An even "newer" animal phylogeny

Rob DeSalle[1]* and Bernd Schierwater[2]

Summary
Metazoa are one of the great monophyletic groups of organisms. They comprise several major groups of organisms readily recognizable based on their anatomy. These major groups include the Bilateria (animals with bilateral symmetry), Cnidaria (jellyfish, corals and other closely related animals), Porifera (sponges), Ctenophores (comb jellies) and a phylum currently made up of a single species, the Placozoa. Attempts to systematize the relationships of these major groups as well as to determine relationships within the groups have been made for nearly two centuries. Many of the attempts have led to frustration, because of a lack of resolution between and within groups. Other attempts have led to "a new animal phylogeny". Now, a study by Dunn et al.,⁽¹⁾ using the expresssed sequence tag (EST) approach to obtaining high-throughput large phylogenetic matrices, presents an "even newer" animal phylogeny. There are two major aspects of this study that should be of interest to the general biological community. First, the methods used by the authors to generate their phylogenetic hypotheses call for close examination. Second, the relationships of animal taxa in their resultant trees also prompt further discussion. *BioEssays* 30:1043–1047, 2008. © 2008 Wiley Periodicals, Inc.

Introduction

How animals are related to each other has been a major preoccupation of biologists for centuries. Some bold ideas about animal relationships have populated the literature, caused controversy and, in at least one case,—Cuvier and Geoffroy—destroyed a collegial and personal relationship. Most early anatomical systems of classification at this level suffered because of difficulties in establishing homology of anatomical parts. Two historical examples of this problem involve (1) the long-standing issue of establishing homology of dorsal–ventral polarity (an argument that caused the main riff between Cuvier and Geoffroy and resulted in their great confrontation in Paris that started in 1830;(2) and (2) the longstanding doubts about the existence of a monophyletic Protostomia (Box 1).

Anatomy has had and continues to have a predominant role in classifying animals into higher groups; for example, the three major animals groups—Bilateria, Cnidaria, Porifera—can easily be distinguished from each other using morphology. The relationships of several enigmatic groups of animals within these phyla causes much more confusion. In addition, the relationships of the three major groups to each other, and the placement of two other important taxa (Ctenophora and Placozoa) have caused problems with this part of the tree of life.

Molecules to the rescue?

It was only 20 years ago that molecular analysis claimed to address the major problems in animal phylogeny. At this time, microbiologists had started to successfully unravel the relationships of the difficult to decipher Bacteria and Archaea using 16S rRNA sequences as characters for systematic analysis. In line with these studies in microbial systematics, Field et al.,(5) used eukaryotic18S rRNA genes as a source of characters from several metazoan taxa, and proposed one of the first molecular hypotheses for this part of the tree of life. This landmark paper is remembered more for its prophetic focus on the problems associated with animal phylogenetics, than for the overall phylogenetic hypothesis proffered by the analysis. The prophecies of the Field et al.,(5) paper include pointing to difficulties in determining robustness of inference. Bootstrapping was only beginning to be developed for phylogenetic analysis at the time, and both jackknifing and Bayesian phylogenetic inference were chocolate bars in some systematists' back pockets. Another issue prophesized by Field et al.,(5) was a dutiful concern for how to treat molecular information in light of morphological patterns. The authors of this paper also pointed to varying phylogenetic incongruence as a result of the choice of phylogenetic method (at that time parsimony versus distance). Finally, the paper was also rather clairvoyant with respect to some of the major controversial hypotheses about animals and added to the number of hypotheses for this part of the tree of life. The Field et al.,(5) paper predated high-throughput sequencing techniques, sophisticated likelihood analyses and hence a strong understanding of long branch attraction, data combination techniques, the impact of taxon sampling and several other

[1] American Museum of Natural History, New York, Sackler Institute for Comparative Genomics, NY.
[2] ITZ, Ecology & Evolution, TiHo Hannover, Bünteweg Germany.
*Correspondence to: Rob DeSalle, American Museum of Natural History, New York, Sackler Institute for Comparative Genomics, 79 St at Central Park West, New York, NY 10024. E-mail: desalle@amnh.org
DOI 10.1002/bies.20842

Box 1. Definitions

Cuvier Geoffroy debate: Georges Cuvier and Geoffroy St. Hillaire were colleagues at the Paris museum in the early 1800s. Cuvier actually was instrumental in hiring Geoffroy into a curatorial position at the museum. Cuvier had very stringent ideas about animal forms and refused to see connections between the major animal groups that he had erected as a part of his life's work. These major groups were his famous four embranchments—Vertebrata, Articulata, Radiata and Mollusca. Geoffroy challenged this way of thinking by proposing that there was a connection between vertebrates and invertebrates via developmental observations. His famous upside down lobster published in 1822 was the beginning of his thinking in this direction. In early 1830, a significant debate between the two colleagues ensued and became the focus of news in Europe. The debate raged for several years, causing a major rift between the two colleagues. For a detailed description of the two men and their debate see Appels.(2)

Long branch attraction (LBA): LBA is a well recognized phenomenon in molecular systematics that grew out of the work of Joseph Felsenstein at the University of Washington. The premise of this phenomenon is based on the observation that some genes in certain taxa have an inordinate amount of evolutionary change. This large amount of sequence change produces evolutionary trees with taxa with extremely long branches. When this phenomenon occurs a taxon with a long branch will have a very high probability of converging randomly with other such taxa with long branches. The two taxa are attracted to each other because of their long branches and will appear as sister taxa in a phylogenetic tree when in reality they have little or no phylogenetic affinity. Molecular systematists have argued that the way to overcome this problem of LBA, is to impose a model of sequence evolution that takes into account the inordinately high degree of sequence change in the taxa or to add more taxa to break up the long branch.

Support: One of the tests as to whether or not a data matrix is composed of robust information is to assess the support that the matrix gives to relationships reflected by nodes in the tree. Currently there are several approaches to assessing the robustness of nodes in a tree. The Dunn et al.(1) paper uses two methods to assess support for nodes in the trees produced in their analyses bootstrap and Bayesian statistics.

Bootstrap: A node's robustness in a tree can be assessed with the *bootstrap*, a procedure that involves resampling with replacement of the characters in a phylogenetic matrix to produce a predetermined number of resampled matrices (usually greater than 1000). A related approach called the jackknife (resampling WITHOUT replacement) exists and in general the results of the two approaches are fairly well correlated. In the bootstrap, each resampled matrix is then analyzed using phylogenetic methods (parsimony, neighbor joining, maximum likelihood) and trees generated from the analysis of the resampled matrices are assessed for the number of times that a particular node exists in a tree produced from a resampled matrix. A bootstrap value of, say 80%, at a node in a tree indicates that the node occurs in 80% of the trees produced by resampling. The bootstrap values at a node range from 0 to 100 percent. Some empirical studies have indicated that bootstrap values below 70% indicate a node is poorly supported by the data in a matrix. The second method Dunn et al. use is to perform Bayesian analysis of the matrix.

Bayesian phylogenetic analysis: While bootstrap and jackknife approaches generate a proportion of resampled matrices that are in agreement with nodes in a phylogenetic tree, Bayesian approaches generate a posterior distribution or probability for a parameter that comprises a phylogenetic tree and a model of evolution. Both of these are founded on the prior probability of the parameter and the likelihood of the data. In order to compute the posterior probabilities of nodes, Bayesian methods rely heavily on Markov Chain Monte Carlo simulation algorithms.

Leaf stability metrics: Leaf instability is most often associated with long branches in phylogenetic analysis. A small number of unstable taxa can create low overall support values if a single taxon ''wanders'' between two clades. If a wandering taxon occurs in one clade 50% of the time, then the bootstrap value for that entire clade will be 50%, even if all other relationships in the clade occur 100% of the time. Leaf stability measures are calculated by taking triplets of taxa from a larger data set and calculating the occurrence of such triplets in trees obtained using an optimality criterion such as likelihood. Taxa that are stable leafs will also show stability in triplets. The leaf stability cutoff used by Dunn et al.(1) of 90% percent means that taxa with leaf stabilities less than 90% were involved in unstable triplets an average of 10% of the time.

Models (WAG model and PhyloBayes CAT model): Models are critical for likelihood and Bayesian approaches. Two kinds of models were used in the Dunn et al.(1) study. The first employs a model of residue changed developed by Whelan and Goldman(3) that consists of a transformation matrix derived from empirical observations of amino acid sequence change computed from a database made up of 3,905 sequences split into 182 protein families. The second approach is a Bayesian approach that utilizes a categorization model called CAT. Technically, the CAT model utilizes a Dirichlet process mixture model that specifies a frequency vector over the 20 amino acids in the target sequences that is then used in likelihood calculations (for details see Ref. 4). These frequency vectors can be viewed as profiles of the substitution process involved in sequence change.

important conceptual advances in phylogenetics developed in the last two decades that are relevant to deciphering the relationships of animals specifically, and the tree of life, in general.

A new animal phylogeny

With greater advance in DNA-sequencing technologies, studies of animal phylogeny progressed to include more taxa and a few more genes. Halanych(6) summarized the various molecular studies addressing the animal tree of life up to 2004, and presented ''a new animal phylogeny'' based on a personalized (albeit careful and comprehensive) supertree approach. A mitigating factor in this supertree analysis was that it was based on synthesis of studies using one, or at most, only a few genes as sources of characters. This new animal phylogeny was unique, in that it hypothesized several animal groups that flew in the face of classical animal relationships. Basically, molecular data to that date supported the putting aside of the Protostomia, and instead, the erection of two new clades called the Ecdysozoa and Lophotrochozoa. Deuterostomia remained as in classical anatomical studies, a ''good'' group. More recent analyses of the problem using multiple genes have been depressing with respect to the resolution of the major groups of Metazoa,(7) and have concluded that, because of the rapid divergence of the lineages involved, resolution with high robustness may never be obtained. Closer examination of the problem revealed that this depressing result of poor resolution, however, could be caused by limited taxon sampling.(8)

Better gardening tools (and fertilizer) for the ''newer'' metazoan tree

So what have two decades of molecules taught us about how to approach the metazoan tree of life? An even ''newer'' animal phylogeny generated by Dunn et al.,(1) published in Nature can shed light on this question as well as give us a better view of the overall patterns of divergence of animals. Dunn et al.,(1) have taken precautions to deal with the prophesies of Field et al.(5) These approaches address taxon sampling, deal with long-branch attraction, and use support metrics to measure the robustness of nodes in their data set. They do this in the following ways. (1) *Sampling*—Using expressed sequence tags (ESTs) from a large number of carefully and comprehensively chosen taxa, Dunn et al.(1) constructed the largest matrix to date addressing this difficult part of the tree of life. Their matrix includes 71 metazoan taxa and 6 outgroups for 150 genes and was analyzed with likelihood and Bayesian approaches. The initial analysis included all 77 taxa and added 29 previously unanalyzed taxa. 2) *Long-branch attraction*—Two methods dealing with the vagaries of long-branch attraction were employed. This first approach involves using different models of residue change and also likelihood analyses (RaxML)(9) The second concerns the examination of taxa for their stability in the phylogenetic tree. (3) *Support*—While the initial analyses with 77 taxa recovered monophyletic Metazoa, Bilateria, and Protostomia with robust support, the Lophotrochozoa and Ecdysozoa were recovered with much lower support values. The authors claim this lower degree of support is the result of instability of a small proportion of taxa in the data set. Consequently, Dunn et al.(1) removed thirteen taxa from their data set based on *leaf stability* values calculated for each taxon. Analyses of the 64 taxon matrix then resulted in a topology with monophyletic Metazoa, Bilateria, Protosotmia, Lophotrochozoa and Ecdysozoa at high bootstrap values and Bayesian posteriors.

Pycnogonids, Ctenophores and water bears! Oh my!

There is too much to discuss with respect to the biological ramifications of this ''even newer'' phylogeny in a single report, and so only a few salient points can be mentioned here (Fig. 1). Satisfyingly, all nodes in the monophyletic Arthropoda are supported unambiguously. One interesting relationship that has to date eluded molecular analyses, concerns the sister group relationship of pycnogonids (sea spiders) with chelicerates. The side effects of the placement of pycnogonids as the sister group to chelicerates include the rejection of pycnogonids as the earliest branching arthropod lineage and the rejection of the Mandibulata—a group hypothesized to contain Myriapoda, Crustacea and Hexipoda. Instead, myriapods are the sister group of chelicerates + pycnogonids.

Of great interest is the novel re-organization of the Lophotrochozoa into two large clades and three smaller new monophyletic groupings (clade A, clade B and clade C) from the analysis. The first smaller clade (A), contains Phoroninida, Nermertea and Brachiopoda. The second clade (B) contains Clade A and Annelids. The third clade (C) contains clade B and Mollusca. Sister to clade C are the entoprocts and these two make up one of the two overall larger Lophotrochozoan clades. The rest of the organization of the Lophotrochozoa includes the second larger clade consisting of rotifers, acoels, gastrotrichs, Platyhelminthes, and Myzostomida.

One strange relationship addressed by the authors remains enigmatic. This relationship concerns the placement of the tardigrades. These little animals commonly known as water bears, have classically been allied with arthropods + onychophorans. More recent molecular data have suggested that they are allied to nematodes. The results of the Dunn et al.(1) study are ambiguous with respect to the placement of tardigrades though, as depending on the model employed, the tardigrades can be allied with the arthropod + onychophoran group (the *Phylo-Bayesian CAT model*(4)) or with nematodes (the *WAG model*(3)).

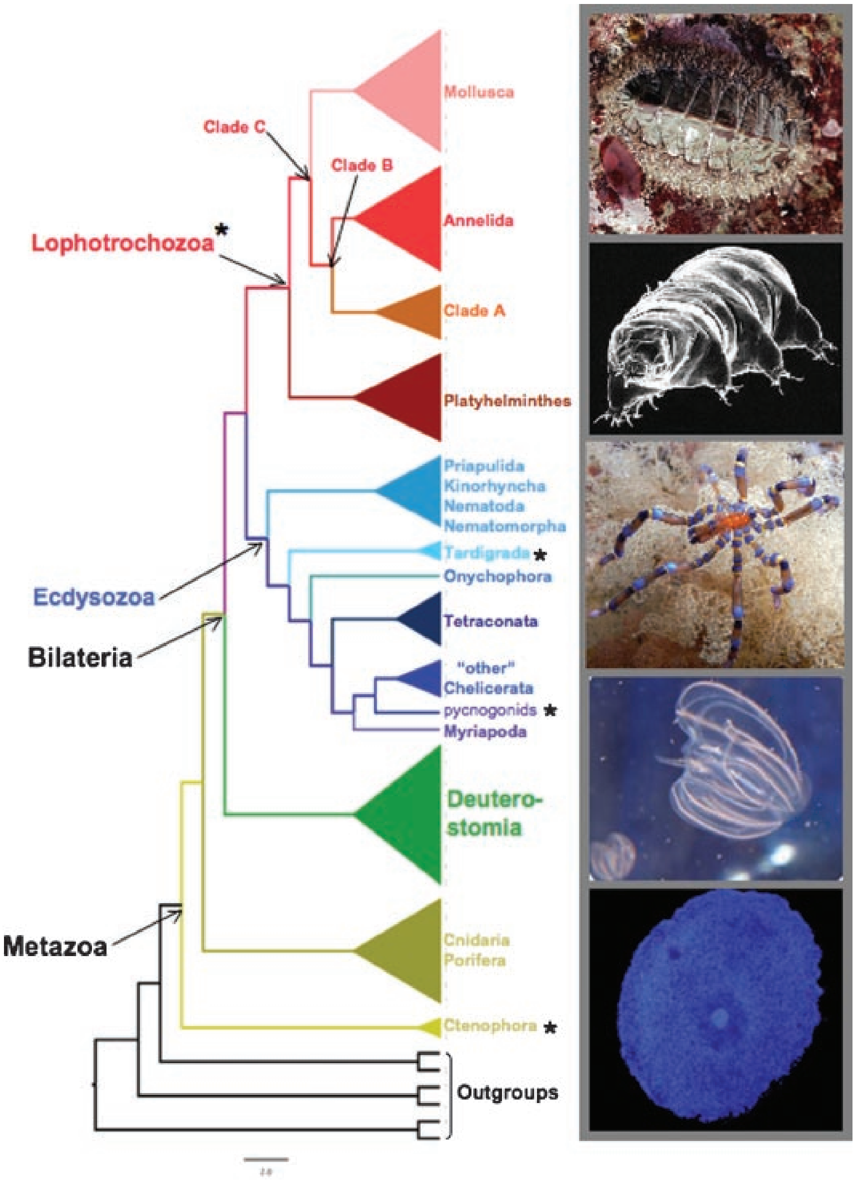

Figure 1. A redrawing of the 64 taxon Phylobase CAT analyzed tree of Dunn et al., (based on Fig. 2 from Ref. 1). The three major groups of Bilateria are indicated by shade of color; green = Deuterstomia, blue = Ecdysozoa, and red = Lophotrochozoa. Outgroups are black and basal animals are pale green. Asterisks mark parts of the tree discussed in the text. For exact taxon names refer to Dunn et al.(1) Photographs on the right represent several of the important results of the "even newer" metazoan tree of life. From top to bottom: A chiton (*Mopalia lignose* from the Big Sur coastline of central California). This animal is a mollusk and a member of the Lophotrochozoa group discussed in the text; photo courtesy of D. Ernisse, California State University at Fullerton. A tardigrade (Tardigrada sp.; a water bear) courtesy of R. Gillis and R.J. Haro, University of Wisconsin—La Crosse at http://serc.carleton.edu/microbelife/topics/tardigrade/. A pycnogonid (*Pseudopallene* sp. from Tasmania); courtesy of. C. Arango & G. Cranitch, Queensland Museum A ctenophore, courtesy of E. Bauer, Smithsonian's National Zoo. A placozoan; this image shows a female placozoan animal growing a large oocyte in the centre of its body. This animal is most likely a new species from an as yet unnamed genus, that shows reproductive isolation; image produced by DAPI staining of nuclei and is the courtesy of the author B.S.

One final result that demands attention is the placement of Ctenophora as the most-basal metazoan. This placement is relatively robust under the conditions of their analysis. But a side effect of the placement of Ctenophora at the base of the Metazoa appears to be the placement of Porifera with Cnidaria in a monophyletic group. We suspect that this arrangement may change with further taxon sampling with taxa such as Placozoa and other kinds of sponges as well as a full representation of the four classes of Cnidaria.

Are we near the end of metazoan phylogenetics with this study? Obviously the tree isn't done yet. But we are nearer to a stable metazoan phylogenetic backbone; at least one where the only prophecy necessary is to claim that things will improve with more sampling.

References

1. Dunn CW, Hejnol A, Matus DQ, Pang K, Brown WE, et al. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452:745–749.
2. Appel T. 1987. The Cuvier-Geoffrey Debate. French Biology in the Decades before Darwin. Oxford: Oxford University Press.
3. Whelan S, Goldman N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol 18:691–699.
4. Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol Biol Evol 21:1095–1109.

5. Field KG, Olsen G, Lane DJ, Giovannoni SJ, Ghiselin MT, et al. 1988. Molecular phylogeny of the animal kingdom. Science 239:748–753.
6. Halanych KM. 2004The new view of animal phylogeny. Ann Rev Ecol Evol Sys 35:229–256.
7. Rokas A, Kruger D, Carroll SB. 2005. Animal evolution and the molecular signature of radiations compressed in time. Science 310:1933–1938.
8. Baurain D, Brinkmann H, Philippe H. 2006. Lack of resolution in the animal phylogeny: closely spaced cladogenesis or undetected systematic errors? Mol Biol Evol 24:6–9.
9. Stamatakis A, Ludwig T, Meier H. 2005. RAxML-III: A Fast Program for Maximum Likelihood-based Inference of Large Phylogenetic Trees. Bioinformatics 21:456–463.