The Opisthokonta and the Ecdysozoa may be clades: An update on Philip et al. (2005)

Marko Prous
Department of Zoology, Institute of Ecology and Earth Sciences, University of Tartu, Vanemuise 46, Tartu 51014, Estonia. E-mail: mprous@ut.ee

Introduction

Reconstructing the deep phylogeny of animals, and eukaryotes in general, is currently a very active and rapidly developing field.

Recently there has been a lively debate regarding some specific phylogenetic hypotheses:

Some genome-scale studies (based on genomes of few model taxa) have questioned the monophyly of Ecdysozoa (and actually also Protostomia) and favoured monophyly of Coelomata instead (fig 1B). On the other hand, Ecdysozoa is supported by studies including large number of taxa, but usually only few genes (fig 1A).

One of those genome-scale studies is of Philip et al. (2005), who also questioned monophyly of Opisthokonta (fungi-animal clade); their results were consistent with the phylogeny shown in fig 1B.

The results of Philip et al. (2005) have not been directly tested yet.

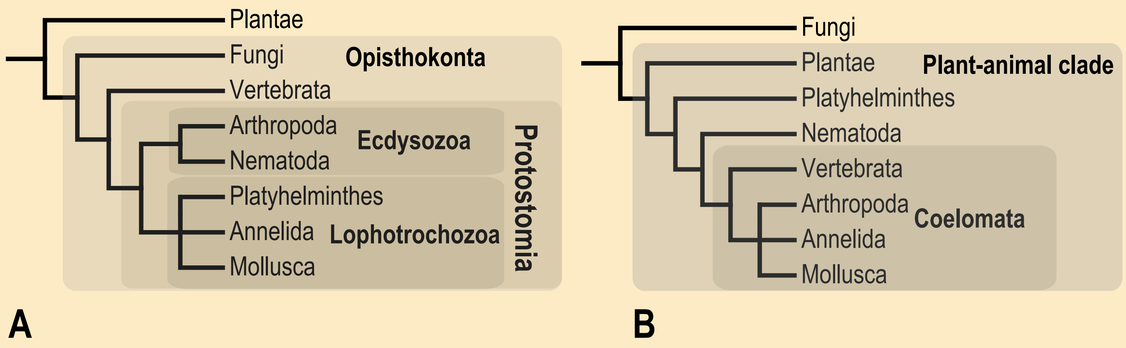

Fig. 1. Two alternative phylogenetic hypotheses tested here. A - topology supported by molecular phylogenetic studies including large number of taxa (see for example Embley & Martin, 2006; Telford, 2006). B - topology consistent with some genome-scale studies using limited number of taxa (e. g. Philip et al., 2005; but other genome-scale studies usually support Opisthokonta instead of plant-animal grouping).

The study of Philip et al. (2005)

Philip et al. (2005) used sequence data from 10 eukaryotic genomes (fig. 2), from which single-gene families were identified.

Phylogenetic analyses using those single-gene families recovered the topology as in fig. 2.

When short proteins (less than 300 aligned positions) and those unable to recover uncontroversial parts of the tree were removed, the resulting tree did not change (fig. 2). Parts of the tree considered uncontroversial were: the monophyly of the mammals (*Homo sapiens* and *Mus musculus*), vertebrates (mammals and *Takifugu*), arthropods (*Drosophila* and *Anopheles*), animals (arthropods, vertebrates, and the nematode *Caenorhabditis elegans*) and fungi (*Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*).

Only 14 single-gene families were found to be universally distributed (present in genomes of all 10 taxa).

From those 14 genes, 5 (table 1) were capable of recovering all the uncontroversial parts of the tree.

When those 5 supposedly reliable genes were concatenated and analysed, the resulting topology was the same as in previous analyses (fig. 2).

Table 1. Five single-copy protein-coding genes, which Philip et al. (2005) found in their data set to be universally distributed and capable of recovering all the uncontroversial clades.

Protein	NCBI accession of the protein sequence of *Homo sapiens*
Prefoldin subunit 2	NP_036526
Mitochondrial import inner membrane translocase (Tim22)	NP_037469
Small nuclear ribonuclear protein U6	NP_036454
MAK16	NP_115898
Autophagocytosis protein	NP_071933

The aim of the current study is to test the results of Philip et al. (2005) using greatly improved taxon sampling. This has been made possible thanks to numerous genome and EST projects completed since 2005. Do the results of Philip et al. (2005) hold up?

Material and methods

I searched the five genes shown in table 1 from different databases (NCBI, Joint Genome Institute, and Broad Institute) using BLAST.

I added total of 33 taxa to the dataset of Philip et al. (2005).

One OTU is chimerical: *Phytophthora*, composing of *Phytophthora capsici* (U6 protein) and *Phytophthora sojae* (all other genes).

After the alignment (using Muscle 3.6; Edgar, 2004) and removal of poorly aligned regions the final 5-gene data matrix contained 706 aligned positions. I aligned the sequences also with ClustalW (Thompson et al., 1994) without removing ambiguous regions, to compare the results obtained here more directly to those of Philip et al. (2005).

I reconstructed phylogenies using PhyML (Guindon & Gascuel, 2003) and PhyML-mixture (Le et al., 2008).

Results

I found that the autophagocytosis protein of *Caenorhabditis elegans* in the dataset of Philip et al. (2005) is actually mitochondrial ribosomal protein S25. Here, I replaced it with the correct protein.

Maximum likelihood analysis (using LG matrix, Le & Gascuel, 2008) of 10 taxa recovered plant-animal and arthropod-vertebrate (Coelomata) clade (fig. 2).

Phylogenetic analysis including additional 33 taxa recovered Opisthokonta, Ecdysozoa, Lophotrochozoa, and Protostomia (fig. 3).

Ecdysozoa and Lophotrochozoa were not recovered using JTT, WAG and Dayhoff amino-acid replacement matrices (but log-likelihood in those cases was lower than using LG matrix). However, Protostomia was always monophyletic (approximate likelihood branch support (aLRT) varied from 91-93%), as well as Opisthokonta (aLRT 92-98%).

Analyses using two (EX2, UL2) and three matrix (EX3, EHO, UL3) models of amino-acid replacement (implemented in PhyML-mixture) all recovered Opisthokonta (aLRT support 98-99%), Ecdysozoa (aLRT 83-96%), Lophotrochozoa (aLRT 96-99%), and Protostomia (aLRT 73-92%).

Coelomata was recovered with weak bootstrap support (41%) only if arthropods and flatworms (Platyhelminthes) were excluded from the analyses. Interestingly, Protostomia (nematodes + molluscs + annelids) was recovered instead (BS 51%), when *Trichinella* was also excluded. More slowly evolving nematodes (e. g. *Xiphinema*) are needed to further test the phylogenetic position of nematodes.

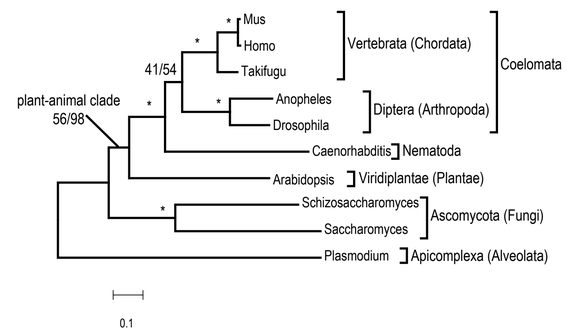

Fig. 2. Maximum likelihood tree of 10 taxa obtained with PhyML (LG + G8 + I model). Bootstrap proportions (100 replicates) were computed based on 2 datasets: poorly aligned regions excluded (left) or included (right);* - equal or more than 95%. Root is placed arbitrarily on the branch leading to *Plasmodium*.

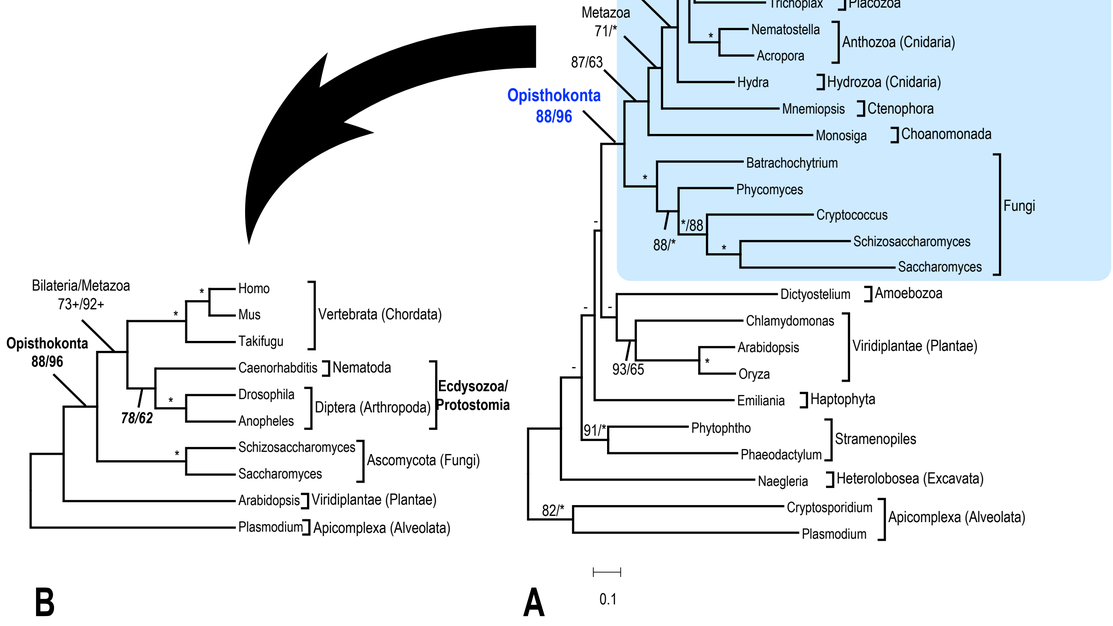

Fig. 3. Phylogeny reconstructed using 43 taxa. A - Maximum likelihood tree obtained with PhyML (LG + G8 + I model). Bootstrap proportions (100 replicates) were computed based on 2 datasets: poorly aligned regions excluded (left) or included (right); * - equal or more than 95%, hyphen (-) - less than 50%. B - The same phylogeny as in A, but 33 taxa missing in fig. 2 have been removed. Bootstrap proportions for Ecdysozoa/Protostomia clade are summed up from replicates supporting Ecdysozoa but not Protostomia and replicates supporting Protostomia but not Ecdysozoa. Root is placed in both figures arbitrarily on the branch leading to Apicomplexa.

Discussion

Some taxa seem to be misplaced (fig. 3A): *Branchiostoma* and *Hydra*, causing non-monophyly of Chordata and Cnidaria, respectively. It could be due to poor character sampling (only 706 aligned amino acids), but addition of certain taxa could also be helpful (for example *Xenoturbella*, cyclostomes, Scyphozoa, Cubozoa, and more slowly evolving Hydrozoa).

While bootstrap support for the monophyly of Ecdysozoa, Lophotrochozoa and Protostomia is quite low (fig. 3), only 19% or 18% of the replicates (in those two analysis mentioned in the text under fig. 3) showed arthropods to be closer to vertebrates than to nematodes (in several of those replicates bilateria was not monophyletic, which clearly suggests that the basal placement of nematodes could be an artefact due to highly divergent nature of their sequences).

While bootstrap support for the monophyly Opisthokonta was not maximal (fig. 3), only 0% or 1% of the replicates (in those two analysis mentioned in the text under fig. 3) showed plants to be closer to animals than to fungi. Based on improved taxon sampling, the plant-animal clade seems highly unlikely.

Conclusions

The results of Philip et al. (2005) do not hold up after improved taxon sampling.

Large taxon sampling and more realistic models of sequence evolution are extremely important for reconstructing deep phylogeny of eukaryotes (including animals).

The 5 genes analysed here and by Philip et al. (2005) are no different from the genes used in other studies (for example widely used rRNA genes): as long as the there is reasonable taxon sampling and likelihood based methods are used, molecular data tends to support Opisthokonta, Ecdysozoa, Lophotrochozoa, and Protostomia (as in fig. 1A, not 1B).

However, the 5 genes are too short (even if concatenated) to draw definite conclusions, and should be combined with other genes (for example those remaining 9 genes identified by Philip et al. (2005) to be universally distributed and in single copy).

References

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32: 1792-1797.

Embley, T. M. & Martin, W. 2006. Eukaryotic evolution, changes and challenges. Nature 440: 623-630.

Guindon, S. & Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52: 696-704.

Le, S. Q. & Gascuel, O. 2008. An Improved General Amino Acid Replacement Matrix. Molecular Biology and Evolution 25: 1307-1320.

Le, S. Q., Lartillot, N., & Gascuel, O. 2008. Phylogenetic Mixture Models for Proteins. Philosophical Transactions of the Royal Society B 363: 3965-3976.

Philip, G. K., Creevey, C. J. & McInerney, J. O. 2005. The Opisthokonta and the Ecdysozoa May Not Be Clades: Stronger Support for the Grouping of Plant and Animal than for Animal and Fungi and Stronger Support for the Coelomata than Ecdysozoa. Molecular Biology and Evolution 22: 1175-1184.

Telford, M. J. 2006. Animal phylogeny. Current Biology 16: R981-R985.

Thompson, J. D., Higgins, D. G., & Gibson, T. J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through