

# **Lack of resolution in the animal phylogeny: closely spaced cladogeneses or undetected systematic errors?**

## **Supplementary material**

Denis Baurain, Henner Brinkmann, Hervé Philippe  
Université de Montréal

### **1. Data assembly**

#### ***1.1 Construction of individual gene alignments***

Each of the protein alignments used in previous studies (Philippe et al. 2004; Philippe, Lartillot, and Brinkmann 2005; Delsuc et al. 2006) was updated with the newly available sequences downloaded from the Trace Archive (<http://www.ncbi.nlm.nih.gov/Traces/>) and the EST Database (<http://www.ncbi.nlm.nih.gov/dbEST/>) of GenBank at the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>) using new features of the program ED from the MUST package (Philippe 1993). Ambiguously aligned regions were automatically deleted with GBLOCKS (Castresana 2000).

#### ***1.2. Gene selection and concatenation***

Slow-evolving orthologous sequences for each individual protein were selected and concatenated into a supermatrix using SCAFOS (Roure, Rodriguez-Ezpeleta, and Philippe 2006). Only proteins represented for at least 37 of a total of 57 species were considered. The resulting alignment consisted of 133 genes (Table S1) with 57 species and 31,089 unambiguously aligned amino acid positions. To reduce the computational burden, only 12,942 positions determined in at least 75% of the 57 species were conserved. This dataset is available upon request from HP.

#### ***1.3. Chimerical Operational Taxonomic Units (OTUs)***

To increase the amount of data, we created chimerical sequences by merging sequences from closely related taxa using SCAFOS. Chimeric OTUs have been named after the inclusive species that was most represented (shown in bold in Table S2).

#### ***1.4. Minimization of missing data***

The construction of a supermatrix containing a reasonable number of taxa implies a certain amount of missing data. In our concatenated dataset, the number of amino acid residues available for the most incomplete species is nevertheless already large with 5,299 positions for *Fasciola hepatica*. The complete dataset comprises 12,942 unambiguously aligned positions with a mean of 11,426 (88%) amino acid residues per taxa (Table S3). Under these conditions, the impact of missing data on the phylogenetic inference can be considered negligible (Wiens 2003; Philippe et al. 2004; Wiens 2005).

### 1.5. Generation of four variants of the dataset

In order to explore the effect of taxon sampling on the phylogenetic resolution, we generated four variants of our dataset: (A) a 23-species alignment including the nematode *Caenorhabditis* (Ce-23); (B) a version of this alignment expanded to 56-species (Ce-56); (C) a variation of the 23-species alignment in which *Xiphinema* had been substituted for *Caenorhabditis* (Xi-23); and (D) the identically substituted version of the 56-species alignment (Xi-56).

## 2. Phylogenetic analyses

### 2.1. Tree reconstruction

For such large datasets, inferring the ML tree and especially estimating its reliability *via* bootstrap analyses requires a tremendous amount of computational time. In addition, each of the four variants of our dataset was analyzed with three complementary approaches.

First, trees were inferred under a concatenated WAG+ $\Gamma$ 4 model using three different heuristics: TREEFINDER version of June 2005 (Jobb, von Haeseler, and Strimmer 2004), PHYML version 2.4 (Guindon and Gascuel 2003), and PHYML version 2.4.3 with SPR moves (subtree pruning and regrafting (Hordijk and Gascuel 2005)). ML bootstrap support values were obtained after 100 pseudo-replicates generated with SEQBOOT [from the PHYLIP package version 3.6 (Felsenstein 2001)].

Second, we performed an exhaustive analysis with constraints since a fully exhaustive search is not possible (23 species lead to  $>10^{25}$  trees; 56 species to  $>10^{86}$  trees). We used the results from the heuristic approaches to constrain the relationships within the six protostomian groups (annelids, arthropods, mollusks, nematodes, platyhelminthes, tardigrades) and an outgroup (composed of chordates, cnidarians, poriferans, choanoflagellates, and fungi). All 945 possible topologies were generated using PROTML version 2.3b3 (Adachi and Hasegawa 1996). The likelihood of each topology and the corresponding branch lengths were computed under a concatenated WAG+ $\Gamma$ 4 model using TREE-PUZZLE (Schmidt et al. 2002). The likelihood of each position for each tree was then computed using CODEML (from the PAML package version 3.13 (Yang 1997)). Site-wise likelihood values were used to compute the RELL bootstrap values of each topology based on 1,000 replicates. The RELL bootstrap support values for each node were finally computed using CONSENSE [from the PHYLIP package version 3.6 (Felsenstein 2001)] as described in Supplementary Material online of Philippe, Lartillot, and Brinkmann (2005).

Third, we used the CAT+ $\Gamma$  model (Lartillot and Philippe 2004) in a MCMC framework as implemented in PHYLOBAYES version 1.1c (<http://www.lirmm.fr/mab/>). First, 100 pseudo-replicates were generated using SEQBOOT. Then, the 100 independent chains were run for a total number of cycles equal to three times the number of cycles required for the convergence of likelihood values and other parameters (3,000 cycles for Ce-23 and 1,000 cycles for Ce-56). Trees were collected after the initial burnin period (1000 cycles and 300 cycles, respectively) and fed to CONSENSE to compute the bootstrap support values for each node. Two more independent chains were run on Ce-23 and Ce-56 to infer the topology and the branch lengths (fig. S3).

## 2.2. Performance of heuristic approaches

Using PHYML nodes as references, bootstrap support values of the 146 (20+53+20+53) nodes present in the four trees inferred from the four variants of our dataset were compared on a pairwise basis for each heuristic, as well as for the exhaustive analysis (fig. S1). When 146 nodes are considered, all three heuristics achieve a similar support, with correlation coefficients varying from 0.859 to 0.935. However, when the heuristics are compared to the exhaustive analysis for the 16 (4x4) internal nodes within protostomes, PHYML performs markedly worse ( $r = 0.407$ ), whereas TREEFINDER ( $r = 0.812$ ) and PHYML-SPR ( $r = 0.896$ ) perform similarly and quite well. For a detailed comparison of the bootstrap support values obtained for the four nodes within protostomes as well as five alternative groupings often retrieved in replicates, see Table S4.

For three out of four variants (Ce-23, Xi-23 and Xi-56), all heuristics found the ML tree identified in the exhaustive analysis. In contrast, all heuristics failed to infer the ML tree from the Ce-56 variant. Instead, they retrieved the second-best tree (fig. S2) identified in the exhaustive analysis, in which platyhelminthes are sisters to nematodes within Ecdysozoa. The artefactual second-best tree differs from the ML tree (fig. 1B) by a tiny  $\Delta\log L$  of 22.9 ( $-439,000.53$  vs  $-438,977.63$ ). This shows that heuristics perform generally well, except when numerous species with heterogeneous evolutionary rates are considered, hence justifying the exhaustive analysis with constraints presented here.

## Literature Cited

- Adachi, J., and M. Hasegawa. 1996. MOLPHY version 2.3: programs for molecular phylogenetics based on maximum likelihood. *Comput. Sci. Monogr.* **28**:1-150.
- Castresana, J. 2000. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Mol Biol Evol* **17**:540-552.
- Delsuc, F., H. Brinkmann, D. Chourrout, and H. Philippe. 2006. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* **439**:965-968.
- Felsenstein, J. 2001. PHYLIP (Phylogene Inference Package). Distributed by the author, Department of Genetics, University of Washington, Seattle.
- Guindon, S., and O. Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**:696-704.
- Hordijk, W., and O. Gascuel. 2005. Improving the efficiency of SPR moves in phylogenetic tree search methods based on maximum likelihood. *Bioinformatics* **21**:4338-4347.
- Jobb, G., A. von Haeseler, and K. Strimmer. 2004. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol. Biol.* **4**:18.
- Lartillot, N., and H. Philippe. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* **21**:1095-1109.
- Philippe, H. 1993. MUST, a computer package of Management Utilities for Sequences and Trees. *Nucleic Acids Res.* **21**:5264-5272.
- Philippe, H., N. Lartillot, and H. Brinkmann. 2005. Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. *Mol Biol Evol* **22**:1246-1253.

- Philippe, H., E. A. Snell, E. Baptiste, P. Lopez, P. W. Holland, and D. Casane. 2004. Phylogenomics of eukaryotes: impact of missing data on large alignments. *Mol Biol Evol* **21**:1740-1752.
- Roure, B., N. Rodriguez-Ezpeleta, and H. Philippe. 2006. SCaFoS: a tool for selection, concatenation and fusion of sequences for phylogenomics. *BMC Evol. Biol.* **in press**.
- Schmidt, H. A., K. Strimmer, M. Vingron, and A. von Haeseler. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**:502-504.
- Wiens, J. J. 2003. Missing data, incomplete taxa, and phylogenetic accuracy. *Syst Biol* **52**:528-538.
- Wiens, J. J. 2005. Can incomplete taxa rescue phylogenetic analyses from long-branch attraction? *Syst Biol* **54**:731-742.
- Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* **13**:555-556.

**Table S1. List of the 133 gene names and number of conserved amino-acid positions for each protein alignment**

| file name | complete gene name   | # of AA positions |
|-----------|--|-------------------|
| ar21      | Actin-related protein 2/3 complex subunit 3                    | 130               |
| arc20     | Actin-related protein 2/3 complex subunit 4                    | 92                |
| arp23     | Actin-related protein 2/3 complex subunit 1b                   | 190               |
| cct-A     | T complex protein 1 alpha subunit                              | 511               |
| cct-B     | T complex protein 1 beta subunit                               | 482               |
| cct-D     | T complex protein 1 delta subunit                              | 492               |
| cct-E     | T complex protein 1 epsilon subunit                            | 516               |
| cct-G     | T complex protein 1 gamma subunit                              | 456               |
| cct-N     | T complex protein 1 eta subunit                                | 478               |
| cct-T     | T complex protein 1 theta subunit                              | 392               |
| cct-Z     | T complex protein 1 ? subunit                                  | 484               |
| cpn60-mt  | Heat shock protein HSP 60kDa mitochondrial                     | 498               |
| crfg      | Nucleolar GTP binding protein 1                                | 382               |
| ef1-EF1   | Elongation factor EF1a   | 435               |
| ef2-EF2   | Elongation factor EF2  | 776               |
| ef2-U5    | Elongation factor Tu family U5 snRNP specific protein          | 676               |
| eif5a     | Eukaryotic initiation factor 5a                                | 119               |
| fibri     | Fibrillarin  | 218               |
| grc5      | 60S ribosomal protein L10 QM protein                           | 209               |
| hsp70-E   | Heat shock 70kDa protein form E                                | 551               |
| hsp70-mt  | Heat shock 70kDa protein, mitochondrial form                   | 513               |
| hsp70-SSE | Heat shock 70kDa protein subfamily SSE1                        | 299               |
| if1a      | Eukaryotic translation initiation factor 1a                    | 52                |
| if2b      | Eukaryotic translation initiation factor 2b                    | 173               |
| if2g      | Eukaryotic translation initiation factor 2g                    | 435               |
| if6       | Eukaryotic translation initiation factor 6                     | 234               |
| l12e-A    | 40S ribosomal Protein S12                                      | 115               |
| l12e-C    | High mobility group like nuclear protein 2 NHP2-like protein 1 | 116               |
| l12e-D    | 60S ribosomal Protein L7a                                      | 218               |
| mcm-B     | Minichromosome family maintenance protein 2                    | 483               |
| metk      | S-adenosyl-methionine synthetase                               | 343               |
| nsf1-G    | 26S proteasome AAA-ATPase regulatory subunit 8                 | 358               |
| nsf1-I    | Putative 26S proteasome ATPase regulatory subunit 7            | 383               |
| nsf1-J    | 26S proteasome AAA-ATPase regulatory subunit 6                 | 372               |
| nsf1-K    | 26S proteasome AAA-ATPase regulatory subunit 6a                | 391               |
| nsf1-L    | 26S proteasome AAA-ATPase regulatory subunit 6b                | 369               |
| nsf1-M    | 26S proteasome AAA-ATPase regulatory subunit 4                 | 411               |
| nsf2-A    | Transitional endoplasmic reticulum ATPase TER ATPase           | 711               |
| orf2      | Putative 28 kDa protein  | 160               |
| pace5     | Schwachman-Bodian-Diamond like-protein                         | 143               |
| psma-A    | 20S proteasome beta subunit macropain zeta chain               | 205               |
| psma-B    | 20S proteasome alpha 1a chain                                  | 199               |
| psma-C    | 20S proteasome alpha 1b chain                                  | 203               |
| psma-D    | 20S proteasome alpha 2 chain                                   | 216               |
| psma-E    | 20S proteasome alpha 1c chain                                  | 195               |
| psma-F    | 20S proteasome alpha 3 chain                                   | 191               |

|         |   |     |
|---------|---|-----|
| psma-G  | 20S proteasome alpha 6 chain                      | 222 |
| psmb-H  | 20S proteasome beta 2 chain                       | 168 |
| psmb-I  | 20S proteasome beta 3 chain                       | 197 |
| psmb-J  | 20S proteasome beta 1 chain                       | 181 |
| psmb-K  | 20S proteasome beta 7 chain                       | 197 |
| psmb-L  | 20S proteasome beta 6 chain                       | 166 |
| psmb-M  | 20S proteasome beta 5 chain                       | 182 |
| psmb-N  | 20S proteasome beta 4 chain                       | 127 |
| rad23   | UV excision repair protein RAD23                  | 131 |
| rf1     | Eukaryotic peptide chain release factor subunit 1 | 376 |
| rfa2-A  | 60S acidic ribosomal protein P2                   | 74  |
| rfa2-B  | 60S acidic ribosomal protein P1                   | 66  |
| rpl1    | 60S ribosomal Protein 1                           | 202 |
| rpl2    | 60S ribosomal Protein 2                           | 248 |
| rpl3    | 60S ribosomal Protein 3                           | 375 |
| rpl4B   | 60S ribosomal Protein 4b                          | 282 |
| rpl5    | 60S ribosomal Protein 5                           | 245 |
| rpl6    | 60S ribosomal Protein 6                           | 112 |
| rpl7-A  | 60S ribosomal Protein 7a                          | 199 |
| rpl9    | 60S ribosomal Protein 9                           | 154 |
| rpl11b  | 60S ribosomal Protein 11b                         | 166 |
| rpl12b  | 60S ribosomal Protein 12b                         | 163 |
| rpl13   | 60S ribosomal Protein 13                          | 135 |
| rpl14a  | 60S ribosomal Protein 14a                         | 111 |
| rpl15a  | 60S ribosomal Protein 15a                         | 204 |
| rpl16b  | 60S ribosomal Protein 16b                         | 162 |
| rpl17   | 60S ribosomal Protein 17                          | 164 |
| rpl18   | 60S ribosomal Protein 18                          | 180 |
| rpl19a  | 60S ribosomal Protein 19a                         | 180 |
| rpl20   | 60S ribosomal Protein 20                          | 148 |
| rpl21   | 60S ribosomal Protein 21                          | 149 |
| rpl22   | 60S ribosomal Protein 22                          | 83  |
| rpl23a  | 60S ribosomal Protein 23a                         | 133 |
| rpl24-A | 60S ribosomal Protein 24a                         | 112 |
| rpl24-B | 60S ribosomal Protein 24b                         | 121 |
| rpl25   | 60S ribosomal Protein 25                          | 123 |
| rpl26   | 60S ribosomal Protein 26                          | 125 |
| rpl27   | 60S ribosomal Protein 27                          | 132 |
| rpl30   | 60S ribosomal Protein 30                          | 100 |
| rpl31   | 60S ribosomal Protein 31                          | 100 |
| rpl32   | 60S ribosomal Protein 32                          | 121 |
| rpl33a  | 60S ribosomal Protein 33a                         | 94  |
| rpl34   | 60S ribosomal Protein 34                          | 108 |
| rpl35   | 60S ribosomal Protein 35                          | 116 |
| rpl37a  | 60S ribosomal Protein 37a                         | 74  |
| rpl38   | 60S ribosomal Protein 38                          | 64  |
| rpl39   | 60S ribosomal Protein 39                          | 51  |
| rpl42   | 60S ribosomal Protein 42                          | 102 |
| rpl43b  | 60S ribosomal Protein 43b                         | 91  |
| rpp0    | 60S acidic ribosomal protein P0 L10E              | 285 |

|        |  |     |
|--------|--|-----|
| rps1   | 40S ribosomal Protein 1                                  | 240 |
| rps2   | 40S ribosomal Protein 2                                  | 208 |
| rps3   | 40S ribosomal Protein 3                                  | 211 |
| rps4   | 40S ribosomal Protein 4                                  | 257 |
| rps5   | 40S ribosomal Protein 5                                  | 187 |
| rps6   | 40S ribosomal Protein 6                                  | 206 |
| rps8   | 40S ribosomal Protein 8                                  | 198 |
| rps10  | 40S ribosomal Protein 10                                 | 87  |
| rps11  | 40S ribosomal Protein 11                                 | 140 |
| rps13a | 40S ribosomal Protein 13a                                | 150 |
| rps14  | 40S ribosomal Protein 14                                 | 135 |
| rps15  | 40S ribosomal Protein 15                                 | 138 |
| rps16  | 40S ribosomal Protein 16                                 | 136 |
| rps17  | 40S ribosomal Protein 17                                 | 102 |
| rps18  | 40S ribosomal Protein 18                                 | 152 |
| rps19  | 40S ribosomal Protein 19                                 | 129 |
| rps20  | 40S ribosomal Protein 20                                 | 100 |
| rps22a | 40S ribosomal Protein 22a                                | 130 |
| rps23  | 40S ribosomal Protein 23                                 | 142 |
| rps25  | 40S ribosomal Protein 25                                 | 90  |
| rps26  | 40S ribosomal Protein 26                                 | 98  |
| rps27  | 40S ribosomal Protein 27                                 | 82  |
| rps28a | 40S ribosomal Protein 28a                                | 59  |
| rps29  | 40S ribosomal Protein 29                                 | 54  |
| sap40  | 40S ribosomal protein SA 40kDa laminin receptor 1        | 198 |
| srp54  | Signal recognition particle 54 kDa protein               | 430 |
| srs    | Seryl tRNA synthetase                                    | 373 |
| suca   | Succinyl-CoA ligase alpha chain mitochondrial precursor? | 287 |
| tftid  | TATA box binding protein related factor 2                | 175 |
| topo1  | DNA topoisomerase I, mitochondrial precursor             | 373 |
| trs    | Threonyl-tRNA synthetase                                 | 425 |
| vata   | Vacuolar ATP synthase catalytic subunit A                | 539 |
| vatb   | Vacuolar ATP synthase catalytic subunit B                | 459 |
| vate   | Vacuolar ATP synthase catalytic subunit C                | 228 |
| vate   | Vacuolar ATP synthase catalytic subunit E                | 194 |
| w09c   | TGF beta inducible nuclear protein                       | 248 |
| wrs    | Tryptophanyl-tRNA synthetase                             | 356 |

---

**Table S2. List of chimerical Operational Taxonomic Units (OTUs)**

*Glomus intraradices* : **Glomus intraradices**, *Glomus versiforme*  
*Hydra magnipapillata* : **Hydra magnipapillata**, *Hydra vulgaris*  
*Hydractinia echinata* : **Hydractinia echinata**, *Podocoryne carnea*  
*Acropora millepora* : **Acropora millepora**, *Acropora palmata*, *Montastraea faveolata*  
*Molgula tectiformis* : **Molgula tectiformis**, *Halocynthia roretzi*  
*Eptatretus burgeri* : **Eptatretus burgeri**, *Myxine glutinosa*, *Eptatretus stoutii*  
*Xenopus tropicalis* : **Xenopus tropicalis**, *Xenopus laevis*  
*Homo sapiens* : **Homo sapiens**, *Mus musculus*, *Bos taurus*, *Canis familiaris*, *Rattus norvegicus*  
*Crassostrea virginica* : **Crassostrea virginica**, *Crassostrea gigas*  
*Biomphalaria glabrata* : **Biomphalaria glabrata**, *Aplysia californica*, *Lymnaea stagnalis*  
*Argopecten irradians* : **Argopecten irradians**, *Pecten maximus*  
*Helobdella robusta* : **Helobdella robusta**, *Helobdella stagnalis*, *Haementeria depressa*  
*Lumbricus rubellus* : **Lumbricus rubellus**, *Eisenia andrei*, *Eisenia fetida*  
*Platynereis dumerilii* : **Platynereis dumerilii**, *Nereis virens*  
*Dugesia ryukyuensis* : **Dugesia ryukyuensis**, *Dugesia japonica*  
*Echinococcus granulosus* : **Echinococcus granulosus**, *Echinococcus multilocularis*  
*Boophilus microplus* : **Boophilus microplus**, *Rhipicephalus appendiculatus*  
*Acanthoscurria gomesiana* : **Acanthoscurria gomesiana**, *Dysdera erythrina*  
*Ixodes scapularis* : **Ixodes scapularis**, *Ixodes pacificus*  
*Daphnia pulex* : **Daphnia pulex**, *Daphnia magna*  
*Litopenaeus vannamei* : **Litopenaeus vannamei**, *Litopenaeus setiferus*, *Penaeus monodon*, *Marsupenaeus japonicus*  
*Homarus americanus* : **Homarus americanus**, *Pacifastacus leniusculus*  
*Locusta migratoria* : **Locusta migratoria**, *Diploptera punctata*, *Gryllus bimaculatus*  
*Hypsibius dujardini* : **Hypsibius dujardini**, *Macrobiotus islandicus*, *Richtersius coronifer*, *Milnesium tardigradum*

**Table S3. Summary of the frequency of missing data per taxa**

| OTU                              | # of AA present | % of AA missing |
|----------------------------------|-----------------|-----------------|
| <i>Acanthoscurria gomesiana</i>  | 9770            | 24.5            |
| <i>Acropora millepora</i>        | 7998            | 38.2            |
| <i>Apis mellifera</i>            | 12812           | 1.0             |
| <i>Argopecten irradians</i>      | 11431           | 11.7            |
| <i>Biomphalaria glabrata</i>     | 11096           | 14.3            |
| <i>Blastocladiella emersonii</i> | 12415           | 4.1             |
| <i>Bombyx mori</i>               | 12935           | 0.1             |
| <i>Boophilus microplus</i>       | 11542           | 10.8            |
| <i>Caenorhabditis elegans</i>    | 12936           | 0.0             |
| <i>Capitella</i> sp.             | 11718           | 9.5             |
| <i>Ciona intestinalis</i>        | 12879           | 0.5             |
| <i>Ciona savignyi</i>            | 12878           | 0.5             |
| <i>Crassostrea virginica</i>     | 11893           | 8.1             |
| <i>Cryptococcus neoformans</i>   | 12911           | 0.2             |
| <i>Danio rerio</i>               | 12795           | 1.1             |
| <i>Daphnia pulex</i>             | 12901           | 0.3             |
| <i>Dugesia ryukyuensis</i>       | 11778           | 9.0             |
| <i>Echinococcus granulosus</i>   | 11242           | 13.1            |
| <i>Eptatretus burgeri</i>        | 12502           | 3.4             |
| <i>Euprymna scolopes</i>         | 10346           | 20.1            |



|                                   |              |             |
|-----------------------------------|--------------|-------------|
| <i>Fasciola hepatica</i>          | 5299         | 59.1        |
| <i>Gallus gallus</i>              | 11371        | 12.1        |
| <i>Glomus intraradices</i>        | 6210         | 52.0        |
| <i>Helobdella robusta</i>         | 11631        | 10.1        |
| <i>Homarus americanus</i>         | 10339        | 20.1        |
| <i>Homo sapiens</i>               | 12942        | 0.0         |
| <i>Hydractinia echinata</i>       | 11868        | 8.3         |
| <i>Hydra magnipapillata</i>       | 12938        | 0.0         |
| <i>Hypsibius dujardini</i>        | 10910        | 15.7        |
| <i>Ixodes scapularis</i>          | 10476        | 19.1        |
| <i>Litopenaeus vannamei</i>       | 12471        | 3.6         |
| <i>Locusta migratoria</i>         | 12336        | 4.7         |
| <i>Lottia gigantea</i>            | 8195         | 36.7        |
| <i>Lumbricus rubellus</i>         | 12422        | 4.0         |
| <i>Molgula tectiformis</i>        | 12715        | 1.8         |
| <i>Monosiga ovata</i>             | 12441        | 3.9         |
| <i>Monosiga brevicollis</i>       | 11081        | 14.4        |
| <i>Nasonia vitripennis</i>        | 10284        | 20.5        |
| <i>Nematostella vectensis</i>     | 12030        | 7.0         |
| <i>Neocallimastix patriciarum</i> | 9841         | 24.0        |
| <i>Petromyzon marinus</i>         | 12035        | 7.0         |
| <i>Platynereis dumerilii</i>      | 6567         | 49.3        |
| <i>Proterospongia sp.</i>         | 7293         | 43.6        |
| <i>Reniera sp.</i>                | 11671        | 9.8         |
| <i>Rhizopus oryzae</i>            | 12888        | 0.4         |
| <i>Saccharomyces cerevisiae</i>   | 12910        | 0.2         |
| <i>Schistosoma mansoni</i>        | 12770        | 1.3         |
| <i>Schistosoma japonicum</i>      | 12364        | 4.5         |
| <i>Schizosaccharomyces pombe</i>  | 12906        | 0.3         |
| <i>Schmidtea mediterranea</i>     | 12781        | 1.2         |
| <i>Spodoptera frugiperda</i>      | 10467        | 19.1        |
| <i>Suberites domuncula</i>        | 9773         | 24.5        |
| <i>Tribolium castaneum</i>        | 12934        | 0.1         |
| <i>Ustilago maydis</i>            | 12908        | 0.3         |
| <i>Xenopus tropicalis</i>         | 12942        | 0.0         |
| <i>Xiphinema index</i>            | 11010        | 14.9        |
| <i>Yarrowia lipolytica</i>        | 12504        | 3.4         |
| <b>mean</b>                       | <b>11426</b> | <b>11.7</b> |

**Table S4. Comparison of the RELL bootstraps of the exhaustive analysis with the values obtained for the three heuristics and the Bayesian inference for nine protostomian nodes frequently found in the trees of the replicates in the four datasets**

A. 23 species including *Caenorhabditis* (Ce-23)

| node | description                                       | exh. analysis | PHYML | PHYML-SPR | TREEFINDER | PHYLOBAYES |
|------|---|---------------|-------|-----------|------------|------------|
| A    | tardigrades+nematodes                             | 45            | 28    | 29        | 34         | 74         |
| B    | tardigrades+nematodes+arthropods (Ecdysozoa)      | 29            | 2     | 11        | 1          | 72         |
| C    | annelids+mollusks                                 | 93            | 94    | 97        | 97         | 32         |
| D    | annelids+mollusks+platyhelminths (Lophotrochozoa) | 29            | 3     | 11        | 1          | 84         |
| E    | tardigrades+arthropods (Panarthropoda)            | 47            | 47    | 59        | 52         | 26         |
| F    | platyhelminths+nematodes                          | 56            | 71    | 71        | 66         | 8          |
| G    | platyhelminths+nematodes+annelids+mollusks        | 37            | 22    | 39        | 34         | 25         |
| H    | platyhelminths+nematodes+tardigrades              | 24            | 48    | 30        | 46         | 0          |
| I    | platyhelminths+nematodes+tardigrades+arthropods   | 28            | 56    | 49        | 56         | 1          |

B. 56 species including *Caenorhabditis* (Ce-56)

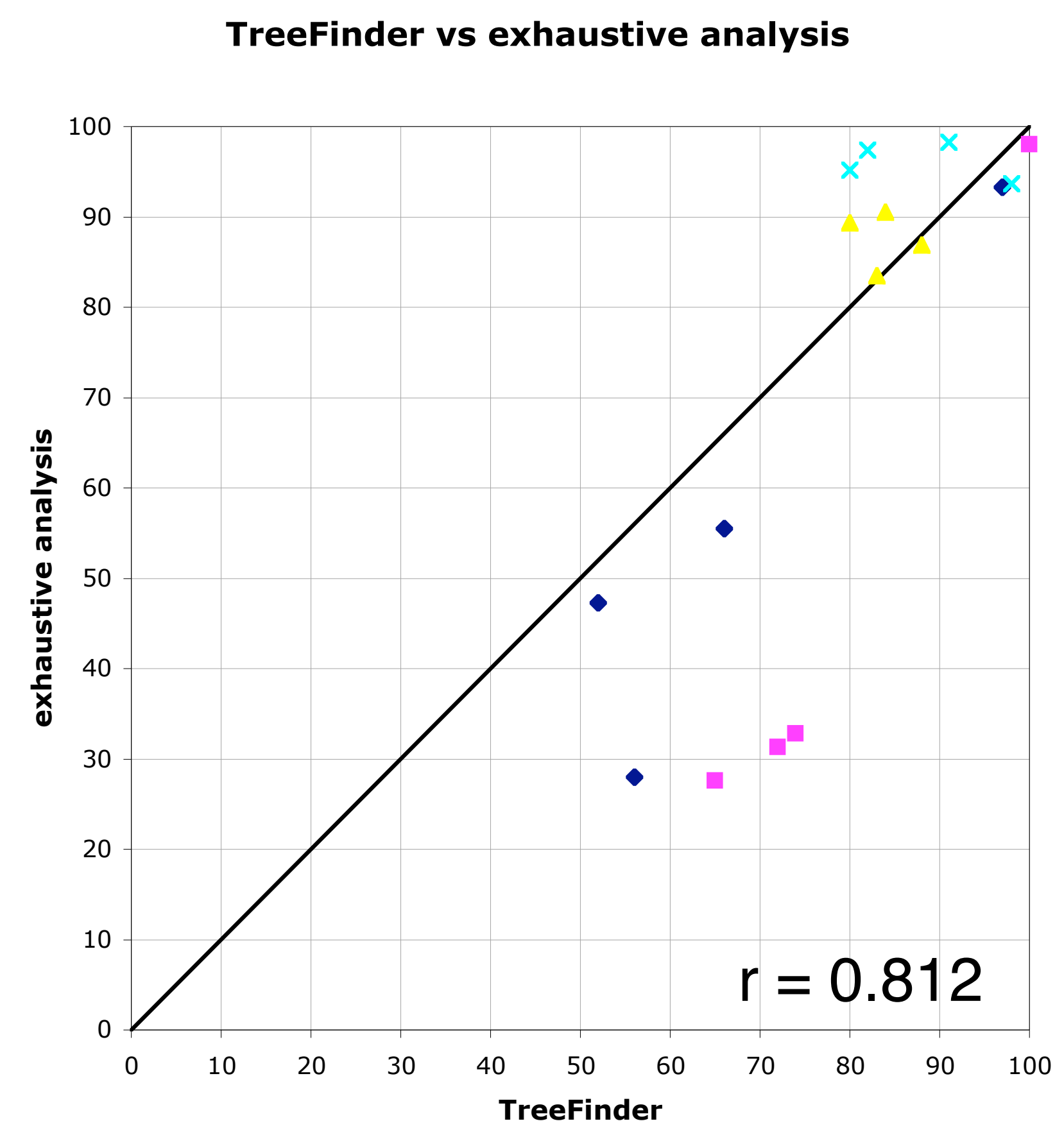
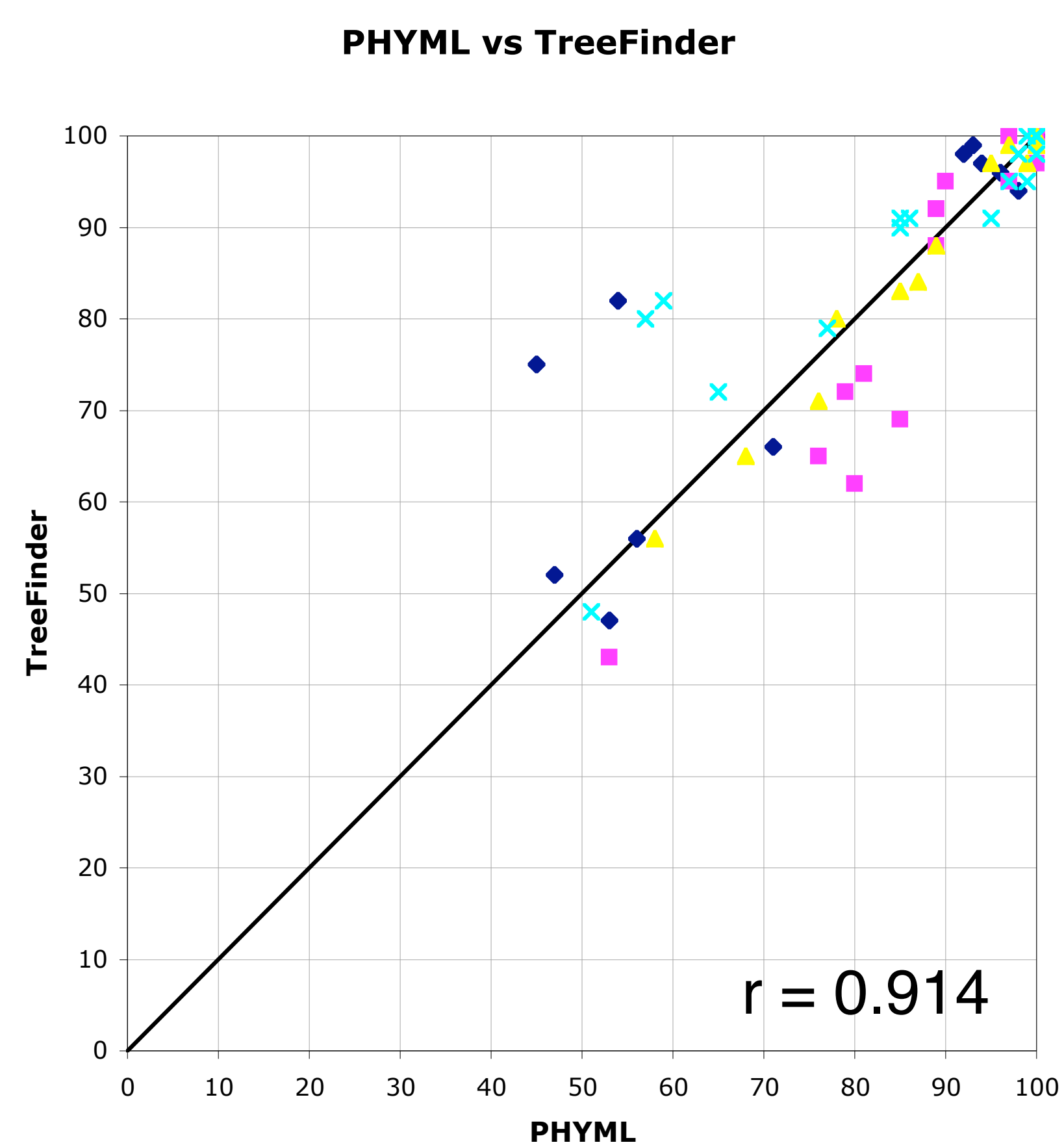
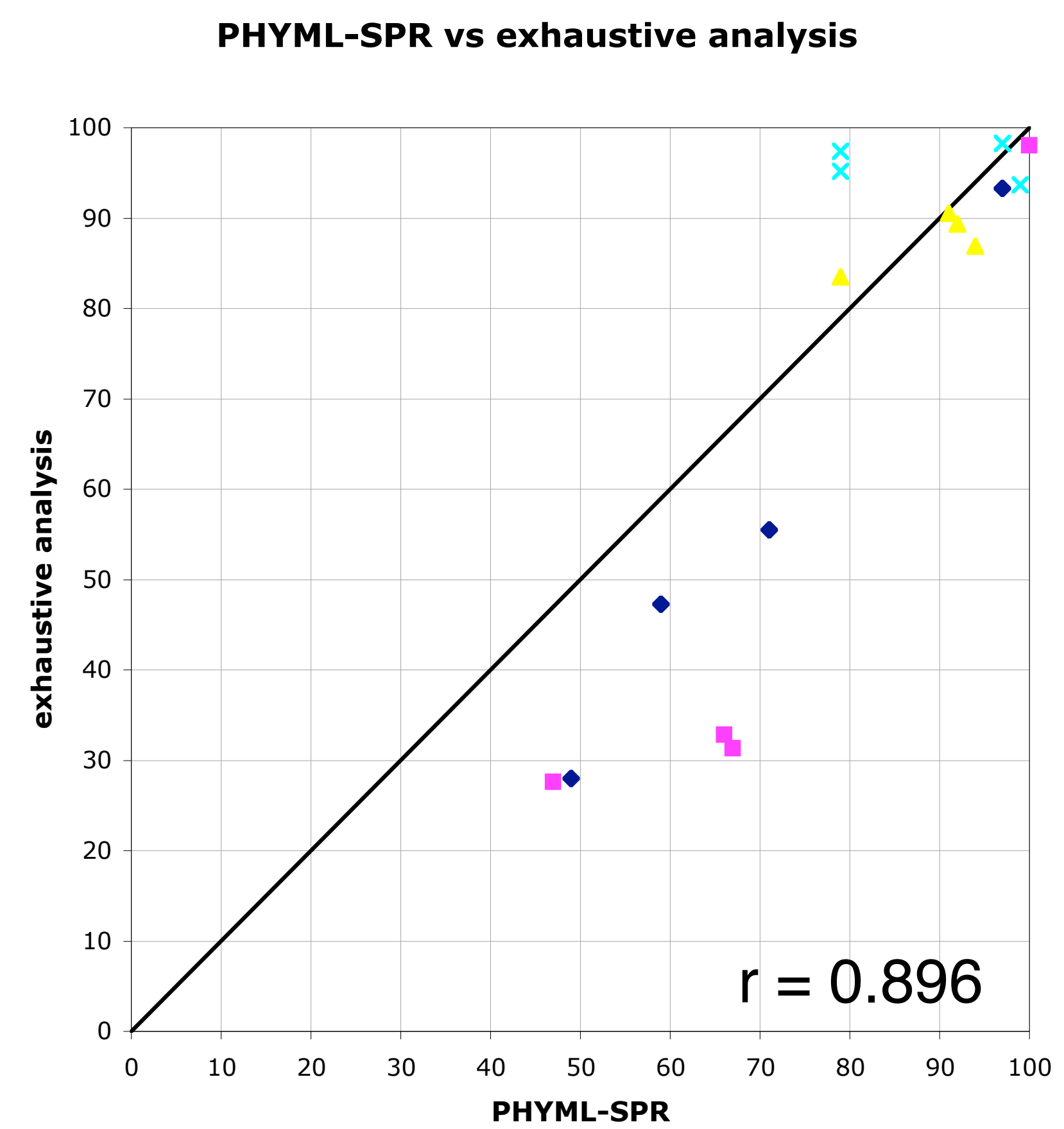
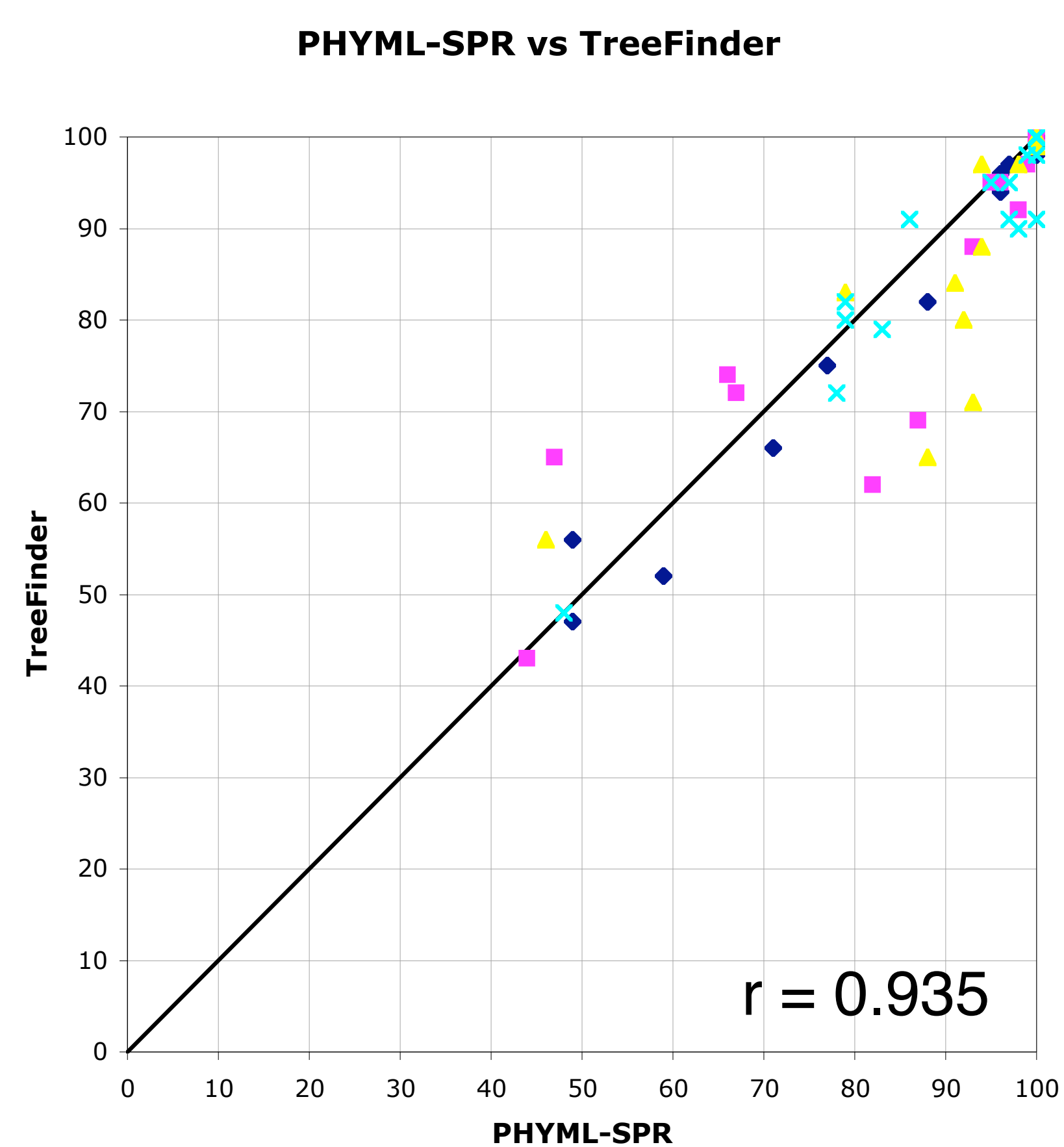
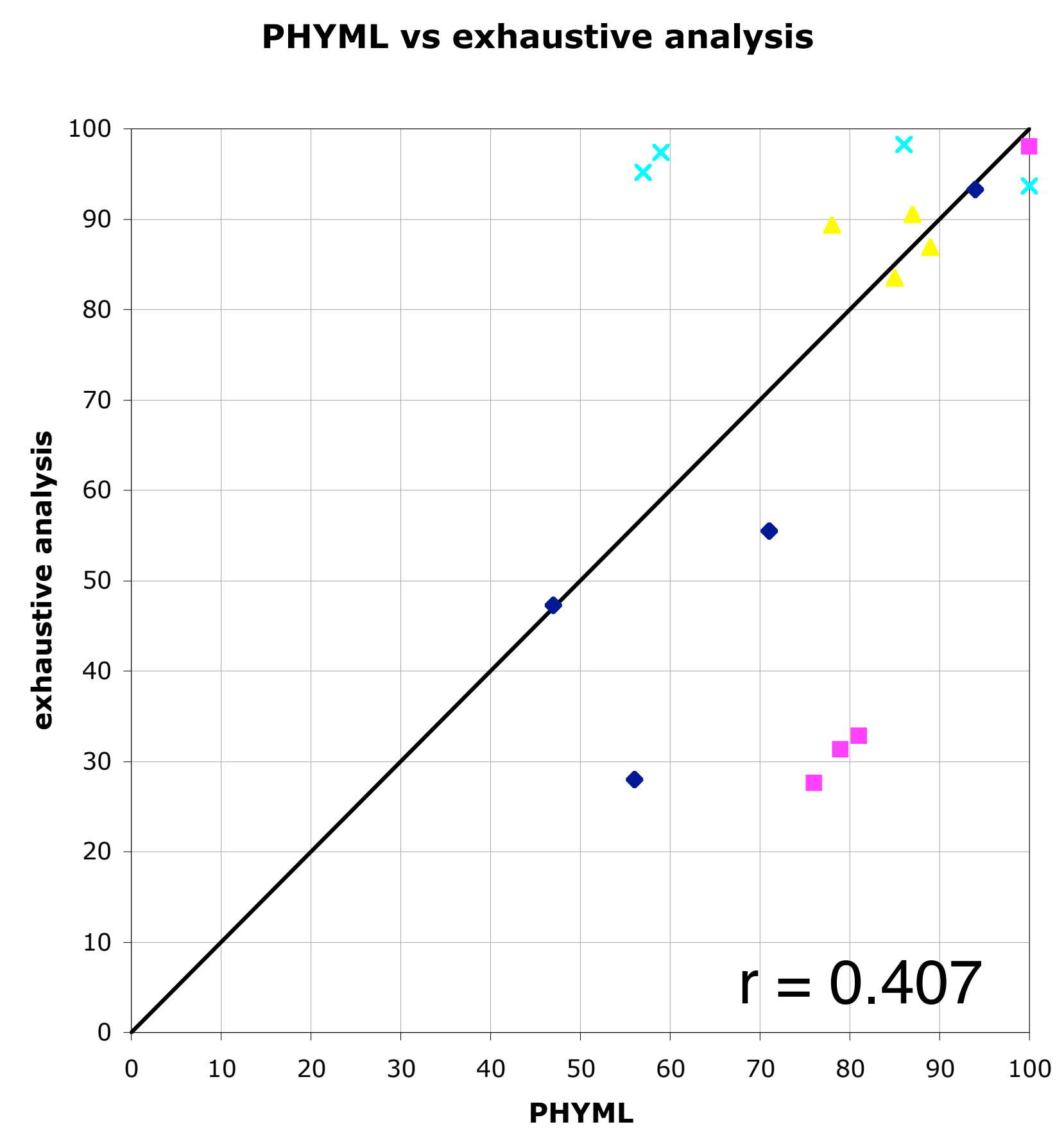
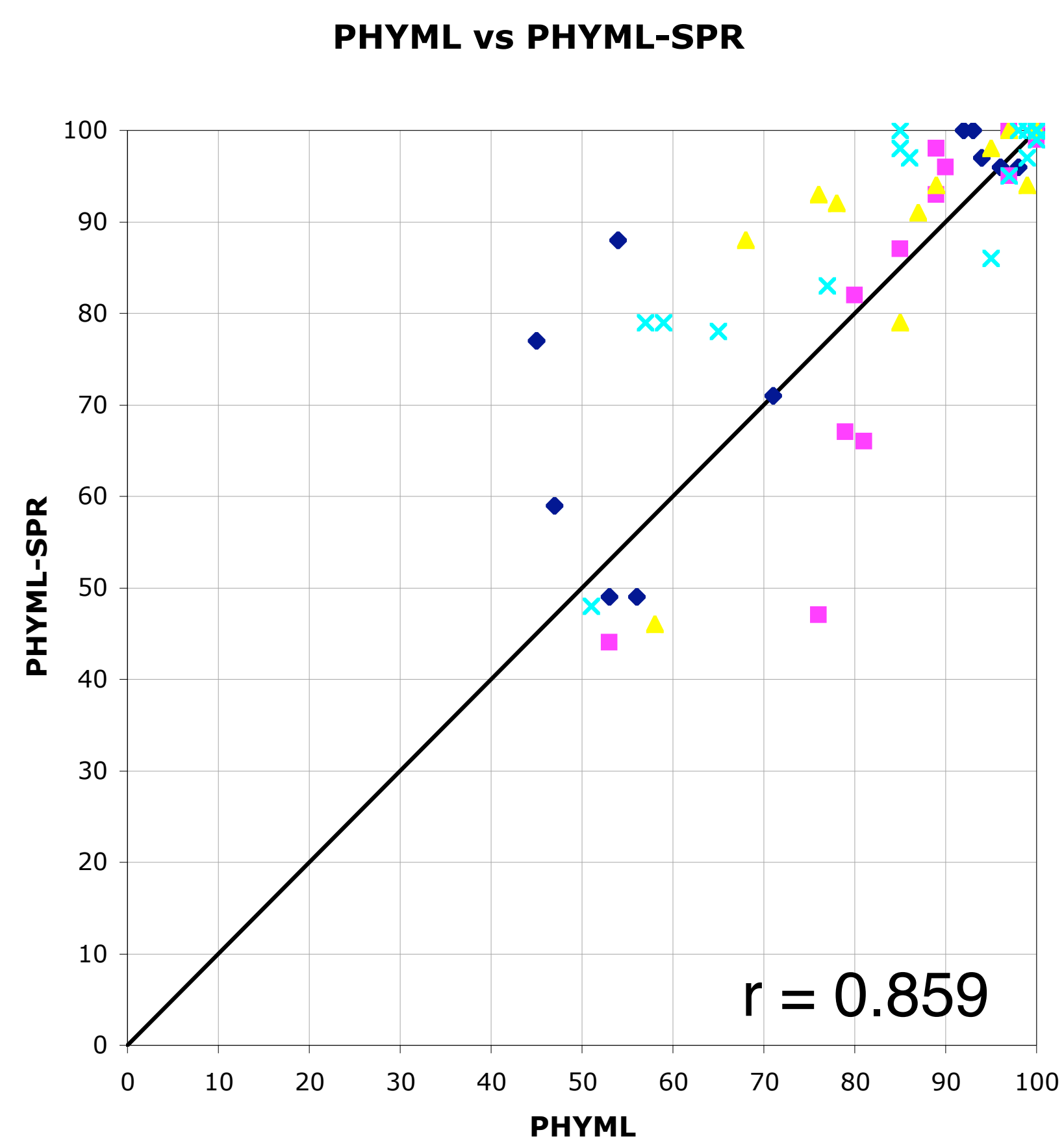
| node | description                                       | exh. analysis | PHYML | PHYML-SPR | TREEFINDER | PHYLOBAYES |
|------|---|---------------|-------|-----------|------------|------------|
| A    | tardigrades+nematodes                             | 72            | 23    | 53        | 35         | 100        |
| B    | tardigrades+nematodes+arthropods (Ecdysozoa)      | 66            | 3     | 29        | 16         | 100        |
| C    | annelids+mollusks                                 | 98            | 100   | 100       | 100        | 65         |
| D    | annelids+mollusks+platyhelminths (Lophotrochozoa) | 66            | 3     | 29        | 16         | 100        |
| E    | tardigrades+arthropods (Panarthropoda)            | 1             | 13    | 5         | 10         | 0          |
| F    | platyhelminths+nematodes                          | 28            | 76    | 47        | 65         | 0          |
| G    | platyhelminths+nematodes+annelids+mollusks        | 1             | 3     | 2         | 3          | 0          |
| H    | platyhelminths+nematodes+tardigrades              | 33            | 81    | 66        | 74         | 0          |
| I    | platyhelminths+nematodes+tardigrades+arthropods   | 31            | 79    | 67        | 72         | 0          |

C. 23 species including *Xiphinema* (Xi-23)

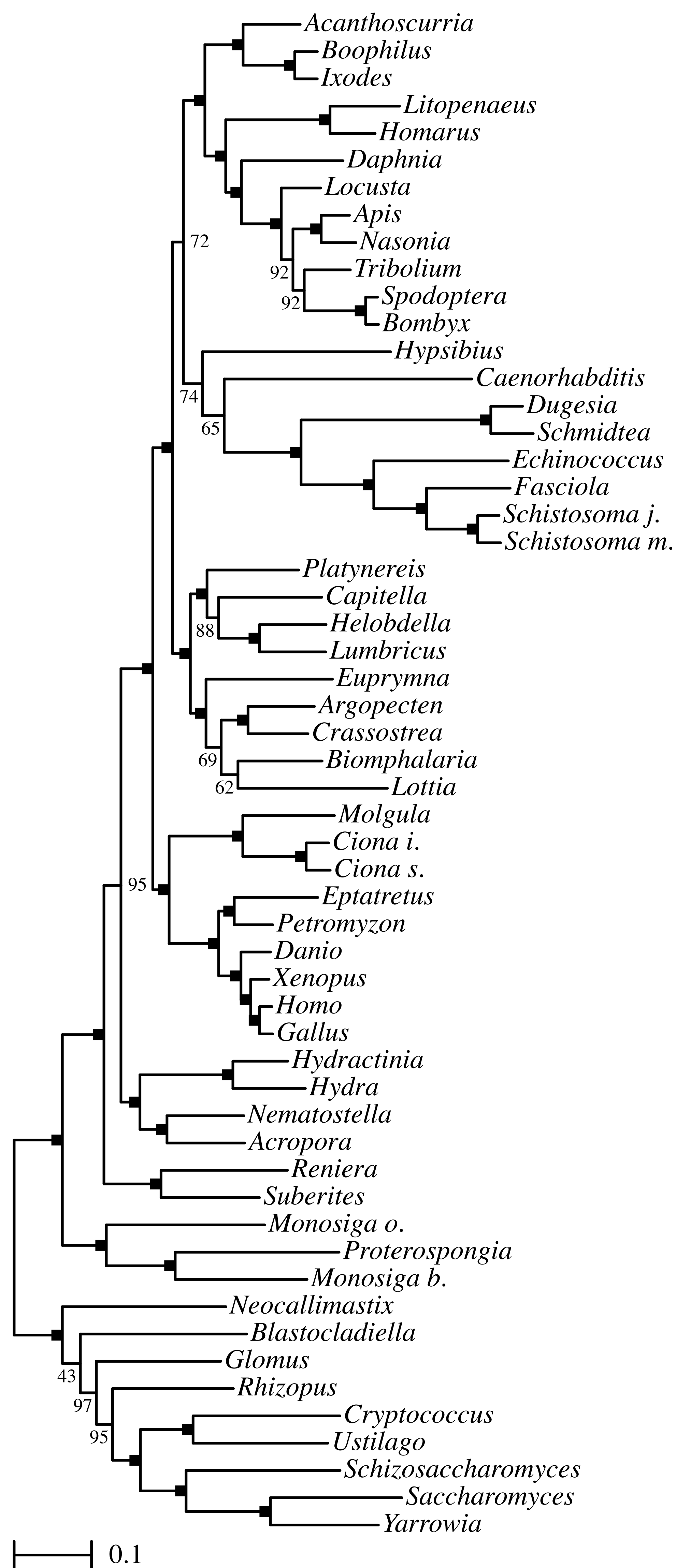
| node | description                                       | exh. analysis | PHYML | PHYML-SPR | TREEFINDER | PHYLOBAYES |
|------|---|---------------|-------|-----------|------------|------------|
| A    | tardigrades+nematodes                             | 84            | 85    | 79        | 83         | —          |
| B    | tardigrades+nematodes+arthropods (Ecdysozoa)      | 91            | 87    | 91        | 84         | —          |
| C    | annelids+mollusks                                 | 87            | 89    | 94        | 88         | —          |
| D    | annelids+mollusks+platyhelminths (Lophotrochozoa) | 89            | 78    | 92        | 80         | —          |
| E    | tardigrades+arthropods (Panarthropoda)            | 15            | 10    | 20        | 12         | —          |
| F    | platyhelminths+nematodes                          | 7             | 2     | 4         | 2          | —          |
| G    | platyhelminths+nematodes+annelids+mollusks        | 7             | 2     | 5         | 1          | —          |
| H    | platyhelminths+nematodes+tardigrades              | 1             | 7     | 3         | 13         | —          |
| I    | platyhelminths+nematodes+tardigrades+arthropods   | 3             | 9     | 4         | 7          | —          |

D. 56 species including *Xiphinema* (Xi-56)

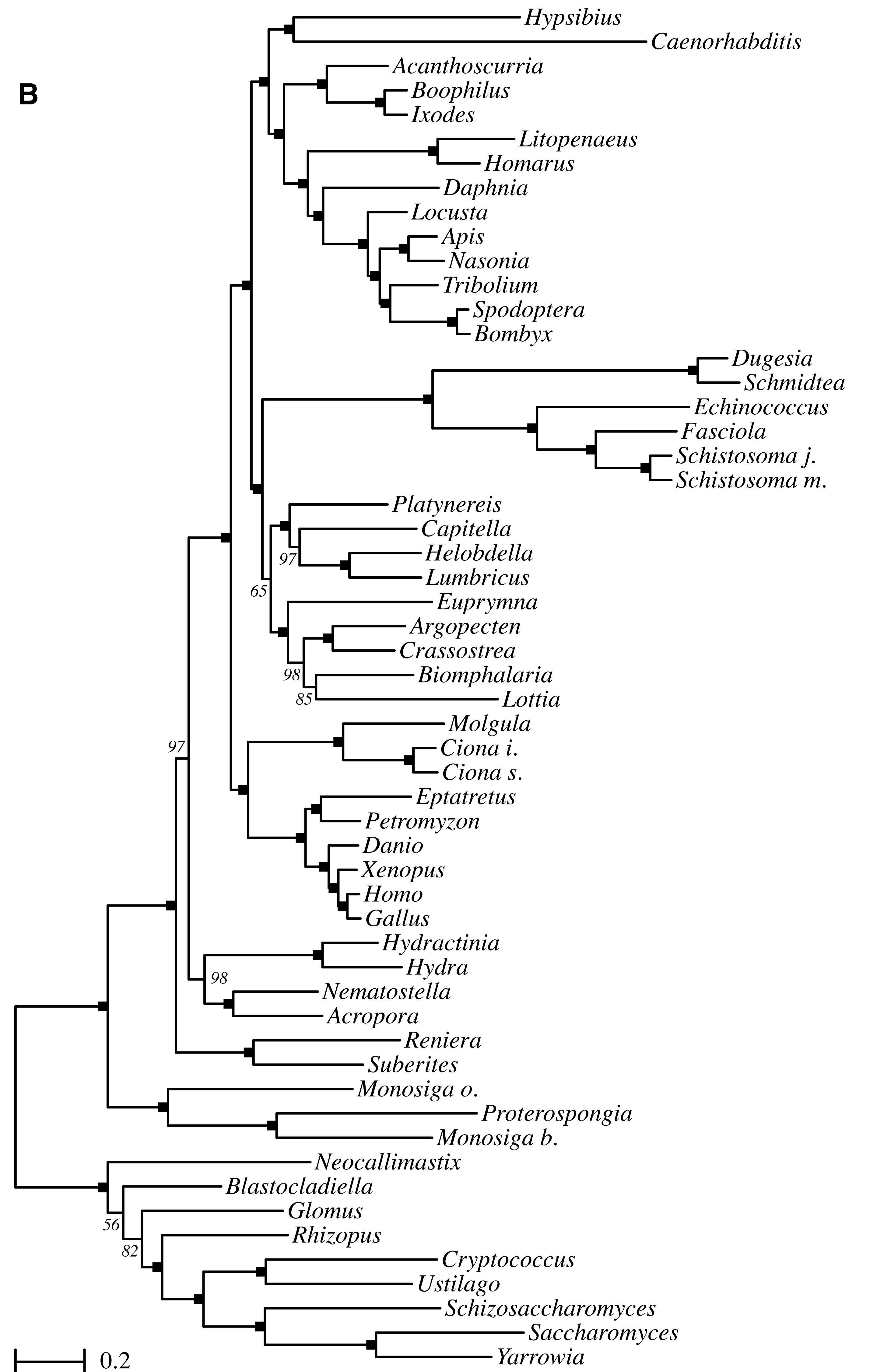
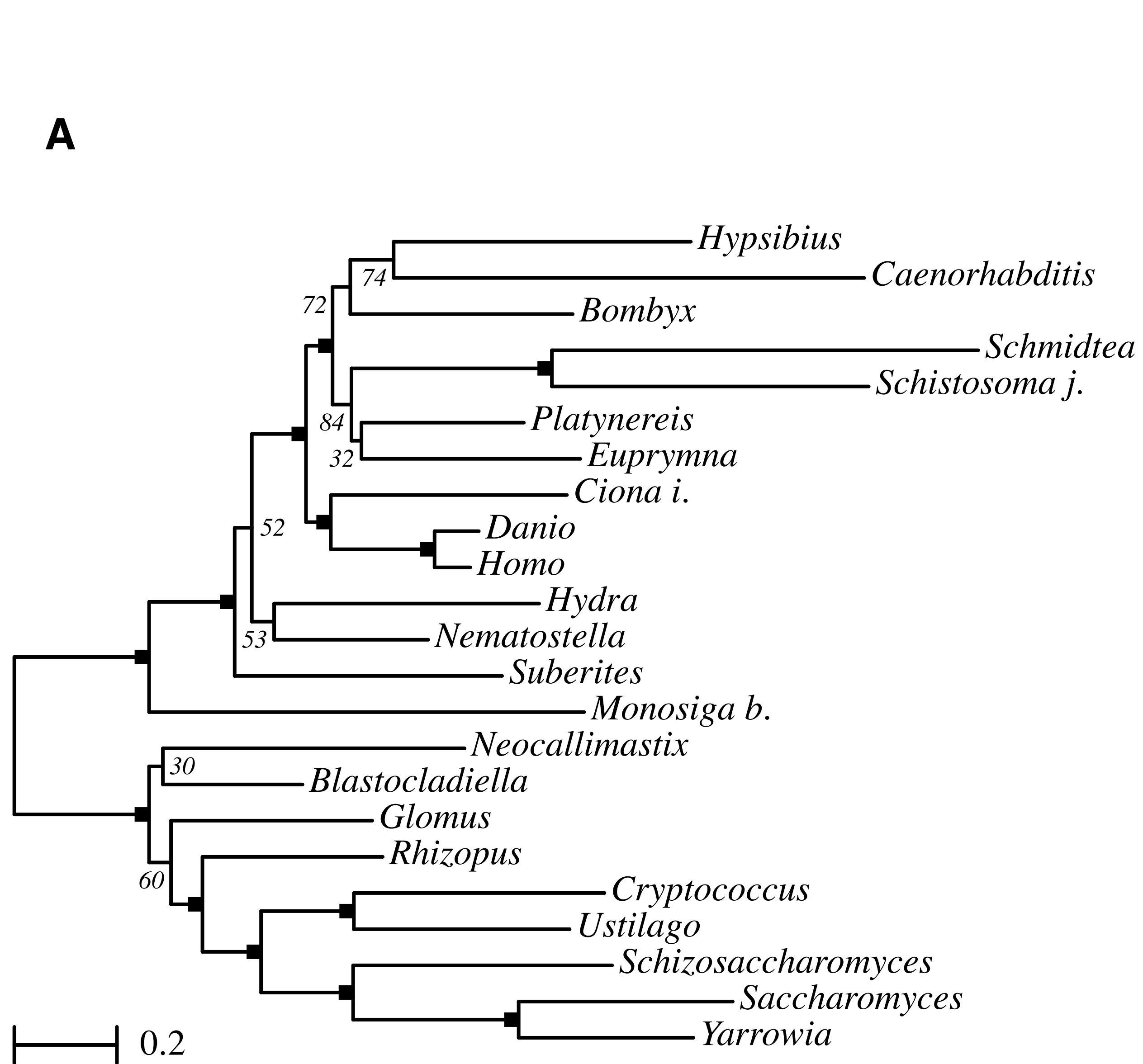
| node | description                                       | exh. analysis | PHYML | PHYML-SPR | TREEFINDER | PHYLOBAYES |
|------|---|---------------|-------|-----------|------------|------------|
| A    | tardigrades+nematodes                             | 98            | 86    | 97        | 91         | —          |
| B    | tardigrades+nematodes+arthropods (Ecdysozoa)      | 97            | 59    | 79        | 82         | —          |
| C    | annelids+mollusks                                 | 94            | 100   | 99        | 98         | —          |
| D    | annelids+mollusks+platyhelminths (Lophotrochozoa) | 95            | 57    | 79        | 80         | —          |
| E    | tardigrades+arthropods (Panarthropoda)            | 0             | 0     | 0         | 0          | —          |
| F    | platyhelminths+nematodes                          | 1             | 6     | 0         | 3          | —          |
| G    | platyhelminths+nematodes+annelids+mollusks        | 0             | 0     | 0         | 0          | —          |
| H    | platyhelminths+nematodes+tardigrades              | 3             | 41    | 20        | 18         | —          |
| I    | platyhelminths+nematodes+tardigrades+arthropods   | 3             | 39    | 17        | 17         | —          |



**Fig. S1. Dot-plots of bootstrap support values for heuristics and the exhaustive analysis.**  
Ce-23: blue diamonds; Ce-56: magenta squares; Xi-23: yellow triangles; Xi-56: cyan crosses.



**Fig. S2.** Tree inferred by TreeFinder, PHYML and PHYML-SPR from the Ce-56 dataset. Nodes supported by 100% bootstrap values in the TreeFinder analysis are denoted by black squares while lower values are given in plain style. This tree is not the ML tree.



**Fig. S3. Bayesian trees inferred under the CAT+ $\Gamma$  model from the Ce-23 (A) and the Ce-56 (B) datasets.** Nodes supported by 100% bootstrap values are denoted by black squares while lower values are given in italic style.