Improved Phylogenomic Taxon Sampling Noticeably Affects Nonbilaterian Relationships

K.S. Pick,†,1 H. Philippe,†,3 F. Schreiber,4 D. Erpenbeck,1 D.J. Jackson,2 P. Wrede,5 M. Wiens,5 A. Alié,6 B. Morgenstern,4 M. Manuel,6 and G. Wörheide*,1

1Department of Earth and Environmental Sciences, Palaeontology and Geobiology & GeoBio-CenterLMU, Ludwig-Maximilians-Universität München, München, Germany
2Department of Geobiology, Courant Research Center Geobiology, Georg-August Universität Göttingen, Göttingen, Germany
3Centre Robert-Cedergren, Département de Biochimie, Université de Montréal, Montréal, Québec, Canada
4Abteilung Bioinformatik, Institut für Mikrobiologie und Genetik, Georg-August-Universität Göttingen, Göttingen, Germany
5Department of Applied Molecular Biology, Institute for Physiological Chemistry and Pathobiochemistry, Mainz, Germany
6University Pierre & Marie Curie (UPMC), Centre National de la Recherche Scientifique, Muséum National d'Histoire Naturelle, Department UMR7138 Systématique, Adaptation, Evolution. UPMC, Paris, France
†These authors contributed equally to the present study.
*Corresponding author: E-mail: woerheide@lmu.de.
Associate editor: Manolo Gouy

Abstract

Despite expanding data sets and advances in phylogenomic methods, deep-level metazoan relationships remain highly controversial. Recent phylogenomic analyses depart from classical concepts in recovering ctenophores as the earliest branching metazoan taxon and propose a sister-group relationship between sponges and cnidarians (e.g., Dunn CW, Hejnol A, Matus DQ, et al. (18 co-authors). 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452:745–749). Here, we argue that these results are artifacts stemming from insufficient taxon sampling and long-branch attraction (LBA). By increasing taxon sampling from previously unsampled nonbilaterians and using an identical gene set to that reported by Dunn et al., we recover monophyletic Porifera as the sister group to all other Metazoa. This suggests that the basal position of the fast-evolving Ctenophora proposed by Dunn et al. was due to LBA and that broad taxon sampling is of fundamental importance to metazoan phylogenomic analyses. Additionally, saturation in the Dunn et al. character set is comparatively high, possibly contributing to the poor support for some nonbilaterian nodes.

Key words: multigene analysis, EST, Metazoa, Ctenophora, Porifera, long-branch attraction, saturation.

Resolving the relationships of deep branching metazoan lineages is critical if we are to understand early animal evolution. Unraveling these relationships through the analysis of large-scale molecular data sets has recently given birth to the field of phylogenomics (e.g., Philippe et al. 2005). Despite significant advances in this field, recent studies have generated contradictory results regarding relationships within and between early diverging metazoan lineages: cnidarians, ctenophores (comb jellies), sponges, placozoans (anatomically the simplest extant metazoans), and bilaterians. Placozoans have historically been regarded by some as relics of the metazoan ancestor (see summary by Schierwater 2005), and some recent analyses place Placozoa at the base of a group of nonbilaterian animals (Dellaporta et al. 2006; Schierwater et al. 2009). However, recent whole-genome (Srivastava et al. 2008) and phylogenomic (Philippe et al. 2009) analyses including *Trichoplax* recovered sponges as the sister group to all other metazoans in accordance with morphological analyses (Ax 1996). Such contradictory hypotheses regarding nonbilaterian metazoan relationships prevent a consensus view of metazoan evolution, a goal that is of fundamental importance if we hope to fully understand the early evolution of animals (for an overview see Erpenbeck and Wörheide 2007).

A recent phylogenomic analysis adds further controversy to this debate (Dunn et al. 2008) (c.f., Hejnol et al. 2009). Their outcome is highly unusual as sponges form a clade with the Cnidaria, while the ctenophores (despite being morphologically derived) are proposed to be the earliest branching metazoan taxon. As suggested by Philippe et al. (2009), we hypothesized that a long-branch attraction (LBA) artifact was responsible for these controversial findings due to insufficient ingroup sampling and an inappropriate choice of outgroup taxa. Furthermore, the Placozoa are conspicuously absent from the Dunn et al. (2008) data set, and sponges are represented by only one Demospongiae and one Homoscleromorpha with no representatives of the remaining two extant sponge classes: Calcarea (Calcispongiae or calcareous sponges) and Hexactinellida (glass sponges). Sparse taxon sampling is a common pitfall of phylogenetic analyses (Lecointre et al. 1993) and is largely responsible for the lack of a robustly supported nonbilaterian metazoan phylogeny (Erpenbeck and Wörheide 2007). With a largely different gene set (only 45 genes in common

Open Access

Letter

with the 150 gene set of Dunn et al. 2008) and an increased sampling of nonbilaterian species, Philippe et al. (2009) obtained monophyletic sponges as the first-diverging metazoan lineage and a sister-group relationship between the Cnidaria and the Ctenophora.

To test whether insufficient sampling of nonbilaterian taxa and inappropriate outgroup choice adversely influenced the analyses performed by Dunn et al. (2008), we reanalyzed their 64-taxon matrix cleared of instable taxa (leaf stability <90%) and with the following major modifications (cf. Baurain et al. 2007):

1) Ingroup taxon sampling was increased by the addition of nonbilaterian expressed sequence tag and genomic sequences. These included: 12 additional sponge taxa representing all four major sponge lineages; 1 additional ctenophore; 5 additional cnidarians (see supplementary table S1, Supplementary Material online), and *Trichoplax adhaerens* (Placozoa).

2) We removed outgroup taxa with long branches. Long branches in the outgroup can strongly influence the topology of early branching ingroups (Philippe and Laurent 1998; Rota-Stabelli and Telford 2008). The long branches of the fungal outgroup are not visible in the cladogram of the PhyloBayes analysis (CAT + Γ4) of Dunn et al. (see their fig. 2) but are evident in their supplementary figure S1 (Supplementary Material online). Consequently, we analyzed our data set with two sets of outgroups. First, using only choanoflagellates, the most likely sister group to all Metazoa (Carr et al. 2008), consisting of *Monosiga ovata* (shortest branch of outgroup taxa of Dunn et al.), *Monosiga brevicollis* (complete genome data), and *Proterospongia* sp.. Second, with more distant outgroups, such as those used by Dunn et al. (2008) (see supplementary fig. S1 and supplementary data, Supplementary Material online, for a detailed taxon list and methods used).

Furthermore, we eliminated errors (e.g., frameshifts) and refined the alignment of Dunn et al. (2008), for example, by reducing missing data and removing 2,150 ambiguously aligned positions (see supplementary data, Supplementary Material online, for detailed procedures). Our extended data set with the choanoflagellate-only outgroup consists of 80 taxa and 19,002 characters. Using this data set we performed Bayesian phylogenetic analyses under the CAT + Γ4 model (Lartillot and Philippe 2004) and subsequent nonparametric bootstrapping (cf. Philippe et al. 2009).

Contrary to Dunn et al. (2008), and also Hejnol et al. (2009), we recover sponges as the sister group to all other metazoan taxa (fig. 1). This is in congruence with earlier morphological (Ax 1996) and phylogenomic analyses (Philippe et al. 2009). In accordance with the latter study, we also recover sponges as a monophyletic group. The Homoscleromorpha, a taxon previously assigned to the Demospongiae (see Hooper and Van Soest 2002), are found to be the sister group to Calcarea as suggested by van Soest (1984) and Grothe (1989) based on morphology and subsequently by Dohrmann et al. (2008) based on ribosomal RNA (rRNA) data. Similarly, Hexactinellida and the

remaining Demospongiae sensu stricto form a monophyletic group (Silicea sensu stricto).

The basal position of ctenophores proposed by Dunn et al. (2008) was probably caused by the attraction of ctenophores to distant outgroup species, particularly fungi. In comparison, our reanalysis of the updated Dunn et al. (2008) data set with increased ingroup taxon sampling and a refined alignment indicates that LBA is reduced, independent of whether we use the choanoflagellate-only outgroup or more distant outgroups (see fig. 1; supplementary fig. S1, Supplementary Material online). This indicates that in-goup taxon sampling and probably to a lesser extent data refinement are the most important parameters affecting nonbilaterian relationships.

Results of our analyses indicate that sponges are the sister group to the remaining Metazoa, and Placozoa are sister group to the Bilateria. We also recover both monophyletic Ctenophores and Cnidaria, but they are paraphyletic with respect to Placozoa + Bilateria (fig. 1). This is in contrast to the findings of Philippe et al. (2009) that supported the "Coelenterata hypothesis" (cf. Haeckel 1866), that is, a monophyletic Cnidaria + Ctenophora clade and a sister-group relationship between Coelenterata and Bilateria. However, support values for the position of Ctenophora, Cnidaria, and Placozoa in our analysis are either not significant (posterior probabilities <0.9) or low (bootstrap support <70%). We suspected that character set of Dunn et al. contains a substantial amount of nonphylogenetic signal due to multiple substitutions. To test this, we conducted a saturation analysis of inferred substitutions against observed amino acid differences (fig. 2). This revealed a higher saturation in the original Dunn et al. (2008) character set (slope = 0.38×) compared with the character set of Philippe et al. (2009) (slope = 0.46×) (fig. 2). From this we conclude that despite increasing the number of nonbilaterian taxa by a factor of 3 (from 9 to 27), multiple substitutions have partly masked phylogenetic signal contributing to the incongruent results reported here with those of Philippe et al. (2009). However, with the expanded and refined data set reported here, none of these incongruencies are statistically significant, indicating that nonphylogenetic signal has been reduced with respect to the original character set of Dunn et al. (2008). Furthermore, Dunn et al. (2008) recovered high support for the sister-group relationship of ctenophores to the remaining Metazoa—based on our analyses here, this hypothesis should be rejected (with a bootstrap value of 91%).

The inclusion of additional taxa has little influence on the relationships within and between bilaterian crown groups. Three of the four differences between the findings of Dunn et al. (2008) and our results affect the relationships of a single sequence within their well-defined clades (*Euprymna* within Mollusca, *Paraplanocera* within Platyhelminthes, and *Anoplodactylus* among the chelicerate arthropods). None of these splits were strongly supported in the original Dunn et al. (2008) analysis. Additionally, we do not recover Panarthropoda due to a difference in
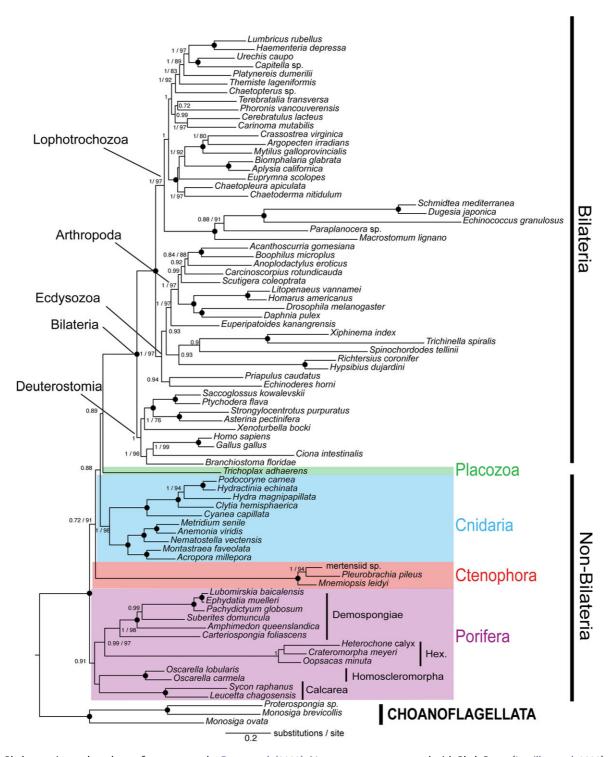
FIG. 1. Phylogenetic tree based on refinements to the Dunn et al. (2008) 64-taxon set reconstructed with PhyloBayes (Lartillot et al. 2009) under the CAT + Γ4 model. Choanoflagellates were set as outgroup and an additional 18 nonbilaterian taxa included. Posterior probabilities >0.7 are indicated followed by bootstrap support values >70. A large black dot indicates maximum support in posterior probabilities and Bayesian bootstraps (=1/100).

the position of Tardigrada. Panarthropoda was also weakly supported in the Dunn et al. (2008) analysis (posterior probability values under WAG and CAT models were 0 and 0.86, respectively, and RAxML bootstrap support under the WAG model with 64 and 77 taxa was 4% and 2%, respectively).

Our results highlight the sensitivity of phylogenomic studies to ingroup taxon sampling and demonstrate the need for great care in the analysis and interpretation of large data sets. Character-rich analyses are thought to outperform character-poor analyses and have been suggested to be of greater importance than increased taxon sampling
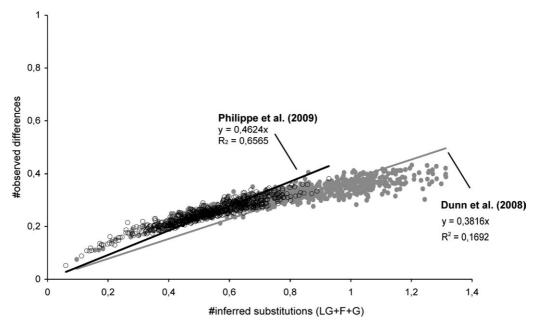
Fig. 2. Saturation plot of character sets. See Supplementary Material for method details. Gray line and filled dots: Dunn et al. (2008). Black line and open dots: Philippe et al. (2009).

with regard to recovering robust metazoan phylogenies (Rokas and Carroll 2005). However, our analyses demonstrate the strong influence of taxon sampling, even though nonbilaterian taxa still remain underrepresented (Cnidaria: no Octocorallia, Cerianthuria, Cubozoa, or Staurozoa; Ctenophora: no Platyctenida, Beroida, or Cestida; just one placozoan strain etc.). The phylogenomic approach promises to reveal a well-resolved consensus metazoan tree, but it should not be assumed that a large data set will automatically produce a strong or correct phylogenetic signal (Jeffroy et al. 2006). A wide range of factors, such as saturation, LBA, the best fitting evolutionary model, and appropriate outgroup choice (Philippe et al. 2005), need to be carefully addressed before a fully resolved and robust animal tree of life will be realized.

Supplementary Material

Supplementary figure S1, supplementary table S1, and supplementary data are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

References

Ax P. 1996. Das System der Metazoa. ein Lehrbuch der Phylogenetischen Systematik. Stuttgart (Germany): Gustav Fischer Verlag.

Baurain D, Brinkmann H, Philippe H. 2007. Lack of resolution in the animal phylogeny: closely spaced cladogeneses or undetected systematic errors? *Mol Biol Evol.* 24:6–9.

Carr M, Leadbeater BS, Hassan R, Nelson M, Baldauf SL. 2008. Molecular phylogeny of choanoflagellates, the sister group to Metazoa. *Proc Natl Acad Sci U S A.* 105:16641–16646.

Dellaporta SL, Xu A, Sagasser S, Jakob W, Moreno MA, Buss LW, Schierwater B. 2006. Mitochondrial genome of *Trichoplax adhaerens* supports placozoa as the basal lower metazoan phylum. *Proc Natl Acad Sci U S A.* 103:8751–8756.

Dohrmann M, Janussen D, Reitner J, Collins A, Wörheide G. 2008. Phylogeny and evolution of glass sponges (Porifera: Hexactinellida). *Syst Biol.* 57:388–405.

Dunn CW, Hejnol A, Matus DQ, et al. (18 co-authors). 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* 452:745–749.

Erpenbeck D, Wörheide G. 2007. On the molecular phylogeny of sponges (Porifera). In: Zhang Z-Q, Shear WA, editors. Linnaeus Tercentenary: Progress in Invertebrate Taxonomy. Zootaxa 1668. Auckland (New Zealand): Magnolia Press. p. 107–126.

Grothe F. 1989. On the phylogeny of homoscleromorphs. *Berl Geowiss Abh A.* 106:155–164.

Haeckel EH. 1866. Generelle Morphologie der Organismen. Berlin (Germany): G. Reimer.

Hejnol A, Obst M, Stamatakis A, et al. (17 co-authors). 2009. Assessing the root of bilaterian animals with scalable phylogenomic methods. *Proc R Soc Lond B Biol Sci.* 276:4261–4270.

Hooper JNA, Van Soest RWM. 2002. Systema Porifera. Guide to the Supraspecific Classification of Sponges and Spongiomorphs (Porifera). New York: Plenum.

Jeffroy O, Brinkmann H, Delsuc F, Philippe H. 2006. Phylogenomics: the beginning of incongruence? *Trends Genet.* 22:225–231.

Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25:2286.

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.

Lecointre G, Philippe H, Van Le HL, Le Guyader H. 1993. Species sampling has a major impact on phylogenetic inference. *Mol Phylogenet Evol.* 2:205–224.

Philippe H, Delsuc F, Brinkmann H, Lartillot N. 2005. Phylogenomics. *Ann Rev Ecol Syst.* 36:541–562.

Philippe H, Derelle R, Lopez P, et al. (20 co-authors). 2009. Phylogenomics restores traditional views on deep animal relationships. *Curr Biol.* 19:706–712.

Philippe H, Laurent J. 1998. How good are deep phylogenetic trees? *Curr Opin Genet Dev.* 8:616–623.

Rokas A, Carroll SB. 2005. More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. *Mol Biol Evol.* 22:1337–1344.

Rota-Stabelli O, Telford MJ. 2008. A multi criterion approach for the selection of optimal outgroups in phylogeny: recovering some support for Mandibulata over Myriochelata using mitogenomics. *Mol Phylogenet Evol.* 48:103–111.

Schierwater B. 2005. My favorite animal, Trichoplax adhaerens. *BioEssays* 27:1294–1302.

Schierwater B, Eitel M, Jakob W, Osigus H, Hadrys H, Dellaporta S, Kolokotronis S, Desalle R, Penny D. 2009. Concatenated analysis sheds light on early metazoan evolution and fuels a modern "urmetazoon" hypothesis. *PLoS Biol.* 7:e20.

Srivastava M, Begovic E, Chapman J, et al. (21 co-authors). 2008. The *Trichoplax* genome and the nature of placozoans. *Nature* 454:955.

van Soest RWM. 1984. Deficient *Merlia normani* from the Curaçao reefs, with a discussion on the phylogenetic interpretation of sclerosponges. *Bijdr Dierkd.* 54:211–219.