SUPPLEMENTAL MATERIAL

Increased taxon sampling reveals thousands of hidden orthologs in flatworms

José M. Martín-Durán, Joseph F. Ryan, Bruno C. Vellutini, Kevin Pang, Andreas Hejnol

TABLE OF CONTENTS

Supplemental Methods

Macrostomum lignano *transcriptome*

Adult and juveniles of *M. lignano* were kept under laboratory conditions as described elsewhere (Rieger et al. 1988). Animals starved for four days were homogenized and used as source material to isolate total RNA with the TRI Reagent (Life Technologies) following the manufacturer's recommendations. A total of 1 µg was used for Illumina paired-end library preparation at the GeneCore facilities (EMBL) and sequencing in a HiSeq 2000 platform. Paired-end reads were cleaned for adaptors using Trimmomatic v.0.35 (Bolger et al. 2014) and resulting reads were assembled *de novo* with Trinity v.r20140717 using default settings (Grabherr et al. 2011).

Data set preparation

We downloaded the Human RefSeq FASTA file from the NCBI FTP site last updated on March 25, 2015 (ftp://ftp.ncbi.nlm.nih.gov/refseq/H_sapiens/H_sapiens/protein/protein.fa.gz). We also downloaded the gene2accession data file from NCBI, which was last updated on July 3, 2015 (ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/gene2accession.gz). We then used the reduce_refseq script (Supplemental File 1) to generate a non-redundant Human RefSeq FASTA file with the following command: (reduce_refseq --fasta=protein.fa.gz --gene2accession=gene2accession.gz > HumRef2015.fa). This script prints only the first isoform for each Gene ID in the RefSeq FASTA file. The resulting file (available in Supplemental File 2) will be hereafter referred to as HumRef2015. Additionally, we assembled *de novo* a transcriptome for *M. lignano* (Supplemental Methods), and downloaded the 28 RNA-seq *de novo* assemblies from (Laumer et al. 2015) and six additional *S. mediterranea* datasets from PlanMine v1.0 (Brandl et al. 2016) on May 29,

2015. On July 14, 2015 we downloaded *Schistosoma mansoni, Hymenolepis microstoma,* and *Girardia tigrina* gene models from the Sanger FTP sites (*S. mansoni*: ftp://ftp.sanger.ac.uk/pub/project/pathogens/Schistosoma/mansoni/; *H. microstoma*: ftp://ftp.sanger.ac.uk/pub/project/pathogens/Hymenolepis/microstoma/; *G. tigrina*: https://drive.google.com/open?id=0B-oXGfLmcaXQUUhVS29IbC1CeHM). Further details on datasets are available in Supplemental Table 1. All flatworm transcriptomes were constructed from adult tissue, except for *S. ellipticus*, *P. vittatus*, *G. applanata* and *Prorhynchus* sp. I, which included embryonic stages.

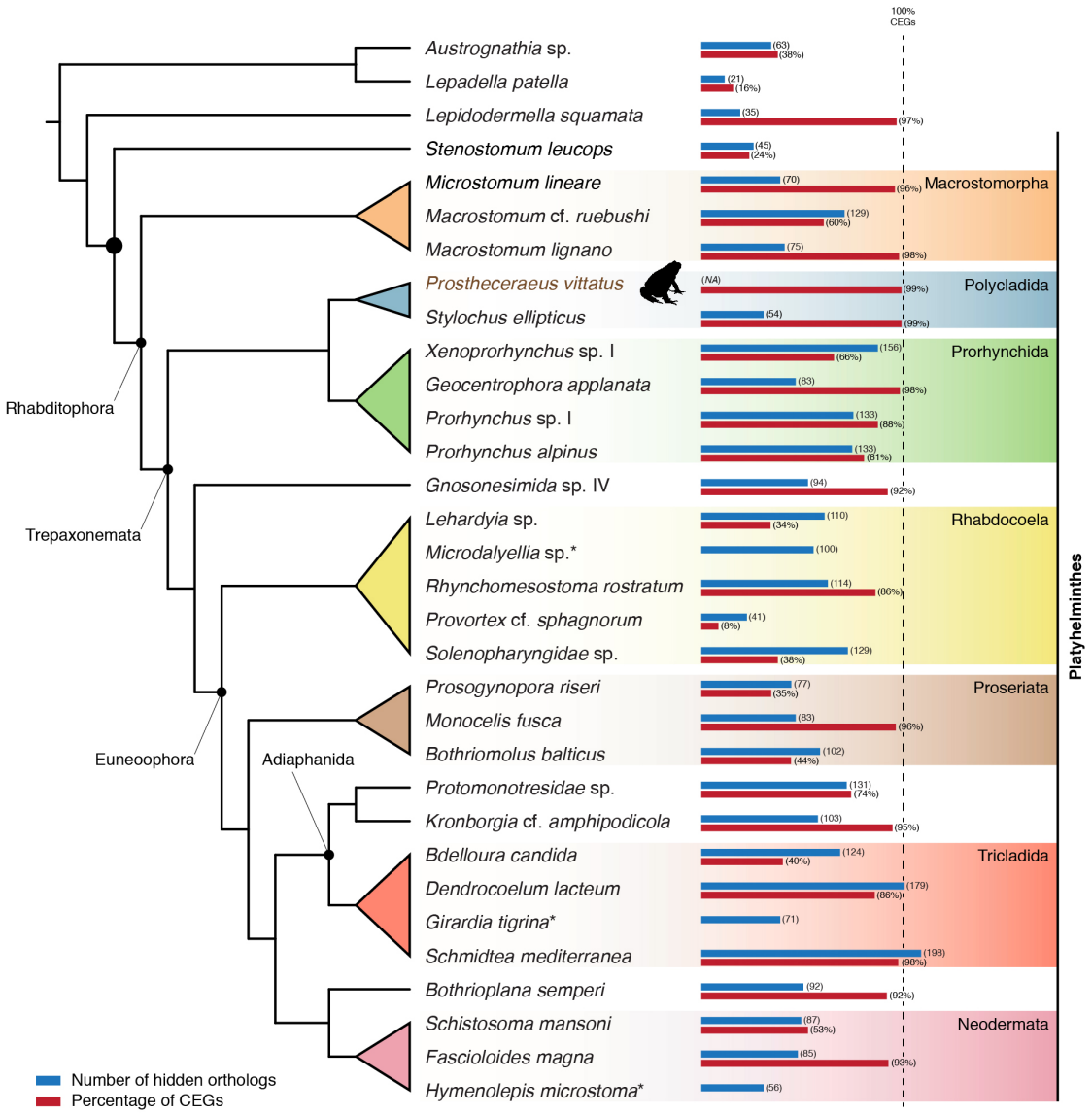
CEGMA analysis, transcriptome quality assessment, and statistics

Transcriptome completeness was evaluated with CEGMA (Parra et al. 2007; Parra et al. 2009). We could not run the CEGMA pipeline in the transcriptomes of *G. tigrina*, *Microdalyellia* sp. and *H. microstoma* due to an untraceable error. We calculated the contig metrics for each transcriptome assembly with TransRate (Smith-Unna et al. 2015). Principal component analysis was performed in R (R Core Team 2015) and plotted using the ggplot2 package (Wickham 2009). For this analysis, branch lengths were the root-to-tip distances inferred from a maximum likelihood phylogeny previously reported (Laumer et al. 2015).

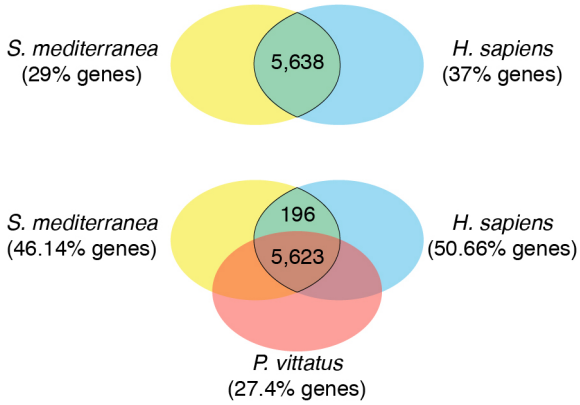
Smed-SDCCAG8 and Smed-CEP192 primer sequences

The following gene specific primers were used to clone *Smed-SDCCAG8* and *Smed-CEP192*: *Smed-SDCCAG8-F* 5’ TTGACCAGGAACCGTACGAA 3’, *Smed-SDCCAG8-R* 5’ CACATTGCTCGAATCTGGCA 3’; *Smed-CEP192-F* 5’ CAACAGGTCCGATTCCAACC 3’, *Smed-CEP192-R* 5’ AACGCAACAGGAACCAGAAC 3’.

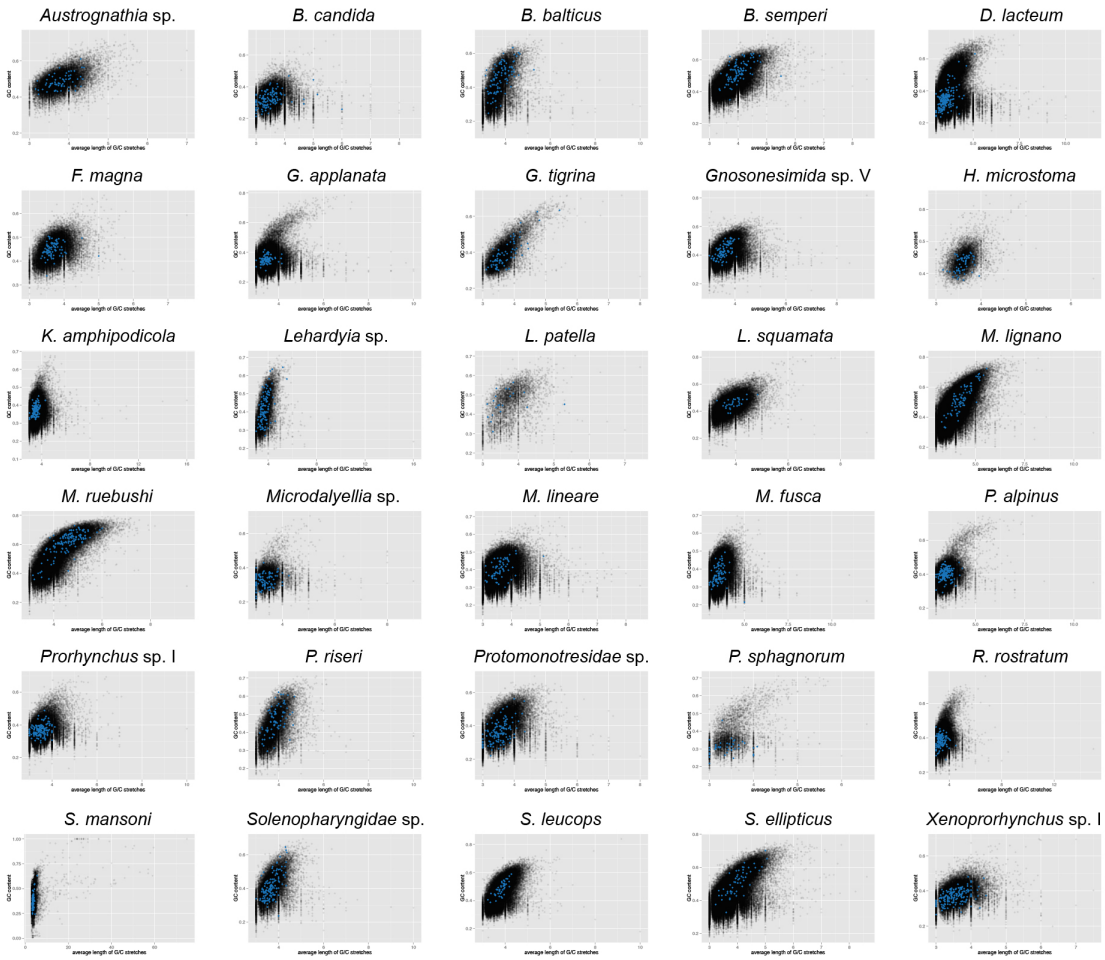
Supplemental Figures



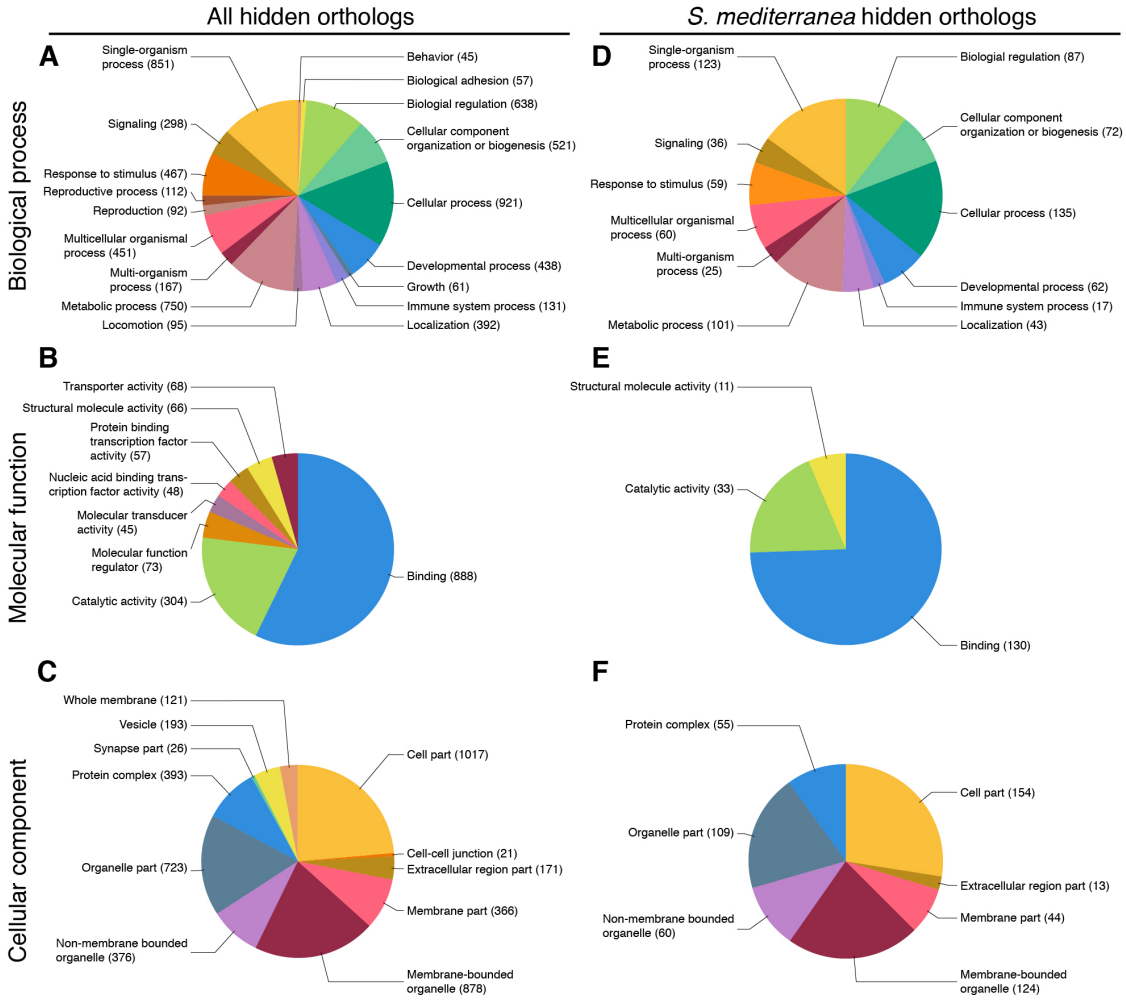
Supplemental Figure S1. Distribution of hidden orthologs in the analyzed flatworm transcriptomes. The figure shows the total number of hidden orthologs in the analyzed transcriptomes in a phylogenetic context and with respect to their completeness (percentage of recovered core eukaryote genes, CEGs). The quality of the transcriptomes seems to be a limitation for the recovery of hidden orthologs in some flatworm lineages (e.g. *Provortex* cf. *sphagnorum*). However, the number of hidden orthologs is very species-specific.



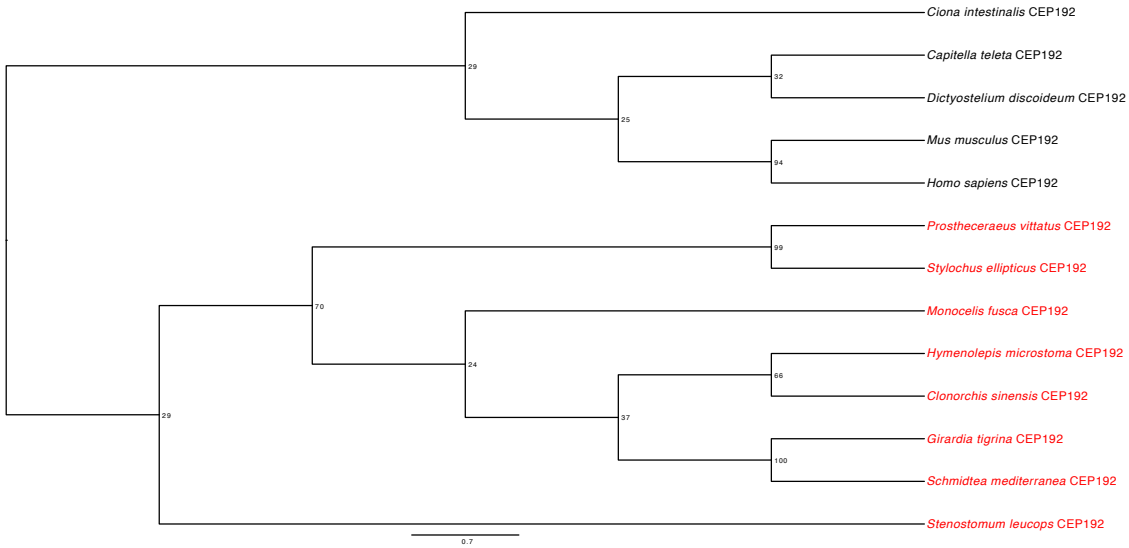
Supplemental Figure S2. Impact of a 'bridge' species in OrthoFinder. Number of orthogroups identified by OrthoFinder when only *S. mediterranea* and human are considered (top) and when a 'bridge' transcriptome is added (bottom). The inclusion of an intermediate species increases the number of sequences assigned to orthogroups from 29% to 46.14% of the transcriptome, and the number of orthogroups between *S. mediterranea* and human from 5,638 to 5,819. The inclusion of the *P. vittatus* 'bridge' transcriptome in an OrthoFinder analysis recovered only 62 out of the 75 hidden orthologs (82.7%) identified by the Leapfrog pipeline in the same *S. mediterranea* transcriptome with similar E-value cutoffs. For *S. mediterranea*, we used the transcriptome sequenced by J. Rink's lab at the MPI in Dresden (Brandl et al. 2016).



Supplemental Figure S3. 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.

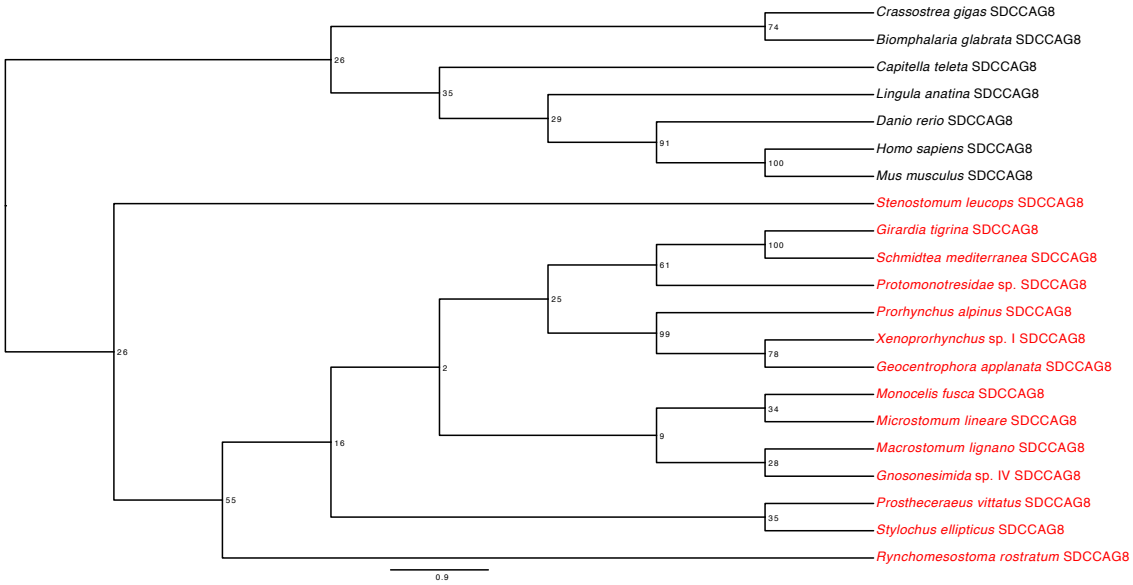


Supplemental Figure S4. Gene Ontology (GO) characterization of hidden orthologs. Distribution of GO terms for all recovered hidden orthologs (**A**–**C**) and for the hidden orthologs identified in *S. mediterranea* (**D**–**F**). Hidden orthologs include a great diversity of GO categories, with a big proportion of binding and catalytic activity. The number of GO nodes in each category is indicated in parentheses.

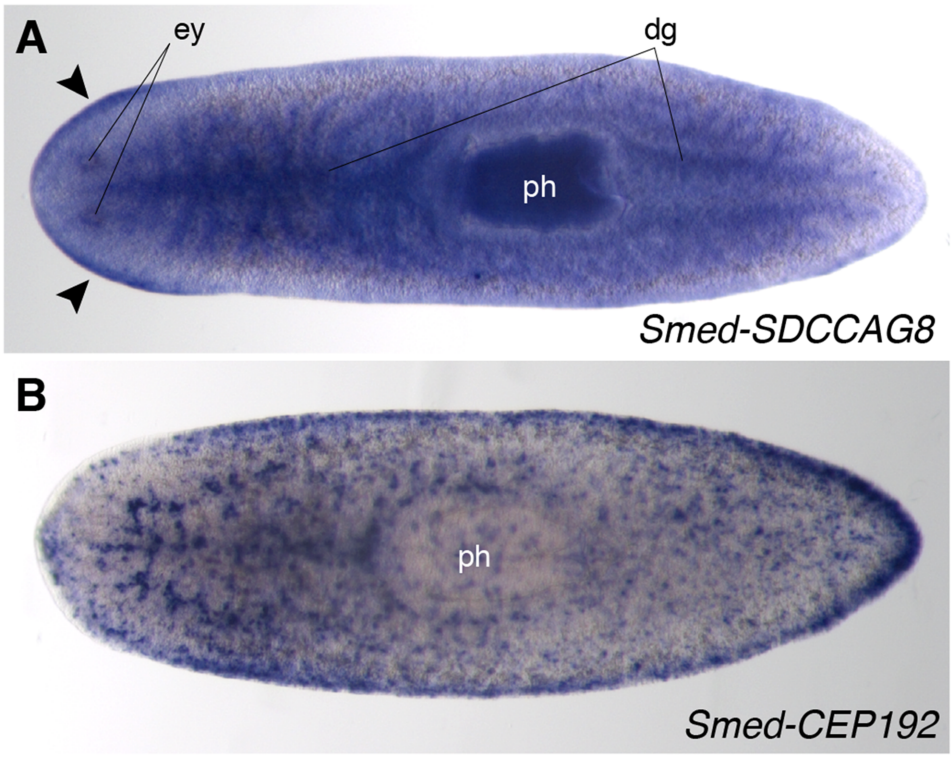


Supplemental Figure S5. Orthology analysis of the centrosomal CEP192 protein.

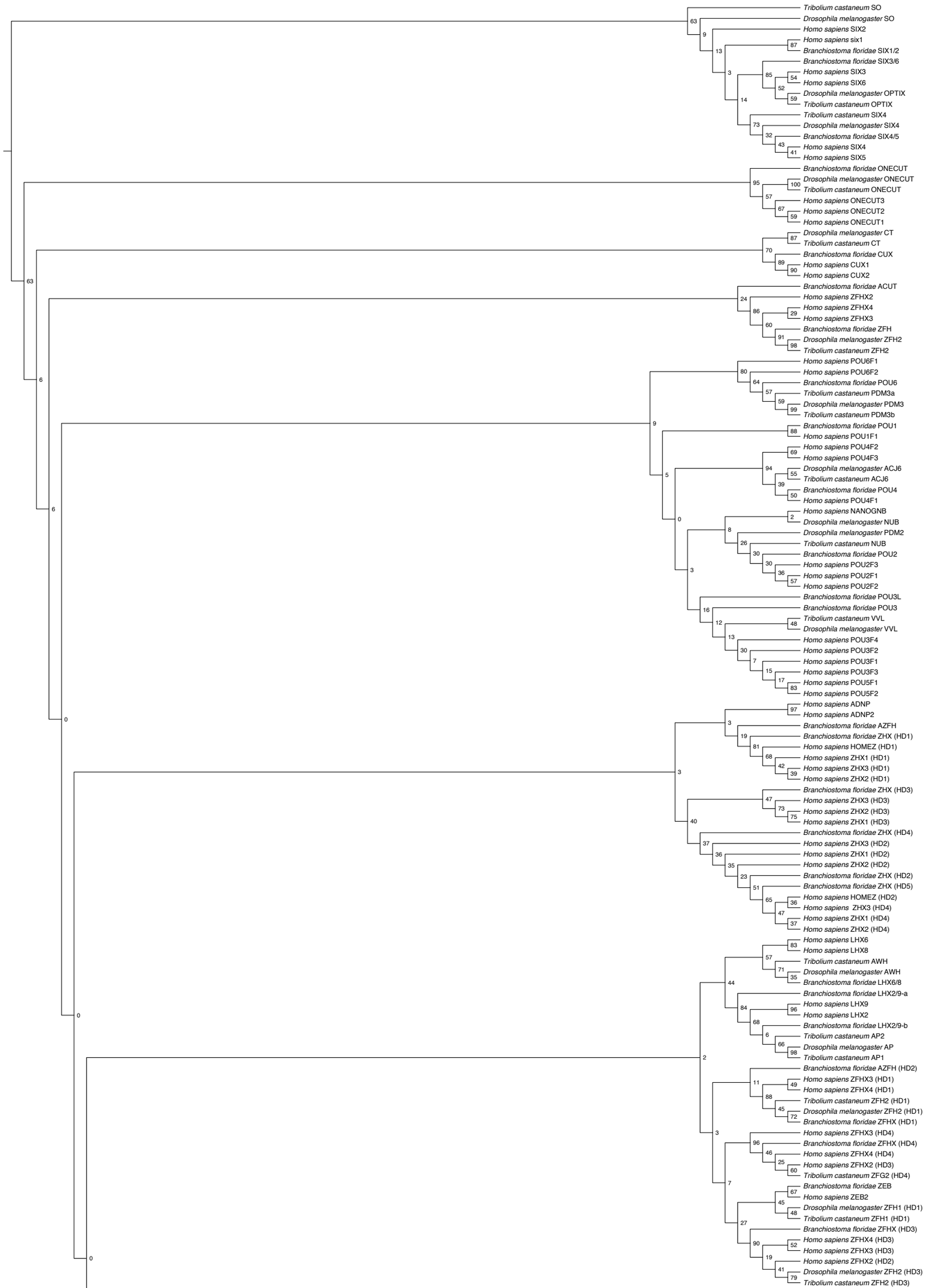
CEP192 proteins do not contain any identifiable protein domain, and there is no known related protein that can help root the tree. Flatworm sequences are highlighted in red. Model of protein evolution: RtRev+I+G+F.

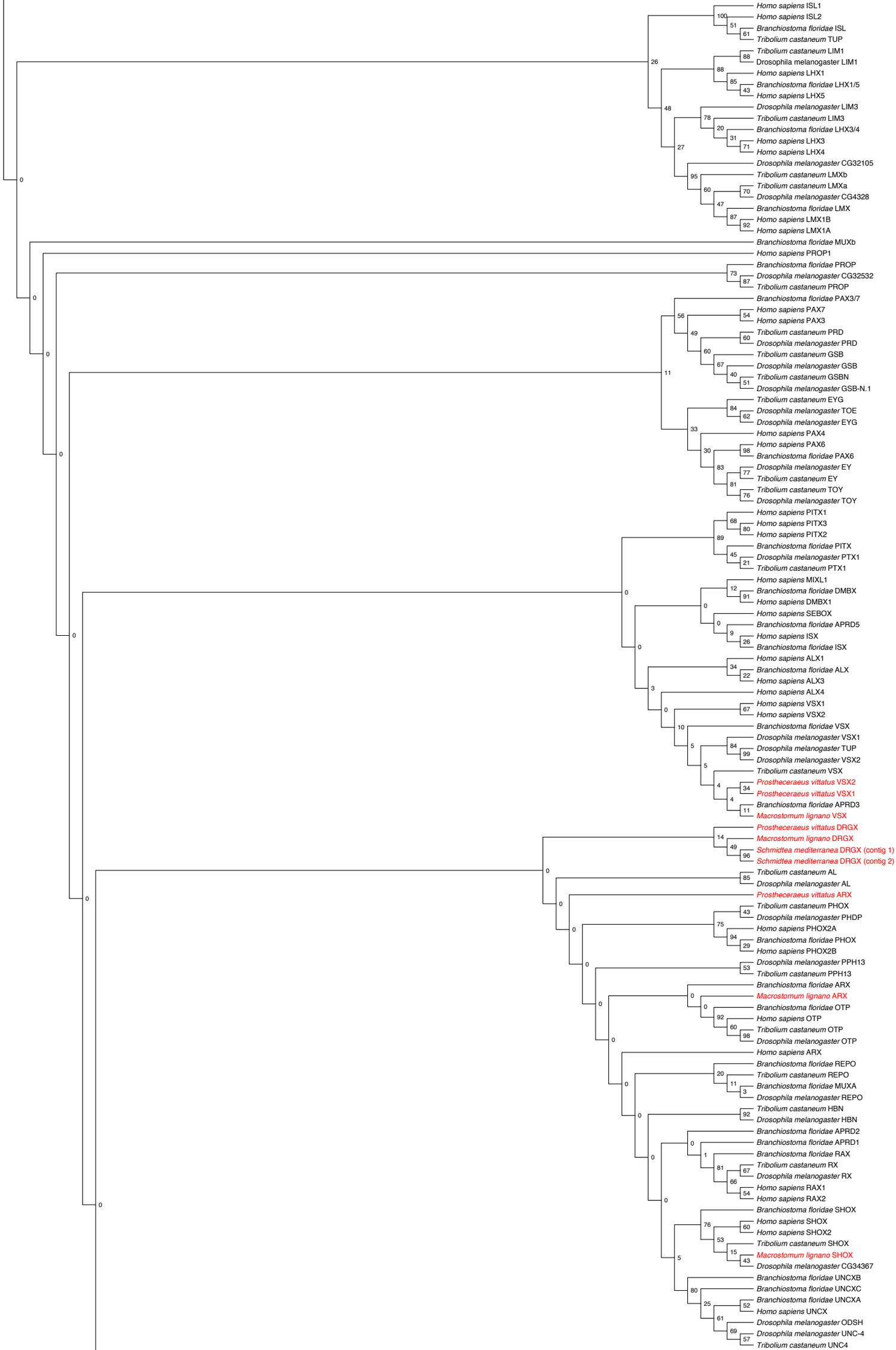


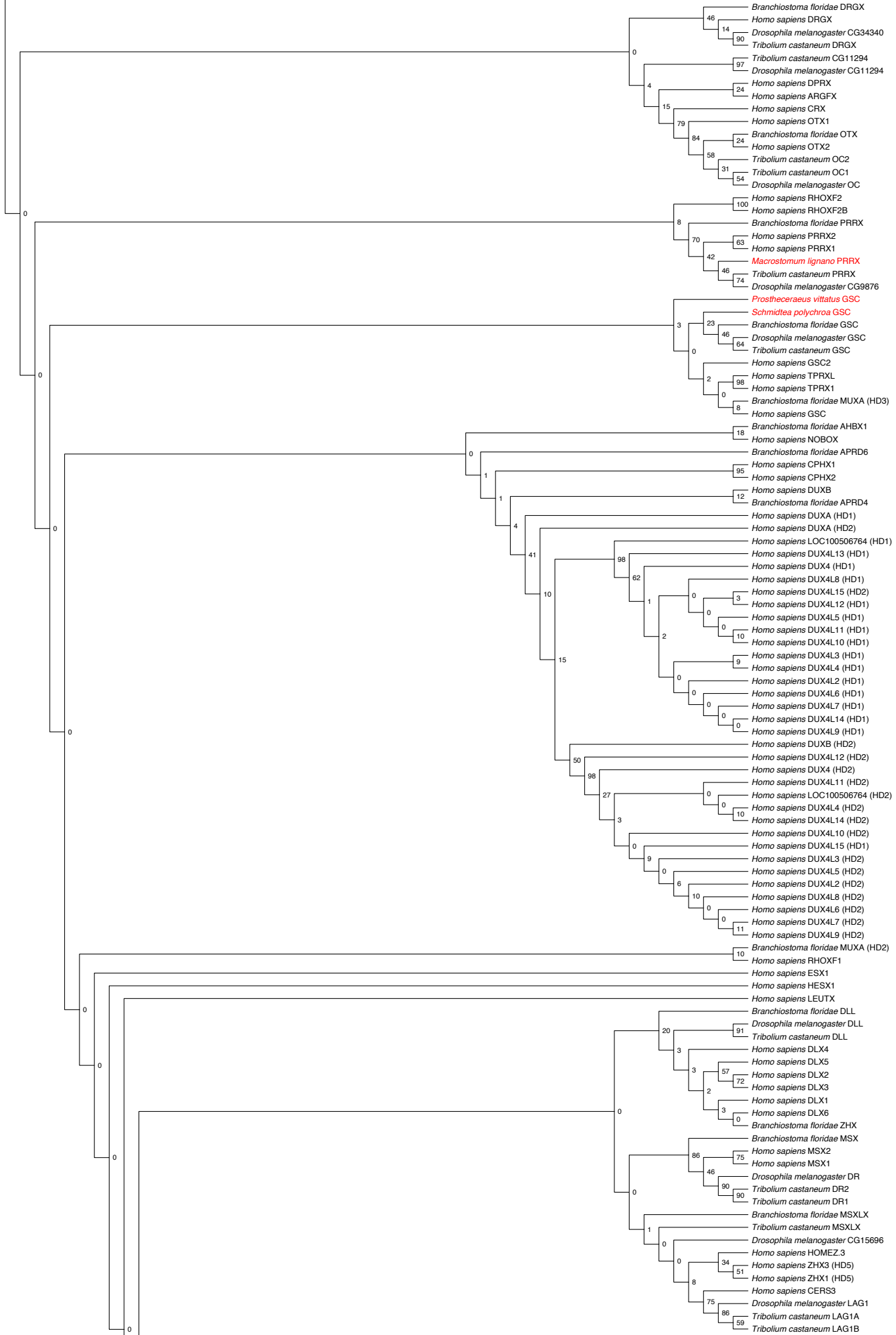
Supplemental Figure S6. Orthology analysis of the centrosomal SDCCAG8 protein. SDCCAG8 proteins contain a SDCCAG8 domain (PFAM: PF15964), which is exclusive of these proteins. The domain is clearly recognizable in all flatworm sequences except *P. alpinus* (fragment too short) and the triclads *G. tigrina* and *S. mediterranea* (too divergent). Flatworm sequences are highlighted in red. Model of protein evolution: JTT+G+F.

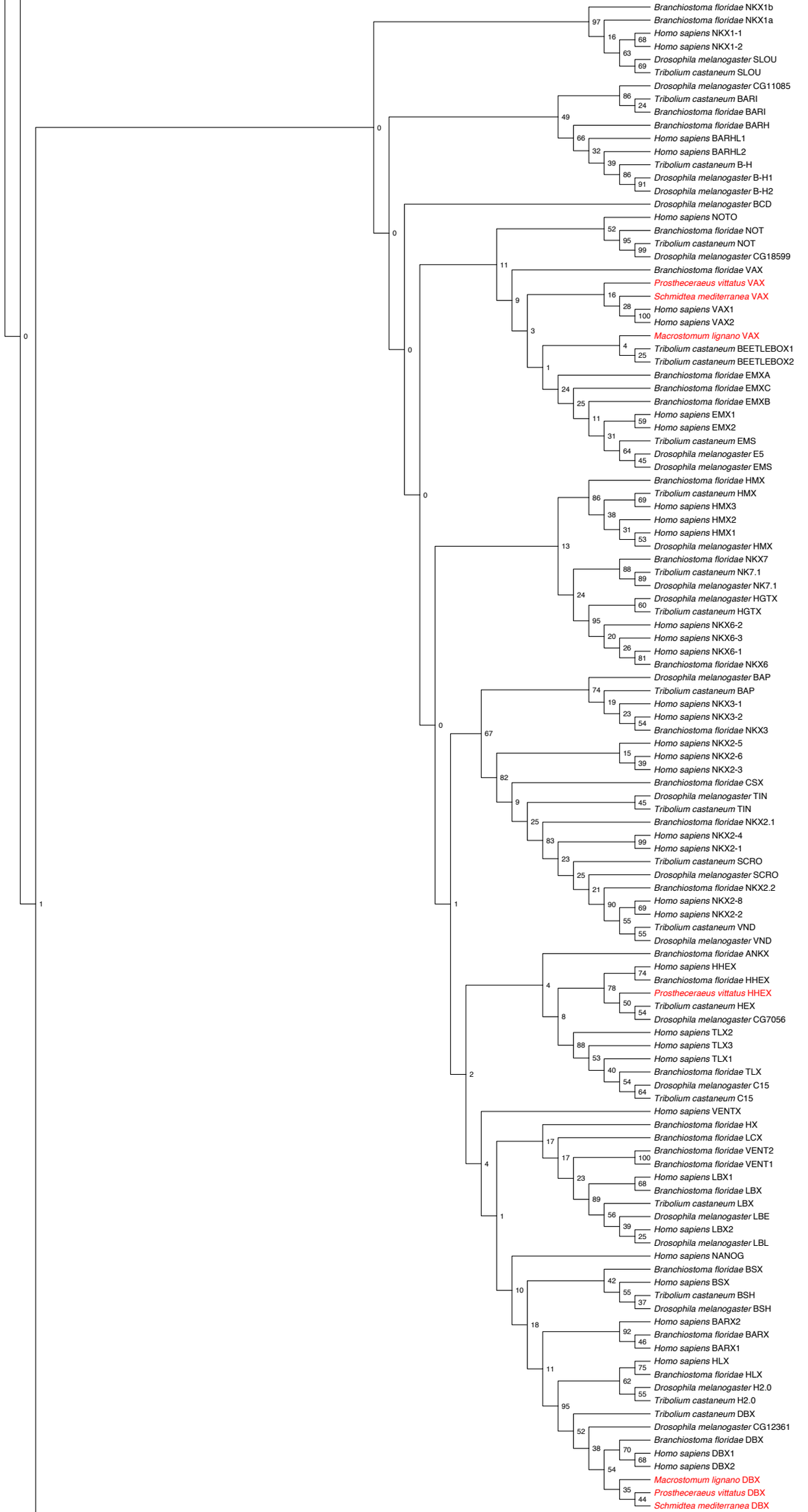


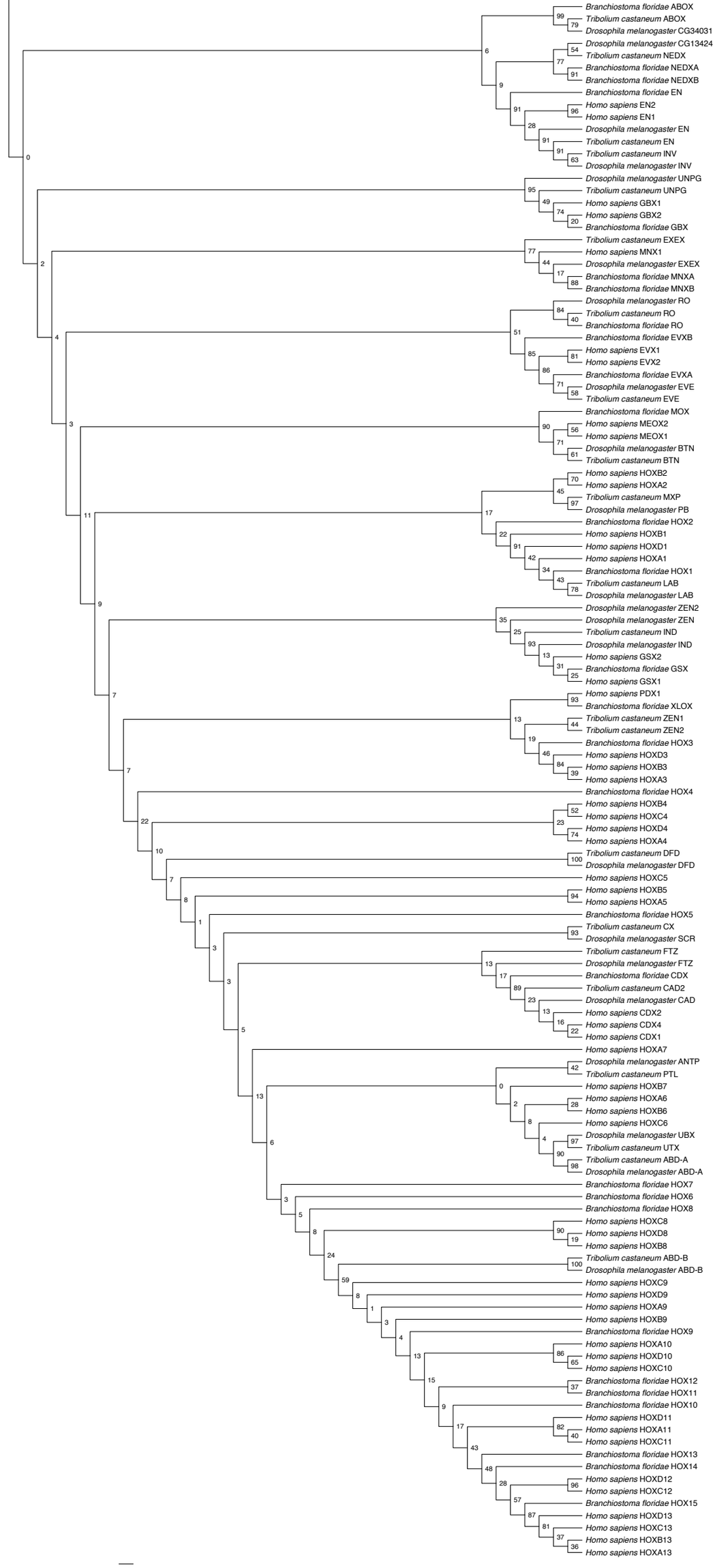
Supplemental Figure S7. Expression of *Smed- SDCCAG8* and *Smed-CEP192* in adult planarians. (**A**–**B**) Whole mount *in situ* hybridization of *Smed-SDCCAG8* (n=6) and *Smed-CEP192* (n=6) in adult intact *S. mediterranea*. (**A**) *Smed-SDCCAG8* is expressed in the epidermis, anterior margin (black arrowheads), in the digestive system (dg) and pharynx, and faintly in the parenchyma. (**B**) *Smed-CEP192* is expressed in a salt-and-pepper manner in isolated cells of the epidermis, posterior margin of the animal, and parenchyma.


Tribolium castaneum SO
Drosophila melanogaster SO
Homo sapiens SIX2
Homo sapiens six1
Branchiostoma floridae SIX1/2
Branchiostoma floridae SIX3/6
Homo sapiens SIX3
Homo sapiens SIX6
Drosophila melanogaster OPTIX
Tribolium castaneum OPTIX
Tribolium castaneum SIX4
Drosophila melanogaster SIX4
Branchiostoma floridae SIX4/5
Homo sapiens SIX4
Homo sapiens SIX5
Branchiostoma floridae ONECUT
Drosophila melanogaster ONECUT
Tribolium castaneum ONECUT
Homo sapiens ONECUT3
Homo sapiens ONECUT2
Homo sapiens ONECUT1
Drosophila melanogaster CT
Tribolium castaneum CT
Branchiostoma floridae CUX
Homo sapiens CUX1
Homo sapiens CUX2
Branchiostoma floridae ACUT
Homo sapiens ZFHX2
Homo sapiens ZFHX4
Homo sapiens ZFHX3
Branchiostoma floridae ZFH
Drosophila melanogaster ZFH2
Tribolium castaneum ZFH2
Homo sapiens POU6F1
Homo sapiens POU6F2
Branchiostoma floridae POU6
Tribolium castaneum PDM3a
Drosophila melanogaster PDM3
Tribolium castaneum PDM3b
Branchiostoma floridae POU1
Homo sapiens POU1F1
Homo sapiens POU4F2
Homo sapiens POU4F3
Drosophila melanogaster ACJ6
Tribolium castaneum ACJ6
Branchiostoma floridae POU4
Homo sapiens POU4F1
Homo sapiens NANOGNB
Drosophila melanogaster NUB
Drosophila melanogaster PDM2
Tribolium castaneum NUB
Branchiostoma floridae POU2
Homo sapiens POU2F3
Homo sapiens POU2F1
Homo sapiens POU2F2
Branchiostoma floridae POU3L
Branchiostoma floridae POU3
Tribolium castaneum VVL
Drosophila melanogaster VVL
Homo sapiens POU3F4
Homo sapiens POU3F2
Homo sapiens POU3F1
Homo sapiens POU3F3
Homo sapiens POU5F1
Homo sapiens POU5F2
Homo sapiens ADNP
Homo sapiens ADNP2
Branchiostoma floridae AZFH
Branchiostoma floridae ZHX (HD1)
Homo sapiens HOMEZ (HD1)
Homo sapiens ZHX1 (HD1)
Homo sapiens ZHX3 (HD1)
Homo sapiens ZHX2 (HD1)
Branchiostoma floridae ZHX (HD3)
Homo sapiens ZHX3 (HD3)
Homo sapiens ZHX2 (HD3)
Homo sapiens ZHX1 (HD3)
Branchiostoma floridae ZHX (HD4)
Homo sapiens ZHX3 (HD2)
Homo sapiens ZHX1 (HD2)
Homo sapiens ZHX2 (HD2)
Branchiostoma floridae ZHX (HD2)
Branchiostoma floridae ZHX (HD5)
Homo sapiens HOMEZ (HD2)
Homo sapiens ZHX3 (HD4)
Homo sapiens ZHX1 (HD4)
Homo sapiens ZHX2 (HD4)
Homo sapiens LHX6
Homo sapiens LHX8
Tribolium castaneum AWH
Drosophila melanogaster AWH
Branchiostoma floridae LHX6/8
Branchiostoma floridae LHX2/9-a
Homo sapiens LHX9
Homo sapiens LHX2
Branchiostoma floridae LHX2/9-b
Tribolium castaneum AP2
Drosophila melanogaster AP
Tribolium castaneum AP1
Branchiostoma floridae AZFH (HD2)
Homo sapiens ZFHX3 (HD1)
Homo sapiens ZFHX4 (HD1)
Tribolium castaneum ZFH2 (HD1)
Drosophila melanogaster ZFH2 (HD1)
Branchiostoma floridae ZFHX (HD1)
Homo sapiens ZFHX3 (HD4)
Branchiostoma floridae ZFHX (HD4)
Homo sapiens ZFHX4 (HD4)
Homo sapiens ZFHX2 (HD3)
Tribolium castaneum ZFG2 (HD4)
Branchiostoma floridae ZEB
Homo sapiens ZEB2
Drosophila melanogaster ZFH1 (HD1)
Tribolium castaneum ZFH1 (HD1)
Branchiostoma floridae ZFHX (HD3)
Homo sapiens ZFHX4 (HD3)
Homo sapiens ZFHX3 (HD3)
Homo sapiens ZFHX2 (HD2)
Drosophila melanogaster ZFH2 (HD3)
Tribolium castaneum ZFH2 (HD3)

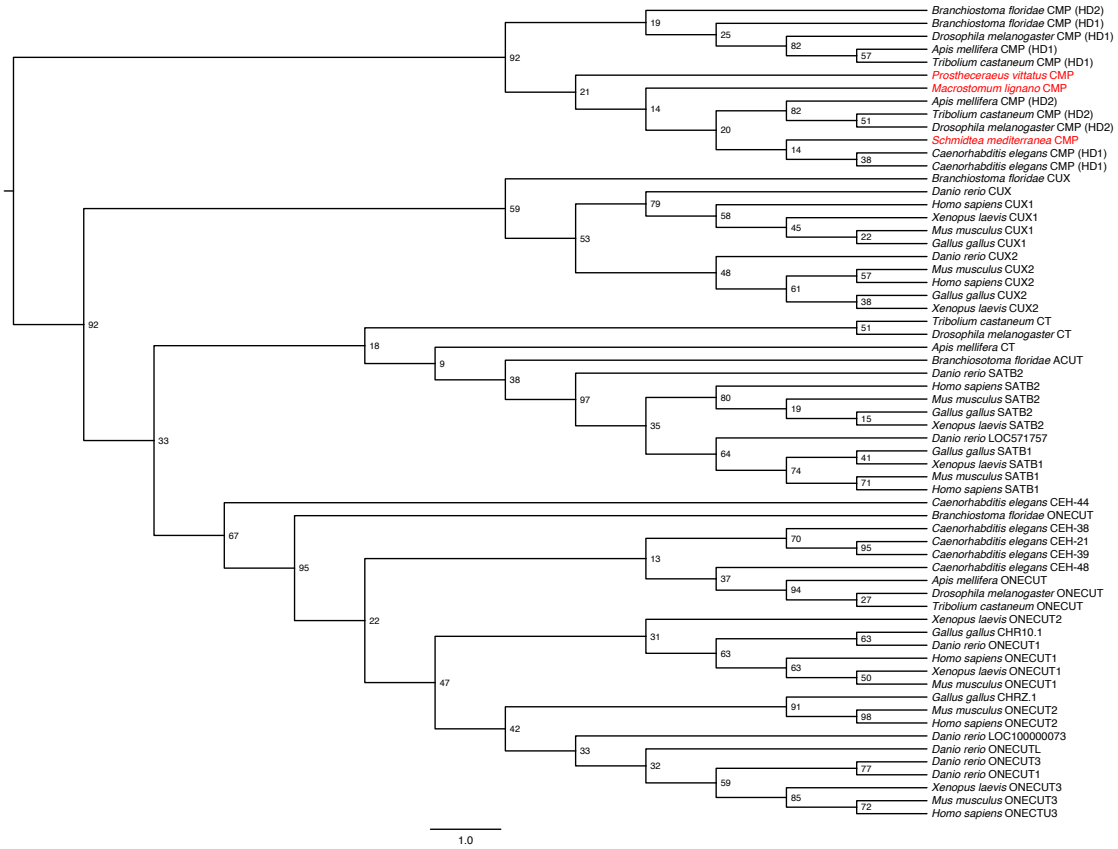

Homo sapiens ISL1
Homo sapiens ISL2
Branchiostoma floridae ISL
Tribolium castaneum TUP
Tribolium castaneum LIM1
Drosophila melanogaster LIM1
Homo sapiens LHX1
Branchiostoma floridae LHX1/5
Homo sapiens LHX5
Drosophila melanogaster LIM3
Tribolium castaneum LIM3
Branchiostoma floridae LHX3/4
Homo sapiens LHX3
Homo sapiens LHX4
Drosophila melanogaster CG32105
Tribolium castaneum LMXb
Tribolium castaneum LMXa
Drosophila melanogaster CG4328
Branchiostoma floridae LMX
Homo sapiens LMX1B
Homo sapiens LMX1A
Branchiostoma floridae MUXb
Homo sapiens PROP1
Branchiostoma floridae PROP
Drosophila melanogaster CG32532
Tribolium castaneum PROP
Branchiostoma floridae PAX3/7
Homo sapiens PAX7
Homo sapiens PAX3
Tribolium castaneum PRD
Drosophila melanogaster PRD
Tribolium castaneum GSB
Drosophila melanogaster GSB
Tribolium castaneum GSBN
Drosophila melanogaster GSB-N.1
Tribolium castaneum EYG
Drosophila melanogaster TOE
Drosophila melanogaster EYG
Homo sapiens PAX4
Homo sapiens PAX6
Branchiostoma floridae PAX6
Drosophila melanogaster EY
Tribolium castaneum EY
Tribolium castaneum TOY
Drosophila melanogaster TOY
Homo sapiens PITX1
Homo sapiens PITX3
Homo sapiens PITX2
Branchiostoma floridae PITX
Drosophila melanogaster PTX1
Tribolium castaneum PTX1
Homo sapiens MIXL1
Branchiostoma floridae DMBX
Homo sapiens DMBX1
Homo sapiens SEBOX
Branchiostoma floridae APRD5
Homo sapiens ISX
Branchiostoma floridae ISX
Homo sapiens ALX1
Branchiostoma floridae ALX
Homo sapiens ALX3
Homo sapiens ALX4
Homo sapiens VSX1
Homo sapiens VSX2
Branchiostoma floridae VSX
Drosophila melanogaster VSX1
Drosophila melanogaster TUP
Drosophila melanogaster VSX2
Tribolium castaneum VSX
Prostheceraeus vittatus VSX2
Prostheceraeus vittatus VSX1
Branchiostoma floridae APRD3
Macrostomum lignano VSX
Prostheceraeus vittatus DRGX
Macrostomum lignano DRGX
Schmidtea mediterranea DRGX (contig 1)
Schmidtea mediterranea DRGX (contig 2)
Tribolium castaneum AL
Drosophila melanogaster AL
Prostheceraeus vittatus ARX
Tribolium castaneum PHOX
Drosophila melanogaster PHDP
Homo sapiens PHOX2A
Branchiostoma floridae PHOX
Homo sapiens PHOX2B
Drosophila melanogaster PPH13
Tribolium castaneum PPH13
Branchiostoma floridae ARX
Macrostomum lignano ARX
Branchiostoma floridae OTP
Homo sapiens OTP
Tribolium castaneum OTP
Drosophila melanogaster OTP
Homo sapiens ARX
Branchiostoma floridae REPO
Tribolium castaneum REPO
Branchiostoma floridae MUXA
Drosophila melanogaster REPO
Tribolium castaneum HBN
Drosophila melanogaster HBN
Branchiostoma floridae APRD2
Branchiostoma floridae APRD1
Branchiostoma floridae RAX
Tribolium castaneum RX
Drosophila melanogaster RX
Homo sapiens RAX1
Homo sapiens RAX2
Branchiostoma floridae SHOX
Homo sapiens SHOX
Homo sapiens SHOX2
Tribolium castaneum SHOX
Macrostomum lignano SHOX
Drosophila melanogaster CG34367
Branchiostoma floridae UNCXB
Branchiostoma floridae UNCXC
Branchiostoma floridae UNCXA
Homo sapiens UNCX
Drosophila melanogaster ODSH
Drosophila melanogaster UNC-4
Tribolium castaneum UNC4


Branchiostoma floridae DRGX
Homo sapiens DRGX
Drosophila melanogaster CG34340
Tribolium castaneum DRGX
Tribolium castaneum CG11294
Drosophila melanogaster CG11294
Homo sapiens DPRX
Homo sapiens ARGFX
Homo sapiens CRX
Homo sapiens OTX1
Branchiostoma floridae OTX
Homo sapiens OTX2
Tribolium castaneum OC2
Tribolium castaneum OC1
Drosophila melanogaster OC
Homo sapiens RHOXF2
Homo sapiens RHOXF2B
Branchiostoma floridae PRRX
Homo sapiens PRRX2
Homo sapiens PRRX1
Macrostomum lignano PRRX
Tribolium castaneum PRRX
Drosophila melanogaster CG9876
Prostheceraeus vittatus GSC
Schmidtea polychroa GSC
Branchiostoma floridae GSC
Drosophila melanogaster GSC
Tribolium castaneum GSC
Homo sapiens GSC2
Homo sapiens TPRXL
Homo sapiens TPRX1
Branchiostoma floridae MUXA (HD3)
Homo sapiens GSC
Branchiostoma floridae AHBX1
Homo sapiens NOBOX
Branchiostoma floridae APRD6
Homo sapiens CPHX1
Homo sapiens CPHX2
Homo sapiens DUXB
Branchiostoma floridae APRD4
Homo sapiens DUXA (HD1)
Homo sapiens DUXA (HD2)
Homo sapiens LOC100506764 (HD1)
Homo sapiens DUX4L13 (HD1)
Homo sapiens DUX4 (HD1)
Homo sapiens DUX4L8 (HD1)
Homo sapiens DUX4L15 (HD2)
Homo sapiens DUX4L12 (HD1)
Homo sapiens DUX4L5 (HD1)
Homo sapiens DUX4L11 (HD1)
Homo sapiens DUX4L10 (HD1)
Homo sapiens DUX4L3 (HD1)
Homo sapiens DUX4L4 (HD1)
Homo sapiens DUX4L2 (HD1)
Homo sapiens DUX4L6 (HD1)
Homo sapiens DUX4L7 (HD1)
Homo sapiens DUX4L14 (HD1)
Homo sapiens DUX4L9 (HD1)
Homo sapiens DUXB (HD2)
Homo sapiens DUX4L12 (HD2)
Homo sapiens DUX4 (HD2)
Homo sapiens DUX4L11 (HD2)
Homo sapiens LOC100506764 (HD2)
Homo sapiens DUX4L4 (HD2)
Homo sapiens DUX4L14 (HD2)
Homo sapiens DUX4L10 (HD2)
Homo sapiens DUX4L15 (HD1)
Homo sapiens DUX4L3 (HD2)
Homo sapiens DUX4L5 (HD2)
Homo sapiens DUX4L2 (HD2)
Homo sapiens DUX4L8 (HD2)
Homo sapiens DUX4L6 (HD2)
Homo sapiens DUX4L7 (HD2)
Homo sapiens DUX4L9 (HD2)
Branchiostoma floridae MUXA (HD2)
Homo sapiens RHOXF1
Homo sapiens ESX1
Homo sapiens HESX1
Homo sapiens LEUTX
Branchiostoma floridae DLL
Drosophila melanogaster DLL
Tribolium castaneum DLL
Homo sapiens DLX4
Homo sapiens DLX5
Homo sapiens DLX2
Homo sapiens DLX3
Homo sapiens DLX1
Homo sapiens DLX6
Branchiostoma floridae ZHX
Branchiostoma floridae MSX
Homo sapiens MSX2
Homo sapiens MSX1
Drosophila melanogaster DR
Tribolium castaneum DR2
Tribolium castaneum DR1
Branchiostoma floridae MSXLX
Tribolium castaneum MSXLX
Drosophila melanogaster CG15696
Homo sapiens HOMEZ.3
Homo sapiens ZHX3 (HD5)
Homo sapiens ZHX1 (HD5)
Homo sapiens CERS3
Drosophila melanogaster LAG1
Tribolium castaneum LAG1A
Tribolium castaneum LAG1B


Branchiostoma floridae NKX1b
Branchiostoma floridae NKX1a
Homo sapiens NKX1-1
Homo sapiens NKX1-2
Drosophila melanogaster SLOU
Tribolium castaneum SLOU
Drosophila melanogaster CG11085
Tribolium castaneum BARI
Branchiostoma floridae BARI
Branchiostoma floridae BARH
Homo sapiens BARHL1
Homo sapiens BARHL2
Tribolium castaneum B-H
Drosophila melanogaster B-H1
Drosophila melanogaster B-H2
Drosophila melanogaster BCD
Homo sapiens NOTO
Branchiostoma floridae NOT
Tribolium castaneum NOT
Drosophila melanogaster CG18599
Branchiostoma floridae VAX
Prostheceraeus vittatus VAX
Schmidtea mediterranea VAX
Homo sapiens VAX1
Homo sapiens VAX2
Macrostomum lignano VAX
Tribolium castaneum BEETLEBOX1
Tribolium castaneum BEETLEBOX2
Branchiostoma floridae EMXA
Branchiostoma floridae EMXC
Branchiostoma floridae EMXB
Homo sapiens EMX1
Homo sapiens EMX2
Tribolium castaneum EMS
Drosophila melanogaster E5
Drosophila melanogaster EMS
Branchiostoma floridae HMX
Tribolium castaneum HMX
Homo sapiens HMX3
Homo sapiens HMX2
Homo sapiens HMX1
Drosophila melanogaster HMX
Branchiostoma floridae NKX7
Tribolium castaneum NK7.1
Drosophila melanogaster NK7.1
Drosophila melanogaster HGTX
Tribolium castaneum HGTX
Homo sapiens NKX6-2
Homo sapiens NKX6-3
Homo sapiens NKX6-1
Branchiostoma floridae NKX6
Drosophila melanogaster BAP
Tribolium castaneum BAP
Homo sapiens NKX3-1
Homo sapiens NKX3-2
Branchiostoma floridae NKX3
Homo sapiens NKX2-5
Homo sapiens NKX2-6
Homo sapiens NKX2-3
Branchiostoma floridae CSX
Drosophila melanogaster TIN
Tribolium castaneum TIN
Branchiostoma floridae NKX2.1
Homo sapiens NKX2-4
Homo sapiens NKX2-1
Tribolium castaneum SCRO
Drosophila melanogaster SCRO
Branchiostoma floridae NKX2.2
Homo sapiens NKX2-8
Homo sapiens NKX2-2
Tribolium castaneum VND
Drosophila melanogaster VND
Branchiostoma floridae ANKX
Homo sapiens HHEX
Branchiostoma floridae HHEX
Prostheceraeus vittatus HHEX
Tribolium castaneum HEX
Drosophila melanogaster CG7056
Homo sapiens TLX2
Homo sapiens TLX3
Homo sapiens TLX1
Branchiostoma floridae TLX
Drosophila melanogaster C15
Tribolium castaneum C15
Homo sapiens VENTX
Branchiostoma floridae HX
Branchiostoma floridae LCX
Branchiostoma floridae VENT2
Branchiostoma floridae VENT1
Homo sapiens LBX1
Branchiostoma floridae LBX
Tribolium castaneum LBX
Drosophila melanogaster LBE
Homo sapiens LBX2
Drosophila melanogaster LBL
Homo sapiens NANOG
Branchiostoma floridae BSX
Homo sapiens BSX
Tribolium castaneum BSH
Drosophila melanogaster BSH
Homo sapiens BARX2
Branchiostoma floridae BARX
Homo sapiens BARX1
Homo sapiens HLX
Branchiostoma floridae HLX
Drosophila melanogaster H2.0
Tribolium castaneum H2.0
Tribolium castaneum DBX
Drosophila melanogaster CG12361
Branchiostoma floridae DBX
Homo sapiens DBX1
Homo sapiens DBX2
Macrostomum lignano DBX
Prostheceraeus vittatus DBX
Schmidtea mediterranea DBX


Branchiostoma floridae ABOX
Tribolium castaneum ABOX
Drosophila melanogaster CG34031
Drosophila melanogaster CG13424
Tribolium castaneum NEDX
Branchiostoma floridae NEDXA
Branchiostoma floridae NEDXB
Branchiostoma floridae EN
Homo sapiens EN2
Homo sapiens EN1
Drosophila melanogaster EN
Tribolium castaneum EN
Tribolium castaneum INV
Drosophila melanogaster INV
Drosophila melanogaster UNPG
Tribolium castaneum UNPG
Homo sapiens GBX1
Homo sapiens GBX2
Branchiostoma floridae GBX
Tribolium castaneum EXEX
Homo sapiens MNX1
Drosophila melanogaster EXEX
Branchiostoma floridae MNXA
Branchiostoma floridae MNXB
Drosophila melanogaster RO
Tribolium castaneum RO
Branchiostoma floridae RO
Branchiostoma floridae EVXB
Homo sapiens EVX1
Homo sapiens EVX2
Branchiostoma floridae EVXA
Drosophila melanogaster EVE
Tribolium castaneum EVE
Branchiostoma floridae MOX
Homo sapiens MEOX2
Homo sapiens MEOX1
Drosophila melanogaster BTN
Tribolium castaneum BTN
Homo sapiens HOXB2
Homo sapiens HOXA2
Tribolium castaneum MXP
Drosophila melanogaster PB
Branchiostoma floridae HOX2
Homo sapiens HOXB1
Homo sapiens HOXD1
Homo sapiens HOXA1
Branchiostoma floridae HOX1
Tribolium castaneum LAB
Drosophila melanogaster LAB
Drosophila melanogaster ZEN2
Drosophila melanogaster ZEN
Tribolium castaneum IND
Drosophila melanogaster IND
Homo sapiens GSX2
Branchiostoma floridae GSX
Homo sapiens GSX1
Homo sapiens PDX1
Branchiostoma floridae XLOX
Tribolium castaneum ZEN1
Tribolium castaneum ZEN2
Branchiostoma floridae HOX3
Homo sapiens HOXD3
Homo sapiens HOXB3
Homo sapiens HOXA3
Branchiostoma floridae HOX4
Homo sapiens HOXB4
Homo sapiens HOXC4
Homo sapiens HOXD4
Homo sapiens HOXA4
Tribolium castaneum DFD
Drosophila melanogaster DFD
Homo sapiens HOXC5
Homo sapiens HOXB5
Homo sapiens HOXA5
Branchiostoma floridae HOX5
Tribolium castaneum CX
Drosophila melanogaster SCR
Tribolium castaneum FTZ
Drosophila melanogaster FTZ
Branchiostoma floridae CDX
Tribolium castaneum CAD2
Drosophila melanogaster CAD
Homo sapiens CDX2
Homo sapiens CDX4
Homo sapiens CDX1
Homo sapiens HOXA7
Drosophila melanogaster ANTP
Tribolium castaneum PTL
Homo sapiens HOXB7
Homo sapiens HOXA6
Homo sapiens HOXB6
Homo sapiens HOXC6
Drosophila melanogaster UBX
Tribolium castaneum UTX
Tribolium castaneum ABD-A
Drosophila melanogaster ABD-A
Branchiostoma floridae HOX7
Branchiostoma floridae HOX6
Branchiostoma floridae HOX8
Homo sapiens HOXC8
Homo sapiens HOXD8
Homo sapiens HOXB8
Tribolium castaneum ABD-B
Drosophila melanogaster ABD-B
Homo sapiens HOXC9
Homo sapiens HOXD9
Homo sapiens HOXA9
Homo sapiens HOXB9
Branchiostoma floridae HOX9
Homo sapiens HOXA10
Homo sapiens HOXD10
Homo sapiens HOXC10
Branchiostoma floridae HOX12
Branchiostoma floridae HOX11
Branchiostoma floridae HOX10
Homo sapiens HOXD11
Homo sapiens HOXA11
Homo sapiens HOXC11
Branchiostoma floridae HOX13
Branchiostoma floridae HOX14
Homo sapiens HOXD12
Homo sapiens HOXC12
Branchiostoma floridae HOX15
Homo sapiens HOXD13
Homo sapiens HOXC13
Homo sapiens HOXB13
Homo sapiens HOXA13
1.0

Supplemental Figure S8. 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. Model of protein evolution: LG+G.



Supplemental Figure S9. 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. Model of protein evolution: LG+G.

Supplemental Tables

Supplemental Table 1. Transcriptomes analyzed in this study.

Supplemental Table 2. Recovered hidden orthologs. Hidden orthologs (as in HumRef2015) recovered in each transcriptome after running Leapfrog with the transcriptome of the polyclad *P. vittatus* used as the 'bridge'.

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

Supplemental Table 4. PFAM domain composition of 130 human queries and their ortholog 'bridge' proteins in *P. vittatus*.

Supplemental Table 5. Number of hidden orthologs recovered with iterated 'bridge' transcriptomes.

Supplemental Table 6. Data set used for principal component analysis.

Supplemental Table 7. Length of hidden orthologs and ORFs in flatworm transcriptomes.

Supplemental Table 8. PFAM domains identified in the hidden orthologs.

Supplemental Table 9. Significantly enriched GO terms in the hidden orthologs recovered in *S. mediterranea*.

Supplemental References

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114-2120.

Brandl H, Moon H, Vila-Farre M, Liu SY, Henry I, Rink JC. 2016. PlanMine - a mineable resource of planarian biology and biodiversity. *Nucleic Acids Res* **44**: D764-773.

Grabherr M, Haas BJ, Yassour M, Levin J, Thompson D, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q et al. 2011. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat Biotechnol* **29**: 644-652.

Laumer CE, Hejnol A, Giribet G. 2015. Nuclear genomic signals of the 'microturbellarian' roots of platyhelminth evolutionary innovation. *Elife* **4**.

Parra G, Bradnam K, Korf I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**: 1061-1067.

Parra G, Bradnam K, Ning Z, Keane T, Korf I. 2009. Assessing the gene space in draft genomes. *Nucleic Acids Res* **37**: 289-297.

R Core Team. 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rieger R, Gehlen M, Haszprunar G, Holmlund M, Legniti A, Salvenmoser W, Tyler S. 1988. Laboratory cultures of marine Macrostomida (Turbellaria). *Progr Zool* **36**: 523.

Smith-Unna RD, Boursnell C, Patro R, Hibberd JM, Kelly S. 2015. TransRate: reference free quality assessment of de-novo transcriptome assemblies. *BioRxiv* **021626**.

Wickham H. 2009. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York.