The LINNEAN SOCIETY of London ZOOLOGICAL Journal of the Linnean Society

Zoological Journal of the Linnean Society, 2013, **167**, 82–92. With 3 figures

Rerooting the rDNA gene tree reveals phoronids to be 'brachiopods without shells'; dangers of wide taxon samples in metazoan phylogenetics (Phoronida; Brachiopoda)

BERNARD L. COHEN*

College of Medical, Veterinary and Life Sciences, University of Glasgow, Urquhart Building, Room 203, Garscube Estate, Glasgow, 611 1QH, Scotland, UK

Received 22 May 2012; revised 24 August 2012; accepted for publication 28 August 2012

Molecular phylogenetics has resulted in conflicting accounts of the relationship between phoronids and brachiopods. Taxonomically comprehensive analyses of brachiopod and phoronid ribosomal DNA sequences (rDNAs) rooted with short-branched mollusc sequences uniformly find that phoronids nest *within* brachiopods as the sister of the three extant inarticulate lineages. Here, this is called the 'alternate' topology because it does not match traditional, morphology-based ideas. Many other analyses of protein-coding genes and/or rDNAs place phoronids elsewhere, often as the sister group of all brachiopods, better matching 'traditional' ideas. However, these analyses generally are based on data from small selections of brachiopods and phoronids, include data from a wide range of other metazoan taxa, and are rooted with distant outgroups. Here, I show that outgroup rooting of brachiopods and phoronid rDNAs is unreliable, and instead find the root position with procedures that are free from all distortions caused by distantly related taxa, i.e. by Bayesian and maximum likelihood relaxed-clock analyses of a purely ingroup alignment. All such analyses confirm the 'alternate' topology: phoronids belong *within* the Brachiopoda as the sister group of the inarticulates. In addition, nine factors are identified that (singly or in combination) can cause misreporting of the phylogenetic signal in wide taxon-range analyses of both rDNA and amino acid sequence data.

doi: 10.1111/j.1096-3642.2012.00869.x

ADDITIONAL KEYWORDS: lophophorates – Lophotrochozoa – Metazoa – systematics.

INTRODUCTION

Because of their distinctive ontogeny and shell-less, lophophorate morphology, phoronids were generally considered a separate (lophotrochozoan) phylum distinct from, but somehow closely related to, bivalve-shelled brachiopods. A close relationship has been confirmed by rDNA-based molecular systematics originally based on quasi-complete small subunit (SSU or 18S) sequences alone, but with the later addition of partial or complete large subunit (LSU, 28S) data (e.g. Cohen & Gawthrop, 1996; Cohen, Gawthrop & Cavalier-Smith, 1998; Cohen & Weydmann, 2005; Santagata & Cohen, 2009). The nature of the relationship between brachiopods and phoronids has become controversial. Many analyses based on either or both rDNAs and protein-coding gene sequences report what is here called the 'traditional' topology, in which phoronids exhibit various relationships with other taxa, but are never placed *within* the phylum Brachiopoda (e.g. Dunn *et al.*, 2008; Paps, Baguna & Riutort, 2009; Hausdorf *et al.*, 2010; Nesnidal *et al.*, 2010; Sperling, Pisani & Peterson, 2011). Other analyses (all of nuclear-encoded rDNAs, rooted with objectively chosen molluscan outgroups)

*E-mail: b.l.cohen@bio.gla.ac.uk, b.l.cohen_home@btinternet.com

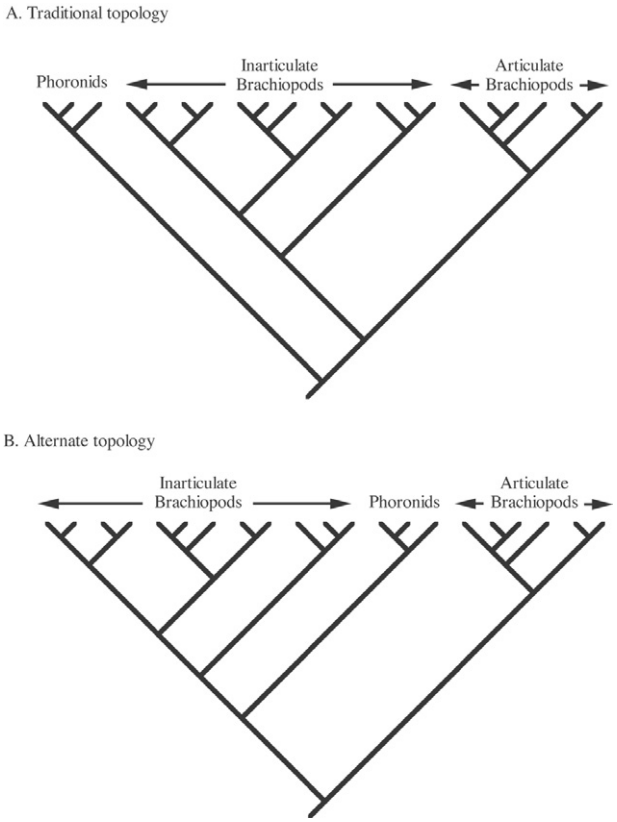

Figure 1. Ribosomal DNA phylogeny of brachiopods and phoronids. Diagrams contrasting the 'traditional' (A) and alternate (B) topologies, drawn with an assumed ur-lophotrochozoan root.

have consistently found weak support for an 'alternate' topology (Fig. 1), in which phoronids are placed *within* brachiopods as the sister group of the inarticulates (e.g. Cohen, 2000; Cohen & Weydmann, 2005; Santagata & Cohen, 2009). Analyses that result in the alternate topology are all from B. L. C.'s laboratory and, unlike those giving the traditional topology, were based on taxonomically representative samples of brachiopod and phoronid rDNAs, i.e. with data from members of all main extant lineages. The aim in this report is to further exclude extraneous influences by re-analysis of these rDNAs in the absence of all distantly related taxa, the root of the tree being located by phylogenetic reconstructions that invoke the molecular clock.

Brachiopods fall into two clades, inarticulates and articulates. The shell valves of extant inarticulates (subphyla Craniiformea and Linguliformea) are linked by soft tissues but not by mineralized hinges. Two inarticulate genera, *Lingula* spp. and *Glottidia* spp. (both with chitino-phosphatic shells), are familiar to many biologists, but atypical amongst brachiopods because they burrow. Other extant inarticulates (craniiforms with tabulate calcitic shells, and discinoid linguliforms with chitino-phosphatic shells) are typically limpet-like and cryptic, and are less familiar. Articulates (subphylum Rhynchonelliformea), by contrast, have calcite valves (in which a secondary layer is fibrous) with mineralized hinges. Of ~100 articulate genera a few sometimes occur in the low intertidal of rocky shores or at shallow depths and may be familiar (e.g. *Calloria*, *Notosaria*, *Terebratalia, Terebratulina*), but most are confined to deeper waters. Fossils demonstrate Cambro-Ordovician origins for all these forms (Williams *et al*., 1996). In molecular analyses of rDNAs from taxonomically representative brachiopod samples, articulates and inarticulates are sister clades (Cohen & Gawthrop, 1996, 1997; Cohen *et al*., 1998); i.e. in these analyses extant brachiopods are monophyletic, as long inferred from morphology.

Disagreement about the relationship between brachiopods and phoronids hinges on the position of the root of the tree (Sperling *et al*., 2011: 291), the placing of which is '. . . frequently the most precarious step in any phylogenetic analysis' (Hillis, Moritz & Mable, 1996: 478). Two rooting methods are generally available: (1) to use outgroups as proxies for ancestral character states (Maddison, Donoghue & Maddison, 1984; Nixon & Carpenter, 1993; Smith, 1994; Huelsenbeck, Bollback & Levine, 2002), and (2) to rely on molecular clock reconstructions because they automatically yield rooted trees (Felsenstein, 2004). In this report I will show that outgroup rooting of brachiopod and phoronid rDNAs fails both empirical and theoretical justifications and will adopt molecular clock-based rooting as the main approach. Molecular clock-based rooting has not previously been used in this context. Very recently a novel, invariant-based rooting method has been described (EP rooting, Sinsheimer, Little & Lake, 2012). As with molecular clock rooting this method can be used with an ingroup-only alignment, and the results reported here serve as a prediction to be tested when this method comes to be applied.

The empirical failure of outgroup rooting in this case arises thus: in earlier work analyses were rooted with chiton and bivalve mollusc SSU sequences chosen when few relevant data were available (Cohen & Gawthrop, 1996; Cohen *et al*., 1998; Cohen & Weydmann, 2005; Santagata & Cohen, 2009). These outgroups were selected to satisfy two prior criteria: (1) uncontroversially, that the branch from the outgroup to the ingroup should be as short as possible (Rosenfeld, Payne & DeSalle, 2012) and (2) more controversially, that the brachiopod tree should preserve the sister-group relationship of chitino-phosphatic linguloids and discinoids (Cohen & Gawthrop, 1996; Cohen *et al*., 1998; Cohen & Weydmann, 2005; Santagata & Cohen, 2009) that is strongly inferred from morphology and the fossil record (Williams, Carlson &

Brunton, 1997a). This firmly established relationship was disrupted by other, more remote, outgroups (Cohen & Weydmann, 2005). Today, however, the selected outgroup sample is obsolete because many more potential outgroup sequences (both SSU and LSU) are now available, and because the topology that results is outgroup-specific. This is clear from preliminary analyses (not shown) in which rooting with the chiton *Acanthopleura* gave the alternate topology (as before, e.g. Cohen & Gawthrop, 1996; Cohen *et al*., 1998; Santagata & Cohen, 2009) whereas rooting with the chiton *Chaetopleura* (present together with *Lepidochiton* in the data of Sperling *et al*., 2011) gave the traditional topology. Furthermore, no current molecular or morphological data securely identify the sister phylum of Brachiopoda, yet that information is needed for reliable outgroup choice (Watrous & Wheeler, 1981; Maddison *et al*., 1984; Nixon & Carpenter, 1993; Smith, 1994). Outgroup rooting is empirically unsafe in these conditions.

A more theoretical reason for distrusting outgroup rooting arises when morphological difference and molecular divergence between phyla is *either* so large that substantial homoplasy is likely *or* (as here) so small that rooting is not robust. Given both empirical and theoretical defects of outgroup rooting in this case, the root will be located by relaxed molecular-clock reconstructions, in which the likelihoods of all possible rooted trees are examined. Relaxed-clock rooting has been found to be robust and effective (Huelsenbeck *et al*., 2002). Moreover, comparable molecular clock reconstructions with similar, virtual roots have been widely used to construct time-trees for distantly related taxa in which variations of evolutionary mechanism are inevitable (e.g. Douzery *et al*., 2004; Sanderson *et al*., 2004; Drummond *et al*., 2006); it is therefore a reasonable approach for analysis of the taxonomically narrow and relatively uniform alignment employed here. The exclusion of all outgroups is also helpful because it removes invisible, potential influences of distant taxa on tree topology. Relaxed-clock analysis is advantageous because it minimizes effects of variation in rates of sequence evolution over time and across lineages, although it is likely that these have been consistently minor in phoronids and inarticulate brachipods.

For assurance that the alternative topologies were not caused by trivial sequence differences or aberrations, the alignment contained five sequences (marked *) from an analysis that gave the 'traditional' topology (Sperling *et al*., 2011). These data were combined with data produced by B. L. C., both new and previously published. The complete alignment includes sequences from all the well-established extant inarticulate genera (linguliforms *Discina*, *Discinisca*, *Discradisca*, *Pelagodiscus*, *Glottidia*, and *Lingula*; craniiforms *Novocrania* and *Neoancistrocrania*), together with data from all major lineages of articulate brachiopods and phoronids. Relaxed (and strict) molecular clock analyses of these data uniformly recovered the alternate topology, i.e. in the absence of outgroups, all tested methods found that phoronids nest *within* brachiopods as sister of the inarticulates. The analyses also located the root of extant craniids and identified the phylogenetic interrelations and root of extant discinoid inarticulates.

MATERIAL AND METHODS

Data compilation

The sequences used by Sperling *et al*. (2011) were from a source in which they had been shortened by the omission of nonhelical secondary structure regions (Mallatt, Craig & Yoder, 2010, 2012). For this paper the complete sequences were downloaded from GenBank, combined with B. L. C.'s pre-existing sequences, and realigned. In Table 1 and the figures they are labelled with *. The data set was realigned in ClustalX (Larkin *et al*., 2007) and pruned with GBlocks (Castresana, 2000), to give the alignment used (length = 3831 sites before and 3429 sites after GBlocks treatment). See Supporting Information Files S1 and S2 for this alignment in Nexus and html formats respectively, the latter showing the regions excluded by GBlocks. Early versions of the Mallatt alignment (Mallatt *et al*., 2010) contained an erroneous sequence (from Passamaneck & Halanych, 2006, *Neocrania*, now believed to belong to the polychaete *Chaetopterus*). *Neocrania* was also an invalid name, replaced by *Novocrania* (Lee & Brunton, 2001). All SSU sequences are quasi full-length, but the LSU sequences vary in completeness (see Files S1 and S2) and those used by Sperling *et al*. (2011) were shortened to match the longest ones from B. L. C.'s laboratory (2072 nucleotides). Six of B. L. C.'s sequences are shorter (1099 nucleotides), missing data being represented by Ns.

Alignment procedures

In ClustalX, various parameter settings and procedural approaches for multiple alignment were explored: gap open penalty, 10 (range explored 5, 10, 15); gap extend penalty, 1 (range 1, 3, 5, 10); transition weight, 0.1 (range 0.1, 0.5, 0.9). The low gap extend penalty and transition weight were chosen because they improved the initial alignment of the more divergent regions without disturbing conserved blocks. Next, an iterative step was employed in which the divergent regions with bounding conserved blocks were realigned at least once. This was found to further improve the alignment, often with some reduction in length. GBlocks was then used

Table 1. Ribosomal DNA phylogeny of brachiopods and phoronids. GenBank accession numbers of aligned sequences.

Taxa (collection localities)	LSU	SSU
Basiliolella (formerly *Eohemithiris*) (Loyalty Ridge)	AY839242	AF025936
Discina (The Gambia, West Africa)	JQ414037	U08333
Discinisca (Namibia)	AY839247	AF202444
Discradisca (formerly *Discinsca*) (Panama)	AY839248	AY842020
Glottidia (eastern North America)	AY839249	U12647
*Glottidia**	AY210459	U12647
*Laqueus**	AY210460	U08323
Lingula (New Caledonia)	AY839250	U08329
Lingula (New Caledonia)	AY839250	U08331
Neoancistrocrania (Norfolk and Chesterfield Ridges)	HQ852083, JX575602	AY842019
Neoancistrocrania (Japan)	HQ852075	HQ852057
Notosaria New Zealand)	AY839243	U08335
*Novocrania**	DQ279949	U08328
Pelagodiscus (north of Galapagos)	JQ414035	JQ414033
Pelagodiscus (Bellinghausen Sea, Antarctica)	JQ414034	JQ414032
Phoronis hippocrepia (Mediterranean)	AY839251	U08325
Phoronis ovalis (Irish Sea)	EU334115	EU334126
Phoronis spp.*	AF342797	U36271
Terebratalia (western North America)	JF509729	AF025945
*Terebratalia**	AF342802	U12650
Terebratulina (Scotland)	AY839244	U08324

*used by Sperling *et al*., 2011.
LSU, large subunit; SSU, small subunit.

to identify and remove remaining regions of potential misalignment and all gapped sites. In GBlocks the minimum length of a retained block was kept at 10, rather than 5 (recommended for RNA sequences) to make selection against potentially misaligned nucleotides more stringent, resulting in removal of ~11% of 3831 sites, as shown in File S2.

Rooting procedures

To find rooted trees, relaxed-clock methods were implemented in BEAST 1.4.7, using the general time reversible model with invariant site and gamma shape parameter model (GTR + I + Γ) and in MrBayes 3.2.0 with the Thorne–Kishino (TK02) model (Thorne & Kishino, 2002; Ronquist & Huelsenbeck, 2003; Drummond *et al*., 2006; Ronquist, Huelsenbeck & Teslenko, 2011). A strict-clock analysis was also performed in PAUP*4 (Swofford, 1998).

Phylogenetic analyses

Base composition homogeneity was tested on the complete data and on taxon subsets using the chi-squared heterogeneity test in PAUP*4. Best-fitting nucleotide-substitution models for phylogenetic analyses were identified with MODELTEST 3.06 under the Akaike information criterion (Posada & Crandall, 1998; Posada & Buckley, 2004). Network analysis was performed in SPLITSTREE 4.10 (Huson, 1998) using p and GTR + I + Γ distances (parameters estimated by MODELTEST) and displayed as an unrooted neighbour-net graph. In PAUP*4 maximum likelihood (ML) heuristic search used neighbour-joining or ten cycles of random taxon addition, followed by tree bisection reconnection or subtree pruning and regrafting branch exchange, with and without the (strict) molecular clock, which was used with 'factory default' settings. A likelihood ratios (chi-squared) test based on the resulting no-clock and strict-clock trees was used to test for clock-like evolutionary behaviour. Node support was calculated by jack-knifing with ML heuristic search with neighbour-joining taxon addition, 37% site deletion, and Jac emulation (37% deletion and the Jac program are described by Farris *et al*., 1996). The jackknife was preferred because it does not distort the primary data whereas bootstrap pseudoreplication reweights characters (Freudenstein & Davis, 2010). Alternate tree topologies were constructed with the tree editor in MacClade 4.02 (Maddison & Maddison, 2001), starting from the tree constructed by PAUP*4 ML + clock heuristic search.

After preliminary analyses showed that partitioning by gene did not alter the main outcome, the data for Bayesian relaxed-clock analyses with MrBayes 3.2.0 and BEAST 1.4.7 were treated as a single partition (Drummond *et al*., 2006). In MrBayes, theTK02

model was used (Thorne & Kishino, 2002), with default parameter settings, run for 1×10^6 and 5×10^6 generations. Convergence was identified by plots of all parameters and by the values of the standard deviation of split frequencies and the potential scale reduction factor, as recommended (Ronquist *et al*., 2011). The MrBayes consensus tree is reported together with its clade-credibility values and is displayed using FIGTREE 1.3.2 to provide a scale in which root position = 1.0, making it equivalent to an uncalibrated time-tree. In BEAST, with the GTR + I + Γ model, normal-distribution priors were set for clade origins (= mean node heights) with a likely error-range (95% highest posterior density) of ~50 Myr. TRACER 1.4 was used to follow analyses and the maximum clade-credibility tree was found with TreeAnnotator 1.5.3. The influence of priors was checked by runs without data.

To remove homoplasious sequence sites a topology-dependent form of the 'Slow-Fast' method was applied (Brinkmann & Philippe, 1999). This test was applicable because most taxon relationships are uncontroversial. Unambiguous site changes (1–10 changes/site) in the alignment were charted in MacClade after loading the matrix and ML + clock trees representing the traditional or alternate topologies. The alignment positions of sites with > three changes were noted and they were excluded, following which the best-fitting substitution models and ML + clock trees were re-estimated in PAUP*4. In these (and other, comparable) data the overall distribution of changes over sites was a close fit to exponential ($r^2 \sim 0.95$). One good explanation of this power-law fit (Stumpf & Porter, 2012) is that multiply changed sites are not intrinsically fast-evolving but result from the accumulation of rare, chance events occurring with a constant, low probability. If so, 'Slow-Fast' is a misnomer.

Saturation analysis

The alignment was checked for mutational saturation by plotting transition and transversion uncorrected '*p*' distances against corrected ML distances, all estimated in PAUP*4. Distances were rounded to three significant figures and charted as scatter-plots ($N = 210$ data points) in CRICKET GRAPH III (CA-Cricket Graph III for Macintosh. Computer Associates Plc., 183-187 Bath Rd., Slough, Berkshire, SL1 4AA, UK). Regression coefficient $r^2 = 1.00$ indicates a perfect fit of the tested regression equation to the data.

RESULTS

Analysis of 21 aligned, concatenated SSU and partial LSU rDNAs

The neighbour-net graph in Figure 2 shows that the alignment contains little or no phylogenetic signal conflict. Thus, any disagreement over brachiopod : phoronid relations is not attributable to discrepancies in the sequence data. Craniids and discinids are evidently on the shortest branches.

In the saturation analysis (Supporting Information File S3) scatter-plots of both transitions ($r^2 = 0.976$) and transversions ($r^2 = 0.982$) versus ML distances closely fitted almost-linear power curves, indicating uniform, almost clock-like, change with little dispersion. Maximum ML distances were: transitions ~ 0.055; transversions ~ 0.040. This analysis suggests that the resulting relaxed-clock tree may reasonably be read as an uncalibrated time-tree.

In addition to its role in root-finding, phylogenetic reconstructions that rely on relaxed-clock analyses will avoid effects caused by between-lineage differences in rates of evolution, especially the faster rate in articulate brachiopods (Cohen & Weydmann, 2005). Figure 3 shows the relaxed-clock consensus tree from MrBayes together with clade credibility (%, equivalent to clade posterior probability, 100% indicated by •). The same topology was obtained in other analyses (not shown) using BEAST and in MrBayes with slightly different priors and parameters, as well as in the ML strict-clock tree from PAUP*4 (shown in Supporting Information File S4). All analyses gave strong support for terebratulide and rhynchonellide articulate brachiopods as the sister group of a clade of inarticulates + phoronids, within which craniid, discinoid, and linguloid clades are each strongly supported. Resolution *within* each of the discinoid, linguloid, phoronid, and articulate clades is broadly consistent with background information and past results (Cohen *et al*., 1998; Cohen & Weydmann, 2005; Cohen, 2007; Kaesler, 1997–2007; Santagata & Cohen, 2009). The *trprobs* output from Mr Bayes comprised one tree with high probability ($P = 0.907$) and six low-probability trees (not shown, $P < 0.062$) with various reconstructions of relationships within the inarticulate clade. Given this, the inter-relationship of craniids, discinoids, and linguloids is provisionally interpreted as unresolved, although as noted above a sister-group relationship of discinoids and linguloids is strongly predicted from morphology. By contrast, phoronids were sister to inarticulates in all relaxed and strict clock trees, i.e. with posterior probability = 1.00; no tree showed the sister-group relationship of brachiopods and phoronids that has been reported elsewhere (e.g. Sperling *et al*., 2011).

The alignment showed neither base-compositional heterogeneity ($P = 1.0$) nor saturation (as noted above). A likelihood ratio test comparing ML no-clock and strict-clock trees from PAUP*4 rejected the strict-clock hypothesis ($P < 0.001$, details not shown), despite which the strict-clock tree had the same basic topology as Figure 3 with strong jackknife support, as

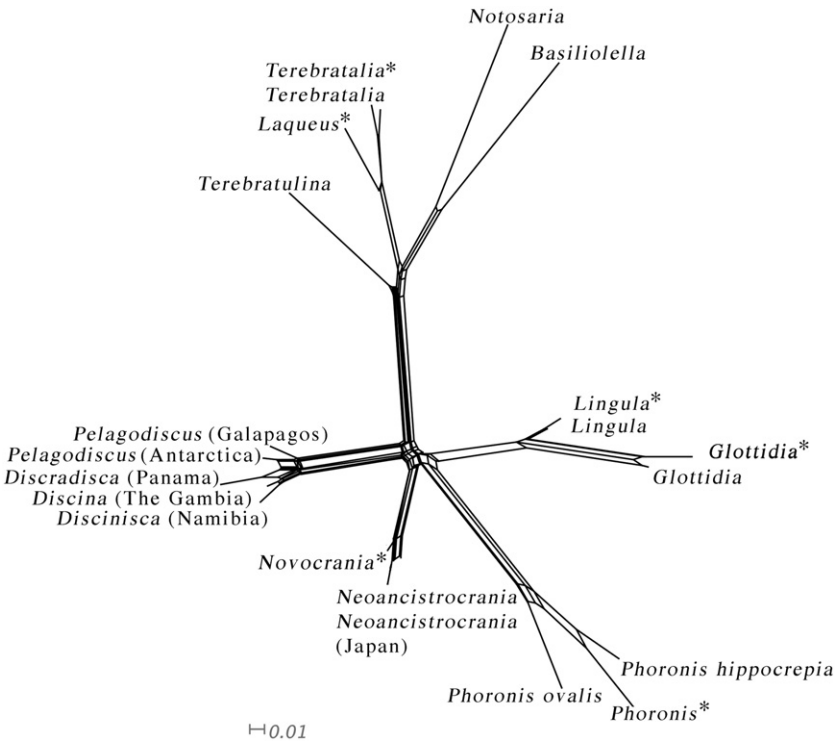

Figure 2. Ribosomal DNA phylogeny of brachiopods and phoronids. Neighbour net graph (general time reversible model with invariant site and gamma shape parameter model maximum likelihood distances) from SPLITSTREE 4.10. All reticulations reflect variation amongst terminals; the data contain no conflicting phylogenetic signal, and the central node, which represents Early Cambrian divergences, is consistent with weak but nonconflicting signal. Discinoids and craniids are on the shortest branches (at lower left). Sequences used by Sperling *et al*. (2011) are marked *.

also did a MrBayes analysis with unlinked 6st (substitution types) models partitioned by gene (not shown). The well-supported sister-group relationship of phoronids and inarticulate brachiopods persisted after exclusion of 69 'fast' sites (identified on the alternate topology ML tree), and of 75 'fast' sites (on the 'traditional topology' tree). Thus, the result shown in Figure 3 is robust; when outgroup(s) are absent, reconstructions with three distinct clock-based tree-building algorithms, evolutionary models, and parameter sets all place phoronids *within* the brachiopod clade as the sister group of the inarticulates, i.e. the alternate topology applies.

As expected, missing data have no phylogenetic effect

All SSU sequences in the alignment are quasi full-length, but the LSU sequences have missing data as described above (and see Files S1 and S2). As has been noted (Wiens & Morrill, 2011), such missing data need not disturb phylogenetic reconstruction and here they do not; every short sequence clusters as expected with its longer homologues and no distinction among sequences of differing lengths can be seen in Figure 2.

DISCUSSION

Data reliability

Ribosomal RNA sequences (as rDNAs) provided the first morphology-independent framework for metazoan evolution and continue to underpin much phylogeny. Their well-known systematic problems such as homoplasy are also common to most other molecular markers (reviewed by, e.g. Philippe, Delsuc & Brinkmann, 2005; Telford & Littlewood, 2009; Mallatt *et al*., 2010). Known, nonsystematic problems affecting brachiopods and phoronids include the mislabelling, referred to above, of *Chaetopterus* LSU as *Neocrania* (in Passamaneck & Halanych, 2006, GenBank accession AY210463), and the misreporting of a brachiopod : phoronid SSU chimaera as an authentic phoronid sequence (Halanych, 1995, GenBank accession U12648; see illustration in supplementary file 2 of Santagata & Cohen, 2009). Sequences used in this analysis have been validated by replication from independent specimens identified by specialists and/or obtained by independent laboratories.

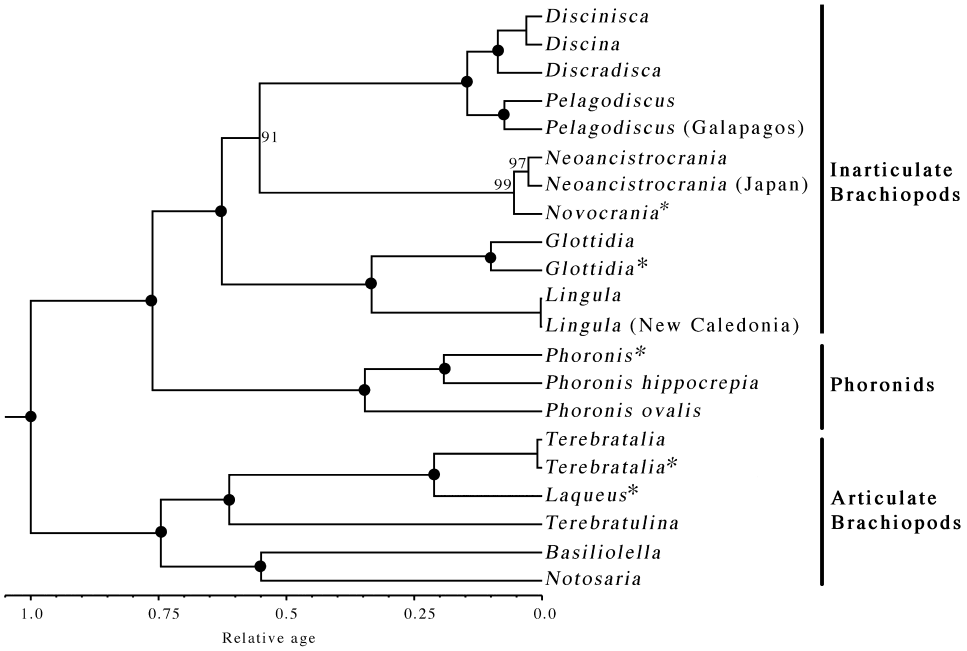

Figure 3. Ribosomal DNA phylogeny of brachiopods and phoronids. Consensus relaxed-clock tree from Mr Bayes (Thorne–Kishino clock model, 10^6 generations) with clade support values of 100% indicated by •. The node with 91% clade support should be considered to be collapsed, forming an inarticulate brachiopod polytomy (craniids, discinoids, linguloids). The same topology of phoronids and brachiopods was obtained in comparable relaxed-clock analyses using BEAST and, with similar (jackknife) support, using strict-clock analysis in PAUP*4. Sequences used by Sperling *et al*. (2011) are marked *.

Advantages of narrow, and drawbacks of wide taxon-range alignments

Why have this and our earlier analyses recovered the alternate relationship when analyses from other laboratories instead find (*inter alia*) phoronids and brachiopods to be sister groups? Features common to the latter analyses include: (1) the ingroup is more-or-less sparsely sampled, (2) the ingroup is a minor component of a wide taxon-range alignment, (3) alignments are often of expressed sequence tag (EST) amino acid alignment data, and (4) alignments contain, or are rooted with, phylogenetically distant taxa. Moreover, (5) a particularly harmful feature of wide-range alignments is an inverse relationship between the number of taxa and the number of reliably aligned nucleotide sites that determine the phylogenetic signal. For example, Paps *et al*. (2009) bravely aligned rDNAs from 564 metazoans, but found it necessary to exclude from analysis 42% of the SSU sequence (36% of the combined SSU + LSU data). (An additional defect of these data is referred to below.) By contrast, in the narrow-range, outgroup-free alignment analysed here, only ~ 5% of SSU was excluded (~ 11% of the combined SSU + LSU data). Thus, wide taxon-range alignments (of rDNAs in particular) are predisposed to misreport close relationships, especially of slowly evolving taxa. Some additional limitations of EST data have been noted (e.g. Siddall, 2010). Supporting Information File S5 presents accounts of nine factors that can disturb wide taxon-range analyses of rDNAs and ESTs, thereby throwing doubt on reported relationships.

Wide taxon-range alignments of amino acid sequence that include data from very divergent lineages are at particular risk of error caused by unrecognized codon homoplasy (Simmons, 2000; Simmons & Freudenstein, 2002). One possible example of this is the morphologically surprising clade that unites nemertines with a phoronid and a brachiopod (first reported in Dunn *et al*., 2008: fig. 2, clade 'A'). This clade is neither present in the most comprehensive multigene account yet published of nemertean molecular phylogeny (Andrade *et al*., 2012), nor does it receive significant support in an alternative analysis of the original data (Siddall, 2010). Synthetic taxa that combine sequences of articulate and inarticulate brachiopods are an additional, possibly unique, drawback of one wide-range alignment (Paps *et al*., 2009: supplementary information). They would, of course, completely preclude the correct reconstruction of phoronid : brachiopod relations.

Brachiopod-specific miRNAs and the early evolution of brachiopods and phoronids

Two miRNAs of unknown functions are expressed in multiple tissues of articulate and inarticulate brachiopods but not in phoronids (only some members of the non-*ovalis* lineage have been tested) nor in any other tested metazoan (Sperling *et al*., 2011: fig. 3). These miRNAs therefore appear to be synapomorphies of Brachiopoda, and they may be thought to exclude the alternate topology. However, that is not a necessary conclusion: miRNAs have largely undefined regulatory roles in post-transcriptional and/or post-translational processing, and the complexity of these processes is such that loss of these miRNAs from phoronids at or after the split from inarticulates cannot be excluded, especially because of an association between organismal simplification and miRNA loss (Erwin *et al*., 2011). Simplification of phoronids relative to brachiopods is possible, but not established, e.g. hermaphroditism and regeneration occur in some phoronids but are rare or unknown amongst brachiopods and may be relevant (Zimmer, 1991; Williams *et al*., 1997).

Wider evolutionary implications of phoronids as 'brachiopods without shells' are unclear; only speculative inferences are available so long as we know nothing of phoronid body-fossils, nothing of pre-Devonian phoronid-candidate trace fossils, and nothing of the ontogeny of any Palaeozoic phoronid; the discussion that follows will therefore be limited. The early fossil record is now known to house a wide and growing diversity of somewhat brachiopod-like forms (e.g. Conway Morris & Peel, 1995; Holmer, Skovsted & Williams, 2002; Balthasar, 2004; Skovsted *et al*., 2008; Balthasar & Butterfield, 2009; Holmer *et al*., 2011), suggesting that early evolution of the brachiopod clade (s.l.) should not be thought of as constrained by simple, dichotomous or parsimonious models, but may have involved novel combinations of ancestral gene regulatory networks (Davidson & Levin, 2005) that generated ontogenies and functional organisms very different from those we readily envisage. However, any hypothesis of phoronid origins must acknowledge established constraints, and I (again) show here that according to rDNAs the split between phoronid and inarticulate lineages occurred *before* the latter diverged into the various subphyla recognized as fossils (see also Cohen *et al*., 1998; Cohen & Weydmann, 2005). Although the disjunction between rDNA and morphological evolution must not be overlooked, this timing alone is enough to rule out the idea that a soft-shelled linguliform, *Lingulosacculus*, could also be a stem-group phoronid (Balthasar & Butterfield, 2009).

Glimpses of discinoid and craniid phylogeny

Hitherto, the root positions in the radiations of extant discinoids and craniids have not been located. Discinoids are linguliforms with shells of apatite-reinforced, organic polymer and are notable for the presence of silica tablets on the larval shell (Williams *et al*., 1998). Adults are generally sessile on sheltered, hard substrates, but juveniles are planktotrophic and long-lived, giving high dispersal potential, exemplified by the cosmopolitan distribution of *Pelagodiscus*. Despite similarly high juvenile dispersal potential, some geographical differentiation has been demonstrated amongst *Lingula* spp. around the Pacific (Endo, Ozawa & Kojima, 2001) and perhaps similar genetic diversity would be found amongst discinoids if more samples and data from faster-evolving genes were available (attempts to amplify mitochondrial rDNAs with standard primers have failed; B. L. C., unpubl. data). Discinoids are small, inconspicuous, and rarely collected, and I have too few samples for a robust or sensitive analysis, but I have found here that samples of *Pelagodiscus* from western Pacific Ocean and Antarctic (Scotia Sea) localities differ appreciably, and differ more from members of the *Discina/Dicinisca/Discradisca* complex, and that the root of the radiation lies, appropriately, on the branch between *Pelagodiscus* and the other genera. A more sensitive analysis of their inter-relations would be required to reliably separate *Discina* and *Discinisca*, both of which were from the West African coast, but *Discradisca* from Panama has slightly more divergent rDNAs, consistent with vicariance associated with opening of the Atlantic Ocean. The generic separation of *Discinisca* from *Discradisca* (Cooper, 1977) relies, however, on a weak morphological character and may need reconsideration.

Craniids are limpet-like inarticulate brachiopods with a distinctive, tabulate, calcitic shell fabric that testifies to the independent origins of calcitic mineralization in the articulate and inarticulate lineages. The analysis presented here provides evidence that the root of the extant craniid radiation lies on the branch joining the two main extant genera, *Novocrania* and *Neoancistrocrania*, rather than where previously placed (Cohen, Long & Saito, 2008). This result is consistent with analyses based on more comprehensive sequence data (B. L. C., unpubl. data, 2012) and as this is also the earliest divergence in the craniid gene tree it seems likely to be a reliable root position.

CONCLUSIONS

Robust ML molecular-clock analyses based on rDNAs from taxonomically representative and comprehen-

sive samples undisturbed by outgroups and other distant taxa confirm earlier conclusions that, so far as rDNA evolution is concerned, phoronids belong within the brachiopod clade as the sister of craniiform and linguliform (inarticulate) brachiopods, rather than outside the clade as the sister of all brachiopods or even further away. This outcome is as conclusive as any single-gene analysis can be, and more conclusive than any other analysis yet published, those based on wide taxon-range analyses being liable to misreport relationships by the operation of any combination of nine factors that have been referred to here. Implications of the phoronid + inarticulate clade for the evolutionary history of phoronid and brachiopod morphology and development are potentially substantial, but remain unsettled.

ACKNOWLEDGEMENTS

I am indebted to Dr J. Mallatt (Washington State University, Pullman) for extensive help and grateful to several colleagues for comments on earlier versions of this paper. An anonymous reviewer alerted me to the inappropriate combinations of articulate and inarticulate brachiopod sequences reported in a cited paper's supplementary information.

REFERENCES

Andrade SCS, Strand M, Schwartz M, Chend H, Kajihara H, von Döhren J, Sung S, Junoy J, Thiel M, Norenburg JL, Turbeville JM, Giribet G, Sundberg P. 2012. Disentangling ribbon worm relationships: multi-locus analysis supports traditional classification of the phylum Nemertea. *Cladistics* **28:** 141–159.

Balthasar U. 2004. Shell structure, ontogeny and affinities of the Lower Cambrian bivalved problematic fossil *Mickwitzia muralensis* Walcott, 1913. *Lethaia* **37:** 381–400.

Balthasar U, Butterfield NJ. 2009. Early Cambrian 'soft-shelled'brachiopods as possible stem-group phoronids. *Acta Palaeontolgica Polonica* **54:** 307–314.

Brinkmann H, Philippe H. 1999. Archea sister group of bacteria? Indications from tree reconstruction artifacts in ancient phylogenies. *Molecular Biology and Evolution* **16:** 817–825.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* **17:** 540–552.

Cohen BL. 2000. Monophyly of brachiopods and phoronids: reconciliation of molecular evidence with Linnaean classification (the subphylum Phoroniformea nov.). *Proceedings of the Royal Society, London, Series B* **267:** 225–231.

Cohen BL. 2007. The brachiopod genome. In: Selden PA, ed. *Treatise on invertebrate paleontology. Part H6 (Supplement), Brachiopoda (revised)*. Lawrence, KS: Geological Society of America and University of Kansas Press, 2356–2372.

Cohen BL, Gawthrop AB. 1996. Brachiopod molecular phylogeny. In: Copper P, Jin J, eds. *Brachiopods: proceedings of the third international brachiopod congress, Sudbury, Ontario. 1995*. Rotterdam: Balkema, 73–80.

Cohen BL, Gawthrop AB. 1997. The brachiopod genome. In: Kaesler RL, ed. *Treatise on invertebrate paleontology. Part H1, Brachiopoda, revised*. Lawrence: Geological Society of America and University of Kansas Press, 189–211.

Cohen BL, Gawthrop AB, Cavalier-Smith T. 1998. Molecular phylogeny of brachiopods and phoronids based on nuclear-encoded small subunit ribosomal RNA gene sequences. *Philosophical Transactions of the Royal Society, B* **353:** 2039–2061.

Cohen BL, Long SL, Saito M. 2008. Living craniids: preliminary molecular evidence of their inter-relationships. *Fossils and Strata* **54:** 283–287.

Cohen BL, Weydmann A. 2005. Molecular evidence that phoronids are a subtaxon of brachiopods (Brachiopoda: Phoronata) and that genetic divergence of metazoan phyla began long before the Early Cambrian. *Organisms, Diversity & Evolution* **5:** 253–273.

Conway Morris S, Peel JS. 1995. Articulated halkierids from the Lower Cambrian of North Greenland and their role in early protostome evolution. *Philosophical Transactions of the Royal Society, B* **347:** 305–358.

Cooper GA. 1977. *Brachiopods from the Caribbean Sea and adjacent waters*. Coral Gables: University of Miami Press, 1–211.

Davidson E, Levin M. 2005. Gene regulatory networks. *Proceedings of the National Academy of Sciences, Washington* **102:** 20063–20066.

Douzery EJP, Snell EA, Bapteste E, Delsuc F, Philippe H. 2004. The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils. *Proceedings of the National Academy of Sciences, USA* **101:** 15386–15391.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *Plos Biology* **4:** e88.

Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, Smith SA, Seaver E, Rouse GW, Obst M, Edgecombe GD, Sorensen MV, Haddock SHD, Schmidt-Rhaesa A, Okusu A, Kristensen RM, Wheeler WC, Martindale MQ, Giribet G. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* **452:** 745–750.

Endo K, Ozawa T, Kojima S. 2001. Nuclear and mitochondrial gene sequences reveal unexpected genetic heterogeneity among northern Pacific populations of the brachiopod *Lingula anatina*. *Marine Biology* **139:** 105–112.

Erwin DH, Laflamme M, Tweedt SM, Sperling EA, Pisani D, Peterson KJ. 2011. The Cambrian conundrum: early divergence and later ecological success in the early history of animals. *Science* **334:** 1091–1097.

Farris JS, Albert VA, Källersjö M, Lipscomb D, Kluge AG. 1996. Parsimony jackknifing outperforms neighbor-joining. *Cladistics* **12:** 99–124.

Felsenstein J. 2004. *Inferring phylogenies*. Sunderland: Sinauer Associates, MA1–664.

Freudenstein JV, Davis JI. 2010. Branch support via resampling: an empirical study. *Cladistics* **26:** 1–14.

Halanych K. 1995. The phylogenetic position of the pterobranch hemichordates based on 18S rDNA sequence data. *Molecular Phylogenetics and Evolution* **4:** 72–76.

Hausdorf B, Helmkampf M, Nesnidal MP, Bruchhaus I. 2010. Phylogenetic relationships within the lophophorate lineages (Ectoprocta, Brachiopoda and Phoronida). *Molecular Phylogenetics and Evolution* **55:** 1121–1127. 10.1016/j.ympev.2009.12.022.

Hillis DM, Moritz C, Mable BK, eds. 1996. *Molecular systematics, 1–655*. Sunderland, MA: Sinauer Associates.

Holmer LE, Skovsted CB, Larsson C, Brock GA, Zhang Z. 2011. First record of a bivalved larval shell in Early Cambrian tomotiids and its phylogenetic significance. *Palaeontology* **54:** 235–239.

Holmer LE, Skovsted CB, Williams A. 2002. A stem group brachiopod from the Lower Cambrian–support for a *Micrina* (halkieriid) ancestry. *Palaeontology* **45:** 875–882.

Huelsenbeck JP, Bollback JP, Levine AM. 2002. Inferring the root of a phylogenetic tree. *Systematic Biology* **51:** 32–43.

Huson DH. 1998. SplitsTree: analysing and visualizing evolutionary data. *Bioinformatics* **14:** 68–73.

Kaesler RL, ed. **1997–2007.** *Treatise on invertebrate paleontology. Brachiopoda, Parts H1-5, revised*. Boulder, CO and Lawrence, KS: Geological Society of America and University of Kansas.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23:** 2947–2948.

Lee DE, Brunton CHC. 2001. *Novocrania*, new name for the genus *Neocrania* Lee and Brunton, 1986 (Brachiopoda, Craniida); preoccupied by *Neocrania* Davis, 1978 (Insecta, Lepidoptera). *Bulletin of the Natural History Museum (Geology Series)* **57:** 1–5.

Maddison DR, Maddison WP. 2001. *MacClade 4: Analysis of phylogeny and character evolution, version 4.02*. Sunderland, MA: Sinauer Associates.

Maddison WP, Donoghue MJ, Maddison DR. 1984. Outgroup analysis and parsimony. *Systematic Zoology* **33:** 83–103.

Mallatt J, Craig CW, Yoder MJ. 2010. Nearly complete rRNA genes assembled from across the metazoan animals: effects of more taxa, a structure-based alignment, and paired-sites evolutionary models on phylogeny reconstruction. *Molecular Phylogenetics and Evolution* **55:** 1–17.

Mallatt J, Craig CW, Yoder MJ. 2012. Nearly complete rRNA genes from 371 Animalia: updated structure-based alignment and phylogenetic analysis. *Molecular Phylogenetics and Evolution* **63:** 604–617.

Nesnidal MP, Helmkampf M, Bruchhaus I, Hausdorf B. 2010. Compositional heterogeneity and phylogenomic inference of metazoan relationships. *Molecular Biology and Evolution* **27:** 2095–2104.

Nixon KC, Carpenter JM. 1993. On outgroups. *Cladistics* **9:** 413–426.

Paps J, Baguna J, Riutort M. 2009. Lophotrochozoa internal phylogeny: new insights from an up-to-date analysis of nuclear ribosomal genes. *Proceedings of the Royal Society, B* **276:** 1245–1254.

Passamaneck Y, Halanych KM. 2006. Lophotrochozoan phylogeny assessed with LSU and SSU data: evidence of lophophorate polyphyly. *Molecular Phylogenetics and Evolution* **40:** 20–28.

Philippe H, Delsuc F, Brinkmann H. 2005. Phylogenomics. *Annual Review of Ecology and Systematics* **36:** 541–562.

Posada D, Buckley TR. 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology* **53:** 793–808.

Posada D, Crandall KP. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14:** 817–818.

Ronquist FR, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19:** 1572–1574.

Ronquist FR, Huelsenbeck J, Teslenko M. 2011. Draft MrBayes version 3.2 manual: tutorials and model summaries, version 15 November 2011. Distributed with the software. Available at: http//:brahms.biology.rochester.edu/software.html

Rosenfeld JA, Payne A, DeSalle R. 2012. Random roots and lineage sorting. *Molecular Phylogenetics and Evolution* **64:** 12–20.

Sanderson MJ, Thorne JL, Wikstrom N, Bremer K. 2004. Molecular evidence on plant divergence times. *American Journal of Botany* **9:** 1656–1665.

Santagata S, Cohen BL. 2009. Phoronid phylogenetics (Brachiopoda; Phoronata): evidence from morphological cladistics, small and large subunit rDNA sequences, and mitochondrial cox1. *Zoological Journal of the Linnean Society* **157:** 34–50. 10.1111/j.1096-3642.2009.00531.x.

Siddall ME. 2010. Unringing a bell: metazoan phylogenomics and the partition bootstrap. *Cladistics* **26:** 444–452. 10.1111/j.1096-0031.2009.00295.x.

Simmons MP. 2000. A fundamental problem with amino-acid sequence characters for phylogenetic analysis. *Cladistics* **16:** 274–282.

Simmons MP, Freudenstein JV. 2002. Artifacts of coding amino acids and other composite characters for phylogenetic analysis. *Cladistics* **18:** 354–365.

Sinsheimer JS, Little RJA, Lake JA. 2012. Rooting trees without outgroups: EP rooting. *Genome Biology and Evolution*. doi: 10.1093/gbe/evs047.

Skovsted CB, Brock GA, Paterson JR, Holmer LE, Budd GE. 2008. The scleritome of *Eccentrotheca* from the Lower Cambrian of South Australia: lophophorate affinities and implications for tommotiid phylogeny. *Geology* **36:** 171–174.

Smith AB. 1994. Rooting molecular trees: problems and strategies. *Biological Journal of the Linnean Society* **51:** 279–292.

Sperling E, Pisani D, Peterson KJ. 2011. Molecular palaeobiological insights into the origin of the Brachiopoda. *Evolution & Development* **13:** 290–303.

Stumpf MPH, Porter MA. 2012. Critical truths about power laws. *Science* **335:** 665–666.

Swofford DL, 1998. *PAUP*: Phylogenetic Analysis Using Parsimony (and Other Methods), version 4.0*. Sinauer Associates, Sunderland Massachusetts.

Telford ML, Littlewood DTJ, eds. 2009. *Animal evolution: genomes, fossils, and trees, 1–245*. Oxford: Oxford University Press.

Thorne JL, Kishino H. 2002. Divergence time and evolutionary rate estimation with multilocus data. *Systematic Biology* **51:** 689–702.

Watrous LE, Wheeler QD. 1981. The out-group comparison method of character analysis. *Systematic Zoology* **30:** 1–11.

Wiens JJ, Morrill MC. 2011. Missing data in phylogenetic analysis: reconciling results from simulations and empirical data. *Systematic Biology* **60:** 719–731. doi: 10.1093/sybio/syr025.

Williams A, Carlson SJ, Brunton CHC. 1997. Linguliformea. In: Kaesler RL, ed. *Treatise on invertebrate paleontology. Part H2, Brachiopoda, revised*. Lawrence: Geological Society of America and University of Kansas Press, 30–157.

Williams A, James MA, Emig CC, Mackay S, Rhodes MC. 1997. Anatomy. In: Kaesler RL, ed. *Treatise on invertebrate paleontology. Part H1, Brachiopoda, revised*. Lawrence: Geological Society of America and University of Kansas Press, 7–188.

Williams A, Carlson SJ, Brunton CHC, Holmer L, Popov L. 1996. A supra-ordinal classification of the Brachiopoda. *Philosophical Transactions of the Royal Society, B* **351:** 1171–1193.

Williams A, Cusack M, Buckman JO, Stachel T. 1998. Siliceous tablets in the larval shells of apatitic discinid brachiopods. *Science* **279:** 2094–2096.

Zimmer RL. 1991. Phoronida. In: Giese AC, Pearse JS, Pearse VB, eds. *Reproduction of marine invertebrates*. Pacific Grove, CA: Boxwood Press, 2–45.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

File S1. Alignment in Nexus format.
File S2. Alignment after GBlocks treatment.
File S3. Saturation analysis.
File S4. Maximum likelihood, strict clock tree with jackknife support.
File S5. Nine factors that can distort wide taxon-range analyses.