Supplementary Tables for

Near Intron Pairs and the Metazoan Tree.

Jörg Lehmann∗, Peter F. Stadler, and Veiko Krauss

∗joe@bioinf.uni-leipzig.de

Taxon name	Abbrev.	Ortholog prediction method	Genome version	Annotation/Gene build
Ixodes scapularis	Isca			
Trichoplax adhaerens	Tadh			
Nematostella vectensis	Nvec			
Strongylocentrotus purpuratus	Spur			
Schistosoma mansoni	Sman			
Acyrthosiphon pisum	Apis			
Drosophila melanogaster	Dmel			
Caenorhabditis elegans	Cele			
Aedes aegypti	Aaeg			
Amphimedon queenslandica	Aque			
Anopheles gambiae	Agam		from Ensembl Metazoa release 14 (May 2012)	
Apis mellifera	Amel			
Bombyx mori	Bmor			
Caenorhabditis brenneri	Cbre			
Caenorhabditis briggsae	Cbri	Ensembl Compara DB		
Caenorhabditis japonica	Cjap			
Caenorhabditis remanei	Crem			
Culex quinquefasciatus	Cqui			
Daphnia pulex	Dpul			
Pediculus humanus	Phum			
Pristionchus pacificus	Ppac			
Tribolium castaneum	Tcas			
Trichinella spiralis	Tspi			
Ciona intestinalis	Cint			
Ciona savignyi	Csav			
Danio rerio	Drer		from Ensembl Core release 67 (May 2012)	
Homo sapiens	Hsap			
Monodelphis domestica	Mdom			
Takifugu rubripes	Trub			
Branchiostoma floridae	Bflo		JGI, Brafl2	JGI, Brafl1_mapped_to_Brafl2
Brugia malayi	Bmal		WORMBASE, WS231	WORMBASE, WS231
Capitella teleta	Ctel		JGI, Capca1	JGI, FilteredModels v1.0
Coprinopsis cinerea	Ccin		BROAD	transcripts.gtf
Dictyostelium purpureum	Dpur		JGI, Dicpu1	Dicpu1_best_genes
Helobdella robusta	Hrob		JGI, Helro1	FilteredModels v3
Heterorhabditis bacteriophora	Hbac	targeted ortholog search with HaMStR	WORMBASE, WS231	WORMBASE, WS231
Lottia gigantea	Lgig		JGI, Lotgi1	FilteredModels v1
Meloidogyne hapla	Mhap		WORMBASE, WS231	WORMBASE, WS231
Meloidogyne incognita	Minc		WORMBASE, WS232	WORMBASE, WS232; mapping via Scipio [1]
Monosiga brevicollis	Mbre		JGI, Monbr1	Monbr1_best_models
Nasonia vitripennis	Nvit		NCBI RefSeq (based on Nvit1.0)	RefSeq of nasonia_automated_gene_model_v1
Schistosoma japonicum	Sjap		CHGC GeneDB v3	CHGC GeneDB_Sjaponicum.v3
Aplysia californica	Acal		NCBI Aplcal2.0	
Hydra magnipapillata	Hmag		NCBI Hydra_RP_1.0	
Mnemiopsis leidyi	Mlei	tblastn	NCBI mmle_1.0	
Rhodnius prolixus	Rpro		GCWU Rpro-1.0.1	
Saccoglossus kowalevskii	Skow		NCBI Skow1.1 (unplaced scaffolds)	
Schmidtea mediterranea	Smed		GCWU Smed-3.1	

Table S1: Species data sources and abbreviations, and used method for ortholog predictions.

Target taxon	Reference species for HaMStR-search
Bflo	Spur,Csav,Drer,Hsap,Dmel,Cele,Tadh
Bmal	Ppac,Cele,Tspi,Dmel,Sman,Spur
Ctel	Sman,Dmel,Cele,Tadh,Nvec,Spur
Ccin	Aque,Tadh,Nvec,Dmel,Cele,Sman,Spur
Dpur	Aque,Tadh,Nvec,Dmel,Cele,Sman
Hrob	Sman,Dmel,Cele,Tadh,Nvec,Spur
Hbac	Ppac,Cele,Tspi,Dmel,Sman,Spur
Lgig	Sman,Dmel,Cele,Tadh,Nvec,Spur
Mhap	Ppac,Cele,Tspi,Dmel,Sman,Spur
Minc	Ppac,Cele,Tspi,Dmel,Sman,Spur
Mbre	Aque,Tadh,Nvec,Dmel,Cele,Sman,Spur
Nvit	Amel,Tcas,Dmel,Apis,Cele,Tadh,Spur
Sjap	Sman,Dmel,Cele,Tadh,Nvec,Spur

Table S2: Subset of reference species (ranked list) used for the targeted ortholog search with HaMStR.

Target taxon	Ref.species for BLAST-search	#final candidates/groups (#targets)	#refined CDS
Acal	Sman,Cele,Dmel,Tadh	885 (904)	9
Hmag	Nvec,Tadh,Dmel,Cele	1,988 (2,027)	1,179
Mlei	Tadh,Nvec,Dmel,Cele	1,847 (1,871)	4
Rpro	Apis,Isca,Dmel,Cele	2,900 (2,921)	32
Skow	Spur,Csav,Dmel,Cele	1,990 (2,007)	1,435
Smed	Sman,Cele,Dmel,Tadh	1,730 (2,023)	23

Table S3: Identified ortholog candidates for BLAST-based taxa. Reference species (ranked order) used for the tblastn/blastp searches are shown. The total numbers of identified top-scoring genes (1 per group) fulfilling the (reciprocity) constraints are shown (with the number in brackets including also secondary hits), as well as the numbers of cases with refined CDS annotation by spliced-alignment with a mapped target protein.

References

[1] K. Hatje, O. Keller, B. Hammesfahr, H. Pillmann, S. Waack, M. Kollmar, Cross-species protein sequence and gene structure prediction with fine-tuned Webscipio 2.0 and Scipio., BMC research notes 4 (1) (2011) 265. doi:10.1186/1756-0500-4-265.
URL http://www.ncbi.nlm.nih.gov/pubmed/21798037