Supplemental data

Table S1. Hypothesis testing regarding hypotheses about the placement of interstitial annelid taxa suggested based on morphological data. Difference in ln L values to the best tree are shown for each hypothesis as well as the probabilities based on the AU test and the posterior probability. Significant values equal to or below 0.05 are in bold. Table S1 is related to Figure 1.

Hypothesis	Δ ln L	AU	PP
best tree		1.000	1.000
Nerillidae within or sister to Errantia	527.8	**$8*e^{-35}$**	**$6*e^{-230}$**
Apharyngtus within or sister to Eunicida	1460.3	**$8*e^{-71}$**	**0**
Dinophilidae with or sister to Eunicida	1845.2	**$7*e^{-07}$**	**0**
Diurodrilidae within or sister to Eunicida	404.5	**$9*e^{-08}$**	**$2*e^{-176}$**
Apharyngtus, Dinophilidae & Diurodrilidae within or sister to Eunicida	2777.2	**$4*e^{-10}$**	**0**
Nerillidae within or sister to Eunicida	746.7	**$2*e^{-07}$**	**0**
Polygordiidae within or sister to Opheliidae	4603.5	**$5*e^{-42}$**	**0**
Protodrilida within or sister to Spionida	4102.6	**$2*e^{-39}$**	**0**

Table S2. Partitions containing potentially paralogous or cross-contaminated sequences for interstitial annelid species. Carefully screening individual gene partitions revealed three partitions apparently affected by cross-contamination and six by paralogous sequences. In two cases, blast searches for paralogy screening did not return a hit. These 14 sequences were deleted from the dataset d01 and phylogenetic reconstructions based on this reduced dataset d02 were conducted. The bootstrap values as well as the species of the detected clade in the single-partition analysis are shown. The sequences of these species in these partitions were excluded from the dataset d01 resulting in dataset d02. Table S2 is related to Table 1.

Partition ID	**Bootstrap value**	**Affected species**	
Potential contamination			
23087	99	*Diopatra cuprea*	*Protodriloides symbioticus*
22003	100	*Trilobodrilus axi*	*Apharyngtus punicus*
23675	100	*Saccocirrus burchelli*	*Polygordius sp.*
Potential paralogy			
22410	97	*Arabella sp.*	*Saccocirrus burchelli*
22161	99	*Polygordius lacteus*	*Nepthys caeca*
22026	100	*Lysidice sp.*	*Protodrilus adherens*
23095	100	*Polygordius lacteus*	*Marphysa bellii*
23683	100	*Polygordius lacteus*	*Protodorvillea kefersteinii*
23761	100	*Protodrilus adherens*	*Ophryotrocha globopalpata*
No hits			
22026	100	*Polygordius lacteus*	*Protodriloides symbioticus*
24126	100	*Saccocirrus burchelli*	*Polygordius lacteus*

Table S3. Bootstrap support for the different positions of *Diurodrilus subterraneus* (Diurodrilidae) in the analyses of the three datasets generated with different d values in MARE (i.e., datasets d01, d03 & d04). Bootstrap values at the significance level of 70 and higher are in italic. Table S3 is related to Table 1.

***Diurodrilus subterraneus* (Diurodrilidae) grouping together with**	**Datasets**		
	d01	**d03**	**d04**
Apharnygtus+Orbiniidae+Parergodrilidae	65	44	*80*
Apharnygtus	65	46	*82*
Dinophilidae	35	54	18

Table S4. Results of the ancestral state reconstructions. Proportional likelihood values *p* are shown below the reconstructed state. n. a. = not applicable. Table S4 is related to Figure 3.

Character	Sedentaria	Orbiniida	Errantia	Protodriliformia
Segmentation	Present	Present	Present	Present
	1.00	1.00	1.00	1.00
Head structure	one ring	two or more	one ring	one ring
	0.65	0.69[1]	0.62	0.96
Anterior cirri	Absent	Absent	Absent	Absent
	1.00	1.00	1.00	1.00
Palps	Present	Absent	Present	Present
	0.67	0.67[2]	1.00	1.00
Origin of palps	Prostomial	n. a.	Prostomial	Prostomial
	0.93		1.00	1.00
Number of palps	1 pair	n. a.	1 pair	1 pair
	0.98		1.00	1.00
Median antenna	Absent	Absent	Absent	Absent
	1.00	1.00	1.00	1.00
Lateral antennae	Absent	Absent	Absent	Absent
	0.98	0.98	0.81	0.98
Parapodia	Present	Present	Present	Absent
	0.84	0.57	0.83	0.56[3]
Size of notopodial lobes	Small	Small	Small	Small
	0.70	0.82	0.80	n. a.[4]
Size of neuropodial lobes	Small	Small	Prominent	Absent
	0.70	0.68	0.63	n. a.[4]
Tori	Absent	Absent	Absent	Absent
	0.87	0.98	0.99	n. a.[4]
Dorsal cirri	Absent	Absent	Absent	Absent
	1.00	1.00	0.90	n. a.[4]
Ventral cirri	Absent	Absent	Absent	Absent
	1.00	1.00	0.87	n. a.[4]
Branchiae	Absent	Absent	Absent	Absent
	0.95	0.99	0.98	n. a.[4]
Pygidial cirri	Absent	Absent	Present	Present
	0.87	0.86[5]	0.61	0.60
Nuchal organs	Present	Present	Present	Present
	1.00	1.00	1.00	1.00
Eyes	Present	Present	Present	Present
	0.76	0.53[6]	0.81	0.58
Chaetae	Present	Present	Present	Present
	0.98	0.93	0.94	0.89
Internalized supporting chaetae	Absent	Absent	Absent	Absent
	0.95	0.97	0.61	0.98
Compound chaetae	Absent	Absent	Absent	Absent
	1.00	0.95	0.97	1.00
Forked or lyrated chaetae	Absent	Absent	Absent	Absent
	1.00	1.00	1.00	1.00
Uncini/hooks	Absent	Absent	Absent	Absent
	0.97	1.00	1.00	1.00

[1]one ring reappears in Orbiniidae
[2]reappears in Nerillidae
[3]reappear in Saccocirridae
[4]only coded for Saccocirridae
[5]present in both Nerillidae and Orbiniidae
[6]present in both Orbiniidae and Dinophilidae

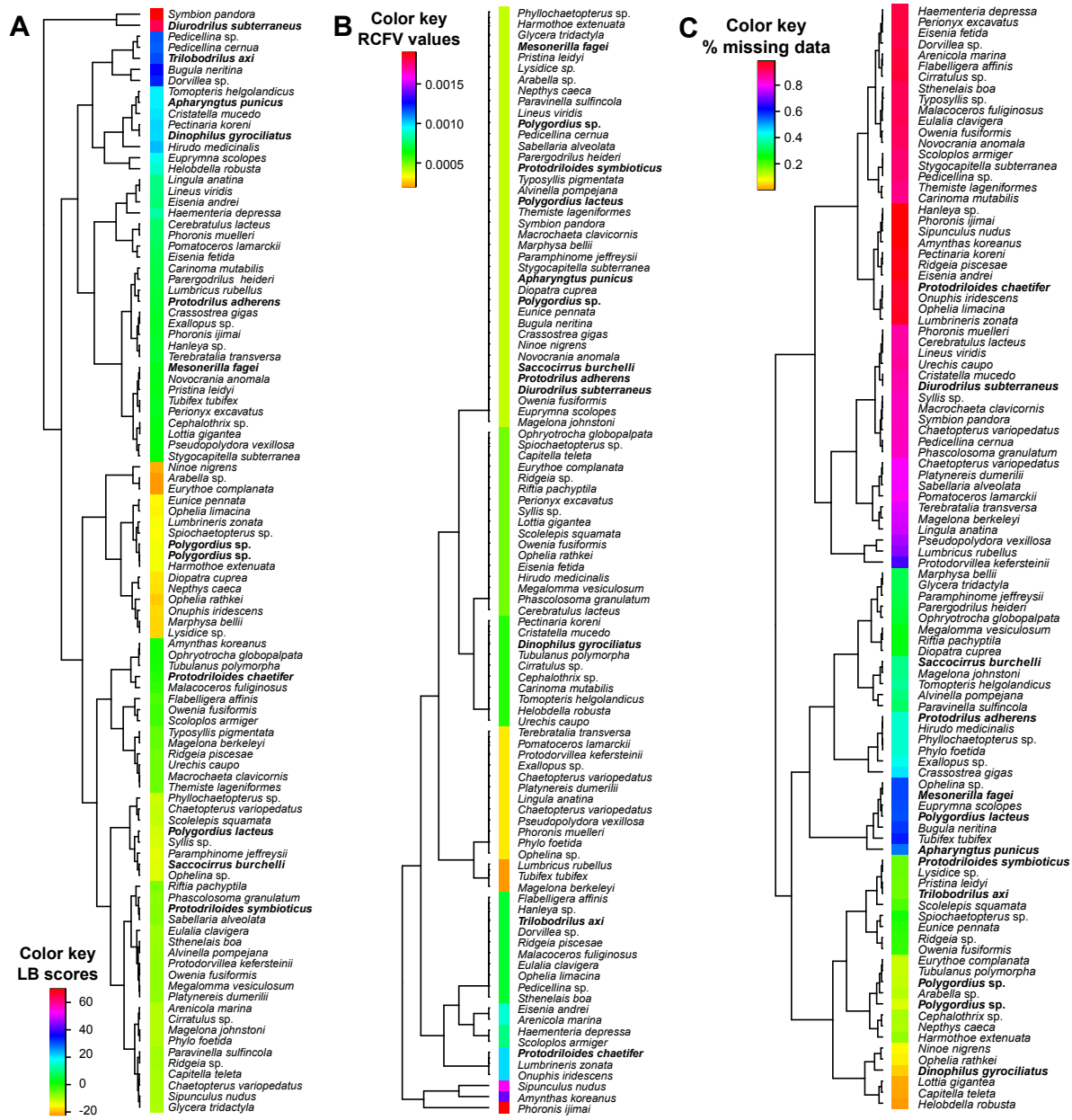

Figure S1. Trees of clustering analyses based on species-specific measurements. Trees were generated using hierarchical clustering analyses based on metric values of potential biases. These analyses showed that in all cases, interstitial species were spread throughout each tree of each clustering analysis conducted. Hence, positions of interstitial species shown in Figure 1 could not merely be attributed to possession of shared biases. (A) Long branch heterogeneity, (B) compositional heterogeneity and (C) degree of missing data. Values of the (A) LB scores, (B) RCFV and (C) degree of missing data of each species are indicated by a color-coded heatmap left to the species name. Species of archiannelid taxa are in bold. Figure S1 is related to Figure 2.

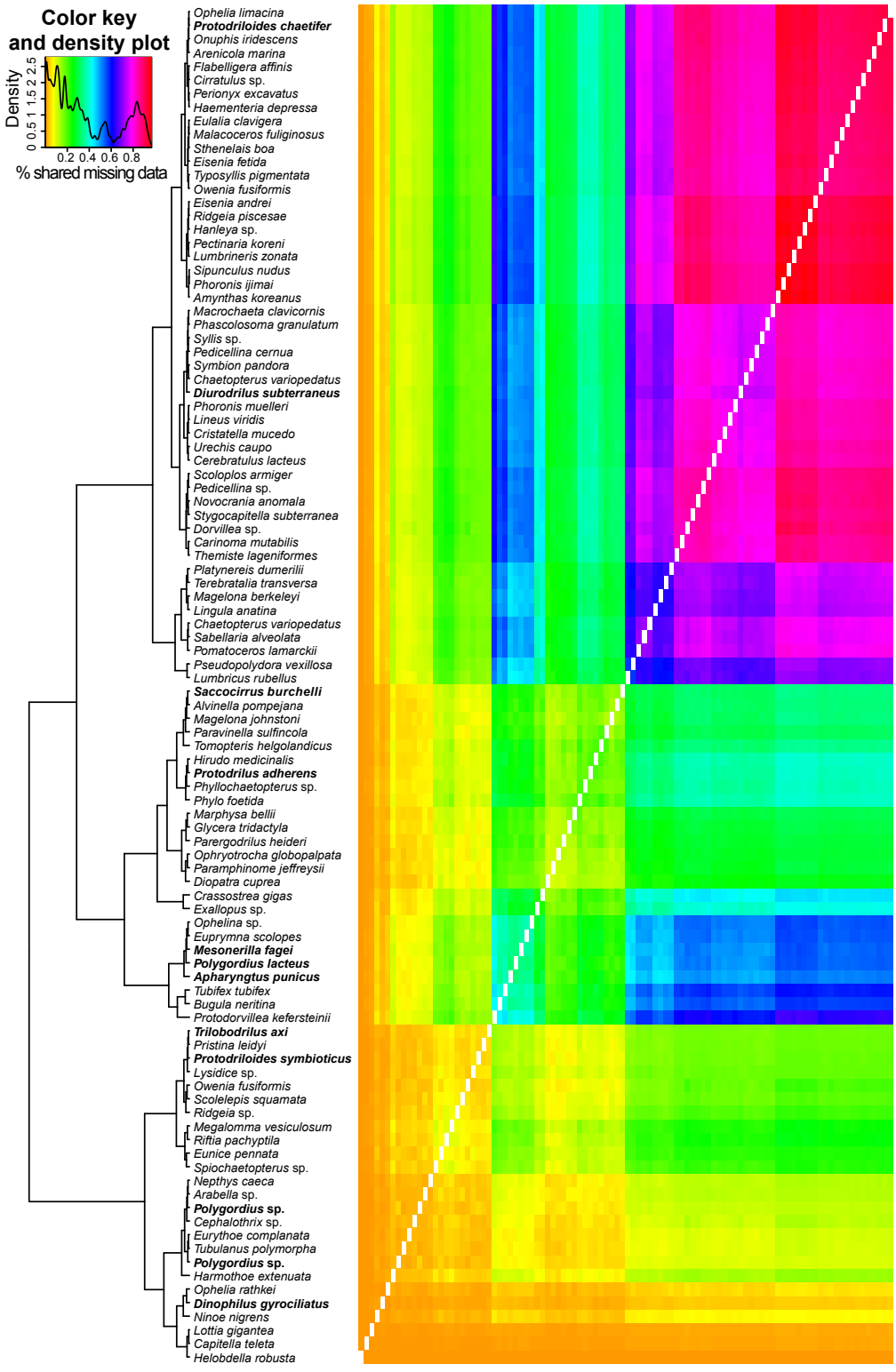

Figure S2. Tree of the clustering and heatmap analysis based on degrees of shared missing data between each species pair. Tree was generated using a hierarchical clustering analysis. Interstitial species were spread throughout the tree of the clustering analysis. Hence, positions of interstitial species shown in Figure 1 could not merely be attributed to possession of shared bias. Degrees of shared missing data are indicated by a color-coded heatmap. The order of the species in the columns of the heatmap from left to right is the same as in the rows from bottom up. The density plot of degrees of shared missing data is also shown in the upper left corner. Species of archiannelid taxa are in bold. Figure S2 is related to Figure 2.

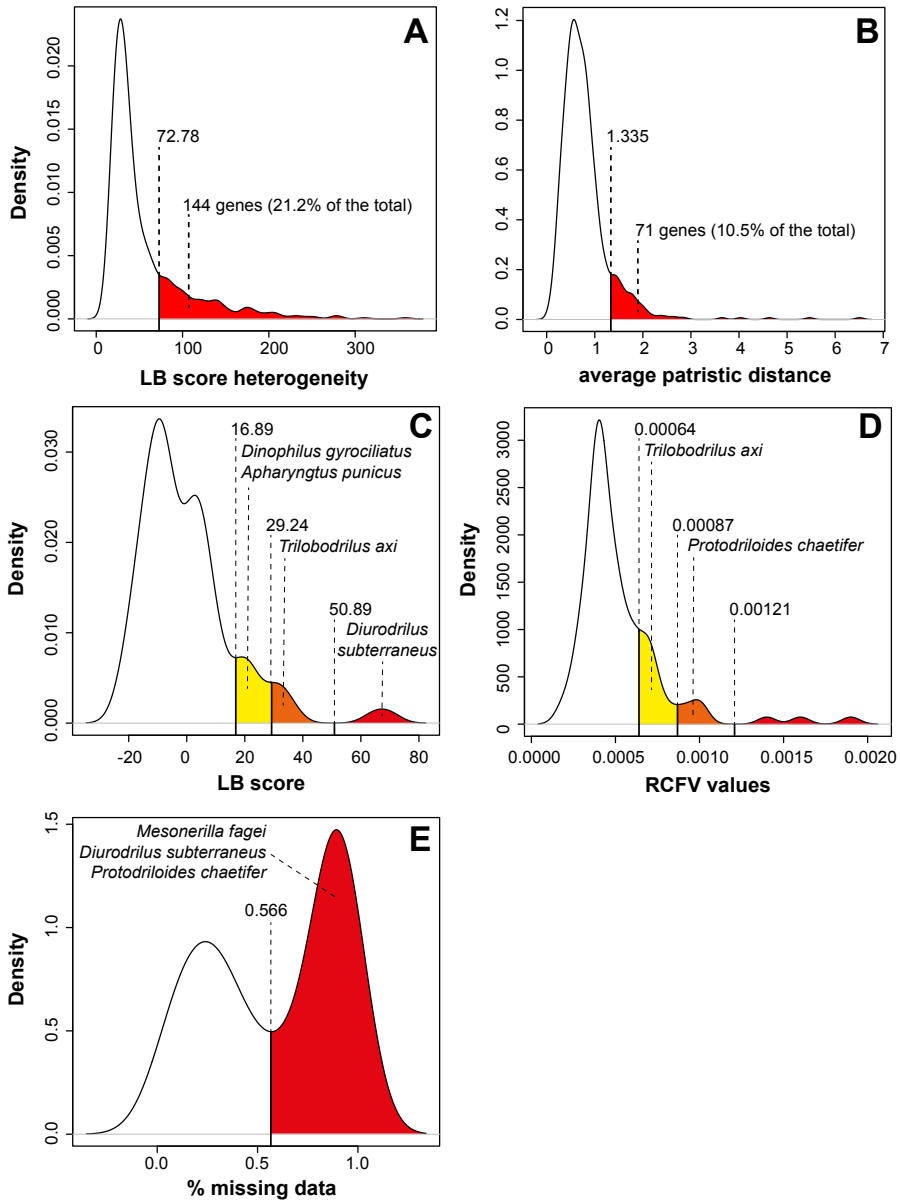

Figure S3. Density plots of the partition-specific as well as species-specific measurements for different biases. Partitions or species, which were part of the skewed right tails of normal distributions, were excluded. (A) branch-length heterogeneity and (B) evolutionary rate for partitions and (C) long branches, (D) compositional heterogeneity and (E) degree of missing data for species. (A & B) Branch-length heterogeneity was assessed using the standard deviation of the species-specific LB scores in each partitions and evolutionary rate based on the average patristic distance between each pair of species within a partition. (C-E) LB scores (A), RCFV values (B) and percentage of missing data (C) of each species were used to assess these biases, respectively. (A-E) Area of the skewed right tail of the normal distribution is indicated as well as the threshold value. (A & B) Number of partitions (i.e, genes), which are part of this area and were excluded (datasets d07 & 08 in Table 1), are also given as well as the percentage of the total number of partitions in the dataset d01. (C-E) Species including the ones from interstitial taxa, which were part of these areas, were successively excluded from the dataset d01 (datasets d09-11, d16-18 & d21 in Table 1). Species of interstitial annelid taxa, which were part of these areas, were then individually re-included into the datasets with the smallest number of taxa (i.e., d11, d18 & d21 in Table 1) resulting in datasets d12-15, d19-20 & d22-24 in Table 1. Figure S3 is related to Figure 1.

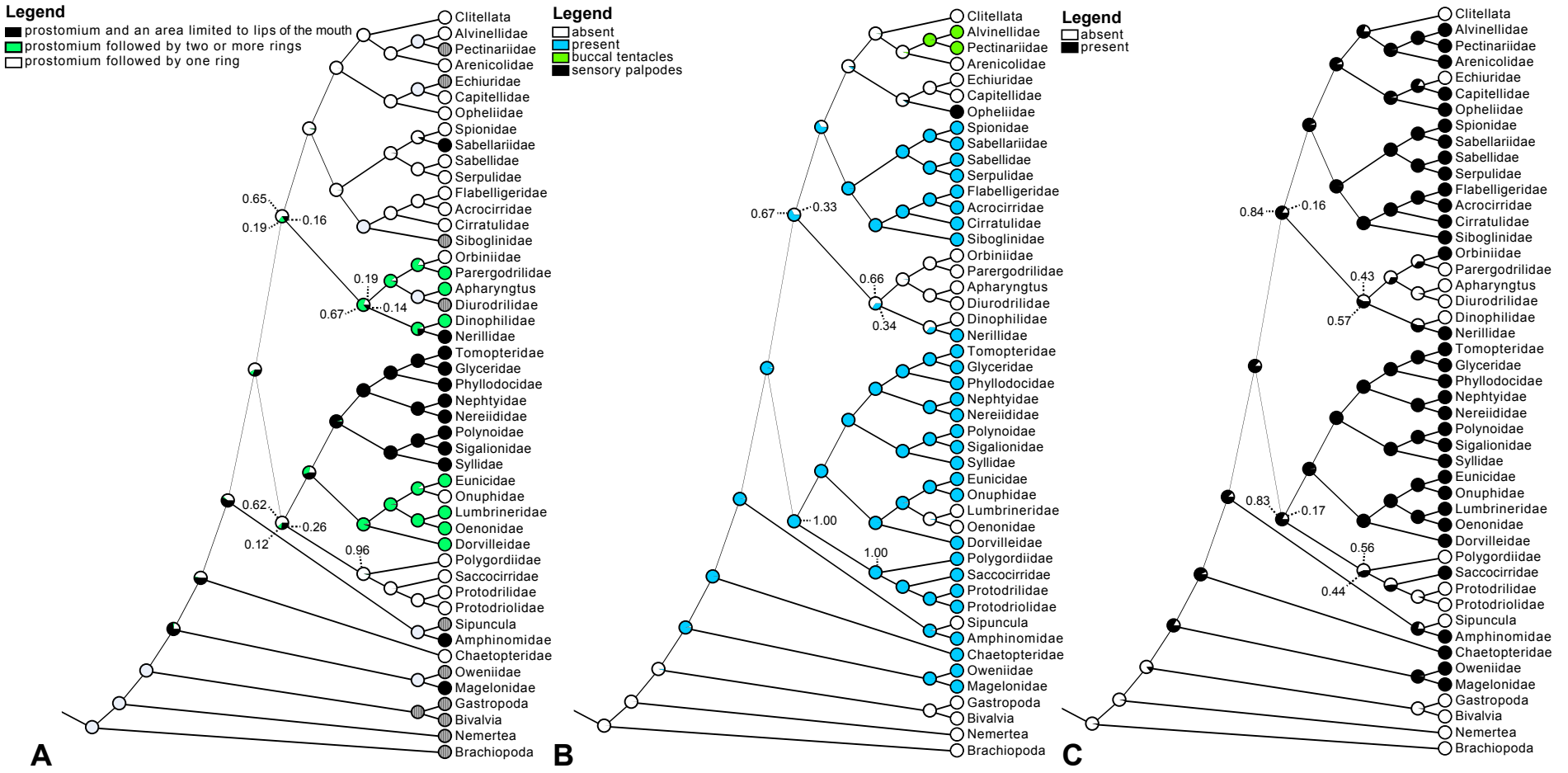

Figure S4. Maximum Likelihood mapping of the character states for three characters on the family-reduced tree of the dataset d01. (A) Head structure, (B) palps, and (C) parapodia. Clades of higher taxonomic units were collapsed. Proportional likelihoods p above 0.05 are indicated for the LCA of Sedentaria, Orbiniida, Erranta and Protodriliformia. Color coding for character states is indicated by the legend in upper left corner of each panel. Figure S4 is related to Figure 3.

Supplemental Experimental Procedures

Table listing information regarding species used in this analysis. Collection locality and tissue used for RNA extraction are provided for each species for which new data were acquired in this study (highlighted in bold). Source of data for additional taxa is also provided. Annelid and outgroup species are in alphabetical order.

Species	Locality	Tissue	Source
Alvinella pompejana	-	-	NCBI dbEST
Amynthas koreanus	-	-	NCBI dbEST
Apharyngtus punicus	**Estuary of Étang de Bigulgia, Corsica (France) N 42°39' 35"/E 009°27' 25"**	**10 specimens**	**NCBI SRA**
Arabella sp.	**Bocas del Toro, Panama**	**1 specimen**	**NCBI SRA**
Arenicola marina	-	-	NCBI dbEST
Capitella teleta	-	-	JGI
Chaetopterus variopedatus US	-	-	NCBI Trace Archive
Chaetopterus variopedatus FR	-	-	NCBI SRA
Cirratulus sp	-	-	NCBI dbEST
Dinophilus gyrociliatus	**Laboratory culture**	**~50 specimens**	**NCBI SRA**
Diopatra cuprea	**Bly Creek, SC (USA) N 33°20' 04.08"/W 079°12' 09.72"**	**few segments**	**NCBI SRA**
Diurodrilus subterraneus	**List (Germany) N 55°00' 04.30"/E 008°23' 19.87"**	**8 specimens**	**NCBI SRA**
Dorvillea sp.	**Bocas del Toro, Panama**	**1 specimen**	**NCBI SRA**
Eisenia andrei	-	-	NCBI dbEST
Eisenia fetida	-	-	NCBI dbEST
Eulalia clavigera	-	-	NCBI dbEST
Eunice pennata	**Massachusetts (USA) N 39°47' 13.8"/W 070°46' 17.7"**	**Few segments**	**NCBI SRA**
Eurythoe complanata	-	-	NCBI SRA
Exallopus sp.	**Northern California (USA) N 40°27' 07.8"/W 124°35 46.8"**	**1 specimen**	**NCBI SRA**
Flabelligera affinis	-	-	NCBI dbEST
Glycera tridactyla	-	-	NCBI SRA
Haementeria depressa	-	-	NCBI dbEST
Harmothoe extenuata	-	-	NCBI SRA
Helobdella robusta	-	-	JGI
Hirudo medicinalis	-	-	NCBI dbEST
Lumbricus rubellus	-	-	NCBI dbEST
Lumbrineris zonata	-	-	NCBI dbEST
Lysidice sp.	**Van Dyks Bay (South Africa) S 34°37' 03.0"/E 019°21' 21.0"**	**1 specimen**	**NCBI SRA**
Macrochaeta clavicornis	-	-	NCBI SRA
Magelona berkeleyi	-	-	NCBI SRA
Magelona johnstoni	-	-	NCBI SRA
Malacoceros fuliginosus	-	-	NCBI dbEST
Marphysa bellii	-	-	NCBI SRA
Megalomma vesiculosum	-	-	NCBI SRA
Mesonerilla fagei	**Roscoff (France) N 48°44' 09.00"/W 004°05' 32.70"**	**50 specimens**	**NCBI SRA**

Species	Locality	Tissue	Source
Nephtys caeca	-	-	NCBI SRA
Ninoe nigrens	**North Caolina (USA) N35°29' 25.2"/W 074°47' 58.8"**	**1 specimen**	**NCBI SRA**
Onuphis iridescens	-	-	NCBI dbEST
Ophelia limacina	-	-	NCBI dbEST
Ophelia rathkei	-	-	NCBI SRA
Ophelina sp.	-	-	NCBI SRA
Ophryotrocha globopalpata	**Washington State (USA) N47°57' 01.2"/W 129°05' 49.8"**	**1 specimen**	**NCBI SRA**
Owenia fusiformis FR	-	-	NCBI SRA
Owenia fusiformis DE	-	-	NCBI SRA
Paralvinella sulfincola	-	-	NCBI SRA
Paramphinome jeffreysii	-	-	NCBI SRA
Parergodrilus heideri	**Ebergötzen (Germany) N 55°34' 51.90"/E 010°04' 26.50"**	**18 specimens**	**NCBI SRA**
Pectinaria koreni	-	-	NCBI dbEST
Perionyx excavatus	-	-	NCBI dbEST
Phascolosoma granulatum	-	-	NCBI SRA
Phyllochaetopterus sp.	-	-	NCBI SRA
Phylo foetida	-	-	NCBI SRA
Platynereis dumerilii	-	-	NCBI dbEST
Polygordius lacteus	**Baie de Morlaix (France) N 48°42' 39"/ W 03°53' 50"**	**1 specimen**	**NCBI SRA**
Polygordius sp.	**Friday Harbour, WA (USA) N 48°32' 40"/ W 122°58' 58"**	**1 specimen**	**NCBI SRA**
Polygordius sp.	**Friday Harbour, WA (USA) N 48°32' 40"/ W 122°58' 58"**	**1 specimen**	**NCBI SRA**
Pomatoceros lamarckii	-	-	NCBI dbEST
Pristina leidyi	-	-	NCBI SRA
Protodorvillea kefersteini	**Roscoff (France)**	**1 specimen**	**NCBI SRA**
Protodriloides chaetifer	**List (Germany) N 55°01' 30.30"/E008°26' 08.80"**	**10 specimens**	**NCBI SRA**
Protodriloides symbioticus	**List (Germany) N 55°01' 30.30"/E008°26' 17.64"**	**5 specimens**	**NCBI SRA**
Protodrilus adhaerens	**Ile Callot (France) N 48°41' 37.62"/W003°55' 13.56"**	**6 specimens**	**NCBI SRA**
Pseudopolydora vexillosa	-	-	NCBI SRA
Ridgeia sp.	-	-	NCBI dbEST
Ridgeia piscesae	-	-	NCBI SRA
Riftia pachyptila	-	-	NCBI SRA
Sabellaria alveolata	-	-	NCBI SRA
Saccocirrus burchelli	**Saldanha bay (South Africa) S 33°00' 25.02"/E017°56' 45.60"**	**1 specimen**	**NCBI SRA**
Scolelepis squamata	-	-	NCBI SRA
Scoloplos armiger	-	-	NCBI SRA
Sipunculus nudus	-	-	NCBI dbEST
Spiochaetopterus sp.	-	-	NCBI SRA
Sthenelais boa	-	-	NCBI dbEST
Stygocapitella subterranea	**Glenancross (UK) N 56°56' 38.94"/W 006°51' 11.64"**	**20 specimens**	**NCBI SRA**

Species	Locality	Tissue	Source
Syllis sp.	-	-	NCBI SRA
Themiste lageniformes	-	-	NCBI Trace Archive
Tomopteris helgolandicus	-	-	NCBI SRA
Trilobodrilus axi	**Rantum (Germany) N54°48' 41.52"/E 008°17' 00.37"**	**50 specimens**	**NCBI SRA**
Tubifex tubifex	-	-	NCBI dbEST
Typosyllis pigmentata	-	-	NCBI dbEST
Urechis caupo	-	-	NCBI Trace Archive
Outgroup taxa			
Bugula neritina	-	-	NCBI SRA
Carinoma mutabilis	-	-	NCBI dbEST
Cephalothrix sp.	-	-	NCBI SRA
Cerebratulus lacteus	-	-	NCBI Trace Archive
Crassostrea gigas	-	-	JGI
Cristatella mucedo	-	-	Struck et al [S1]
Euprymna scolopes	-	-	NCBI dbEST
Hanleya sp.	-	-	NCBI SRA
Lineus viridis	-	-	Struck & Fisse [S2]
Lottia gigantea	-	-	JGI
Lingula anatina	-	-	NCBI SRA
Novocrania anomala	-	-	Helmkampf et al [S3]
Pedicellina cernua	-	-	NCBI Trace Archive/ EST
Pedicellina sp.	-	-	NCBI Trace Archive
Phoronis ijimai	-	-	NCBI Trace Archive
Phoronis muelleri	-	-	Helmkampf et al. [S3]
Symbion pandora	-	-	NCBI Trace Archive
Terebratalia transversa	-	-	NCBI Trace Archive
Tubulanus polymorpha	-	-	NCBI SRA

Data generation. The table above lists the collection localities and tissue material of the 23 annelid species for which new transcriptomic libraries have been determined. Upon collection, samples were either snap-frozen at -80^{o}C or stored in RNAlater. For *Apharyngtus punicus*, *Diurodrilus subterraneus*, *Mesonerilla fagei*, *Parergodrilus heideri*, *Polygordius lacteus*, *Protodriloides chaetifer*, *Protodrilus adhaerens*, *Stygocapitella subterranea* and *Trilobodrilus axi* library reconstruction and multiplex sequencing on a Illumina Hi-Seq 2000 was conducted using the protocol for interstitial species like gastrotrichs and gnathostomulids as described as in Struck et al. [S4]. For the other species the procedure followed the protocols described in Weigert et al. [S5] for multiplex sequencing on an Illumina Hi-Seq 2000. For *Protodorvillea kefersteini* and *Saccocirrus burchelli* the same protocol has been used as for, e.g., *Magelona johnstoni* [S5]. For *Arabella* sp., *Dinophilus gyrociliatus*, *Diopatra cuprea*, *Dorvillea* sp., *Eunice pennata*, *Exallopus* sp., *Lysidice* sp., *Ninoe nigrens*, *Ophryotrocha globopalpata*, *Protodriloides symbioticus* and both *Polygordius* sp. libraries were generated in the laboratory of KMH employing the protocol, which had be described for, e.g., *Magelona berkeleyi* [S5]. All our original sequence data has been deposited in the NCBI short read archive and can be accessed via the BioProject PRJNA282709.

Data assembly. Bases were called with IBIS 1.1.2 [S6], adaptor and primer sequences were removed and reads with low complexity as well as false paired indices were discarded. Raw data of all libraries were trimmed by applying a filter of 15, discarding all reads with more than 5 bases below a quality score of 15. Subsequently all libraries were assembled *de novo* using either the CLC Genomics Workbench 5.1 (CLC bio, Århus, Denmark) or as described elsewhere [S7, 8]. The settings for the CLC assembly were as follows: mismatch cost 3; insertion cost 3; deletion cost 3; length fraction 0.5; similarity fraction 0.8; minimum contig length 200; automatic word size; automatic bubble size; and contig adjustment by mapped reads. The assembled transcriptomic and genomic data of 77 annelid and outgroup taxa of the

analyses of Weigert et al. [S5] complemented these data. For the species information see the table above and for the library information the one below. All datasets used in and results of this study are available at DataDryad.

Orthology prediction was performed using HaMStR version 8b [S9] with the representative option and the core-ortholog set lophotrochozoa_hmmer3 comprising seven primer taxa (*Helobdella robusta*, *Capitella teleta*, *Lottia gigantea*, *Schistosoma mansoni*, *Daphnia pulex*, *Apis mellifera* and *Caenorhabditis elegans*) and 2,339 orthologous genes [S5]. Redundant sequences were eliminated using a custom Perl script, which checked for the presence of redundant sequence identifiers in the assigned orthologous genes for each taxon. Additionally, we checked our putative orthologous sequences for contamination with protist sequences using local BLAST against the Apicomplexa proteome. We chose Apicomplexa for comparisons as gregarine parasites have been described for several annelids. Orthologous sequences which were more similar to Apicomplexa than to the reference-taxon *Helobdella robusta* were discarded.

Each set of orthologous genes was individually aligned using MAFFT-Linsi [S10] followed by the masking of gaps as well as highly diverse amino acid positions with REAP [S11]. All masked single gene alignments were concatenated into a supermatrix using a custom Perl script. The super-matrix was reduced based on the phylogenetic signal in a gene by assessing the tree-likeness by quartet-mapping using extended geometry mapping as implemented in MARE with a *d* value of 1.0 and keeping all taxa [S12]. This resulted in a supermatrix comprising 679 genes and 100 species (dataset d01 in Table 1).

Table listing information on the libraries used in the study. New libraries are highlighted in bold.

Species	Reads/ESTs	Contigs	Assembly	Genes present in dataset d01
Annelida				
Alvinella pompejana	142,322	25,723	MGIBlast/CAP3	583
Amynthas koreanus	1,118	228	MIRA	23
Apharyngtus punicus	**54,004,596**	**153,425**	**CLC**	**525**
Arabella sp.	**52,481,310**	**217,183**	**CLC**	**653**
Arenicola marina	2,191	1,262	MGIBlast/CAP3	92
Capitella teleta	-	20,982	MGIBlast/CAP3	679
Chaetopterus variopedatus FR	604,208	2,198	CLC	170
Chaetopterus variopedatus US	647	453	MGIBlast/CAP3	229
Cirratulus sp.	3,152	1,038	MGIBlast/CAP3	63
Dinophilus gyrociliatus	**58,467,188**	**188,729**	**Trinity**	**664**
Diopatra cuprea	**55,937,358**	**138,779**	**Trinity**	**627**
Diurodrilus subterraneus	**15,564,570**	**8,063**	**CLC**	**259**
Dorvillea sp.	**49,518,260**	**150,550**	**Trinity**	**164**
Eisenia andrei	1,108	710	MGIBlast/CAP3	58
Eisenia fetida	3,041	2,181	MGIBlast/CAP3	128
Eulalia clavigera	1,078	697	MGIBlast/CAP3	91
Eunice pennata	**35,650,614**	**93,814**	**Trinity**	**635**
Eurythoe complanata	-	548,027	Velvet/Oases	668
Exallopus sp.	**24,462,060**	**55,654**	**Trinity**	**520**
Flabelligera affinis	869	669	MGIBlast/CAP3	85
Glycera tridactyla	42,796,389	90,503	CLC	608
Haementeria depressa	890	499	MGIBlast/CAP3	93
Harmothoe extenuata	17,673,128	77,176	CLC	666
Helobdella robusta	-	20,911	MGIBlast/CAP3	679
Hirudo medicinalis	26,833	10,582	MGIBlast/CAP3	498
Lumbricus rubellus	19,888	10,386	MGIBlast/CAP3	328
Lumbrineris zonata	685	364	MGIBlast/CAP3	51
Lysidice sp.	**39,983,000**	**144,726**	**Trinity**	**662**
Macrochaeta clavicornis	540,542	1,883	CLC	194
Magelona berkeleyi	21,950,214	24,038	CLC	362
Magelona johnstoni	9,611,241	38,512	CLC	590
Malacoceros fuliginosus	1,008	679	MGIBlast/CAP3	95

Species	Reads/ESTs	Contigs	Assembly	Genes present in dataset d01
Marphysa bellii	13,950,047	62,151	CLC	618
Megalomma vesiculosum	21,177,922	74,048	CLC	632
Mesonerilla fagei	**52,075,977**	**38,254**	**CLC**	**486**
Nephtys caeca	23,873,733	93,461	CLC	661
Ninoe nigrens	**80,683,422**	**151,183**	**Trinity**	**671**
Onuphis iridescens	665	381	MGIBlast/CAP3	64
Ophelia limacina	368	276	MGIBlast/CAP3	58
Ophelia rathkei	-	215,807	Velvet/Oases	672
Ophelina sp.	647,439	20,068	CLC	531
Ophryotrocha globopalpata	**31,700,000**	**129,450**	**Trinity**	**626**
Owenia fusiformis FR	56,363,524	78,177	CLC	588
Owenia fusiformis DE	947	572	MGIBlast/CAP3	95
Paralvinella sulfincola	2,593,853	96,961	CLC	624
Paramphinome jeffreysii	39,877,015	71,866	CLC	606
Parergodrilus heideri	**11,569,822**	**32,340**	**CLC**	**645**
Pectinaria koreni	672	255	MGIBlast/CAP3	29
Perionyx excavatus	1,159	823	MGIBlast/CAP3	92
Phascolosoma granulatum	699,044	2,261	CLC	193
Phyllochaetopterus sp.	14,375,156	24,856	CLC	533
Phylo foetida	10,452,849	29,965	CLC	569
Platynereis dumerilii	1,487	444	MGIBlast/CAP3	170
Polygordius lacteus	**63,866,525**	**86,444**	**CLC**	**339**
Polygordius sp.	**38,808,702**	**106,549**	**Trinity**	**662**
Polygordius sp.	**64,099,584**	**171,704**	**Trinity**	**663**
Pomatoceros lamarckii	4,132	2,315	MGIBlast/CAP3	204
Pristina leidyi	1,550,174	80,655	CLC	660
Protodorvillea kefersteini	**19,430,629**	**18,923**	**CLC**	**446**
Protodriloides chaetifer	**685,811**	**887**	**CLC**	**74**
Protodriloides symbioticus	**44,840,470**	**189,477**	**Trinity**	**665**
Protodrilus adhaerens	**42,272,870**	**42,096**	**CLC**	**589**
Pseudopolydora vexillosa	288,128	15,505	CLC	329
Ridgeia sp.	1,092,906	33,987	MGIBlast/CAP3	617
Ridgeia piscesae	502	402	CLC	50
Riftia pachyptila	1,333,110	131,978	CLC	645
Sabellaria alveolata	2,714,531	12,258	CLC	264
Saccocirrus burchelli	**58,871,023**	**47,009**	**CLC**	**598**
Scolelepis squamata	9,737,753	37,024	CLC	627
Scoloplos armiger	798,532	1,351	CLC	133
Sipunculus nudus	2,317	1,541	MGIBlast/CAP3	33
Spiochaetopterus sp.	20,317,772	56,616	CLC	616
Sthenelais boa	1,005	497	MGIBlast/CAP3	88
Stygocapitella subterranea	**601,812**	**1,325**	**CLC**	**155**
Syllis sp.	601,004	3,355	CLC	190
Themiste lageniformes	1,690	1,117	MGIBlast/CAP3	142
Tomopteris helgolandicus	11,700,720	22,582	CLC	623
Trilobodrilus axi	**59,807,531**	**31,618**	**CLC**	**656**
Tubifex tubifex	17,009	8,008	MGIBlast/CAP3	385
Typosyllis pigmentata	730	523	MGIBlast/CAP3	93
Urechis caupo	2,003	845	MGIBlast/CAP3	142
Outgroup				
Bugula neritina	3,112	5,142	MGIBlast/CAP3	463
Carinoma mutabilis	3,039	872	MGIBlast/CAP3	125
Cephalothrix sp.	23,104,655	44,041	CLC	619
Cerebratulus lacteus	5,127	1,676	MGIBlast/CAP3	160
Crassostrea gigas	26,571	17,258	MGIBlast/CAP3	558
Cristatella mucedo	2,981	668	MGIBlast/CAP3	151
Euprymna scolopes	77,159	27,036	MGIBlast/CAP3	489
Hanleya sp.	1,096	73	CLC	36
Lineus viridis	4,515	3,235	MGIBlast/CAP3	162
Lottia gigantea	-	23,851	MGIBlast/CAP3	676
Lingula anatina	70,309	5,748	CLC	307
Novocrania anomala	2,247	1,699	MGIBlast/CAP3	117
Pedicellina cernua	7,777	2,348	MGIBlast/CAP3	164
Pedicellina sp.	1,907	648	MGIBlast/CAP3	119
Phoronis ijimai	1,517	376	MGIBlast/CAP3	36
Phoronis muelleri	2,315	1,467	MGIBlast/CAP3	157

Species	Reads/ESTs	Contigs	Assembly	Genes present in dataset d01
Symbion pandora	3,763	949	MGIBlast/CAP3	169
Terebratalia transversa	3,154	1,902	MGIBlast/CAP3	238
Tubulanus polymorpha	13,645,909	37,851	CLC	649

Phylogenetic analyses and testing. The most appropriate substitution model was LG+I+Γ as determined by the ProteinModelSelection script for RAxML [S13]. We conducted several ML analyses as part of the sensitivity analyses. In total, 679 single-gene and 24 supermatrices ML analyses were conducted with RAxML [S13] using 100 and 5*100 bootstrap replicate searches, respectively. Each bootstrap search was followed by a search for the best tree. For supermatrix analyses, the 500 bootstrap replicates were plotted onto the best ML tree of the 5 runs for each supermatrix dataset.

To test if the best ML tree of the largest dataset d01 differed significantly from *a priori* hypotheses based on morphology [S14] (Table S1), per-site log-likelihoods of the best trees from unconstrained and constrained analyses of the largest dataset d01 were computed using RAxML [S13]. For hypothesis testing, an AU-test was conducted as well as posterior probabilities calculated using CONSEL [S15].

Sensitivity analyses. The 679 genes were screened to test if the placements of the interstitial annelid taxa found in the best ML trees were due to paralogous sequences and contamination only [S16]. For this purpose, the *a posteriori*-screening implemented in TreSpEx [S17] (i.e., option -fun b) was used, which specifically searched the trees of the single-gene analyses for groupings with a bootstrap value of 95 or higher and that also comprised taxa of either the clade *Apharyngtus*, Dinophilidae, Diurodrilidae, Nerillidae, Orbiniidae and Parergodrilidae or the clade Protodrilida, Polygordiidae and the other Errantia (Figure 1). However, all groupings congruent with subclades of these two clades, i.e., consisting only of species from either Dinophilidae, Orbiniidae plus Parergodrilidae, Eunicida, Phyllodocida, Protodrilida, or Polygordiidae, were not considered using the masking option -gts. During the screening groupings, which contained no more than three species and a least one species with a branch length below 0.00001, were regarded as indicative of cross-contamination (i.e., using the -possb option). To determine if the detected groupings, which were not indicated as contaminations, indicated the presence of paralogy blast searches of all sequences from the dataset containing the detected grouping were conducted in TreSpEx (i.e., -fun c) against the reference databases of *Bos taurus* and *Branchiostoma floridae* with the upper limit of the e-value set to e^{-20}. The automatic sorting options of TreSpEx were 0.9 for the non-paralogy and 0.1 for paralogy. The blast searches of two datasets did not return hits with e values better than e^{-20} for the sequences of the detected grouping. Hence, we also conducted a search against the reference databases of *Helobdella robusta* and *Capitella teleta*, but could not find different results. Sequences of groupings, which showed signs of cross-contamination or paralogy or had no hits, were pruned from the corresponding single-gene dataset. All 679 single-gene datasets were then concatenated into a supermatrix using FASconCAT [S18] (i.e., dataset d02 in Table 1).

To investigate the influence of missing data within a gene partition we generated two datasets from the largest dataset d01 by increasing the *d* values of MARE to 1.5 and 2.0, respectively, and, hence, reducing the number of genes in the supermatrix with a high degrees of missing data (datasets d03 & d04 in Table 1). To assess the effect of genes with either very heterogeneous evolutionary rates across species or over-all high rates on the reconstruction we determined the standard deviation of all LB scores as well as the average patristic distance of each gene as measurements of branch-length heterogeneity and evolutionary rate, respectively, using TreSpEx [S17] and the best ML trees of the 679 single-gene analyses. Exclusion of partitions was not based on arbitrary thresholds (e.g., 10% or 20% of partitions with the highest degree of missing data), but instead density plots were generated for each

type of potential bias examined (e.g., degree of missing data). The density plots of these values were generated using R. Partitions, which were part of the skewed right tails of normal distributions, were excluded (Figure S3). The remaining genes were concatenated into supermatrices using FASconCAT [S18] (i.e., dataset d07 & d08 in Table 1).

Besides the effect of the gene partitions on the reconstruction we also assess the influence of biased species. Therefore, to assess if a species is long branched we determined the LB score using TreSpEx [S17] and the best ML tree of dataset d01. Additionally, we determined the degree of missing data for each species and the compositional heterogeneity as measured by the taxon-specific RCFV using BaCoCa [S19] and the same dataset. Heatmap and clustering analyses as well as density plots of these species-specific values were generated using R. As for the partitions biased species were excluded from the dataset d01 based on the density plots of the biases (Figure S3) resulting in datasets d09-11, d16-18 and d21 in Table 1. To test if the position of a biased interstitial annelid species is different, when the similarly biased other species are excluded from the analyses, we individually re-included the biased interstitial species to the datasets with the smallest number of species resulting in the additional datasets d12-15, d19-20 and d21-24 in Table 1.

To further explore the stability of the positions of interstitial annelid species in analyses, we calculated the bootstrap trees of ML analyses of the datasets d01, d03 & d04 comprising all species sampled. For these three datasets, we separately summarized the bootstrap for each clade found in at least one of the 500 replicates using TreSpEx [S17] to gain additional insight into the support for different hypotheses of unstable *Diurodrilus subterraneus*.

Ancestral state reconstruction. Ancestral states were reconstructed based on the maximum likelihood tree of the largest dataset d01 shown in Figure 1 with the clades of higher annelid taxonomic units (i.e., polychaete families, Clitellata, Echiura, Sipuncula) collapsed and including Gastropoda, Bivalvia, Nemertea and Brachiopoda as outgroups. The morphological data matrix of Weigert et al [S5] was adapted by adding additional taxa (i.e., the "archiannelid" families, Dorvilleidae and Oenonidae). Ancestral states were computed with Mesquite using the maximum likelihood reconstruction with Markov k-state 1 parameter model [S20]. Character coding of Protodrilida families, Polygordiidae, Dinophilidae, Nerillidae, Oenonidae and Dorvillidae followed Zrzavy et al [S21] and for characters not considered by them based on Westheide [S22] for Protodrilida families, Polygordiidae, Dinophilidae and Nerillidae and Rouse & Pleijel [S23] for Oenonidae and Dorvillidae. Diurodrilidae and *Apharyngtus* were not included in the analyses of Zrzavy et al [S21] and coded based on Worsaae & Rouse [S24] and Westheide [S22] for Diurodrilidae and Westheide [S25] for *Apharyngtus*.

Moreover, a few character were coded differently from Zrzavy et al [S21]. Polygordiidae were coded as possessing a coelom as well as palps and not lateral antennae according to Westheide [S22]. Nerillidae possess a median antennae [S22]. With respect to the character head structure for some interstitial taxa, recent results of Dinophilidae have to be taken into consideration. In dinophilids the prostomium is followed by a ring bearing the mouth and then by two rings, which possess only one pair of ganglia, the suboesophageal ganglion, while all following rings possess one pair each (see Fig. 7 in Müller & Westheide [S26]). This indicates that more than one ring is integrated into the head structure of Dinophilidae. Although investigations of the nervous system of *Apharngtus* are currently lacking, the position of the mouth at the very posterior end of the first ring indicates the suboesophageal ganglion, if present, could only occur in the second ring similar to dinophilids [S25]. Hence, the head structure with a prostomium followed by two rings is also very likely for *Apharyngtus*. In parergodrilids, the prostomium is followed by a ring bearing the mouth and a so-called achaetous segment with two rings and then by several segments bearing chaetae [S22]. As in dinophilids and *Apharyngtus* the mouth is located at the very posterior end of the first ring [S27, 28]. Moreover, the suboesophageal ganglion is not located in the

ring with the mouth, but in the following next two rings of the so-called achaetous segment (Fig. 14 in Reisinger [S28]). Indeed the suboesophageal ganglion is slightly fused with the ganglion of the first chaetiger. In all other chaetigers the ganglion is positioned at the anterior end of the segment similar to the organization observed in dinophilids. Hence, as for dinophilids the head of parergodrilids incorporates the prostomium and more than one ring. For Diurodrilidae, rings or segments are not discernable in the outer body organization [S24] and, thus, it cannot be concluded if one or more rings follow the prostomium as part of the head.

Supplemental References

S1. Struck, T.H., Paul, C., Hill, N., Hartmann, S., Hosel, C., Kube, M., Lieb, B., Meyer, A., Tiedemann, R., Purschke, G., et al. (2011). Phylogenomic analyses unravel annelid evolution. Nature *471*, 95-98.

S2. Struck, T.H., and Fisse, F. (2008). Phylogenetic position of nemertea derived from phylogenomic data. Molecular biology and evolution *25*, 728-736.

S3. Helmkampf, M., Bruchhaus, I., and Hausdorf, B. (2008). Phylogenomic analyses of lophophorates (brachiopods, phoronids and bryozoans) confirm the Lophotrochozoa concept. P Roy Soc B-Biol Sci *275*, 1927-1933.

S4. Struck, T.H., Wey-Fabrizius, A.R., Golombek, A., Hering, L., Weigert, A., Bleidorn, C., Klebow, S., Iakovenko, N., Hausdorf, B., Petersen, M., et al. (2014). Platyzoan paraphyly based on phylogenomic data supports a non-coelomate ancestry of Spiralia. Mol. Biol. Evol. *31*, 1833-1849.

S5. Weigert, A., Helm, C., Meyer, M., Nickel, B., Arendt, D., Hausdorf, B., Santos, S.R., Halanych, K.M., Purschke, G., Bleidorn, C., et al. (2014). Illuminating the base of the annelid tree using transcriptomics. Mol. Biol. Evol. *31*, 1391-1401.

S6. Kircher, M., Stenzel, U., and Kelso, J. (2009). Improved base calling for the Illumina Genome Analyzer using machine learning strategies. Genome Biol *10*.

S7. Moroz, L.L., Kocot, K.M., Citarella, M.R., Dosung, S., Norekian, T.P., Povolotskaya, I.S., Grigorenko, A.P., Dailey, C., Berezikov, E., Buckley, K.M., et al. (2014). The ctenophore genome and the evolutionary origins of neural systems. Nature *510*, 109–114.

S8. Cannon, J.T., Kocot, K.M., Waits, D.S., Weese, D.A., Swalla, B.J., Santos, S.R., and Halanych, K.M. (2014). Phylogenomic resolution of the hemichordate and echinoderm clade. Curr. Biol. *24*, 2827-2832.

S9. Ebersberger, I., Strauss, S., and von Haeseler, A. (2009). HaMStR: Profile hidden markov model based search for orthologs in ESTs. BMC Evol Biol *9*, 157.

S10. Katoh, K., Kuma, K.-i., Toh, H., and Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. *33*, 511-518.

S11. Hartmann, S., and Vision, T.J. (2008). Using ESTs for phylogenomics: Can one accurately infer a phylogenetic tree from a gappy alignment? BMC Evol Biol *8*.

S12. Meusemann, K., von Reumont, B.M., Simon, S., Roeding, F., Strauss, S., Kück, P., Ebersberger, I., Walzl, M., Pass, G., Breuers, S., et al. (2010). A phylogenomic approach to resolve the arthropod Tree of Life. Mol. Biol. Evol. *27*, 2451-2464.

S13. Stamatakis, A. (2006). RAxML-VI-HPC: Maximum Likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics *22*, 2688–2690.

S14. Struck, T.H. (2006). Progenetic species in polychaetes (Annelida) and problems assessing their phylogenetic affiliation. Integr. Comp. Biol. *46*, 558–568.

S15. Shimodaira, H., and Hasegawa, M. (2001). CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics *17*, 1246-1247.
S16. Struck, T.H. (2013). The impact of paralogy on phylogenomic studies – A case study on annelid relationships. PLoS ONE *8*, e62892.
S17. Struck, T.H. (2014). TreSpEx – Detection of misleading signal in phylogenetic reconstructions based on tree information. Evolutionary Bioinformatics *10*, 51-67.
S18. Kück, P., and Meusemann, K. (2010). FASconCAT: Convenient handling of data matrices. Mol. Phylogenet. Evol. *56*, 1115-1118.
S19. Kück, P., and Struck, T.H. (2014). BaCoCa - A heuristic software tool for the parallel assessment of sequence biases in hundreds of gene and taxon partitions. Mol. Phylogenet. Evol. *70*, 94–98.
S20. Maddison, W.P., and Maddison, D.R. (2009). Mesquite: a modular system for evolutionary analysis., 2.71 Edition.
S21. Zrzavy, J., Riha, P., Pialek, L., and Janouskovec, J. (2009). Phylogeny of Annelida (Lophotrochozoa): total-evidence analysis of morphology and six genes. BMC Evol. Biol. *9*, 189.
S22. Westheide, W. (2008). Polychaetes: Interstitial families., Second Edition Edition, (Shrewsbury: Field Studies Council).
S23. Rouse, G.W., and Pleijel, F. (2001). Polychaetes, (Oxford: University Press).
S24. Worsaae, K., and Rouse, G.W. (2008). Is *Diurodrilus* an annelid? J. Morphol. *269*, 1426-1455.
S25. Westheide, W. (1971). *Apharyngtus punicus* nov. gen. nov. spec., ein aberranter Archiannelide aus dem Mesopsammal der tunesischen Mittelmeerküste. Mikrofauna Meeresb. *6*.
S26. Müller, M.C.M., and Westheide, W. (2002). Comparative analysis of the nervous systems in presumptive progenetic dinophilid and dorvilleid polychaetes (Annelida) by immunohistochemistry and cLSM. Acta Zool. *83*, 33-48.
S27. Karling, T.G. (1958). Zur Kenntnis von *Stygocapitella subterranea* Knöllner und *Parergodrilus heideri* Reisinger. Arkiv för Zoologi *1*, 307-324.
S28. Reisinger, E. (1960). Die Lösung des *Parergodrilus*-Problems. Z. Morphol. Ökol. Tiere *48*, 517-544.