Supplementary data

Supplementary Table and Figure Legends

Table S1. Primers used for mitochondrial genome amplification and sequencing.

Table S2. Descriptive statistics of genetic loci. 'All' refers to analyses including data from all seven considered species; 'Mes' refers to analyses limited to *Mesenchytraeus*.

Table S3. Combined-gene datasets and analyses.

Table S4. PartitionFinder recommended partitioning schemes used for concatenated datasets; 'pos_1/2/3' indicates codon position.

Table S5. GenBank accession numbers.

Table S6. Pairwise percent identity matrices for combined datasets. Lower diagonal is nucleotide data; upper (red) diagonal is amino acid. The housekeeping gene dataset excludes 28S rRNA.

Table S7. Bayesian inference gene trees in Newick format.

Table S8. Morphological trait comparison of white snow worms to *Mesenchytraeus hydrius*.

Figure S1. Scanning electron microscope image of *Mesenchytraeus antaeus*. Ventral surface is up in both images. A. Tongue-like organ projecting from mouth and triangular head pore (HP) B. Segments XXIV-XXVII showing the change in orientation of ectal tips of setae from XXV to XXVI.

Figure S2. Morphological features of the white snow worm, *Mesenchytraeus hydrius*. 1. Anterior-dorsal view. 2. Anterior-lateral view. 3. Clitellum: A, dorsal; B, lateral. 4. Spermatheca: A, dorsal; B, lateral. 5. Intestinal region. 6. Penial bulb. 7. Nephridium: A, postclitellar; B, anteclitellar. 8. Sperm funnel. 9. Chaeta. Abbreviations: a, atrium; ag, atrial gland; b, brain; f, female pore; hp, head pore; id, intestinal diverticulum; mco, male copulatory organ; mp, male pore; p, prostomium; pbi, penial bulb invagination; pg, pharyngeal gland; pm, pharyngeal muscles; pp, pharyngeal pad; pr, pristemium; r, rectum; s, spermatheca; sf, sperm funnel; t, testis; vnc, ventral nerve cord.

Table S1. Primers used for mitochondrial genome amplification and sequencing. The approximate location of primers within each gene is based on *M. solifugus* sequence.

Primer	Approx. location	Sequence (5' → 3')	Source
Nad1-F1	5' end	GCYATAGCATTTTACACCYTWATAG	1
Nad1-R1	237 bp from 3' end	AATRTTDGCATATTCWGCTAT	1
Nad4-F1	588 bp from 5' end	GGATTTHTAGTAAAACTWCCWATATTYTC	1
Nad4-R1	1110 bp from 5' end	TTAATDGATGGRGGMGCAGCTAT	1
Nad5-F1	443 bp from 5' end	TACTATCARAAYCCWAAATCTYTDGC	1
Nad5-R1	1044 bp fro 5' end	TTKCCCATTCATCGTARRTCTTG	1
Nad5-F2	530 bp from 5' end	GCHTGAACYTTAAAYCARGGKCA	1
Nad5-R2	1072 bp from 5' end	TGWDGCTACTGGYATTTG	1
Cox2-F1	5' end of COII	TGAGGWCAAYTAAYATTYCAAGACG	1
Cox2-R1	Middle of COII	TAYTCRTATCTTCARTATCATTG	1
Cox2-F2	Middle of COII	CAATGATAYTGAAGATAYGARTA	1
16S-F1	Middle of 16S	GTATCCTAACCGTGCAAAGG	1
16S-R1	3' end of 16S	CTTACGCCGGTCTGAACTCAG	1
16S-R2	Middle of 16S	TACCTTTGCACGGTTAGGATAC	1
Cox3-F1	Middle of COIII	AGGNGTDACHGTWACMTGAGC	1
Cox3-R1	3' end of COIII	ACRTCWACAAARTGTCARTATC	1
Cox3-F2	170 bp 3' of Cox3-F1	TTTGTAGCHACWGGWTTCCAYGG	1
Cox3-R2	170 bp 3' of Cox3-F1	CCRTGGAAWCCWGTDGCTACAAA	1
Nad6-F1	178 bp from 5' end	GGYATATTAGTAATATTTKCWTAYTTYGT	1
Nad6-R2	178 bp from 5' end	GCYACRAARTAWGMAAATATTACTAATATRCC	1
Nad2-F1	163 bp from 5' end	GAAGCAKCARTTAARTAYTTMTTA	1
Nad2-R1	Middle of Nad2	CATCCTATRTGRGTAATWGATG	1
Nad2-F2	150 bp 3' of Nad2-F1	GGWATRGCHCCMTGYCAYYTATG	1
Nad2-R2	150 bp 3' of Nad2-F1	CATARRTGRCAKGGDGCYATWCC	1
Nad3-F1	112 bp from 5' end	CCATTTGARTGTGGATTYGAYCC	1
Nad3-R1	3' end	CADTYTAATGANCCYTWATTTCATTC	1
CytB-F1	370 bp from 5' end	ACHATRGCHACAGCATTTATAGG	1
CytB-R1	3' end	TCTTCTACTGGBCGSCCHCCAATTC	1
CytB-F2	120 bp 3' of CytB-F1	AATWTGAGGHGGNTTTGCHGTAG	1
CytB-R2	120 bp 3' of CytB-F1	CTACDGCAAANCCDCCTCASATTC	1
CytB-F3	307 bp from 3' end	GAGTDTATGCWATTYTACGATCWATTCC	1
CytB-R3	307 bp from 3' end	GGAATSGATCGTARAATSGCATAHACTC	1
Atp6-F496	496 bp from 5' end	GCWAAYATAAGAGCNGGBCAYATYGT	1
Atp-R521	496 bp from 5' end	ACRATRTGVCCNGCTCTTATRTTSGC	1
12S-A1	5' end	AAACTAGGATTAGATACCCTATTAT	2
12S-R1	3' end	GAGAGYGACGGGCGATGTGT	1
LCO1490	5' end of COI	GGTCAACAAATCATAAAGATATTGG	3
HCO2198	Middle of COI	TAAACTTCAGGGTGACCAAAAAATCA	3
COI_r	3' end of COI	CCDCTTAGWCCTARRAARTGTTGNGG	1
Cox5.1	Middle of COI	GATTCTTTGGACATCCAGAAG	4
IW-LCO	5' end of COI	ACTCAACTAATCACAAAGACATTG	1
IW-HCO	Middle of COI	ACTTCTGGATGTCCAAAGAATC	1

Primer sources: 1 = this study; 2 = Borda and Siddall, 2004; 3 = Folmer, *et al.*, 1994; 4 = Hartzell, *et al.*, 2005.

Table S2.

Gene Category	Gene	Alignment positions	Conserved		Variable		Parsimony informative		Parsimony informative	
			All	Mes	All	Mes	All	%	Mes	%
Nuclear-encoded housekeeping genes	*actin*	1131	878	911	253	220	157	13.9	94	8.3
	α-tubulin	1353	1026	1080	327	273	207	15.3	95	7.0
	EF-1α	1398	1041	1155	357	243	216	15.5	77	5.5
	histone3	408	267	305	141	103	80	19.6	47	11.5
	GAPDH	1011	689	770	322	241	226	22.4	87	8.6
	28S rRNA	600	441	542	159	57	128	21.3	13	2.2
Nuclear-encoded ATP synthase subunits	*alpha*	1656	1133	1284	523	372	389	23.5	116	7.0
	beta	1311	958	1019	353	292	229	17.5	105	8.0
	gamma	840	417	599	423	241	198	23.6	83	9.9
	delta	525	319	357	206	153	159	30.3	61	11.6
	epsilon	201	113	140	88	61	33	16.4	13	6.5
	b	927	463	614	461	307	355	38.3	110	11.9
	c	486	296	350	187	124	152	31.3	41	8.4
	d	516	267	360	249	156	199	38.6	52	10.1
	OSCP	633	366	438	267	195	206	32.5	72	11.4
Mitochondrial protein-coding OXPHOS genes	*cox1*	1536	1008	1079	528	457	344	22.4	199	13.0
	Cox2	678	414	460	264	218	183	27.0	95	14.0
	Cox3	777	472	513	305	264	222	28.6	125	16.1
	Cytb	1137	677	754	460	383	323	28.4	170	15.0
	Nad1	936	503	596	433	340	316	33.8	153	16.3
	Nad2	1005	419	532	586	470	440	43.8	200	19.9
	Nad3	354	171	203	183	148	125	35.3	46	13.0
	Nad4	1359	612	777	747	579	534	39.3	226	16.6
	Nad4L	297	141	183	156	111	118	39.7	52	17.5
	Nad5	1746	800	977	946	769	698	40.0	334	19.1
	Nad6	474	197	248	277	220	223	47.0	100	21.1
	Atp6	699	341	450	358	249	284	40.6	113	16.2
	Atp8	165	78	92	87	67	69	41.8	32	19.4
Mitochondrial RNA genes	*12s*	802	508	587	294	195	198	24.7	61	7.6
	16s	1164	744	898	420	266	287	24.7	79	6.8

Table S3.

Dataset	Genes	Size (bp)	Analysis	Framework
HK6	actin, α-tubulin, EF-1α, GAPDH, histone H3, 28S rRNA	5,916	Phylogenetic inference	ML/BI
Mt4	cox1, cytb, 12S rRNA, 16S rRNA	4,600	Phylogenetic inference	ML/BI
MtPCGs	cox1-3, cytb, nad1-6 + 4L, atp6, atp8	11,107	Phylogenetic inference	ML/BI
Mt13+RNAs	cox1-3, cytb, nad1-6 + 4L, atp6, atp8, 12S, tRNA-V, 16S	13,381	Phylogenetic inference	ML/BI
nuASU	alpha, beta, gamma, delta, b, c, d, OSCP	6,894	Phylogenetic inference	ML/BI
Coal_set1	cox1-3, cytb, nad1-6, alpha, beta, b, actin, α-tubulin, EF-1α, GAPDH, 28S	18,429	Multi-species coalescent	BI
Coal_set2	cytb, nad2/4/5, 12S, alpha, gamma, b, c, delta, actin, EF-1α, α-tubulin, histone H3, 28S	14,232	Multi-species coalescent	BI
Diverge	cox1, cytb, nad4, actin, EF-1α, histone H3, 28S	7,478	Divergence dating	BI

Table S4.
PartitionFinder recommended partitioning schemes used for concatenated datasets.

Dataset	Genes	No. of partitions	Partition scheme
HK6	Actin, a-tubulin, EF1a, GAPDH, Histone3, 28S rRNA	5	(EF1a_pos1, H3_pos1, aTub_pos1, actin_pos1, GAPDH_pos1) (EF1a_pos2, H3_pos2, aTub_pos2, actin_pos2, GAPDH_pos2) (EF1a_pos3, aTub_pos3, actin_pos3) (H3_pos3, GAPDH_pos3) (28S)
Mt4	COI, Cytb, 12S rRNA, 16S rRNA	5	(Cytb_pos1, COI_pos1) (COI_pos2) (Cytb_pos3, COI_pos3) (Cytb_pos2) (rRNAs)
MtPCGs	Cox1-3, Cytb, Nad1-6 + 4L, Atp6, Atp8	6	(COI_p1, Cox2_p1, Cox3_p1, CytB_p1) (COI_p2) (Atp6_p3, Atp8_p3, COI_p3, Cox2_p3, Cox3_p3, CytB_p3, Nad1_p3, Nad2_p3, Nad3_p3, Nad4L_p3, Nad4_p3, Nad5_p3, Nad6_p3) (Cox2_p2, Cox3_p2, CytB_p2, Nad1_p2, Nad4L_p2) (Atp6_p1, Atp8_p1, Nad1_p1, Nad2_p1, Nad3_p1, Nad4L_p1, Nad4_p1, Nad5_p1, Nad6_p1) (Atp6_p2, Atp8_p2, Nad2_p2, Nad3_p2, Nad4_p2, Nad5_p2, Nad6_p2);
Mt13+RNAs	Cox1-3, Cytb, Nad1-6 + 4L, Atp6, Atp8, 12S, tRNA-V, 16S	7	Same as above, plus one additional partition for the RNAs
nuASU	Alpha, Beta, Gamma, Delta, b, c, d, OSCP	7	(alpha_pos1, beta_pos1) (alpha_pos2, beta_pos2) (alpha_pos3, delta_pos3, oscp_pos3, syn-b_pos3, syn-c_pos3, syn-d_pos3) (delta_pos1, gamma_pos1, oscp_pos1, syn-b_pos1, syn-c_pos1, syn-d_pos1) (gamma_pos2, oscp_pos2, syn-b_pos2, syn-d_pos2) (beta_pos3, gamma_pos3) (delta_pos2, syn-c_pos2)

Table S5.
GenBank accession numbers.

Gene Category	Gene	*E. albidus*	*E. crypticus*	*M. solifugus*	*M. pedatus*	*M. antaeus*	*M. gelidus*	*M. hydrius*
Nuclear-encoded housekeeping genes	*actin*	KU728801	KU728802	KU728803	KU728804	KU728805	KU728806	KU728807
	α-tubulin	KU728808	KU728809	KU728810	KU728811	KU728812	KU728813	KU728814
	EF-1α	KU728815	KU728816	KU728817	KU728818	KU728819	KU728820	KU728821
	histone3	KU728829	KU728830	KU728831	KU728832	KU728833	KU728834	KU728835
	GAPDH	KU728822	KU728823	KU728824	KU728825	KU728826	KU728827	KU728828
	28s rRNA							
Nuclear-encoded ATP synthase subunits	*alpha*	KU728752	KU728753	KU728754	KU728755	KU728756	KU728757	KU728758
	beta	KU728766	KU728767	KU728768	KU728769	KU728770	KU728771	KU728772
	gamma	No Data	KU728788	KU728789	KU728790	KU728791	KU728792	KU728793
	delta	KU728780	KU728781	KU728782	KU728783	KU728784	KU728785	KU728786
	epsilon							
	b	KU728759	KU728760	KU728761	KU728762	KU728763	KU728764	KU728765
	c	KU728745	KU728746	KU728747	KU728748	KU728749	KU728750	KU728751
	d	KU728773	KU728774	KU728775	KU728776	KU728777	KU728778	KU728779
	OSCP	KU728794	KU728795	KU728796	KU728797	KU728798	KU728799	KU728800
Mitochondrial protein-coding genes	*COI*	KU728850	KU728851	KU728852	KU728853	KU728854	KU728855	KU728856
	Cox2	KU728857	KU728858	KU728859	KU728860	KU728861	KU728862	KU728863
	Cox3	KU728864	KU728865	KU728866	KU728867	KU728868	KU728869	KU728870
	Cytb	KU728843	KU728844	KU728845	KU728846	KU728847	KU728848	KU728849
	Nad1	KU728871	KU728872	KU728873	KU728874	KU728875	KU728876	KU728877
	Nad2	KU728878	KU728879	KU728880	KU728881	KU728882	KU728883	KU728884
	Nad3	KU728885	KU728886	KU728887	KU728888	KU728889	KU728890	KU728891
	Nad4	KU728892	KU728893	KU728894	KU728895	KU728896	KU728897	KU728898
	Nad4L	KU728899	KU728900	KU728901	KU728902	KU728903	KU728904	KU728905
	Nad5	KU728906	KU728907	KU728908	KU728909	KU728910	KU728911	KU728912
	Nad6	KU728913	KU728914	KU728915	KU728916	KU728917	KU728918	KU728919
	Atp6	KU728836	KU728837	KU728838	KU728839	KU728840	KU728841	KU728842
	Atp8	KU639961	KU639962	KU639963	KU639964	KU639965	KU639966	KU639967
Mitochondrial rRNA genes	*12s*	KU728920	KU728921	KU728922	KU728923	KU728924	KU728925	KU728926
	16s	KU728927	KU728928	KU728929	KU728930	KU728931	KU728932	KU728933

Table S6.
Pairwise percent identity matrices for combined datasets. Lower diagonal is nucleotide data; upper (red) diagonal is amino acid. The housekeeping gene dataset excludes 28S rRNA.

Dataset		*E. albidus*	*E. crypticus*	*M. pedatus*	*M. solifugus*	*M. antaeus*	*M. gelidus*	*M. hydrius*
5 nuclear house-keeping genes 5304 bp	*E. albidus*		98.64	94.84	94.61	94.72	95.18	94.90
	E. crypticus	93.1		94.72	94.67	94.72	95.01	95.12
	M. pedatus	82.3	82.6		96.49	96.88	97.17	96.43
	M. solifugus	82.1	82.6	86.7		97.62	98.47	97.17
	M. antaeus	84.0	84.7	85.1	86.0		98.13	96.77
	M. gelidus	83.9	84.5	85.3	86.9	90.0		97.28
	M. hydrius	84.5	84.8	85.9	86.4	89.5	90.0	
8 nuclear ATP synthase subunits 6894 bp	*E. albidus*		93.7	78.03	77.2	77.72	77.76	77.58
	E. crypticus	92.7		80.49	79.31	79.78	80	79.68
	M. pedatus	70.6	72.1		91.58	94.31	93.83	92.64
	M. solifugus	69.9	70.9	82.5		92.15	91.49	92.71
	M. antaeus	70.0	71.3	86.9	83.4		95.89	92.99
	M. gelidus	70.7	71.7	86.6	83.4	89.1		92.64
	M. hydrius	70.4	71.6	83.9	84.0	85.8	85.8	
13 mito - chondrial PCGs 11,160 bp	*E. albidus*		96.75	69.22	69.43	69.97	70.15	70.29
	E. crypticus	88.4		69.43	69.61	70.21	70.14	70.28
	M. pedatus	69.4	69.2		84.27	86.54	87.39	85.16
	M. solifugus	68.3	67.9	75.9		85.5	86.37	88.29
	M. antaeus	70.0	69.5	78.4	76.9		89.58	86.95
	M. gelidus	66.8	66.6	75.7	74.1	77.0		87.77
	M. hydrius	69.5	69.2	77.6	78.1	79.2	75.8	
Mito-chondrial COI-CytB 12s-16s rRNAs 4600 bp	*E. albidus*							
	E. crypticus	91.7						
	M. pedatus	75.2	74.9					
	M. solifugus	75.3	75.0	81.9				
	M. antaeus	75.1	74.6	83.1	83.5			
	M. gelidus	74.3	74.5	83.4	83.7	85.7		
	M. hydrius	73.9	73.9	82.9	84.3	84.8	85.0	

Table S7.

Loci	Bayesian Inference Newick Trees
12s	(Ealb:0.006770898,Ecry:0.03538963,(Mped:0.1356714,(Msol:0.09907861,Mhyd:0.08418686):0.03167925,(Mant:0.07932975,Mgel:0.09419123):0.02030384):0.7611206);
16s	(Ealb:0.006770898,Ecry:0.03538963,(Mped:0.1356714,(Msol:0.09907861,Mhyd:0.08418686):0.03167925,(Mant:0.07932975,Mgel:0.09419123):0.02030384):0.7611206);
28s	(Ealb:0.009272668,Ecry:0.003186291,((Mped:0.01092156,((Msol:0.02990588,Mhyd:0.03602543):0.007846756,Mant:0.01396765):0.008108208):0.007772201,Mgel:0.011
Actin	(Ealb:0.006634816,Ecry:0.01077796,(Mped:0.02552142,(((Msol:0.06971496,Mgel:0.04039868):0.03221018,Mant:0.1219178):0.05155338,Mhyd:0.02191136):0.01205492
Alpha	(Ealb:0.008025507,Ecry:0.03388649,((((Mped:0.08920178,Mant:0.04654193):0.007337236,Mgel:0.05208632):0.04008744,Msol:0.08265423):0.02041718,Mhyd:0.074234
Atp6	(Ealb:0.1006109,Ecry:0.06697589,((Mped:0.5359191,(Msol:0.2669599,Mhyd:0.1365852):0.07045715,Mgel:0.1374161):0.1030408,Mant:0.03734736):1.667381);
Atp8	(Ealb:0.04558744,Ecry:0.05204976,(Mped:0.08935008,(((Msol:0.2350216,Mhyd:0.06556376):0.2291701,Mant:0.04224696):0.05427059,Mgel:0.1921267):0.1337614):0.7
aTub	(Ealb:0.0280425,Ecry:0.05940868,((Mped:0.1844469,Mant:0.047322):0.0229864,((Msol:0.05237005,Mgel:0.06000996):0.01533611,Mhyd:0.05423477):0.02439432):0.10
Beta	(Ealb:0.009431517,Ecry:0.01898405,(((Mped:0.09013813,Msol:0.1703738):0.04445966,Mgel:0.04070362):0.0203751,(Mant:0.06328075,Mhyd:0.08784072):0.0186773):0
COI	(Ealb:0.04967886,Ecry:0.08279515,(Mped:0.1265964,(Msol:0.1823573,(Mant:0.1022228,(Mgel:0.1406486,Mhyd:0.1248605):0.02142548):0.0308885):0.01767774):0.166
Cox2	(Ealb:0.05148498,Ecry:0.07268306,(Mped:0.2101649,((Msol:0.2368364,Mhyd:0.07374255):0.07473834,Mant:0.1347546):0.04838382,Mgel:0.1468044):0.4432648);
Cox3	(Ealb:0.1100225,Ecry:0.08157745,(Mped:0.1309824,(Msol:0.3234226,(Mant:0.2194696,Mhyd:0.2078351):0.04899946,Mgel:0.1683284):0.06326157):0.3822504);
CytB	(Ealb:0.1899507,Ecry:0.1136645,(Mped:0.2216164,(((Msol:0.2761503,Mhyd:0.2795215):0.08509904,Mant:0.2359159):0.06556462,Mgel:0.2144372):0.1312525):0.74179
Delta	(Ealb:0.02311254,Ecry:0.007912509,((((Mped:0.1132083,Mant:0.1007896):0.0255964,Mgel:0.08999871):0.05545259,Msol:0.07418853):0.04720779,Mhyd:0.08393262):0
EF1a	(Ealb:0.06155822,Ecry:0.05248541,(Mped:0.05515498,(((Msol:0.05667427,Mhyd:0.06419409):0.02041683,Mant:0.03704231):0.01661503,Mgel:0.03527733):0.03079627
GAPDH	(Ealb:0.02965163,Ecry:0.01018119,(Mped:0.1023636,(Msol:0.1110167,Mhyd:0.07963731):0.02086502,Mant:0.08317348,Mgel:0.08051307):0.3720227);
His3	(Ealb:0.4618253,Ecry:0.1657169,((Mped:0.09145983,Msol:0.2043889):0.06660217,Mant:0.05458691,Mgel:0.1081008,Mhyd:0.08565743):0.07861968);
Nad1	(Ealb:0.06557506,Ecry:0.1439481,(Mped:0.3475321,((Msol:0.2614244,Mhyd:0.2768651):0.06552739,Mgel:0.2453313):0.05999569,Mant:0.1464659):1.106616);
Nad2	(Ealb:0.1166942,Ecry:0.09196712,(Mped:0.3152666,(Msol:0.4380866,Mhyd:0.269151):0.1388592,(Mant:0.2681369,Mgel:0.2627457):0.07513771):2.38637);
Nad3	(Ealb:0.08377358,Ecry:0.09204989,(Mped:0.2560777,Msol:0.2309416,Mant:0.1874927,Mgel:0.1822523,Mhyd:0.2224673):0.8099003);
Nad4	(Ealb:0.1254377,Ecry:0.06112254,(Mped:0.186308,(Msol:0.2832701,Mhyd:0.2412424):0.1010536,Mant:0.2694843,Mgel:0.2732755):1.464979);
Nad4L	(Ealb:0.09111356,Ecry:0.06359623,(Mped:0.2287165,Msol:0.3784204,(Mant:0.169601,Mhyd:0.2387357):0.1138872,Mgel:0.09100801):1.544231);
Nad5	(Ealb:0.06632153,Ecry:0.1155907,(((Mped:0.2878917,Mgel:0.1724932):0.07616429,Mant:0.2707214):0.06106031,(Msol:0.3768889,Mhyd:0.1754368):0.08591947):1.147
Nad6	(Ealb:0.09655636,Ecry:0.1306967,(Mped:0.5228473,Msol:0.3751322,Mant:0.2812365,Mgel:0.3089747,Mhyd:0.3991824):3.104096);
OSCP	(Ealb:0.007914523,Ecry:0.01891683,(((Mped:0.11301,(Mant:0.07559698,Mgel:0.06912302):0.02215433):0.05354757,Mhyd:0.06046808):0.05404024,Msol:0.1235045):0.
b	(Ealb:0.03214107,Ecry:0.0252717,((Mped:0.1431415,(Mant:0.05179623,Mgel:0.0955153):0.01966341):0.03779533,(Msol:0.1404522,Mhyd:0.09338837):0.05312848):0.9
c	(Ealb:0.01380941,Ecry:0.02547177,(Mped:0.03104475,((Msol:0.07769612,Mhyd:0.08155793):0.04698265,Mant:0.0831281,Mgel:0.09443937):0.06045314):0.6834484);
d	(Ealb:0.01097478,Ecry:0.027533,((((Mped:0.09748302,Mgel:0.0635017):0.02492598,Mant:0.07978162):0.05921604,Mhyd:0.08366598):0.07140724,Msol:0.05010475):0.
Gamma	(Ecry:0.5470455,Mped:0.06583683,(((Msol:0.1688175,Mhyd:0.1053316):0.02367685,Mgel:0.06574272):0.02567122,Mant:0.07674278):0.03466705);
Epsilon	(Ecry:0.4470465,(Mped:0.05858022,(Msol:0.1491025,Mant:0.1020348):0.05211):0.04423441,(Mgel:0.04290376,Myhd:0.09981654):0.04159724);

Table S8.
White snow worm morphological comparison

Traits	Sample 1 (small)	Sample 2 (medium)	*Mesenchytraeus hydrius*	Compare
Shape of Body	cylindrical, elongate, smooth and approximately of the same diameter	cylindrical, elongate, smooth and approximately of the same diameter	Cylindrical, elongate, smooth and approximately of the same diameter throughout.	same
Color	whitish	whitish	whitish	same
Body length	10 mm	16 mm	17-24 mm	Less than *M. hydrius*
Body Width	0.87 mm	0.94 mm	0.76-0.91 mm	Less than *M. hydrius*
Number of segments	70	87	82-97	Less than *M. hydrius*
Head pore	The small, smooth, rounded prostomium carries the head pore near its tip.	The small, smooth, rounded prostomium carries the head pore near its tip.	The small, smooth, rounded prostomium carries the head pore near its tip.	same
Chaetae and number	2-5 lateral 2-6 ventral	2-5 lateral 2-6 ventral	4-7 lateral 5-9 ventral	Less than *M. hydrius*
Clitellum position	12-13 annulus	12-13 annulus	$^{1/3}$11-13 annulus	same
Clitellar glands	Present in XII	Present in XII	presence	same
Spermatheca	attached to between of IV-V annulus, 2 pairs of diverticula, extending to XI.	attached to between of IV-V annulus, 2 pairs of diverticula, extending to XI.	attached to between of IV-V annulus, 2 pairs of diverticula, extending to XI.	same
Genital pore	Male gonopore in between XI-XII, Female gonopore in between XII-XIII,	Male gonopore in between XI-XII, Female gonopore in between XII-XIII,	?	?
Testis	Present	Present	?	?
Sperm funnel	approximately cylindrical	approximately cylindrical	approximately cylindrical	same
Atrial gland	Present: 5	Present: 5	about 5 atrial glands	same
accessory gland	no accessory glands	no accessory glands	no accessory glands	same
sperm sac	no sperm sacs	no sperm sacs	no sperm sacs	same
Vas deferens	present	present	present	same
Ovary	?	?	1 ovisac extending to 31-33	?

Figure S1. *Mesenchytraeus antaeus* SEM images

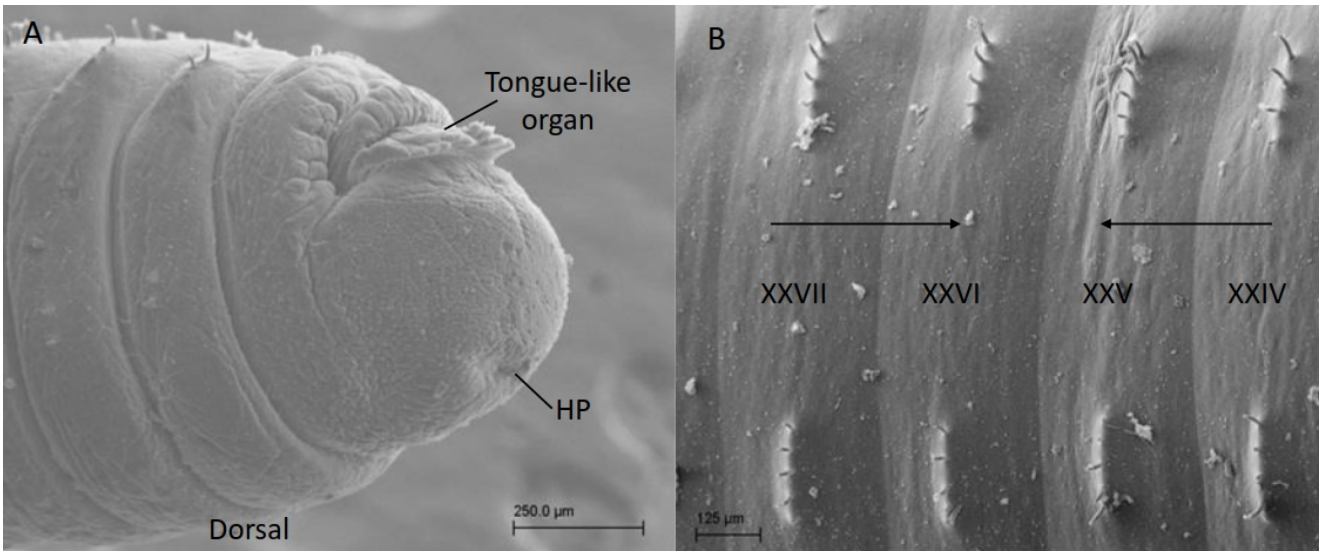
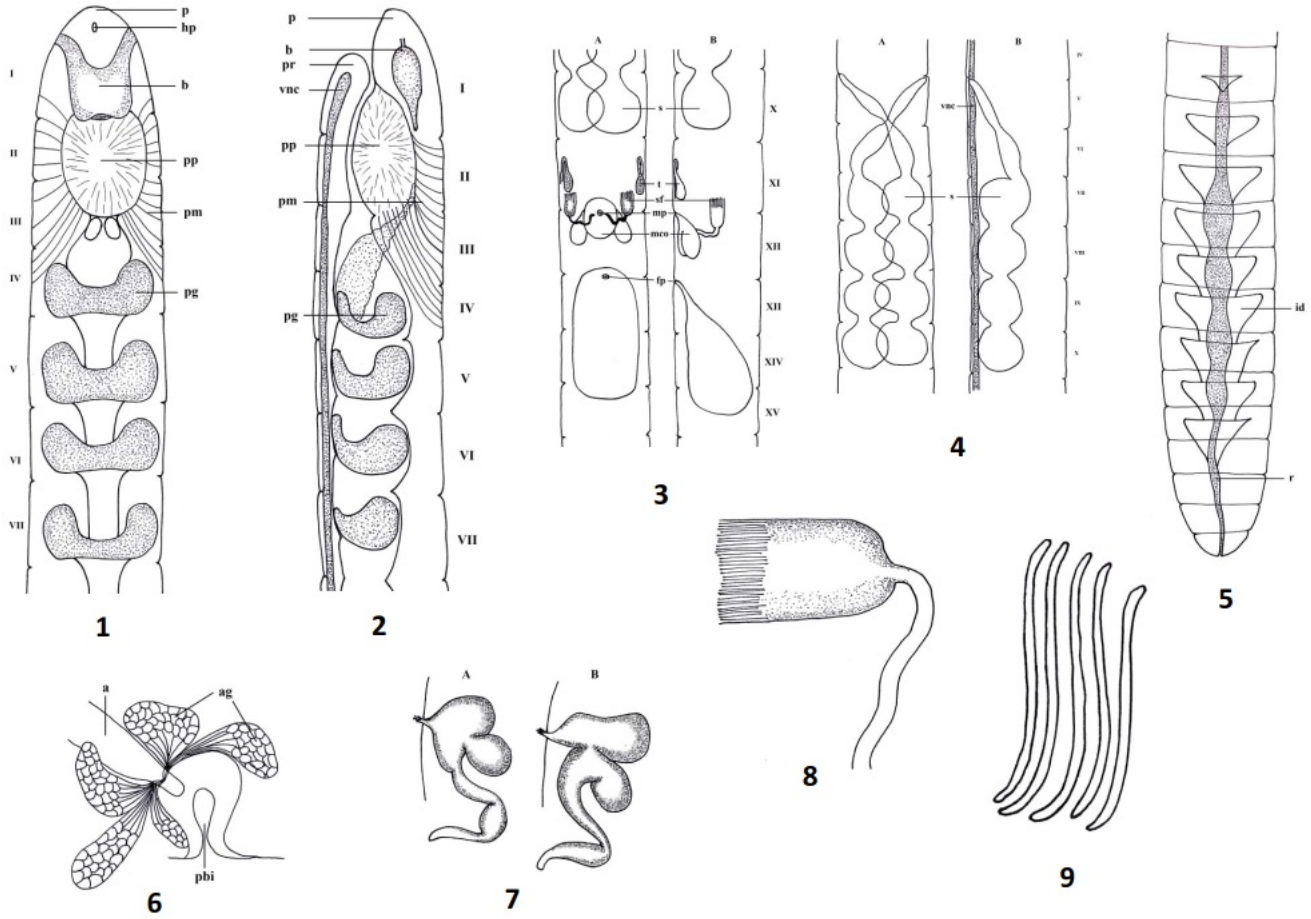


Figure S2. *Mesenchytraeus hydrius* morphology



Supplementary Materials and Methods

Specimens

Enchytraeidae, most of which are <15 mm long and transparent or white, are primarily identified by internal structures using live specimens examined with a light microscope (Schmelz and Collado, 2010). However, this technique is often not practical for *Mesenchytraeus* species due to their size and/or pigmentation. Thus, morphological identifications were performed on freshly sacrificed specimens prepared as follows. Worms were placed in 70% ethanol for 24 hours, then dehydrated through an ethanol series: 80% (10 min), 90% (10min), 100% (2 x 10 min). Dehydrated specimens were transferred to a 1:2 ratio of benzyl alcohol:benzyl benzoate (Sigma-Aldrich) and incubated until transparent (~20 min). Internal structures were sketched and photographed for comparison to published descriptions. In addition, external morphology was imaged with a LEO 1450EP scanning electron microscope as described in Shain et.al (2000).

Species identification

With the exception of *M. hydrius*, the *Mesenchytraeus* species chosen for this investigation were those that have been reliably observed/collected/identified in specific locations by ourselves and colleagues within the past ~25 years. *Mesenchytraeus solifugus* is recognized as the only species of glacier ice worm in the PNW and considerable molecular data (e.g., Hartzell et al., 2005; Dial et al., 2012) has confirmed this. The aquatic species, *M. pedatus*, was collected from the Sacramento River at Redding, CA, identified as site 18 by Healy and Fend (2002), who found *only M. pedatus* at that location. The external features (color, size, setae arrangement, head pore location, penial papillae) were consistent with Healy and Fend's (2002) expanded description, but internal morphology was not examined and molecular population data showed wide variation. Thus, sequence data for this species is designated *Mesenchytraeus cf. pedatus.* Likewise, multiple *M. antaeus* worms were collected from well-rotted logs at the same location described in Rota and Brinkhurst (2000), and external features, in particular their extraordinary length (30-60 mm) and lack of pigmentation, are in complete agreement with those authors' detailed description. Due to their size, attempts to view internal structures were unsuccessful, however, SEM images captured the everted tongue-like structure and anterior-posterior setae orientation switch described by the authors (Fig. S2A, B) and we are confident that our terrestrial species is *M. antaeus*.

The worms identified here as *M. gelidus* were collected from the Mt. Rainier region as were those originally described by Welch (1916). Size and pigmentation prevented viewing of internal structures, but external morphology was consistent with Welch's description, with the exception of fewer anterio-ventral setae (5-7 vs. 7-9). GenBank has two entries for 28S rRNA sequence for *M. gelidus* that differed from ours by 1%, an unaccountably large difference if these are the same species. However, those specimens were collected from Yosemite, CA and it is unclear that a definitive identification was performed. Due to these discrepancies, and the fact that Welch's comparative analyses were primarily based on internal structures, the sequence data for this species has been designated *Mesenchytraeus cf. gelidus*.

The final *Mesenchytraeus* species considered here, *M. hydrius*, was discovered while collecting worms from the transient snow fields in the Mt. Rainier area. We were able to perform in-depth examinations of these worms and are confident they are *M. hydrius* (see Table S8, Fig. S2 for details).

The outgroup species, *Enchytraeus albidus*, collected in Denmark, was verified by B. Christensen (Fisker 2014). In view of the known problems with identification of *E. crypticus/E. variatus*, the molecular data generated for this species was designated *Enchytraeus cf. crypticus*.

Polymerase Chain Reaction

Reactions contained 34.5 µl water, 200 µM dNTPs, 600 nM primers, 10 µl of 5X Phusion HF buffer, 1 U Phusion Hot Start II DNA Polymerase (Thermo Scientific), and 50-200 ng total genomic DNA. Reaction conditions were: denaturation at 98°C for 30 seconds, followed by 35 cycles of 98°C for 10 s, 47-60°C for 20 s, and 72°C for 0.5-2 min; final extensions were 5 min at 72°C. Single-band PCR products were purified using DNA Clean and Concentrator (Zymo Research); multi-band products were electrophoresed on 0.8% agarose gels and desired bands excised. Gel-excised DNA was purified using a Zymoclean Gel DNA Recovery Kit (Zymo Research) according to the manufacturer's protocol. Fragments were cloned into pMiniT vectors (New England Biolabs), transformed into NEB 10-beta chemically-competent *E. coli*, and 3-5 clones/fragment/species were selected for purification using Zyppy Plasmid Miniprep (Zymo Research). Sanger DNA sequencing was performed by GENEWIZ, Inc. (South Plainfield, NJ). Initial sequencing was performed using vector-specific forward and reverse primers. Resequencing was performed as necessary using original and internal primers, to ensure recovery of overlapping sequences. For regions spanning multiple consecutive tRNAS, KAPA2G Robust polymerase (KAPA Biosystems) was used according to the manufacturer's protocol.

Transcriptome generation

Transcriptomes were generated at different facilities, thus preparation details varied and are described separately. For *M. solifugus, M. pedatus* and *E. crypticus,* 10-12 worms/species were pooled and total RNA was extracted using TRIzol (Invitrogen) according to the manufacturer's protocol. The SMART cDNA library construction kit (Clontech) was used for first-strand cDNA synthesis, followed by synthesis and amplification of double-stranded cDNA using the Advantage 2 PCR Kit (Clontech). Illumina sequencing was conducted at the Genomics Services Laboratory at Hudson Alpha Institute for Biotechnology (Huntsville, AL). For *M. pedatus* and *M. solifugus*, sequencing runs were performed on an Illumina HiSeq 2000 platform with 2 x 100 bp paired-end reads using the TruSeq v3 (Illumina) protocol. *Enchytraeus crypticus* sequencing employed the Illumina HiSeq 2500 platform with 2 x 150 bp paired-end reads, also using the TruSeq v3 (Illumina) protocol. Paired-end reads were demultiplexed and *de novo* assembly was performed using the Trinity software package (Grabherr et al., 2011), version October 2012 (*M. pedatus, M. solifugus*) or February 2013 (*E. crypticus*) with default settings. See Waits et al. (2016) for additional details.

For *M. antaeus, M. gelidus,* and *M. hydrius*, total RNA was extracted from single worms using TRIzol Plus RNA Purification Kit (Invitrogen) according to the manufacturer's protocol. RNA samples were express-shipped to the Giribet lab at Harvard University's Department of Organismic and Evolutionary Biology and purified using the Dynabeads mRNA Purification Kit (Invitrogen) following manufacturer's instructions. cDNA libraries were constructed in the Apollo 324 automated system using the PrepX mRNA kit (IntegenX). Further details about the protocols can be found in Fernández et al. (2014). Samples were run using the Illumina HiSeq 2500 platform with

2 x 150 paired-end reads at the FAS Center for Systems Biology at Harvard University. Demultiplexed sequencing results, in FASTQ format, were retrieved, each sample being quality-filtered according to a threshold average quality score of 30 based on a Phred scale and adaptor trimmed using Trim Galore! 0.3.3 (Wu et al., 2011). Ribosomal RNA and mitochondrial DNA were filtered out using Bowtie v. 1.0.0 (Langmead et al., 2009). Strand specific de novo assemblies were done individually in Trinity (Haas et al., 2013) using paired read files, a strand specificity flag and path reinforcement distance enforced to 75. Access to the *E. albidus* transcriptome was provided by Christer Erséus (U. of Gothenburg, Sweden).

Gene extraction and verification from transcriptomes

Multiple loci with varying levels of conservancy were chosen for phylogenetic analyses in an effort to effectively infer the most plausible species tree. Well-conserved nuclear housekeeping genes, *actin*, *α-tubulin*, *EF-1*α, *histone H3*, and *GAPDH* were extracted, as well as core subunits of ATP synthase: *alpha, beta, gamma, delta, epsilon, b, c, d and OSCP*. The latter genes, though nuclear-encoded, function within the mitochondria in conjunction with mitochondrial-encoded genes to produce the bulk of cellular energy. The critical importance of this metabolic process imposes mito-nuclear coevolutionary constraints that are linked to speciation (Schmidt et al., 2001; Gershoni et al., 2009; Bar-Yaacov et al., 2012; Hill, 2016), thus making them good candidates for species delimitation analysis. In addition, all 13 mitochondrial PCGs and *12S* and *16S* ribosomal RNAs were extracted for comparison to PCR-amplified sequences (see Table S2 for gene details). Local BLAST searches were performed using NCBI's BLAST 2.2.26 utilizing the aforementioned Trinity assemblies as databases. Queries for nuclear-encoded genes used annelid gene sequence when available, but otherwise queries were made using mollusc, lancelet, and acorn worm sequence. Extracted transcripts were subjected to GenBank searches to verify gene identity, translated to ensure that they were not pseudogenes, and aligned to each other to assess relatedness. Extracts with questionable or incomplete sequence were re-BLAST using genus-specific queries and/or subjected to raw-read mapping to rule out assembly error.

Mitochondrial sequence alignment and annotation

DNA sequence chromatograms were evaluated and trimmed using Chromas 2.4.3 (Technelysium). Overlapping sequence fragments were aligned in MEGA version 6 (Tamura et al., 2013) and assembled into consensus sequence contigs using Mesquite v.3.04 (Maddison and Maddison, 2015). Protein coding genes (PCGs) and ribosomal RNAs were identified using NCBI's BLASTx and BLASTn algorithms. Individual genome contigs were uploaded to DOGMA (http://dogma.ccbb.utexas.edu) for identification of tRNAs and whole genome annotation using default parameters for invertebrate mitochondrial genomes (tRNA data is discussed in a separate manuscript).

Phylogenetic analyses

Estimation of divergence time was performed in BEAST v1.8.2 (Drummond et al., 2012) and in MrBayes v3.2.6 (Ronquist et al., 2011) using a dataset comprising *cox1, cytb, nad4, actin, EF-1α, histone H3* and *28s* rRNA (Table S3). For these analyses, a minimum age bound for family Enchytraidae was 35 million years before

present based on a fossilized specimen found in Baltic amber (Ulrich and Schmelz, 2001). *Lumbricus terrestris* was used as the outgroup with sequences retrieved from GenBank (*cox1, cytb, nad4*: U24570.1; *28S*:HQ691218.1; *actin*:X96514.1; *EF-1α*:DQ813372.1; *histone H3*:FJ214241.1). For the BEAST analysis, the XML file specified application of the GTR substitution model, 4 gamma categories, proportion invariant (0-1), and an uncorrelated relaxed clock (Drummond et al., 2006) with lognormal distribution. For 'time to most recent common ancestor' (tmrca), a uniform age constraint prior on the root of 35/300 or 44/500 (mya) was applied. The lower bounds represent the youngest (35) and mean (44) estimated ages of Baltic amber (Larsson, 1978; Standke, 1998), while the upper bounds represent the estimated ages of oligochaetes (300) and annelids (500) (Bomfleur et al., 2012; Parry et al., 2014). Each age constraint file had 5 independent runs of 100 million generations, sampling every 10,000, executed via the CIPRES-Portal v.3.3 (Miller et al., 2010). Convergence, ESS, 95% HPD, and ucld.stdev were assessed in Tracer v1.6 (Rambaut et al., 2014).

MrBayes analysis employed the independent gamma rates (IGR) relaxed clock model with a GTR + gamma substitution model and variable rates across partitions (determined with PartitionFinder v1.1.0; Lanfear et al., 2012). A uniform root constraint (35/300 mya) was set for calibration, and a partial backbone constraint was specified to separate *Mesenchytraeus* from *Enchytraeus*. Four independent runs of 5 million generations were performed and convergence was assessed as described previously.

Supplementary References

Bar-Yaacov, D., Blumberg, A., Mishmar, D. (2012) Mitochondrial-nuclear co-evolution and its effects on OXPHOS activity and regulation. *Biochim Biophys Acta* 1819: 1107-1111.

Bomfleur, B., Kerp, H., Taylor, T., Moestrup, O., Taylor, E. (2012) Triassic leech cocoon from Antarctica contains fossil bell animal. PNAS 109: 20971–20974.

Borda, E. and Siddall, M.E. (2004) Arhynchobdellida (Annelida: Oligochaeta: Hirudinida): phylogenetic relationships and evolution. *Mol Phylogenet Evol* 30: 213-25.

Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A. (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biology* 4: e88.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A. (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7 *Mol Biol Evol* 29: 1969-1973.

Fernández, R., Laumer, C.E., Vahtera, V., Libro, S., Kaluziak, S., Sharma, P.P., Pérez-Porro, A.R., Edgecombe, G.D., Giribet, G. (2014) Evaluating topological conflict in centipede phylogeny using transcriptomic data sets. *Mol Biol Evol* 31: 1500–1513. doi:10.1093/molbev/msu108.

Fisker KV. 2014 Physiological and biochemical adaptations to low temperature and copper stress in terrestrial annelids. PhD thesis, Aarhus University, Department of Bioscience, Denmark.

Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R. (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol* 3: 294–299.

Gershoni, M., Templeton, A.R., Mishmar, D. (2009) Mitochondrial bioenergetics as a major motive force of speciation. *Bioessays* 31: 642-650.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A. (2011) Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat Biotechnol*. 29(7): 644-52. doi: 10.1038/nbt.1883.

Hartzell, P.L., Nghiem, J.V., Richio, K.J., Shain, D.H. (2005) Distribution and phylogeny of glacier ice worms (*Mesenchytraeus solifugus* and *Mesenchytraeus solifugus rainierensis*). *Can J Zool* 83(9):1206-1213, 10.1139/z05-116.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., Macmanes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., Leduc, R.D., Friedman, N., Regev, A. (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*. Aug;8(8): 1494-512. doi: 10.1038/nprot.2013.084.

Hill, G.E. (2016) Mitonuclear coevolution as the genesis of speciation and the mitochondrial DNA barcode gap. *Ecol Evol* 6:5831–5842. doi: 10.1002/ece3.2338.

Lanfear, R., Calcott, B., Ho, S., Guindon, S. (2012) PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol Biol Evol* 29: 1695–1701.

Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10. doi:10.1186/Gb-2009-10-3-R25.

Larsson, S.G. (1978) Baltic Amber - A palaeontological study. Entomograph, Vol. I. Scandinavian Scientific Press, Klampenborg.

Maddison, W.P. and Maddison, D.R. (2015) Mesquite: a modular system for evolutionary analysis, Version 3.04. http://mesquiteproject.org.

Miller, M.A., Pfeiffer, W., Schwartz, T. (2010) "Creating the CIPRES Science Gateway for inference of large phylogenetic trees" in Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA pp 1-8.

Parry, L., Tanner, A., Vinther, J. (2014) The origin of annelids. *Palaeontology* 57: 1091-1103.

Rambaut, A., Suchard, M.A., Xie, D., Drummond, A.J. (2014) Tracer v1.6, Available from http://beast.bio.ed.ac.uk/Tracer.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P. (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61(3):539-42. doi: 10.1093/sysbio/sys029.

Schmelz, R.M., and Collado, R. (2010) A guide to European terrestrial and freshwater species of Enchytraeidae (Oligochaeta). *Soil Organisms*, 82: 1-176.

Schmidt, T.R., Wu, W., Goodman, M., Grossman, L.I. (2001) Evolution of nuclear- and mitochondrial-encoded subunit interaction in cytochrome c oxidase. *Mol Biol Evol* 18(4):563-569.

Shain, D.H., Mason, T.A., Farrell, A.H., Michalewicz, L.A. (2001) Distribution and behavior of ice worms (*Mesenchytraeus solifugus*) in south-central Alaska. Canadian Journal of Zoology 79(10): 1813-1821, 10.1139/z01-143.

Standke, G. (1978) Die Tertiarprofile der Samlandischen Bernsteinkuste bei Rauschen. *Schriftenreihe Geowiss* 7: 93-103.

Tamura, K; Stecher, G., Peterson, D., Filipski, A., Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30: 2725-2729.

Ulrich, H. and Schmelz, R.M. (2001) Enchytraeidae as prey of Dolichopodidae, recent and in Baltic amber. *Bonn Zool Beitr* 50: 89-101.

Waits, D.S., Santos, S.R., Thornhill, D.J., Li, Y., Halanych, K.M. (2016) Evolution of sulfur binding by hemoglobin in Siboglinidae (Annelida) with special reference to bone-eating worms, *Osedax. J Mol Evol* 82: 219-229. DOI 10.1007/s00239-016-9739-7.

Wu, Z., Wang, X., Zhang, X. (2011) Using non-uniform read distribution models to improve isoform expression inference in RNA-Seq. *Bioinformatics* 27(4):502-8.
doi: 10.1093/bioinformatics/btq696.