

SUPPLEMENTARY MATERIAL

**Increased taxon sampling reveals thousands of hidden orthologs in flatworms**

José M. Martín-Durán<sup>1§</sup>, Joseph F. Ryan<sup>1,2§</sup>, Bruno C. Vellutini<sup>1</sup>, Kevin Pang<sup>1</sup>,  
Andreas Hejnol<sup>1\*</sup>

CONTENT:

Supplementary Figures 1–6

Supplementary Tables 1–8

## **Supplementary Material**

**Supplementary Figure 1. GC content in flatworm transcriptomes.** GC content of each transcript plotted against its average length of G/C stretches for each flatworm species under study. The transcripts corresponding to hidden orthologs are in blue. Hidden orthologs do not differentiate from the majority of transcripts.

**Supplementary Figure 2. Gene Ontology (GO) characterization of hidden orthologs.** Distribution of GO terms for all recovered hidden orthologs (A–C) and for the hidden orthologs identified in *S. mediterranea* (D–F). Hidden orthologs include a great diversity of GO categories, with a big proportion of binding and catalytic activity. The number of GO nodes in each category is indicated in parentheses.

**Supplementary Figure 3. Orthology analysis of the centrosomal CEP192 protein.** CEP192 proteins do not contain any identifiable protein domain, and there is no known related protein that can help root the tree. Flatworm sequences are highlighted in red.

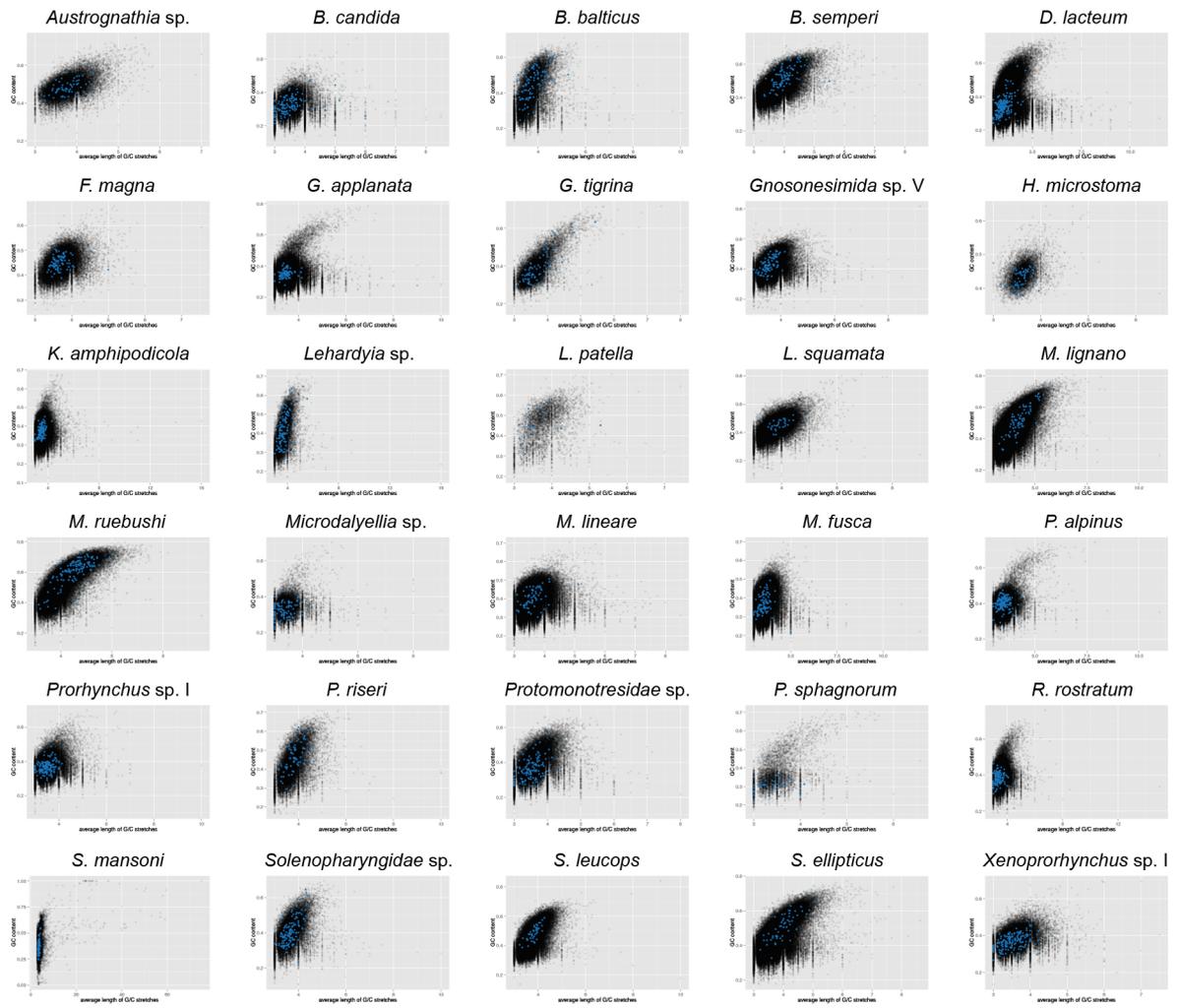
**Supplementary Figure 4. Orthology analysis of the centrosomal CCCAP protein.** CCCAP proteins contain a CCCAP domain (PFAM: PF15964), which is exclusive of these proteins. The domain is clearly recognizable in all flatworm sequences except *P. alpinus* (fragment too short) and the triclads *G. tigrina* and *S. mediterranea* (too divergent). Flatworm sequences are highlighted in red.

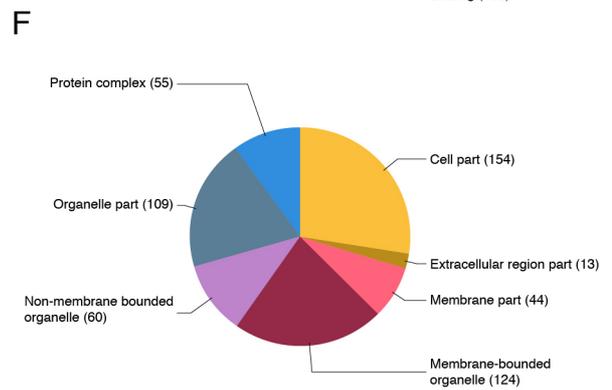
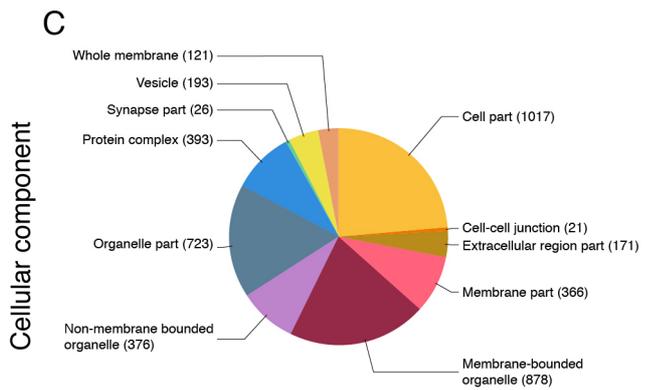
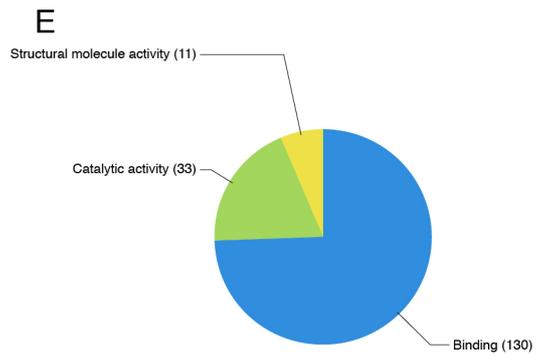
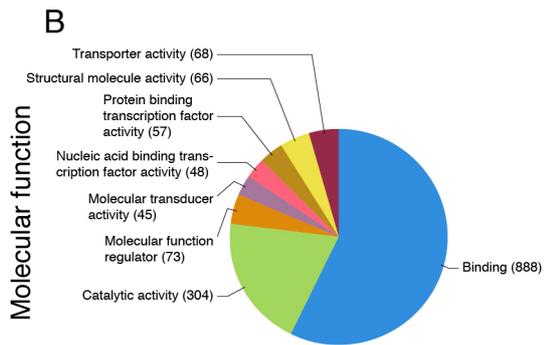
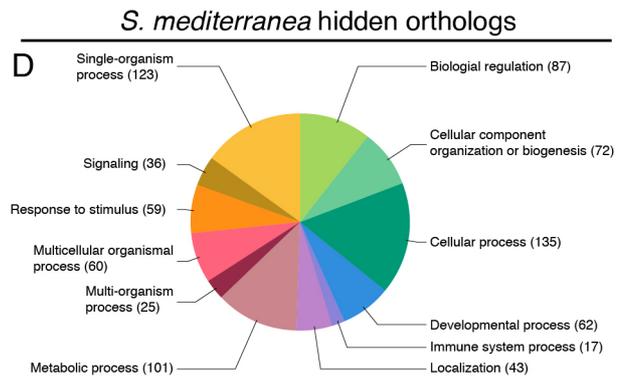
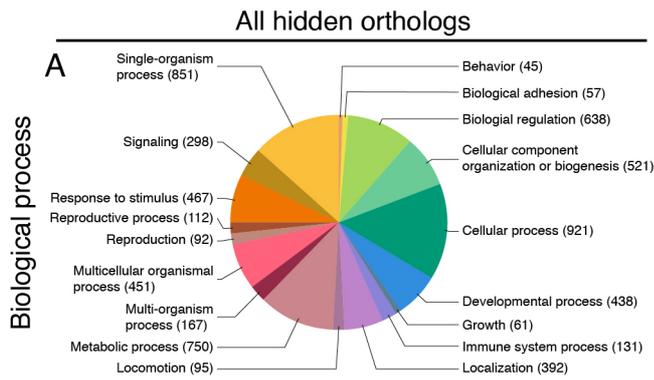
**Supplementary Figure 5. Orthology analysis of the ANTP homeodomain class.**

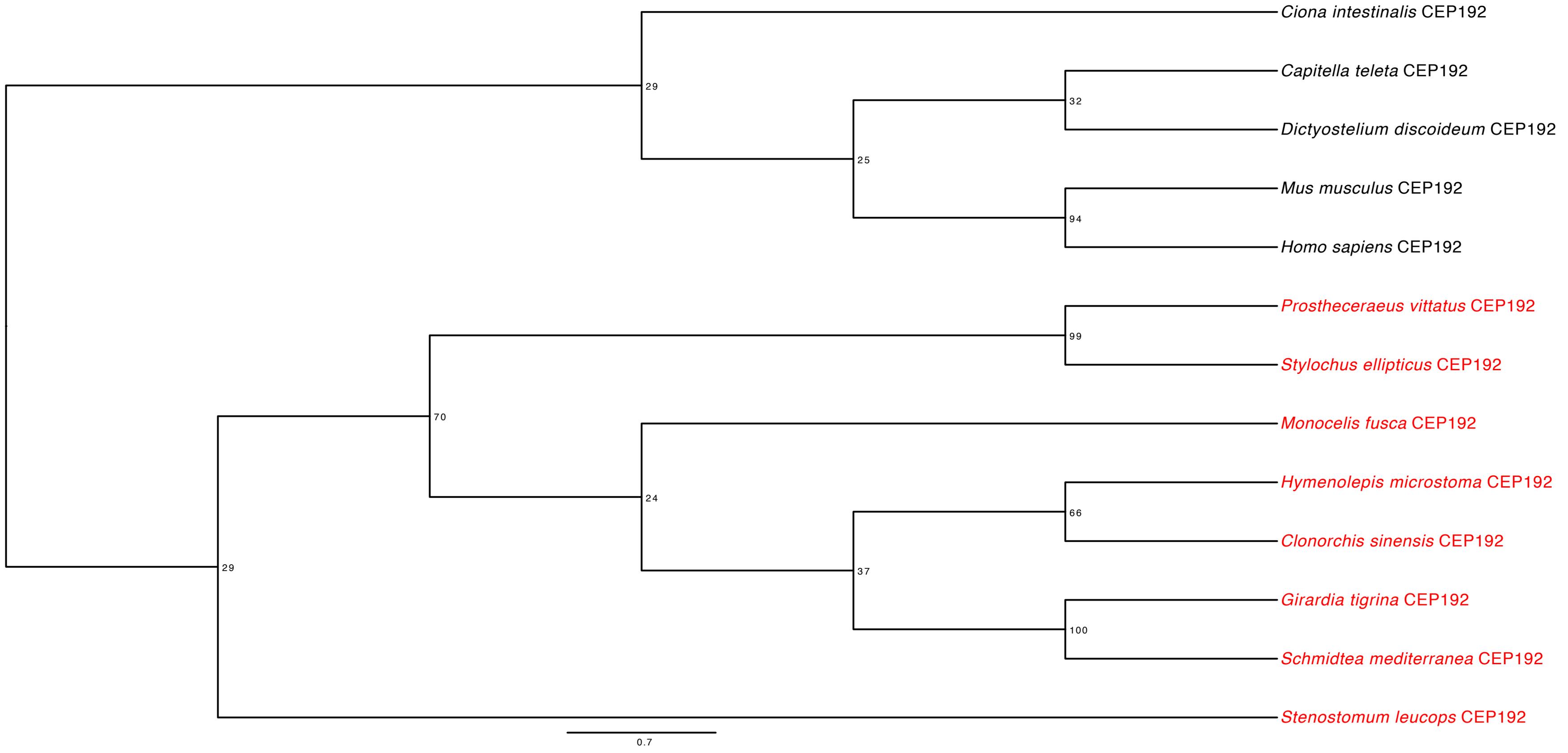
The newly identify sequences in the macrostomid *M. lignano*, the polyclad *P. vittatus* and the triclad *S. mediterranea* are highlighted in red.

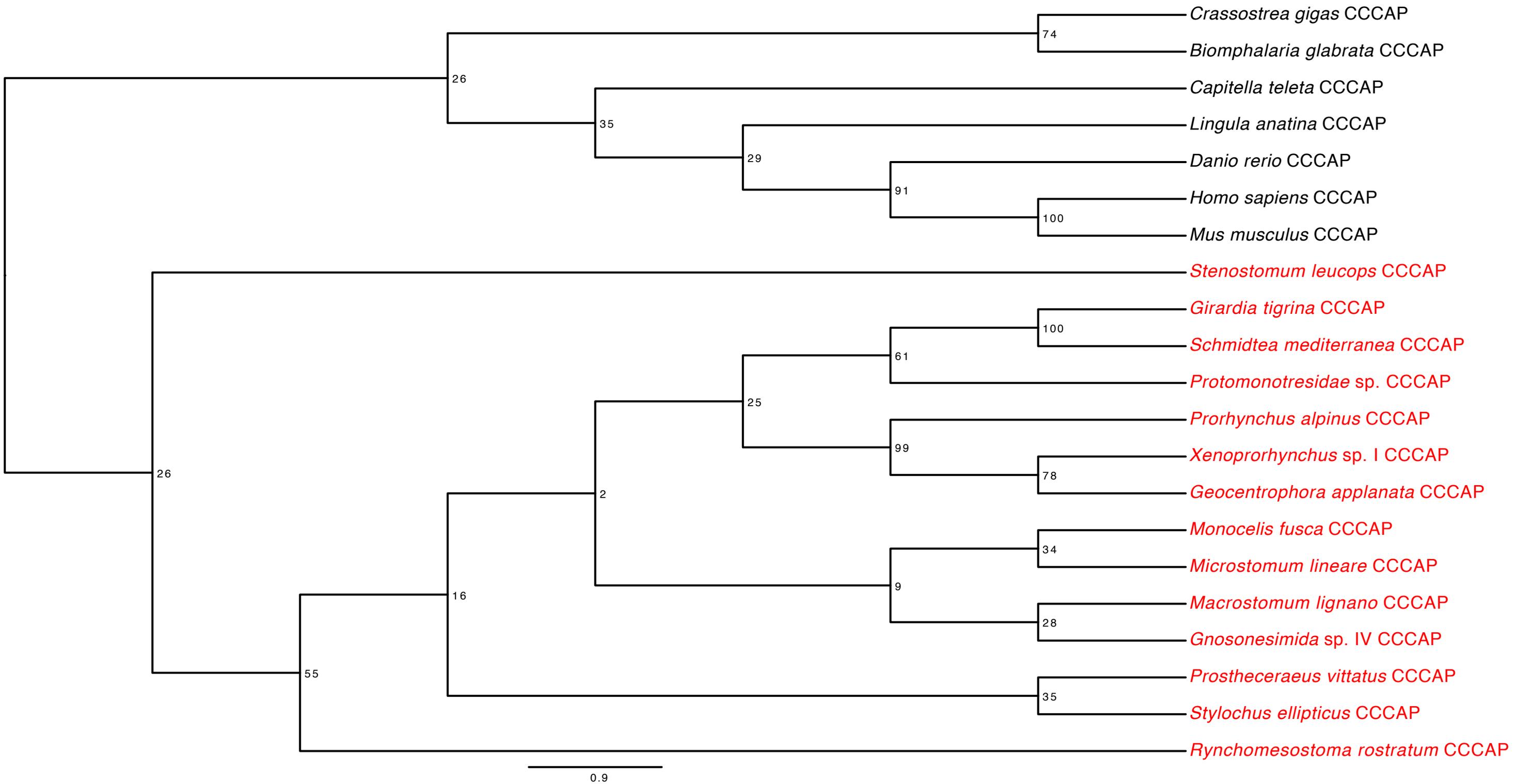
**Supplementary Figure 6. Orthology analysis of the CUT homeodomain class.** The

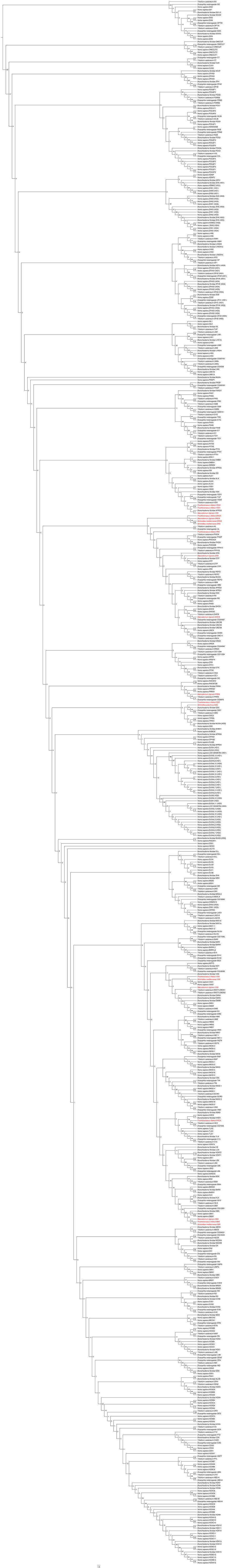
newly identify sequences in the macrostomid *M. lignano*, the polyclad *P. vittatus* and the triclad *S. mediterranea* are highlighted in red.

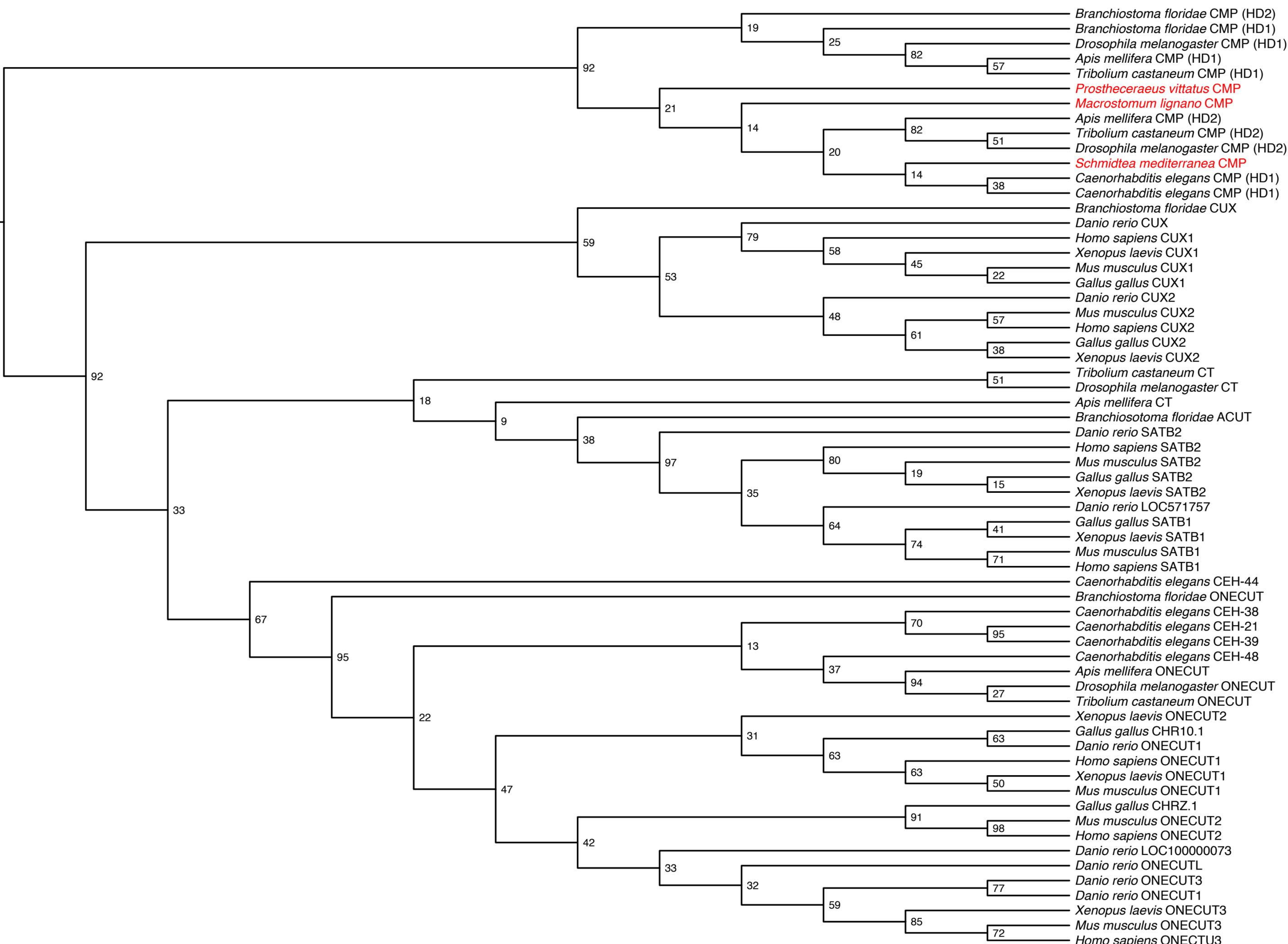












1.0

**Supplementary Table 1. Transcriptomes analyzed in this study**

| <b>Species</b>                  | <b>Group</b>    | <b>Source</b>               | <b>Depth (bp)</b> |
|---------------------------------|-----------------|-----------------------------|-------------------|
| <i>Austrognathia</i> sp.        | Gnathostomulida | Laumer, et al. (2015)       | 12,897,815        |
| <i>B. candida</i>               | Platyhelminthes | Laumer, et al. (2015)       | 22,365,603        |
| <i>B. balticus</i>              | Platyhelminthes | Laumer, et al. (2015)       | 11,257,706        |
| <i>B. semperi</i>               | Platyhelminthes | Laumer, et al. (2015)       | 56,081,914        |
| <i>D. lacteum</i>               | Platyhelminthes | Laumer, et al. (2015)       | 115,254,878       |
| <i>F. magna</i>                 | Platyhelminthes | Laumer, et al. (2015)       | 33,048,065        |
| <i>G. applanata</i>             | Platyhelminthes | Laumer, et al. (2015)       | 84,830,526        |
| <i>G. tigrina</i>               | Platyhelminthes | Laumer, et al. (2015)       | 28,510,283        |
| <i>Gnosonesimida</i> sp. IV     | Platyhelminthes | Laumer, et al. (2015)       | 49,489,839        |
| <i>H. microstoma</i>            | Platyhelminthes | Tsai, et al. (2013)         | 14,805,351        |
| <i>K. amphipodicola</i>         | Platyhelminthes | Laumer, et al. (2015)       | 51,328,359        |
| <i>Lehardyia</i> sp.            | Platyhelminthes | Laumer, et al. (2015)       | 9,340,372         |
| <i>L. patella</i>               | Rotifera        | Laumer, et al. (2015)       | 1,846,131         |
| <i>L. squamata</i>              | Gastrotricha    | Laumer, et al. (2015)       | 61,247,222        |
| <i>M. ruebushi</i>              | Platyhelminthes | Laumer, et al. (2015)       | 32,818,396        |
| <i>M. lignano</i>               | Platyhelminthes | SRA (SRX1343824)            | 146,421,223       |
| <i>Microdalyellia</i> sp.       | Platyhelminthes | Laumer, et al. (2015)       | 15,908,586        |
| <i>M. lineare</i>               | Platyhelminthes | Laumer, et al. (2015)       | 104,592,789       |
| <i>M. fusca</i>                 | Platyhelminthes | Laumer, et al. (2015)       | 107,319,077       |
| <i>P. alpinus</i>               | Platyhelminthes | Laumer, et al. (2015)       | 33,251,738        |
| <i>Prorhynchus</i> sp. I        | Platyhelminthes | Laumer, et al. (2015)       | 35,939,907        |
| <i>P. riseri</i>                | Platyhelminthes | Laumer, et al. (2015)       | 11,833,063        |
| <i>P. vittatus</i>              | Platyhelminthes | Laumer, et al. (2015)       | 361,770,473       |
| <i>Protomonotresidae</i><br>sp. | Platyhelminthes | Laumer, et al. (2015)       | 26,049,265        |
| <i>P. sphagnum</i>              | Platyhelminthes | Laumer, et al. (2015)       | 3,673,073         |
| <i>R. rostratum</i>             | Platyhelminthes | Laumer, et al. (2015)       | 54,514,686        |
| <i>S. mansoni</i>               | Platyhelminthes | Berriman, et al. (2009)     | 120,642,446       |
| <i>S. mediterranea</i>          | Platyhelminthes | Laumer, et al. (2015)       | 37,838,721        |
| <i>S. mediterranea</i> (BE)     | Platyhelminthes | Adamidi, et al. (2011)      | 26,258,087        |
| <i>S. mediterranea</i> (DD)     | Platyhelminthes | Brandl, et al. (2015)       | 51,791,386        |
| <i>S. mediterranea</i> (MU)     | Platyhelminthes | Sandmann, et al. (2011)     | 31,092,796        |
| <i>S. mediterranea</i> (OX)     | Platyhelminthes | Kao, et al. (2013)          | 59,986,795        |
| <i>S. mediterranea</i> (TO)     | Platyhelminthes | BioProject<br>(PRJNA205281) | 30,446,539        |
| <i>Solenopharyngidae</i><br>sp. | Platyhelminthes | Laumer, et al. (2015)       | 15,247,204        |
| <i>S. leucops</i>               | Platyhelminthes | Laumer, et al. (2015)       | 79,082,476        |
| <i>S. ellipticus</i>            | Platyhelminthes | Laumer, et al. (2015)       | 129,598,829       |
| <i>Xenoprorhynchus</i> sp I     | Platyhelminthes | Laumer, et al. (2015)       | 20,991,891        |

**Supplementary Table 2. Recovered hidden orthologs.** Hidden orthologs (as in human RefSeq) recovered in each transcriptome after running ‘Leapfrog’ with the transcriptome of the polyclad *P. vittatus* used as the ‘bridge’.

**Supplementary Table 3. List of hidden orthologs with reciprocal best BLAST hit against multiple human proteins.**

**Supplementary Table 4. PFAM domain composition of 130 human queries and their ortholog ‘bridge’ proteins in *P. vittatus*.**

**Supplementary Table 5. Data set used for principal component analysis.**

**Supplementary Table 6. Length of hidden orthologs and ORFs in flatworm transcriptomes.**

**Supplementary Table 7. PFAM domains identified in the hidden orthologs.**

**Supplementary Table 8. Significantly enriched GO terms in the hidden orthologs recovered in *S. mediterranea*.**