

Supplementary material to:

**Nuclear genome annotation of wheel animals and thorny-headed worms –
inferences about the last common ancestor of Syndermata (Rotifera s.l.)**

Laura Hagemann^{1,*}, Katharina M. Mauer^{1,*}, Thomas Hankeln², Hanno Schmidt¹, Holger Herlyn^{1,§}

¹Institute of Organismic and Molecular Evolution (iomE), Anthropology, Johannes Gutenberg University Mainz, Mainz, Germany

²Institute of Organismic and Molecular Evolution (iomE), Molecular Genetics and Genome Analysis, Johannes Gutenberg University Mainz, Mainz, Germany

*These authors contributed equally.

§Author of correspondence

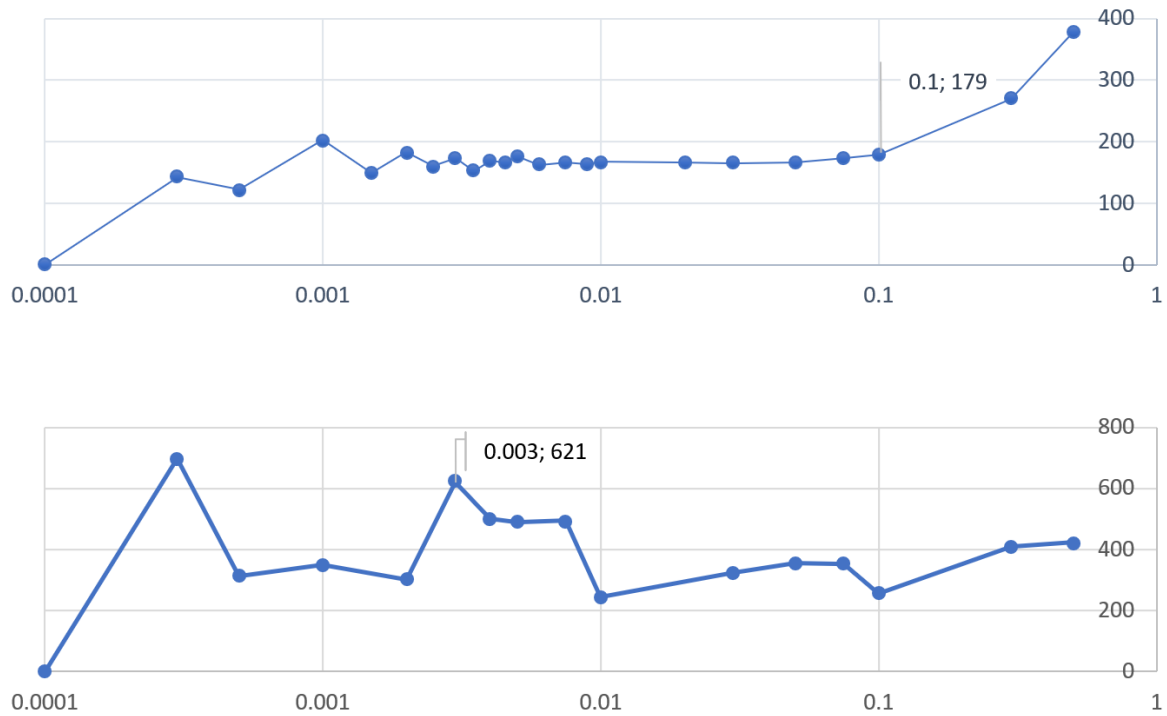
S1 Determination of potential bacterial contamination

To detect potential bacterial contamination, we compared the genome assemblies of *P. laevis* and *S. nebaliae* (Mauer et al., 2020, 2021) to a database of all complete bacterial genome assemblies available at NCBI (as of 2021-06-30) using BLASTN v2.5.0+ (Camacho et al., 2009). If a contig showed a match with an E-value $\leq 1e-20$, all open reading frames with at least 90 nucleotides were determined using the program getorf (Emboss: <https://www.bioinformatics.nl/cgi-bin/emboss/getorf>) and blasted against the SwissProt database (release 2021_03; 1e-10 BLASTP). We kept contigs if > 50% of the ORFs matched best to a eukaryotic SwissProt sequence and/or matches referred to rRNA sequences for which evolutionary conservation across bacteria and eukaryotes appeared possible. Otherwise, a contig was deleted from the assembly.

S2 Generation of a custom repeat database

To create custom repeat databases for *A. vaga* and *B. plicatilis*, we used genomic paired-end reads: SRR6047135 (83,492,774 reads) for *B. plicatilis* (Han et al., 2019) and SRR801068 (32,879,175 reads) as well as SRR801084 (85,564,542 reads) for *A. vaga* (Flot et al., 2013). Reads were checked and processed using fastqc v0.11.9 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and Trimmomatic (Bolger et al., 2014) (ILLUMINACLIP:Adaptor.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:20, HEADCROP:X MINLEN:20; X: Bplic = 7 Avaga 440 bp =11; Avaga 862 bp =11). We removed mitochondrial reads from the datasets with bbduk v38.73 (hdist=1, k=31; <https://sourceforge.net/projects/bbmap/>), utilizing mitogenome sequences AP009408.1 and AP009407.1 (Suga et al., 2008) as template for *B. plicatilis*. For *A. vaga*, we newly assembled the mitogenome employing NOVOPlasty v4.3 (Dierckxsens et al., 2017), starting from the processed dataset SRR801068 and using the published *cox1* sequence JX184001.1 (Lasek-Nesselquist, 2012) as seed. Via the MITOS web server (Bernt et al., 2013), we confirmed that all expected protein-coding genes were included in the resulting 14,033 bp long circular assembly. Processed and filtered reads (80,201,171 in *B. plicatilis*; 32,364,142 and 80,038,348 in *A. vaga*) were starting points for a dnaPipeTE (Goubert et al., 2015) pipeline (for details, see (Schell et al., 2017; Mauer et al., 2020), assuming an assembly size of 217 Mb for *A. vaga* (Flot et al., 2013) and 107 Mb for *B. plicatilis* (Han et al., 2019). To determine the most appropriate coverage for repetitive regions, parallel runs of dnaPipeTE were conducted with corresponding values set from 0.0001 to 0.5, whereby minimum contig size was fixed at 50 bp. By doing so, 0.1 was established as appropriate coverage for *A. vaga* and 0.003 for *B. plicatilis* as these values marked a steady increase or a local maximum in N50 (Supplementary Figure S1).

Taking above coverage values and a minimum contig length of 200 bp, subsequent dnaPipeTE runs (N = 50, each) resulted in 116,772 assembled contigs for *A. vago* and 877 for *B. plicatilis*. These were filtered for duplicates (100% ID, longest sequence was retained) using CD-HIT-EST (Huang et al., 2010) and randomly generated sequences (here defined as contigs showing a coverage of less than three times the average genome contig coverage). Coverage was determined by aligning processed DNA reads to dnaPipeTE contigs with Bowtie 2 (<http://usegalaxy.eu>; Langmead & Salzberg, 2012). In contigs for *B. plicatilis*, average coverage per single base position was 195.4. The respective value was 16.7 for contigs in *A. vago*, whereby contigs larger than 1,000 bp had a higher average per-base coverage value of 54.37. To avoid masking of paralogous genes, protein-coding sequences not belonging to transposons were removed from both repeat databases using NCBI BLASTX + v2.5.0+ (Camacho et al., 2009). We generated additional repeat databases with RepeatModeler v2.0.1 (Smit & Hubley, 2008) based on the genome assemblies GCA_010279815.1 (*B. plicatilis*; Han et al., 2019) and GCA_000513175.1 (*A. vago*; Flot et al., 2013) using default settings. In a final step, we compared our custom databases with the SwissProt database (release 2020_03; 562,878 sequences) via BLASTX. Contigs with matches were excluded (E-value: 1e-05) if the matching sequences in SwissProt lacked matches in the RepBaseRepeatMaskerEdition-20181026 database (TBLASTN; E-value: 1e-05).



Supplementary Figure S1. Determination of appropriate coverage values. The graphs show the set coverage values on the x-axis and the corresponding N50 values as determined by dnaPipeTE on the y-axis *A. vago* (top) and *B. plicatilis* (bottom).

Supplementary Table S1. Annotation of repetitive regions

Repeat	<i>B. plicatilis</i>			<i>A. vaga</i>			<i>S. nebaliae</i>			<i>P. laevis</i>		
classification	Count	bp	%	Count	bp	%	Count	bp	%	Count	bp	%
DNA	13,861	2,911,656	2.72	6504	1,695,485	1.68	4916	1,023,755	2.33	60,642	7,650,096	3.02
LINE	16,911	4,584,756	4.29	21,341	4,601,032	4.56	7103	1,934,130	4.41	219,850	93,689,452	36.99
LTR	1880	832,831	0.78	7779	1,700,854	1.69	2366	973,611	2.22	49,406	14,356,601	5.67
RC	331	172,316	0.16	565	110,397	0.11	15	2221	0.01	5	653	0.00
Retro	188	38,563	0.04	8472	1,548,906	1.54	986	177,364	0.40	53,304	5,260,604	2.08
SINE	304	98,368	0.09	798	101,692	0.10	152	12,961	0.03	11,215	971,932	0.38
Unknown	119,030	20,405,738	19.08	18,699	4,409,336	4.37	10,112	1,627,665	3.71	142,376	36,827,974	14.54
Low complexity	18,052	905,564	0.85	7,250	357,877	0.35	3728	180,264	0.41	12,609	738,995	0.29
Simple repeat	40,161	1,746,978	1.63	39,400	1,487,254	1.48	21,613	959,956	2.19	62,014	3,996,784	1.58
Satellite				14	4348	0.00						
rRNA	2	213	0.00	41	41,537	0.04	63	14,377	0.03	95	380,324	0.15
tRNA	560	47,553	0.04	84	7946	0.01	368	34,492	0.08			
Total	211,280	31,744,536	29.68	110,947	16,066,664	15.93	51,422	6,940,796	15.82	611,516	163,873,415	64.7

Totals have been adjusted if required.

Supplementary Table S2. OrthoVenn2 clustering results

	Cluster Count	Protein Count			
		<i>B. plicatilis</i>	<i>A. vaga</i>	<i>S. nebaliae</i>	<i>P. laevis</i>
	2114	2805	3735	2284	2567
	1403	1922	2589	1521	
	480	670	796		575
	132		273	159	239
	57	136		73	62
	1374	1862	2659		
	212		524		261
	206		566	220	
	187	501		221	
	84	143			106
	41			53	170
	3356		11,778		
	878	3918			
	811				3740
	534			1791	
total	11,869	11,957	22,920	6322	7720
total orthogroups	6290	8039	11,142	4531	3980
total paralogous groups	5579	3918	11,778	1791	3740
not clustered		4843	6812	5180	4353
species present in orthogroups [%]		90.6	94.1	65.8	49.6
sequences per core orthogroup (mean)		1.33	1.77	1.08	1.21
genes in core orthogroups [%]		16.7	12.6	19.9	21.3
genes in other orthogroups [%]		31.2	24.9	19.5	11.7
genes in other paralogues groups [%]		23.3	39.6	15.6	31.0
genes not clustered [%]		28.8	22.9	45.0	36.1

Supplementary Table S3. Gene ontologies of 194 clusters containing syndermatan proteins and lacking homologues in platyhelminths

ID	Protein Count	Swiss-Prot Hit	Gene Ontology (GO) Annotation		
cluster156	34	Q96SG3	GO:0001967	suckling behavior	Ensembl
cluster385	23	Q08180	GO:0046666	retinal cell programmed cell death	FlyBase
cluster430	22	Q60HG0	GO:0035269	protein O-linked mannosylation	UniProtKB
cluster449	22	Q9VUL9	GO:0006486	protein glycosylation	UniProtKB
cluster503	21	Q8BMZ5	GO:0000379	tRNA-type intron splice site recognition and cleavage	GO_Central
cluster1147	15	Q06561	GO:0060298	positive regulation of sarcomere organization	UniProtKB
cluster1624	13	Q3UQ28	GO:0006979	response to oxidative stress	InterPro
cluster1940	12	P10895	GO:0042127	regulation of cell proliferation	Ensembl
cluster1985	12	Q9Y2J7	GO:0007275	multicellular organism development	UniProtKB-KW
cluster2269	11	Q9NBX4	GO:0006313	transposition, DNA-mediated	UniProtKB
cluster2684	10	P24844	GO:0006937	regulation of muscle contraction	ProtInc
cluster2692	10	Q90419	GO:0055002	striated muscle cell development	ZFIN
cluster2694	10	Q28727	GO:0009617	response to bacterium	Ensembl
cluster2771	10	P70496	GO:0043434	response to peptide hormone	RGD
cluster2986	10	Q6DDL7	GO:0016021	integral component of membrane	UniProtKB-KW
cluster3323	9	P54107	GO:0060046	regulation of acrosome reaction	Ensembl
cluster3331	9	Q08174	GO:0007399	nervous system development	ProtInc
cluster4257	8	Q9DE67	GO:0007601	visual perception	InterPro
cluster4405	8	P58353	GO:0009750	response to fructose	UniProtKB
cluster5152	7	Q9NJG9	GO:0022008	neurogenesis	FlyBase
cluster5223	7	Q69KL9	GO:0000381	regulation of alternative mRNA splicing, via spliceosome	GO_Central
cluster5236	7	Q16534	GO:0006366	transcription by RNA polymerase II	ProtInc
cluster5292	7	Q9DFG8	GO:0016567	protein ubiquitination	UniProtKB- UniPathway
cluster5305	7	P46939	GO:0014894	response to denervation involved in regulation of muscle adaptation	Ensembl
cluster5345	7	Q4PM64	GO:0006412	translation	InterPro
cluster5367	7	O54922	GO:0015031	protein transport	UniProtKB-KW
cluster5427	7	Q5JPB2	GO:0003676	nucleic acid binding	InterPro
cluster5494	7	Q0VFX8	GO:0046872	metal ion binding	UniProtKB-KW
cluster6236	6	P34410	GO:0071805	potassium ion transmembrane transport	GO_Central
cluster6244	6	Q09441	GO:0006355	regulation of transcription, DNA-templated	InterPro
cluster6293	6	Q5RJK3	GO:0000049	tRNA binding	GO_Central
cluster6311	6	Q07139	GO:0035023	regulation of Rho protein signal transduction	InterPro
cluster6339	6	Q8R420	GO:0051384	response to glucocorticoid	Ensembl
cluster6341	6	P73481	GO:0003824	catalytic activity	InterPro

cluster6342	6	Q94547	GO:0006351	transcription, DNA-templated	UniProtKB
cluster6354	6	Q8IZS8	GO:0034765	regulation of ion transmembrane transport	UniProtKB-KW
cluster6356	6	P29294	GO:0004687	myosin light chain kinase activity	UniProtKB-EC
cluster6381	6	Q60SZ2	GO:0005743	mitochondrial inner membrane	UniProtKB-SubCell
cluster6429	6	Q7Z418	GO:0006813	potassium ion transport	UniProtKB
cluster6441	6	P29119	GO:1901394	positive regulation of transforming growth factor beta1 activation	UniProtKB
cluster6497	6	N/A	N/A		
cluster6521	6	Q14517	GO:0007156	homophilic cell adhesion via plasma membrane adhesion molecules	InterPro
cluster6524	6	Q2R1Z5	GO:0019722	calcium-mediated signaling	GO_Central
cluster6551	6	N/A	N/A		
cluster6576	6	Q962X7	GO:0045047	protein targeting to ER	UniProtKB
cluster6590	6	N/A	N/A		
cluster6592	6	A2VE04	GO:0016491	oxidoreductase activity	UniProtKB-KW
cluster6623	6	Q2KI69	GO:1901379	regulation of potassium ion transmembrane transport	Ensembl
cluster6628	6	O94632	GO:1903257	selenoneine biosynthetic process	PomBase
cluster7539	5	Q9NZ63	GO:0048024	regulation of mRNA splicing, via spliceosome	GO_Central
cluster7546	5	Q8CWR6	GO:0046467	membrane lipid biosynthetic process	UniProtKB
cluster7547	5	Q2KII4	GO:0006511	ubiquitin-dependent protein catabolic process	GO_Central
cluster7550	5	O46504	GO:0016491	oxidoreductase activity	GO_Central
cluster7570	5	Q505F5	GO:0003723	RNA binding	InterPro
cluster7573	5	Q8CIB6	GO:0048489	synaptic vesicle transport	UniProtKB
cluster7575	5	Q9GP16	GO:0006412	translation	InterPro
cluster7577	5	Q9DED6	GO:0009888	tissue development	GO_Central
cluster7578	5	Q09221	N/A		
cluster7580	5	Q9UI14	GO:0008022	protein C-terminus binding	Ensembl
cluster7586	5	Q8IVI9	GO:0050999	regulation of nitric-oxide synthase activity	Reactome
cluster7588	5	Q9U3F4	GO:0000003	reproduction	WormBase
cluster7590	5	Q99J45	GO:0035556	intracellular signal transduction	GO_Central
cluster7597	5	Q9VXX8	GO:0002181	cytoplasmic translation	FlyBase
cluster7601	5	Q5JUQ0	N/A		
cluster7604	5	Q5CZ37	GO:0008654	phospholipid biosynthetic process	GO_Central
cluster7605	5	P31696	GO:0071340	skeletal muscle acetylcholine-gated channel clustering	AgBase
cluster7615	5	Q9VN03	GO:0006401	RNA catabolic process	GO_Central
cluster7617	5	P23499	GO:0007165	signal transduction	InterPro
cluster7621	5	N/A	N/A		
cluster7622	5	A7MBE0	GO:0015101	organic cation transmembrane transporter activity	Ensembl

cluster7624	5	Q9GM70	GO:2000377	regulation of reactive oxygen species metabolic process	UniProtKB
cluster7627	5	Q6DDT1	GO:0008654	phospholipid biosynthetic process	UniProtKB-KW
cluster7629	5	Q9VZF4	GO:0007419	ventral cord development	FlyBase
cluster7630	5	Q9UBT7	GO:0007266	Rho protein signal transduction	MGI
cluster7632	5	Q07998	GO:0016055	Wnt signaling pathway	UniProtKB-KW
cluster7641	5	Q9Y696	GO:0007035	vacuolar acidification	Ensembl
cluster7654	5	O61397	GO:0006486	protein glycosylation	UniProtKB- UniPathway
cluster7659	5	N/A	N/A		
cluster7664	5	Q54XX5	GO:0004674	protein serine/threonine kinase activity	GO_Central
cluster7677	5	P06581	GO:0009408	response to heat	WormBase
cluster7687	5	Q17QJ5	GO:0051604	protein maturation	Ensembl
cluster7693	5	Q5SUE8	N/A		
cluster7702	5	N/A	N/A		
cluster7708	5	N/A	N/A		
cluster7709	5	Q4H4B6	GO:0001944	vasculature development	ZFIN
cluster7715	5	Q09JK2	N/A		
cluster7718	5	P23731	GO:0005198	structural molecule activity	InterPro
cluster7720	5	Q9BW72	GO:0055114	oxidation-reduction process	UniProtKB-KW
cluster7724	5	Q8K2D0	N/A		
cluster7726	5	Q9M047	GO:0051017	actin filament bundle assembly	TAIR
cluster7729	5	Q21735	GO:0006606	protein import into nucleus	GO_Central
cluster7734	5	Q8NBT9	GO:0007155	cell adhesion	UniProtKB-KW
cluster7736	5	Q5ZJ66	GO:0006424	glutamyl-tRNA aminoacylation	GO_Central
cluster7738	5	Q6DEI8	GO:0022417	protein maturation by protein folding	UniProtKB
cluster7739	5	Q9NZJ4	GO:0006457	protein folding	UniProtKB
cluster7742	5	Q5XIY2	GO:0016021	integral component of membrane	UniProtKB-KW
cluster7749	5	P28048	GO:0006396	RNA processing	InterPro
cluster7752	5	Q9UPV9	GO:0047496	vesicle transport along microtubule	GO_Central
cluster7756	5	Q9W4L1	GO:0006412	translation	FlyBase
cluster7760	5	P61603	GO:0051085	chaperone cofactor-dependent protein refolding	GO_Central
cluster7765	5	Q9H902	GO:0032386	regulation of intracellular transport	Ensembl
cluster7769	5	O88667	GO:0007165	signal transduction	InterPro
cluster7770	5	Q8AVK1	GO:0072321	chaperone-mediated protein transport	InterPro
cluster7785	5	O54829	GO:0060078	regulation of postsynaptic membrane potential	SynGO
cluster7787	5	P38404	GO:0007188	adenylate cyclase-modulating G protein-coupled receptor signaling pathway	InterPro
cluster7796	5	Q24799	N/A		
cluster7808	5	N/A	N/A		

cluster7812	5	Q58DA5	GO:0007155	cell adhesion	UniProtKB-KW
cluster7813	5	D0PRN3	GO:0007155	cell adhesion	UniProtKB-KW
cluster7816	5	Q4VBH4	GO:0006515	protein quality control for misfolded or incompletely synthesized proteins	UniProtKB
cluster7823	5	Q20591	GO:0055085	transmembrane transport	GO_Central
cluster7825	5	Q6NVE9	GO:0004721	phosphoprotein phosphatase activity	UniProtKB-KW
cluster7827	5	P60467	GO:0030433	ubiquitin-dependent ERAD pathway	Ensembl
cluster7829	5	P06601	GO:0006366	transcription by RNA polymerase II	UniProtKB
cluster7835	5	P47236	GO:0061056	sclerome development	AgBase
cluster7845	5	P07311	GO:0006796	phosphate-containing compound metabolic process	ProtInc
cluster7849	5	Q9UJZ5	GO:0009966	regulation of signal transduction	UniProtKB
cluster7854	5	Q3IHW0	GO:0019518	L-threonine catabolic process to glycine	UniProtKB- UniPathway
cluster7861	5	Q5ZK47	GO:0031098	stress-activated protein kinase signaling cascade	GO_Central
cluster7864	5	Q86X52	GO:0051923	sulfation	Ensembl
cluster7867	5	Q5TYP8	GO:0016021	integral component of membrane	UniProtKB-KW
cluster7868	5	O60285	GO:1901796	regulation of signal transduction by p53 class mediator	Reactome
cluster7878	5	Q5E938	GO:0003743	translation initiation factor activity	GO_Central
cluster7880	5	Q3T134	GO:0006465	signal peptide processing	GO_Central
cluster7882	5	N/A	N/A		
cluster7886	5	Q9VTJ8	GO:0030150	protein import into mitochondrial matrix	UniProtKB
cluster7888	5	O35796	GO:0008380	RNA splicing	UniProtKB-KW
cluster7895	5	A0A1D8P NZ7	GO:0046475	glycerophospholipid catabolic process	CGD
cluster7903	5	Q62824	GO:0090522	vesicle tethering involved in exocytosis	InterPro
cluster7910	5	Q8LPN7	GO:0016567	protein ubiquitination	GO_Central
cluster7918	5	N/A	N/A		
cluster7922	5	Q8BXJ2	GO:0045893	positive regulation of transcription, DNA-templated	MGI
cluster7926	5	Q09261	GO:0003735	structural constituent of ribosome	InterPro
cluster7929	5	O94964	GO:0010506	regulation of autophagy	UniProtKB
cluster7936	5	G5EBQ8	GO:0061063	positive regulation of nematode larval development	UniProtKB
cluster7940	5	Q28DE7	GO:0032874	positive regulation of stress-activated MAPK cascade	GO_Central
cluster7942	5	Q9Y3D7	GO:0030150	protein import into mitochondrial matrix	GO_Central
cluster7944	5	Q4R4Q3	GO:0042127	regulation of cell proliferation	InterPro
cluster7946	5	A0A1B0G TI8	GO:0016021	integral component of membrane	UniProtKB-KW
cluster9368	4	Q92888	GO:0007266	Rho protein signal transduction	ProtInc

cluster9393	4	Q3V1V3	GO:0006364	rRNA processing	InterPro
cluster9396	4	Q58FA4	GO:0060707	trophoblast giant cell differentiation	UniProtKB
cluster9397	4	Q8IY18	GO:0000722	telomere maintenance via recombination	UniProtKB
cluster9402	4	Q9V998	GO:0034126	positive regulation of MyD88-dependent toll-like receptor signaling pathway	FlyBase
cluster9409	4	O95299	GO:0032981	mitochondrial respiratory chain complex I assembly	UniProtKB
cluster9414	4	Q9NYG5	GO:0006511	ubiquitin-dependent protein catabolic process	GO_Central
cluster9420	4	Q4KMD7	GO:0016180	snRNA processing	UniProtKB
cluster9445	4	Q6AXT0	GO:0006412	translation	InterPro
cluster9448	4	Q9Y3B1	GO:0015914	phospholipid transport	GO_Central
cluster9485	4	Q5T160	GO:0006418	tRNA aminoacylation for protein translation	Reactome
cluster9488	4	Q9V9W8	GO:0007472	wing disc morphogenesis	FlyBase
cluster9502	4	Q9IA12	GO:0007275	multicellular organism development	UniProtKB-KW
cluster9509	4	Q9UQ90	GO:0046902	regulation of mitochondrial membrane permeability	UniProtKB
cluster9516	4	O86938	GO:0032923	organic phosphonate biosynthetic process	InterPro
cluster9524	4	Q9VLJ9	GO:0006412	translation	UniProtKB
cluster9531	4	Q923Z3	GO:0002098	tRNA wobble uridine modification	MGI
cluster9545	4	Q5VUM1	GO:0003407	neural retina development	UniProtKB
cluster9550	4	Q6JEL2	GO:0007286	spermatid development	Ensembl
cluster9554	4	Q02365	GO:0032981	mitochondrial respiratory chain complex I assembly	UniProtKB
cluster9555	4	N/A	N/A		
cluster9563	4	P56839	GO:0032923	organic phosphonate biosynthetic process	UniProtKB- UniPathway
cluster9569	4	Q5XJP1	GO:0016567	protein ubiquitination	GO_Central
cluster9572	4	Q2HJH0	GO:0009617	response to bacterium	Ensembl
cluster9584	4	Q8TC07	GO:0043087	regulation of GTPase activity	UniProtKB
cluster9585	4	Q9SUM2	GO:0000387	spliceosomal snRNP assembly	InterPro
cluster9593	4	Q55EA2	GO:0000727	double-strand break repair via break-induced replication	GO_Central
cluster9596	4	Q3SZW1	GO:0007286	spermatid development	UniProtKB
cluster9606	4	Q3ZCE0	GO:0016070	RNA metabolic process	GO_Central
cluster9608	4	Q9SR52	GO:0043419	urea catabolic process	UniProtKB- UniPathway
cluster9620	4	P42029	GO:0032981	mitochondrial respiratory chain complex I assembly	UniProtKB
cluster9624	4	Q9GL59	GO:0017187	peptidyl-glutamic acid carboxylation	InterPro
cluster9633	4	Q96RW7	GO:0007601	visual perception	UniProtKB-KW
cluster9661	4	Q6CSP8	GO:0034462	small-subunit processome assembly	EnsemblFungi

cluster9664	4	Q9CV94	GO:0000722	telomere maintenance via recombination	UniProtKB
cluster9665	4	Q2WBH0	GO:0006400	tRNA modification	UniProtKB-UniRule
cluster9675	4	Q80YT9	GO:0008380	RNA splicing	UniProtKB-KW
cluster9692	4	P61282	GO:0006357	regulation of transcription by RNA polymerase II	AgBase
cluster9697	4	Q9VMU8	GO:0072334	UDP-galactose transmembrane transport	FlyBase
cluster9703	4	Q56K13	GO:0000398	mRNA splicing, via spliceosome	UniProtKB
cluster9712	4	Q32N55	GO:0006417	regulation of translation	UniProtKB-KW
cluster9722	4	Q5TM50	GO:0006363	termination of RNA polymerase I transcription	GO_Central
cluster9723	4	Q5REH5	GO:0005762	mitochondrial large ribosomal subunit	UniProtKB
cluster9724	4	Q8TCC3	GO:0070126	mitochondrial translational termination	Reactome
cluster9725	4	N/A	N/A		
cluster9748	4	N/A	N/A		
cluster9749	4	A9X1A9	GO:0006412	translation	InterPro
cluster9754	4	O44424	GO:0007399	nervous system development	UniProtKB
cluster9768	4	Q9H4P2	GO:0042797	tRNA transcription by RNA polymerase III	IC:HGNC
cluster9769	4	Q3SZM1	GO:0003723	RNA binding	UniProtKB-KW
cluster9777	4	Q5XTS1	GO:0016042	lipid catabolic process	UniProtKB-KW
cluster9780	4	Q68EV6	GO:0003735	structural constituent of ribosome	InterPro
cluster9782	4	P93471	GO:0009585	red, far-red light phototransduction	UniProtKB-KW
cluster9788	4	A4L691	GO:0007004	telomere maintenance via telomerase	Ensembl
cluster9793	4	Q9V9Z1	GO:0006412	translation	FlyBase

GOs refer to the largest clusters of syndermatan proteins (*B. plicatilis*, *A. vaga*, *S. nebaliae*, *P. laevis*), for which OrthoVenn2 did not find homologues in two platyhelminths (*M. lignano*, *S. mediterranea*).

References

- Bernt, M., A. Donath, F. Jühling, F. Externbrink, C. Florentz, G. Fritzsche, J. Pütz, M. Middendorf, & P. F. Stadler, 2013. MITOS: Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution* 69: 313–319. <https://doi.org/10.1016/j.ympev.2012.08.023>
- Bolger, A. M., M. Lohse, & B. Usadel, 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, & T. L. Madden, 2009. BLAST+: Architecture and applications. *BMC Bioinformatics* 10:121. <https://doi.org/10.1186/1471-2105-10-421>
- Dierckxsens, N., P. Mardulyn, & G. Smits, 2017. NOVOPlasty: *De novo* assembly of organelle genomes from whole genome data. *Nucleic Acids Research* 45: e18. <https://doi.org/10.1093/nar/gkw955>
- Flot, J. F., B. Hespeels, X. Li, B. Noel, I. Arkhipova, E. G. J. Danchin, A. Hejnol, et al. 2013. Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta vaga*. *Nature* 500: 453–457. <https://doi.org/10.1038/nature12326>
- Goubert, C., L. Modolo, C. Vieira, C. ValienteMoro, P. Mavingui, & M. Boulesteix, 2015. *De novo* assembly and annotation of the Asian tiger mosquito (*Aedes albopictus*) repeatome with dnaPipeTE from raw genomic reads and comparative analysis with the yellow fever mosquito (*Aedes aegypti*). *Genome Biology and Evolution* 7: 1192–1205. <https://doi.org/10.1093/gbe/evv050>
- Han, J., J. C. Park, B.-S. Choi, M.-S. Kim, H.-S. Kim, A. Hagiwara, H. G. Park, B.-Y. Lee, & J.-S. Lee, 2019. The genome of the marine monogonont rotifer *Brachionus plicatilis*: Genome-wide expression profiles of 28 cytochrome P450 genes in response to chlorpyrifos and 2-ethyl-phenanthrene. *Aquatic Toxicology* 214: 105230. <https://doi.org/10.1016/j.aquatox.2019.105230>

- Huang, Y., B. Niu, Y. Gao, L. Fu, & W. Li, 2010. CD-HIT Suite: A web server for clustering and comparing biological sequences. *Bioinformatics* 26: 680–682. <https://doi.org/10.1093/bioinformatics/btq003>
- Langmead, B., & S. L. Salzberg, 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9: 357–359. <https://doi.org/10.1038/nmeth.1923>
- Lasek-Nesselquist, E., 2012. A mitogenomic re-evaluation of the bdelloid phylogeny and relationships among the Syndermata. *PLoS ONE* 7: e43554. <https://doi.org/10.1371/journal.pone.0043554>
- Mauer, K., S. L. Hellmann, M. Groth, A. C. Fröbius, H. Zischler, T. Hankeln, & H. Herlyn, 2020. The genome, transcriptome, and proteome of the fish parasite *Pomphorhynchus laevis* (Acanthocephala). *PLoS ONE* 15: 1–30. <https://doi.org/10.1371/journal.pone.0232973>
- Mauer, K. M., H. Schmidt, M. Dittrich, A. C. Fröbius, S. L. Hellmann, H. Zischler, T. Hankeln, & H. Herlyn, 2021. Genomics and transcriptomics of epizoid Seisonidea (Rotifera, syn. Syndermata) reveal strain formation and gradual gene loss with growing ties to the host. *BMC Genomics* 22: 1–24. <https://doi.org/10.1186/s12864-021-07857-y>
- Schell, T., B. Feldmeyer, H. Schmidt, B. Greshake, O. Tills, M. Truebano, S. D. Rundle, J. Paule, I. Ebersberger, & M. Pfenninger, 2017. An annotated draft genome for *Radix auricularia* (Gastropoda, Mollusca). *Genome Biology and Evolution* 9: 585–592. <https://doi.org/10.1093/gbe/evx032>
- Smit, A., & R. Hubley, 2008. RepeatModeler Open-1.0. 2008-2015.
- Suga, K., D. B. Mark Welch, Y. Tanaka, Y. Sakakura, & A. Hagiwara, 2008. Two circular chromosomes of unequal copy number make up the mitochondrial genome of the rotifer *Brachionus plicatilis*. *Molecular Biology and Evolution* 25: 1129–1137. <https://doi.org/10.1093/molbev/msn058>