Xenacoelomorpha Flatworms are Basal Deuterostome

Yi Wang*

Department of Zoology, Fu Dan University, Shanghai, China

Research Article

ABSTRACT

Background: Whether position Xenacoelomorpha as an early branch of Bilateria (Protostomes+Deuterostomes) has been intensely debated during last several decades. Considering Darwin's "tree of life", with the "Phylogenetic Species Concept", we choose mitochondrial genome as the subject to predict phyloenetic position of Xenacoelomorpha, by genes geneology.

Results: Herein, we sequence *Heterochaerus australis* mitochondrial genome and infer intrinsic relationships of metazoan with *Xenacoelomorpha*. The optimal tree under the popular Maximum Likelihood (ML) and Bayesian phylogenetic reconstructions are consensus with each other being strongly supported. The relationship between Chordates, Ambulacrarians and *Xenoturbella*/Acoelomorph is resolved. To avoid previous query about alignment process, the datasets are alignmented and trimmed automatically. Reducing taxon or cutting outgroups cannot affect the relationship between Xenacoelomorpha and other metazoan. Meanwhile, analysis using CAT model and Dayhoff groups also supporting the prediction made by mtZOA, relaxing the restriction of alignment criteria (MAFFT, strategy G-ins-1, BLOSUM 62, 45, 30) introducing potential misleading signals cannot challenge the tree topology indicating our auto-alignment mitochondrial dataset is not artificially restricted one.

Conclusion: Finally, a repeatable prediction of the genes geneology with reliable statistical support places Xenacoelomorpha as a basal Deuterostome.

Keywords: Xenacoelomorpha; Metazoans; Genome; Lateral gene transfer ; Mitochondria.

INTRODUCTION

The "tree of life" represents the genealogy history of species and the "tree" with no doubt is metaphysically prior to any particular species concept [1]. As there is no consensus concerning the meaning of the term 'species', we choose practical one from the others, for the purpose of inferring relationships of *Xenacoelomorpha* with Metazoans.

There are two commonly used concept ill-named 'Biological Species Concepts (BSC) and 'Phylogenetic Species Concepts' (PSC). BSC defines species as "groups of interbreeding natural populations that are reproductively isolated from other such groups." and PSC -"A species is the smallest diagnosable cluster of individual organisms within which there is a parental pattern of ancestry and descent."[2,3]. The former needs in the filed endeavor, but the later bends us to take advantage of the dramatic progressing genome data size predicting phylogenetic position of Xenacoelomorpha. Choosing PSC to predict species geneology by using gene geneology is the main logic of this article.

The phylogenetic species concept includes three levels: Genes, organisms and species level. Same epistemology reminds us to take coalesced homolog genes geneology history as the representative of both organisms and species geneology once the picked genes is "majority enough" and their "tree" topology is in concordance with each other [1]. Obviously, there is no way currently to get the single objective tree with existing species [4], because of the evolutionary speciation events and genes fundamental but not totally sufficient "bearing alone" roles in species evolution history [5-8]. Practically, the prediction should be compatible with job purpose and the methodology should be self-adapting.

The methodology compatible with PSC includes three steps: 1. Homology assessment to get a data matrix; 2. Inferring species tree by genes trees' topology; 3. Methodological self-adapting [9]. We cannot deal all three steps simultaneously because of computational limitation. Thus, the so called "majority enough" corresponding to the first step claims for enough data source and sufficient number of homolog genes for concatenating, considering biological phenomenon such as Lateral Gene Transfer (LGT), in order to represent species geneology [10-13]. Unfortunately, current computational methods in step one using nuclear genome or transcriptome data is inadequate to fulfil the "majority criteria" finding the position of Xenacoelomorpha in the Metazoans concordance tree, regardless of a hundred or a thousand of genes concatenated, the alignment matrix is far less than the proportion of LGT alone in human nuclear genome (8% of total nuclear genes) [7]. Let alone other non-phylogenetic signals existing and the never proven "common history" of most nuclear genes [13].

Nevertheless, the mitochondrial genome as data source meets the "majority criteria" [1] using present alignment method though not 100% of genes in this kind of plasmon being included. Furthermore, with particularly uniparental reproduction in Metazoan [14], independent adaptive evolution rates [15], conservative biosynthesis function been selected by environment, physical segregated from nuclear genome [16], mitochondrial genes geneology shall represent the geneology of related species at a certain extent, which can be depicted as 'primary concordance tree' of Metazoan, to position Xenacoelomorpha.

The position of Xenacoelomorpha in the tree of life remains an unresolved question

Morphological studies previously suggest *Xenoturbella bocki* as a turbellarian flatworm in 1949 and in 1997 position *Xenoturbella* either as a basal bilateria or with molluscan [17-22]. we don't discuss their result in this article as PSC we chosen.

Molecular studies based primarily on nuclear genes place Xenoturbellida as sister group of the Ambulacraria (echinoderms+hemichordates) [2,3,12], a basal Deuterostomes [12,23], or sister group to all other bilateria [10,11,24]. Using mitochondrial sequences prefer to support a basal Deuterostome placement [12,26].

We briefly summaries their result and find out that the prediction based on nuclear genes cannot congruent with each other. Although some problem has been resolved, such as sequence contaminated [27,28], rRNA compositional bias [10,11,29], or topology misleading by inadequate choice of other Metazoan clades [10,11,12,22] but their limited alignment matrix is not majority enough to represent nuclear genome as we discussed above. Conversely, most mitochondiral genome analyses leads to congruent prediction but being questioned by their unrepeatable results with low statistical support. Both nuclear and mitochondria data suggest to group *Xenoturbella* with Acoelomorph as a clade named Xenacoelomorpha.

Phylogenomic inference from mitochondrial sequences to position Xenacoelomorpha always been complicated by several misleading factors such as compositional heterogeneity [30], Long Branch Attraction (LBA) [9] and accelerated substitution rates, which violate the models of evolution. Each of them have complex source of cause and cannot been absolutely separated from each other. Unfortunately, useful approaches have been used to alleviate system error but barely satisfactory.

Bourlat, et al. aimed to study the relationships among all Deuterostome groups with mitochondrial genome, leading to a result that *Xenoturbella* is a sister group of Ambulacraria but cannot group Cephalochordate with Echinoderm. The global topology error probably caused by lack of phylogenetic signal within 170 nuclear proteins and dissatisfactory choice of species the misplacing of *Branchiostoma* sp. among different datasets analyses [22].

Perseke, et al. expanded Bourlat, et al. dataset and using Phylobayes CAT model to alleviate compositional bias getting a position that *Xenoturbella* is an early branch of Deuterostome with unreliable PP support, which claims for far more suited model and improved alignment matrix either [25].

Later, Bourlat, et al. carried out Bayesian analyse with mtZOA leading to a basal placement of *Xenoturbella* within Deuterostome but the prediction of other clades in the same tree being fiercely criticized–grouping the urochordates with the echinoderms remains enigmatic in his study [26].

Peterson's, et al. team conducting 37 well-chosen species, mainly because of their low sequence saturation, supporting Xenacoelomorpha as a basal Deuterostomes. His arguing about grouping Acoelomorph and *Xenoturbella* together is generally accepted. Meanwhile, the low support of internode between Chordata and Ambulacraria, the compositional heterogenoity of mitochondrial genome and "slightly handmade" alignments make his work unrepeatable [10-12].

Helen E. Robertson inferring phylogenetic tree from mitochondrial genes retrieves Xenacoelomorpha as an early branch of Deuterostome but the statistical support for his prediction is week, probably caused by dissatisfactory choice of species [23].

Here, we have first illustrated the philosophical foundation why it is necessary to predicting Xenacoelomorpha's geneology with mitochondrial genome instead of a bag of nuclear genes. Afterwards, we briefly surveyed the relevant researches related to phylogenetic position. In the next section, our analysis, inheriting the achievements of predecessors, Xenacoelomorpha's using currently more suitable alignment methods to avoid "handmade", building varies matrix with filtered species directly accounts for compositional heterogeneity, excluding outgroups to elaborate LBA, leads to a robust consensus topology which hold the view that Xenacoelomorpha flatworms are basal Deuterostome. In sharp contrast, inferring phylogenetic position of Xenacoelomorpha by nuclear datasets is hazardous with LBA menace and cannot fill the gap between methodology and philosophy.

MATERIALS AND METHODS

Getting mitochondrial sequence

Adult specimens of *Heterochaerus australis* been cultured and asexual reproduced in laboratory. DNA was isolated from adult specimens after starved two weeks totally shaded to avoid contamination by their food and symbiosis algae, extracted total DNA as described in Arimoto, et al. [38]. Sequencing, assembling and annotation are carried out as described in Felix Guerrero, et al. [39].

Phylogenetic analysis

Concatenated amino acid alignments of twelve protein-coding gene from 39 species [12] including our Heterochaerus australis are used to infer all phylogenetic analyses in this article. Alignments are inferred by MAFFT V7 (G-ins-we, Blosom 62) and trimmed by Gblocks 0.91b (b1=default, b2=default, b3= default, b4=5, b5=h.). Maximum likelihood analyses and bootstrap test carried out by RAXML V8.2 ML+BP online platform. Bayesian analyses are executed by PHYLOBAYES V3.2.6. Cross-validation and posterior probably test is conducted by embedded package in PHYLOBAYES. Each Bayesian analyses carries out two separated chain and stops according to maximum difference less than 0.1 criteria.

RESULTS

The sister group of deuterostome

This study uses mitochondrial genome data to position Xenacoelomorpha based on *Heterochaerus australis* sequence and additional four Xenoturbellida [12]. Other species are similar to Kevin J. Peterson previously chosen. As the rate of sequence evolution is high, we choose amino-acid sequence to deal with. The sequences are alignmented by program MAFFT, strategy G-INS-1, scoring matrix for amino acid sequences BLOSUM 62, refined by Gblocks, parameters in detail in online method [31]. All Bayesian phylogenetic reconstructions are conducted by software package, PhyloBayes, and the optimal tree under the popular ML criterion is found by RAXML in this work. Consensus tree from bayesian analysis under the mtZOA model and ML analysis with mtZOA+I+PROTGAMMA, inferred from 13 proteins concatenation alignments including all 37 Metazoans sequences. Their tree topology congruent with each other and supported by much higher statistical support than almost ever been published – monophyly of Xenacoelomorpha including *Xenoturbella* and Acoelomorph (PP=1 and BP=100), Xenacoelomorpha as an early branch of Deuterostome (PP=0.99 and BP=90), Kevin J. Peterson's confusion about the "relationship between chordates and ambulacrarians" being resolved (PP=1 and BP=90) (Figure 1). Thus, we lead to a solid prediction of phylogenetic position of Xenacoelomorpha with auto-alignment methods avoiding "handmade", which concurs with the finding that Xenacoelomorpha is sister group of all remaining bilateria based on nuclear dataset.

Figure 1. Animal phylogeny based on mitochondrial proteins reconstructed using MtZOA under a Bayesian analyses and mtZOA+I+PROTGAMMA under a maximum likelihood analyses. *Xenoturbella* and Acoelomorph constitute a monophyly named Xenacoelomorpha (PP=1 and BP=100), Xenacoelomorpha as a basal deuterostome strongly supported (PP=0.99 and BP=90), placing chordates as a sister group of ambulacrarians (PP=1 and BP=90) and recovering the monophyly of deuterostomia (PP=0.99 and BP=90). Topology of both trees is nearly identical. Analysis carried out on auto-trimmed alignments. Scale units are substitutions per site.
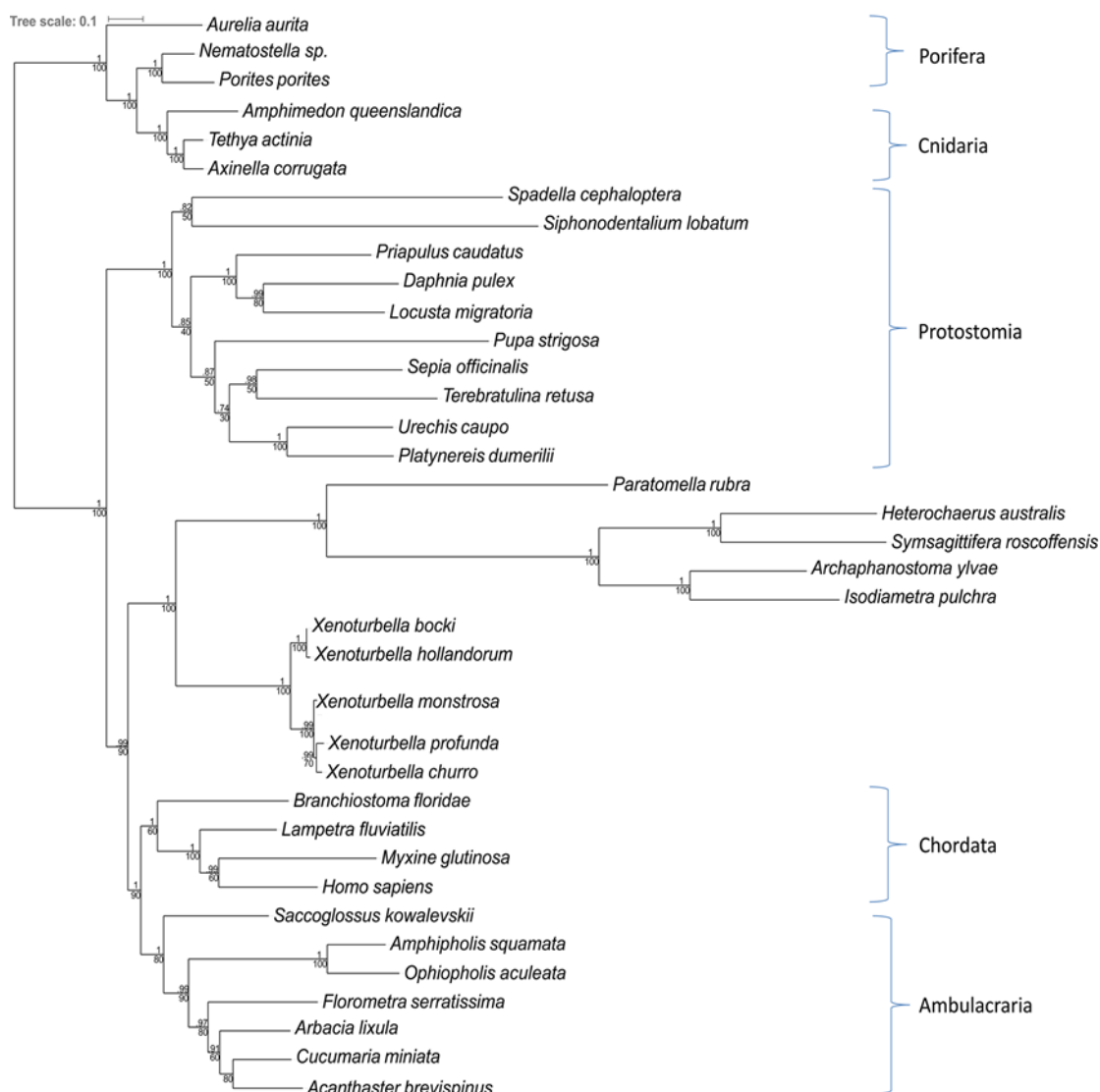
We choose mtZOA from a varies of models by AICc and BIC criteria, which is inferred from a dataset including Lophotrochozoans, non-chordate Deuterostomes and diploblastic metazoans [32]. As CAT under a Bayesian analyses has been proved to be better fitted than GTR model, regarding compositional heterogeneity [12], we also try CAT+GTR+Γ 12 (Figure 2a) and CAT+GTR+Γ4 (Figure 2b) based on 37 species matrix. In spite of they haven't challenged mtZOA's prediction, also positioning Xenacoelomorpha as a basal Deuterostomes, even cross-validation test proving CAT+GTR+Γ has better score than mtZOA, there are some problematic nodes in the CAT prediction which is weekly supported and cannot congruent with the prediction under ML analyses. Thus, we harbor the idea that mtZOA as an empirical transition probability matrix probably has replacement rates and stationary frequencies better fitted to evolutionary process of our matrix leading to a robust result.

Figure 2. Choosing CAT accounts for compositional heterogeneity. (a) Phylogeny of 37 species based on mitochondrial proteins reconstructed using CAT+GTR+Γ12 under a Bayesian analyse, places Xenacoelomorpha as a sister group of chordates with dissatisfactory internode support; (b) Phylogeny of 37 species based on mitochondrial proteins reconstructed inferred with CAT+GTR+Γ4 in a Bayesian analyse, problematically groups Xenacoelomorpha with chordates missing node support either. Cross-validation demonstrates that CAT+GTR+Γ is better fitted model comparing to mtZOA with significantly statistical support (ΔlnL=94 ± 14), but using mtZOA can

lead to a prediction strongly supported globally. As a matter of fact, missing internode support can prove nothing but Xenacoelomorpha can be seen as a clade of Deuterostomes (PP=0.97 and PP=1, respectively. Xenoturbeb is "another *Xenoturbella monstrosa*" illustrated in following section).
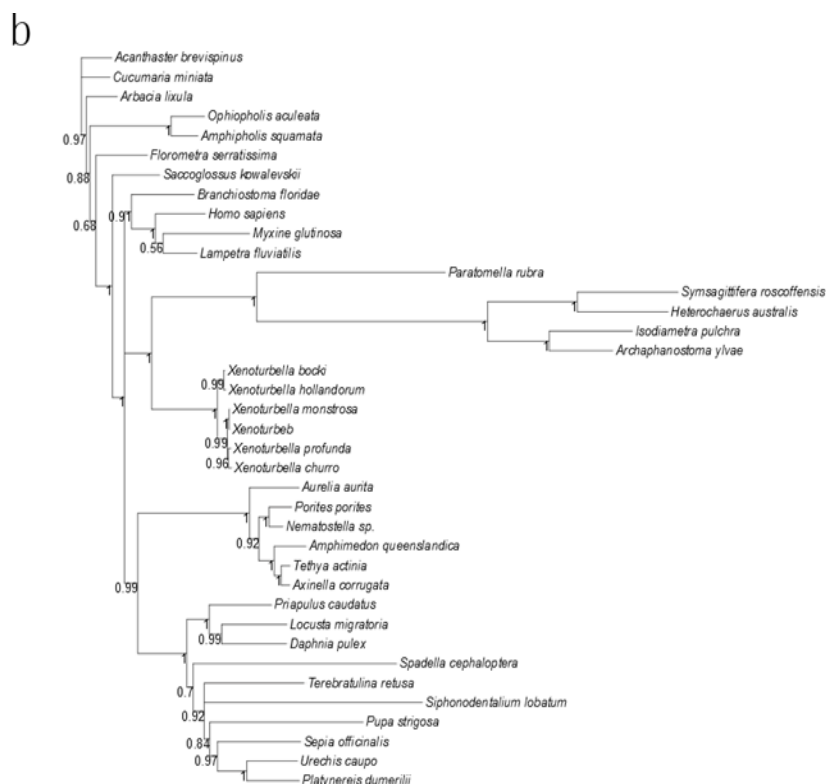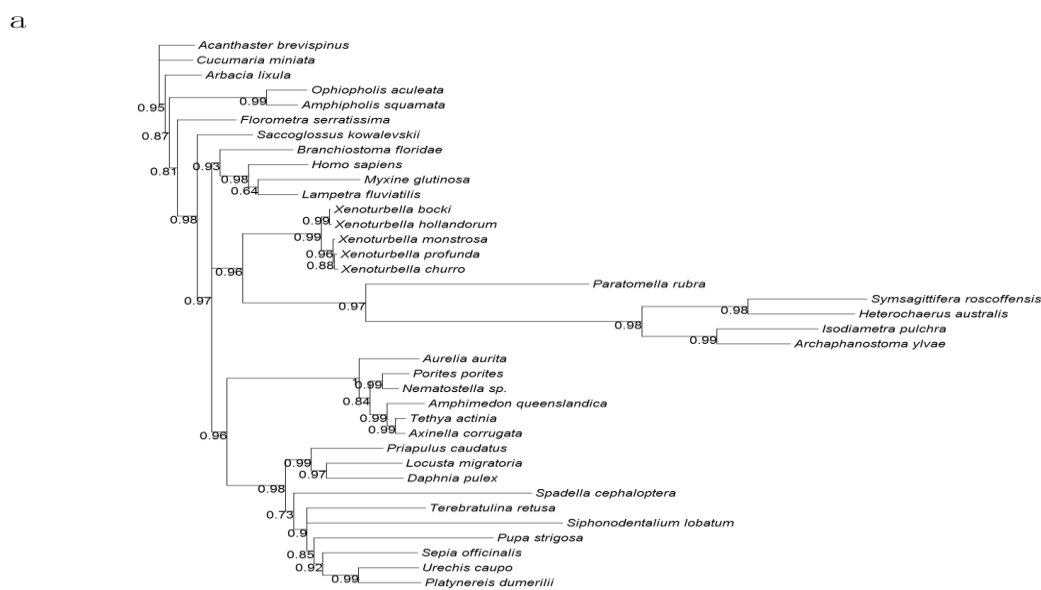

For fear of model over-fitting, we do ML and Bayesian based on varies alignments–building two matrix using MAFFT (strategy G-INS-1, scoring matrix for amino acid sequences choosing BLOSUM45 and the other with BLOSUM30), both refined by Gblocks in the same way detailing in online method. All analyses using mtZOA model based on 37 species matrix or 39 species matrix (additional *Schmidtea mediterranea* and *Dugesia japonica* to exaggerate LBA event) can carry out congruent result under both Bayesian and ML analyses placing Xenacoelomorpha as a basal branch of Deuterostome with reliable statistical support (Figure 1). In contrast, CAT model under Bayesian analyses

cannot lead to a consensus concerning about Xenacoelomorpha's phylogenetic position accrossing different datasets and not congruent with ML analyses. We then take mtZOA as a better fitted model to our 37 species matrix regardless of the possibility of over-fitting and suggest the 37 species alignments with parameter BLOSUM62 remaining the least non-phylogenetic signals inferring the most robust taxonomic placement with the strongest support to predicts Xenacoelomorpha as the basally branching at Deuterostome and illuminate's relationship between chordates, ambulacrarians, effectively (Figure 1).

Inadequate taxon selection can exaggerate the compositional bias and always lead to LBA events claiming for carefully dataset preparing. Considering the compositional heterogeneity among species, we double the *Xenoturbella monstrosa* sequence in step one, and doing Z-score test with randomly selected two training sets of Cross-validation used (each one amounts to 90% of whole alignments statistically) it's about the epistemology that compositional heterogeneity hardly inevitably affects tree topology and we wondering logical and random datasets choice shall bias the tree topology in what extent. Filtering species with less than Z-score 10 (30 species) and Z-score 5 (19 species), at least one acoel retained for obvious reason, building these two alignment matrix, modeling with the same strategy as we have already done to the 37 species matrix. All 37,30 and 19 species datasets inferring an consensus concerning that position Xenoacoelomorpha as a basal Deuterostome under both Bayesian and ML analyse (Figures 4a and 4e), showing robust topology based on varies selections of outgroup-6 outgroups to total 37 taxons, 5 to 30 and 3 to 19. We also wondering if excluded all other Xenoacoelomorphas, with single acoel we sequenced included in our 37 species matrix, can exaggerate LBA event but fail (Figure 3e). Our prediction of Xenoacoelomorpha geneology seems barely been biased by compositional heterogeneity with adequate choice of taxons.

Figure 3. Varies species selection to exaggerate the violation of models. (a) 30 species phylogeny based on mitochondrial proteins reconstructed under a Maximum likelihood analysis with mtZOA+I+PROTOGAMMA, Xenacoelomorpha and five acoel species group together (BP=69) and Xenacoelomorpha within Deuterostome (BP=75); (b) 19 species phylogeny based on mitochondrial proteins reconstructed under a Maximum likelihood analysis using mtZOA+I+PROTOGAMMA, Xenacoelomorpha and five acoel species group together (BP=73) and Xenacoelomorpha as a monophyly within Deuterostome (BP=75). Weak support can be illustrated by lacking phylogenetic signal in less taxons and the tree topology still consensus with our 37 species results; (c) 30 species phylogeny based on mitochondrial proteins reconstructed under a Bayesian analyse using mtZOA, also groups Xenacoelomorpha and five acoel species together (PP=0.98) as an early branch of Deuterostome (PP=0.99); (d) 19 species phylogeny based on mitochondrial proteins reconstructed using mtZOA under a Bayesian analyse, Xenacoelomorpha and five acoel species group together (PP=0.99), Xenacoelomorpha branching at early Deuterostome (PP=0.99); (e) Using MtZOA+I+PROTGAMMA under a Maximum likelihood analysis with our acoel included only, demonstrates no significant LBA event.
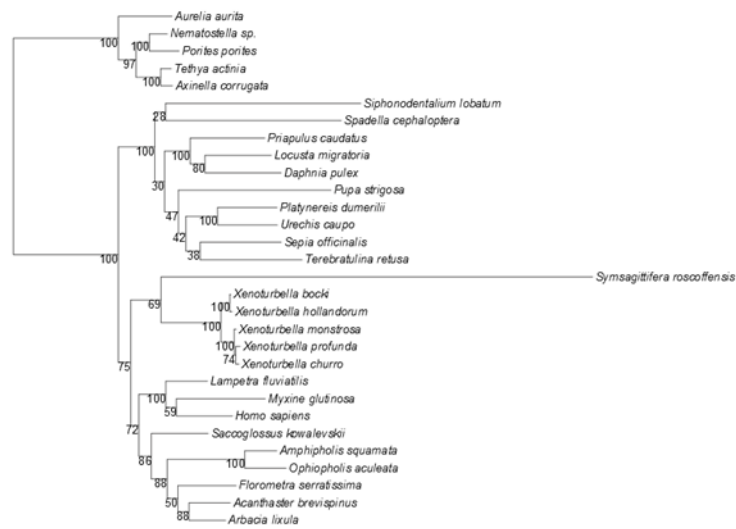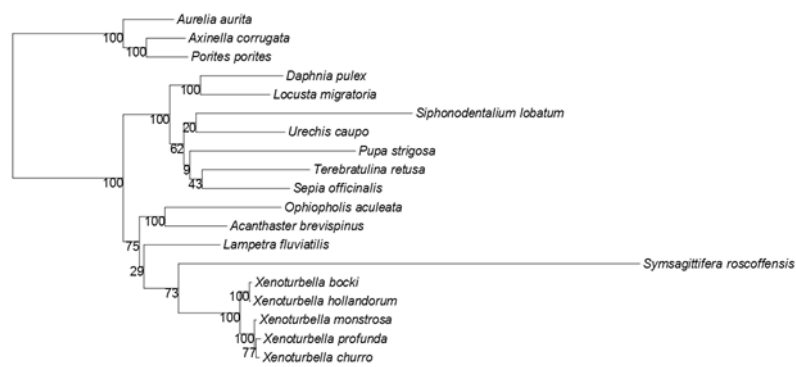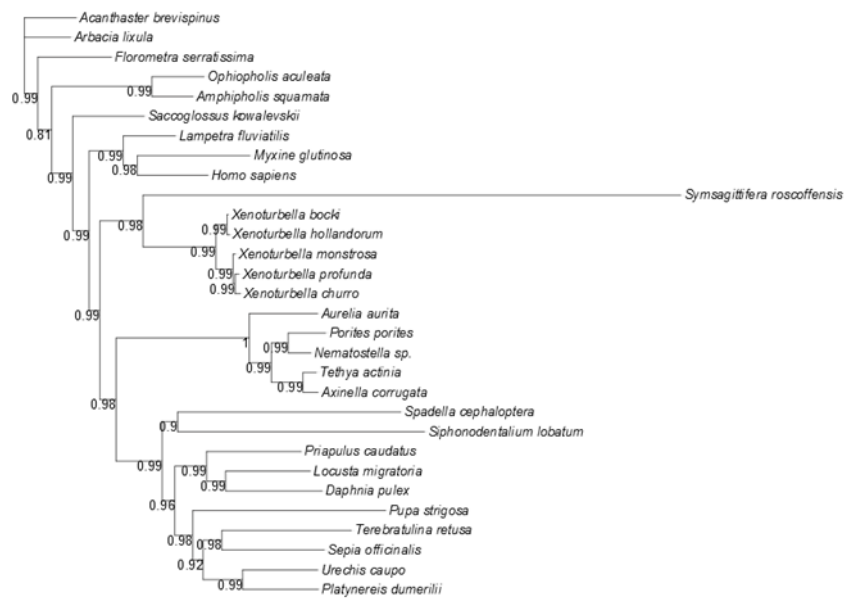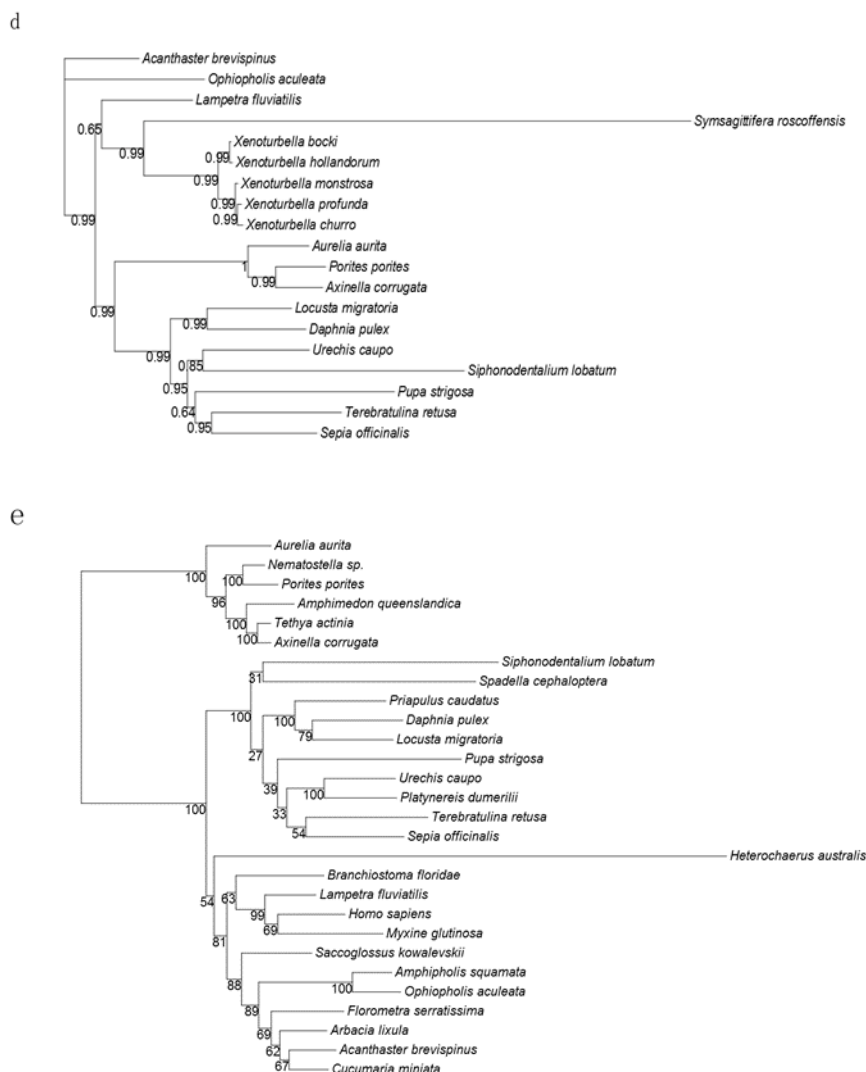
We also try "Dayhoff groups" to reduce the potential influence of compositional heterogeneity on phylogenomic analyses of our datasets [32]. The analyse under Bayesian using CAT+GTR+Γ50+Dayhoff6 (Figure 3a) based on 37 species matrix and the analyse under ML using PROTOGAMMA+I+Dayhoff inferred from 30 species (Figure 3d) concurrently force Acoelomorph to the base of bilateria with significantly less well supported (PP=0.71, BP=100, respectively). In spite of that the 37 species and 19 species analyzed under ML with PROTOGAMMA+I+Dayhoff position *Xenoturbella* as a basal Deuterostome (Figures 4b and 4c), overwhelmingly, different taxon selections cannot lead to a congruent prediction, with problematic nodes and much lower likelihood respectively, indicating that the Dayhoff groups cannot reflects the evolutionary pattern of the given taxonomic sample properly.

Figure 4. Reducing the potential influence of compositional heterogeneity on phylogenomic analyses of our datasets with Dayhoff groups. (a) 37 species phylogeny based on mitochondrial proteins reconstructed using CAT+GTR+Γ50+Dayhoff6 under a Bayesian analyse places Acoelomorph within outgroups (PP=0.71); (b) 37 species being analyzed using PROTOGAMMA+I+Dayhoff under a ML analyse, places Xenoturbella as a basal Deuterostome but with low support (BP=45) and monopoly of Xenoacoelomorpha being challenged (BP=56). (c) 19 species analyzed under ML using PROTOGAMMA+I+dahoff, position Xenoacoelomorpha as Deuterostome (BP=44), constitutes Acoelomorph and *Xenoturbella* as a monophyly (BP=52). (d) The analyse of 30 species under ML, using

PROTOGAMMA+I+Dayhoff, places *Xenoturbella* in Deuterostome (BP=44) but Acoelomorph as a basal bilateria (BP=100).
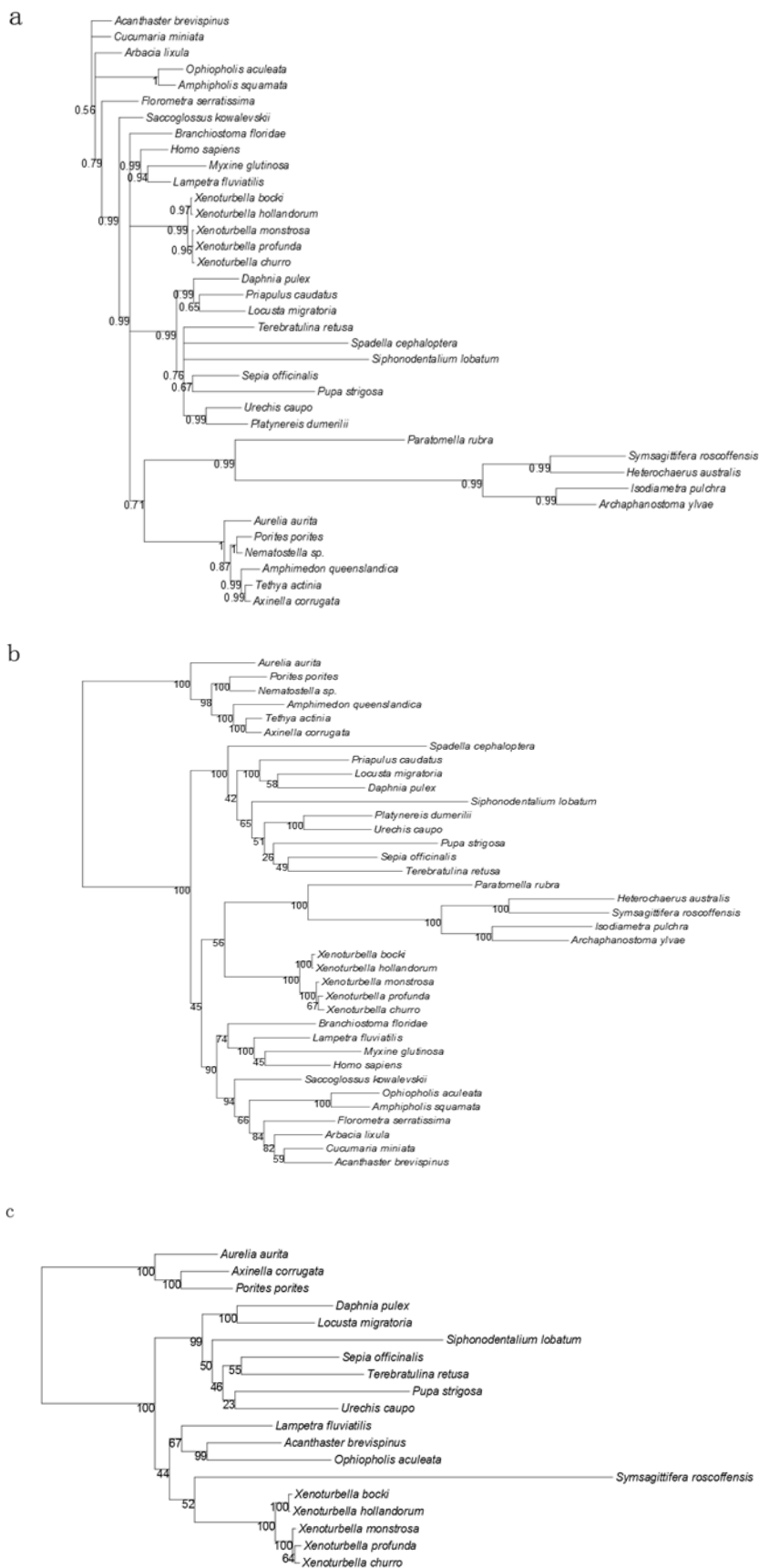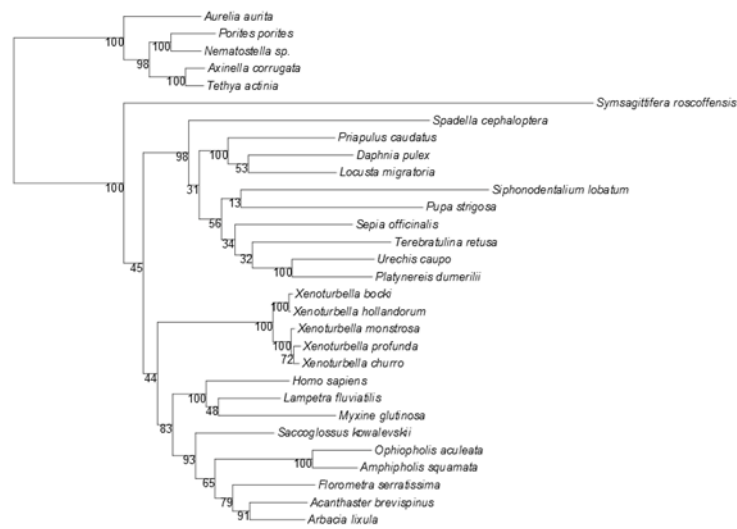
We also reanalyzed the three matrix (37, 30 and 19 taxon selections) after removing all of the outgroup taxa to rule out the possibility that the Xenoacoelomorpha branches at the basal of Deuterostome owing to Long Branch Attraction (LBA). All three analyses congruent to position Xenoacoelomorpha as a basal Deuterostomes (Figures 5a and 5c) BP=99,92,91) and the other metazoan clades being slightly infected. The decipher is obvious that once we simply cut of the outgroup clades in Figure 1, it will lead to the topology of Figure 5a, standing to reason the topology being pretty similar globally suggesting that LBA was not the cause of this prediction.

Figure 5. LBA was not the cause of this prediction. (a-c) All three taxons selections (37, 30, 19 species) without outgroups (remaining 31,25 and 16 species, respectively) position Xenoacoelomorpha as a basal Deuterostomes (BP=99,92,91) and consititution of Xenoacoelomorpha being unchallenged (BP=100,100,100), which can be comprehended as logical results after topological cutting.
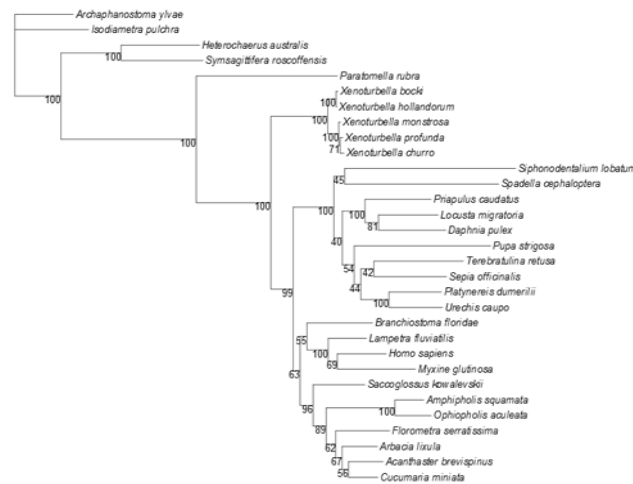
Above all, the use of mtZOA which reflects the evolutionary pattern of our 37 species matrix, results in a much higher likelihood than CAT and Dayhoff groups. Phylogenetic information content of the matrix after auto-alignmenting and trimming merely violate the model of protein evolution. Analyses based on varies kinds of taxon selections indicate a robust backbone tree of metazoan inferring *Xenoacoelomorpha* as an early branch of Deuterostome.

DISCUSSION

Majority enough

Homology or particularly orthology are such kind of congruent character which allow tree thinker to unite practical and conceptual advantages [13]. We want to figure out the species tree of life with PSC concept, underlying this concept is the epistemology gene as the major inherited information carrier bearing the weight of most kind of molecular evolutionary events and being selected by environment all way through geneology history. When we inertially start from "homolog gene" we shall not neglect the true noumenon is either nuclear acid or amino acid "whole sequence" itself. We acknowledged different genes go through distinct evolution events, when we talk about gene level in PSC, we are trying to represent the species geneology with "the majority of whole sequence of related species genome" but not just "a bag of genes" represent themselves suffered from varies kinds of evolutionary selection. If not, why should we use the interpretation that single gene tree inferred from cox1, getting such terrible statistical support because of lacking phylogenetic signal. If not, does it means that over ninety percent of Xenacoelomorpha's nuclear genes have nothing to do with the species evolution? It is commonly acknowledged

Yi Wang conceived the project, designed the experiment and wrote the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

REFERENCES

1. Velasco JD. Species, genes, and the tree of life. Br J Philos Sci. 2010; 61:599-619.

2. Quentin D, et al. Species concepts and phylogenetic theory: a debate. 2000;17-29.

3. Cracraft J, et al. Species concepts and speciation analysis. Curr Ornithol. 1983;159-187.

4. LaPorte J. Is there a single objective, evolutionary tree of life? J Philos. 2005; 102:357-374.

5. Bannert N. et al. Retroelements and the human genome: New perspectives on an old relation.

6. Proc Natl Acad Sci USA. 2001;101: 14572-14579.

7. de Jong P, et al. Initial sequencing and analysis of the human genome. Nature. 2001; 412:565

566.

8. Bapteste E, et al. On the need for integrative phylogenomics, and some steps toward its creation. Biol Philos. 2010; 25: 711-736.

9. Bannert N, et al. Retroelements and the human genome: New perspectives on an old relation. Proc Natl Acad Sci USA. 2004; 101:14572-14579.

10. Rouse GW, et al. New deep-sea species of *Xenoturbella* and the position of Xenacoelomorpha. Nature. 2016; 530:94.

11. Cannon JT, et al. Xenacoelomorpha is the sister group to nephrozoa. Nature. 2016; 530:89.

12. Philippe H, et al. Acoelomorph flatworms are deuterostomes related to *Xenoturbella*. Nature. 2011;470: 255.

13. Bapteste E, et al. Do orthologous gene phylogenies really support tree-thinking? BMC Evolutionary Biology. 2005; 5.

14. Radzvilavicius AL, et al. Mitigating mitochondrial genome erosion without recombination. Genetics. 2017; 207:1079-1088.

15. James JE, et al. The rate of adaptive evolution in animal mitochondria. Molecular Ecology. 2016; 25:67-78.

16. Kaniak-Golik A, et al. Mitochondria-nucleus network for genome stability. Free Radic Biol Med. 2015; 82:73-104.

17. Westblad E. *Xenoturbella bocki* n.g., n. sp. a peculiar, primitive turbellarian type. Arkiv för zoology. 1949; 1:3-29.

18. Hejnol A. Assessing the root of bilaterian animals with scalable phylogenomic methods. Proc Biol Sci. 2009; 276:4261-4270.

19. Srivastava M. Whole-body acoel regeneration is controlled by Wnt and Bmp-Admp signaling. Curr Biol. 2014; 24: 1107-1113.

20. Ehlers U, et al. Ultrastructure of the subepidermal musculature of *Xenoturbella bocki*, the adelphotaxon of the Bilateria. Zoomorphology. 1997;117:71-79.

21. Israelsson O. And molluscan embryogenesis. Nature. 1997; 390:32-32.

22. Bourlat SJ, et al. Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. Nature. 2006; 444:85-88.

23. Robertson HE. The mitochondrial genomes of the acoelomorph worms *Paratomella rubra, Isodiametra pulchra* and *Archaphanostoma ylvae*. Scientific Reports. 2017; 1847: 7.