissection can be a powerful tool to investigate rapid radiations in deep geologic time.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

Both morphological and molecular approaches to phylogenetic reconstruction work well when divergences between taxa are separated by relatively long time intervals, as the accumulation of substantial numbers of derived characters in the stem lineages creates a strong phylogenetic signal. However, when divergence occurred rapidly in deep time and stem lineages are of short duration, accurate phylogenetic reconstruction is difficult. This is because continuing evolution results in convergence and reversals that ultimately overwhelm the weak signal in short internal branches. In such situations, unequal rates of evolution can lead some branches to accumulate a significantly larger number of substitutions leading to the well-known problem of long-branch attraction (LBA: Felsenstein, 1978). While LBA has long been recognised as a problem, how best to identify trees affected by LBA and tease out historical signal from systematic biases remains a major challenge (Brinkmann and Philippe, 1999; Ruitz-Trillo et al., 1999; Pisani, 2004; Lartillot and Philippe, 2008; Jeffroy et al., 2006; Sperling et al., 2009; Rota-Stabellli et al., 2010). Indeed, while the signature of rapid divergence is a phylogenetic tree where branching order cannot be resolved with confidence, LBA can confuse the picture causing the recovery of artefactual groups with very high support (Jeffroy et al., 2006).

One problematic area of the metazoan tree concerns how the five echinoderm classes are related (Smith et al., 2004; Janies et al., 2011). Both morphology and molecular data place crinoids as sister group to the other classes (echinoids, asteroids, ophiuroids, holothurians), and pair echinoids and holothurians together. Yet the interrelationships of asteroids, ophiuroids and the echinoid–holothurian clade remain disputed. Morphological data favours either an asteroid–ophiuroid pairing (Mooi and David, 2000) or an ophiuroid plus echinoid–holothurian pairing (Littlewood et al., 1997), whereas different molecular analyses have found support for all three possible groupings (Field et al., 1988; Littlewood et al., 1997; Janies, 2001; Mallatt and Winchell, 2007; Pereske et al., 2010; Janies et al., 2011; Letsch and Kjer, 2011).

* Corresponding author.
E-mail address: a.smith@nhm.ac.uk (A.B. Smith).

1055-7903/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2011.08.028

These echinoderm clades pose a particularly acute problem for molecular phylogenetic analyses because they underwent crown group diversification long after they had split from one another and all three have long stem groups that cannot be broken up by selective sampling of the modern fauna, making them particularly susceptible to LBA.

2. Materials and methods

2.1. Molecular data assembled

Total RNA was collected from the ophiuroid *Ophiopholis* and a small RNA library constructed and sequenced following Wheeler et al. (2009), resulting in 3804 parsed non-redundant reads. These were then compared with previously published small RNA libraries drawn from an asteroid (*Henricia sanguinolenta*), echinoid (*Strongylocentrotus purpuratus*), hemichordate (*Saccoglossus kowalevskii*) and other metazoans published previously and analysed by miRMiner (Wheeler et al., 2009) for known and potentially novel miRNAs (Table S1).

Six nuclear housekeeping genes (aldolase, methionine adenosyltransferase, ATP synthase beta chain, elongation factor 1 alpha, triosephosphate isomerase and phosphofructokinase) were sequenced from the ophiuroid *Ophiopholis* sp. following the protocol described in Sperling et al. (2009). These sequences have been deposited in Genbank under accession numbers (JN716365–JN716370). Sequences for *Aplysia californica*, *Alvinella pompejana* and *Tubifex tubifex*, as well as three genes for *Carinoma mutabilis*, were downloaded from the NCBI trace archives. Unpublished sequences from *Chaetopleura apiculata* and *Leptochiton asellus* were kindly provided by J. Vinther (Yale University). Sequences for other lophotrochozoan taxa were taken from previously published reports (Peterson et al., 2004), and new sequences were manually added to the pre-existing alignment used, for example in Sperling et al. (2011). Data for ribosomal 5.8S, 18S and 28S ribosomal genes for 22 deuterostome, 35 lophotrochozoan, and 15 ecdysozoan taxa were assembled, either taken directly from Mallatt et al. (2010) or downloaded from the NCBI Genbank website and manually aligned to the Mallatt et al. (2010) sequences. Chimaeras at the generic level were permitted when data for the same species were not available. After the removal of minor indels, the amino acid matrix was 88% complete and the ribosomal matrix was 76% complete. The seven nuclear housekeeping genes (2049 amino acids in total) and three ribosomal genes (4682 nucleotides in total) were concatenated for analysis.

2.2. Sequence analysis

2.2.1. Conventional phylogenetic analysis

The protein and rRNA partitions were first independently analysed to investigate the nature of the principal signal (Pisani and Wilkinson, 2002) in these data sets. Protein analyses were performed using the heterogeneous CAT-GTR model, and rDNA analyses were performed using the GTR + G model, which proved to be the best fitting model (selected using MrModeltest) for our nucleotide data. CAT-GTR analyses were performed in Phylobayes V. 3 (Lartillot and Philippe, 2004). We used posterior predictive analysis as implemented in Phylobayes (see also Sperling et al., 2009) to discover whether the taxa of interest (i.e. the echinoderms) were compositionally homogeneous or heterogeneous.

The rRNA and protein partitions were concatenated and analysed under mixed models using Bayesian and Maximum Likelihood (ML) analyses. Maximum Parsimony (MP) and Neighbour Joining (NJ) (with uncorrected P distances and no gamma correction) were also performed**.** Bayesian analyses were performed using MrBayes 3.1 (Huelsen-beck and Ronquist, 2001), ML analyses were performed using RAxML (Stamatakis, 2006), while MP and NJ analyses were performed using PAUP4b10 (Swofford, 2002). Support for nodes found in the MP, NJ and ML analyses was estimated using the bootstrap, with 500 replicates for MP and NJ (but see Supplementary information) and 5000 replicates for ML.

For all Bayesian mixed models analyses both the rRNA and protein partitions were modelled using GTR + G. Sperling et al. (2009) showed that for this protein data set, GTR + G is the best fitting amongst the homogeneous substitution models implemented in MrBayes, whilst we showed here that GTR + G is the best fitting model for our nucleotides partition. CAT-GTR analyses could not be performed for the concatenated data set because of software limitation (Lartillot, pers. comm.). For the ML analyses the protein partition was modelled using LG + G. The nucleotide partition was modelled using GTR + G.

2.2.2. Phylogenetic signal dissection

Both the rRNA and the Protein data sets were partitioned into sets of "homogeneously evolving" and "heterogeneously evolving" sites using a modification of Brinkmann and Philippe's (1999) slow–fast approach (see Sperling et al., 2011 for justifications). This method assigns rates to characters semi-independent of tree topology. The characters in the rRNA and protein data sets were independently ranked according to their evolutionary rate (estimated as slow–fast parsimony scores) and partitioned into four quartiles. For each data set (proteins and rRNAs) characters were split into two groups: the first containing all the sites in the fourth quartile plus invariant sites, the second contained all the variant sites in the first, second and third quartiles. The characters in the first data partition represent a combination of sites with highly heterogeneous rates (i.e. very fast and constant sites only). This partition included 1247 AA and 3206 NN positions, of which 748 AA and 2332 NN positions were constant and 499 AA and 874 NN where deemed to be fast evolving. Because of the extreme rate variation (including constant and fast evolving sites only), and the high substitution rates and homoplasy levels of the variable characters it includes, this data partition presents a hard phylogenetic problem, and is prone to generate phylogenetic artefacts (e.g. LBA) even when analysed using well-fitting, parameter-rich models. The second data partition is composed of phylogenetically more reliable, rate-homogeneous, characters of slow to intermediate evolutionary rate. This partition includes 811 AA and 1476 NN (all of which are parsimony informative) and is more likely to support relationships that represent historical signal (see Sperling et al., 2009, 2011; Rota-Stabellli et al., 2010).

We then evaluated the strength of the signals supporting the three possible arrangements of asteroids, echinoids and ophiuroids residing in the three data sets (i.e. all sites, rate-heterogeneous sites and rate-homogeneous sites), under three, differently performing, methods – Parsimony, Neighbour Joining and Bayesian analysis. The fit to data of the three topologies (see Fig. 1) into which asteroids, echinoids and ophiuroids can be arranged were compared using Bayes Factors (BF; e.g., Sperling et al., 2010; Holton and Pisani, 2010) as follows. For each data set (homogeneous, heterogeneous and all sites), and each sister-group hypothesis (E + O, A + E and E + A), a constrained tree search (of 2 runs and four chains per run) was performed in MrBayes (Huelsenbeck and Ronquist, 2001). Each constrained tree search was run for 5,000,000 generations and a burn-in of 2,500,000 generations was used. This burn-in period was sufficiently long to allow each analysis to converge, and generated an identical number of data points (per data set and hypothesis) to calculate the BF. For each data set, the MrBayes ".p" file corresponding to the chain of maximal marginal likelihood across all trees (estimated using the harmonic mean) was selected, and used to estimate the BF for each pair of considered hypotheses in Tracer v1.5.1 (Rambaut and

Download English Version:

https://daneshyari.com/en/article/5920335

Download Persian Version:

https://daneshyari.com/article/5920335

Daneshyari.com