Phylogenomic analyses unravel annelid evolution

Torsten H. Struck¹, Christiane Paul², Natascha Hill³, Stefanie Hartmann³, Christoph Hösel¹, Michael Kube⁴, Bernhard Lieb⁵, Achim Meyer³, Ralph Tiedemann³, Günter Purschke¹ & Christoph Bleidorn⁶,‡

Annelida, the ringed worms, is a highly diverse animal phylum that includes more than 15,000 described species and constitutes the dominant benthic macrofauna from the intertidal zone down to the deep sea. A robust annelid phylogeny would shape our understanding of animal body-plan evolution and shed light on the bilaterian ground pattern. Traditionally, Annelida has been split into two major groups: Clitellata (earthworms and leeches) and polychaetes (bristle worms), but recent evidence suggests that other taxa that were once considered to be separate phyla (Sipuncula, Echiura and Siboglinidae (also known as Pogonophora)) should be included in Annelida⁴–⁶. However, the deep-level evolutionary relationships of Annelida are still poorly understood, and a robust reconstruction of annelid evolutionary history is needed. Here we show that phylogenomic analyses of 34 annelid taxa, using 47,953 amino acid positions, recovered a well-supported phylogeny with strong support for major splits. Our results recover chaetopterids, myzostomids and sipunculids in the basal part of the tree, although the position of Myzostomida remains uncertain owing to its long branch. The remaining taxa are split into two clades: Errantia (which includes the model annelid Platynereis), and Sedentaria (which includes Clitellata). Ancestral character trait reconstructions indicate that these clades show adaptation to either an errant or a sedentary lifestyle, with alteration of accompanying morphological traits such as peristaltic movement, parapodia and sensory perception. Finally, life history characters in Annelida seem to be phylogenetically informative.

Annelids are found throughout the world’s terrestrial, aquatic and marine habitats. They represent one of three major animal groups with segmentation, so understanding annelid body-plan evolution is crucial for elucidating aspects of the evolution of Bilateria⁴–⁶. Several annelid taxa have recently emerged as model organisms in various biological disciplines⁷. Surprisingly, the evolution of Annelida is still poorly understood, and it is uncertain how well these model organisms represent the ancestral character traits in Annelida. To rectify this situation, multigene data sets are needed to evaluate the diversity and the relationships of major annelid clades.

Annelida traditionally included Polychaeta and Clitellata. Morphological and molecular data corroborate clitellate monophyly and provide robust phylogenetic hypotheses within this taxon⁸. Polychaetes are classified into approximately 80 family-level taxa that are generally supported as monophyletic; however, arrangement of these taxa into well-supported, more-inclusive nodes is problematic²,⁹–¹⁰. Historically, polychaetes were classified as either Sedentaria or Errantia on the basis of their morphology and mode of life¹¹–¹³. This systematization was dismissed in the 1970s as being arbitrary groupings useful only for practical purposes¹⁴. About 15 years ago, on the basis of morphological cladistic analyses, a monophyletic Polychaeta consisting of two major clades, Scolicida and Palpata, was proposed, with the latter clade divided into Canalipalpata and Aciculata¹⁵. There is increasing molecular evidence, however, that places Clitellata, as well as the non-segmented taxa Echiura and Sipuncula, within polychaetes and thus renders Polychaeta paraphyletic¹⁴. So far, molecular work based on only a few genes has not supported the proposed monophyly¹⁶ of most major polychaete clades. Yet, support for basal nodes in these studies is less than 50 or 0.50 for bootstrap support (BS) or posterior probability (PP), respectively, resulting in a lack of support for alternative hypotheses¹⁷–¹⁹.

To address these major outstanding issues of annelid phylogeny, we used a phylogenomic approach, generating expressed sequence tag (EST) libraries for 17 annelid taxa, which are in addition to the publicly available EST or genomic data from annelids. We reconstructed relationships of major annelid taxa using 47,953 amino acid positions derived from 231 gene fragments that span 20 traditional polychaete ‘families’, Siboglinidae, Myzostomida, Echiura, Clitellata, Sipuncula and five outgroup taxa. This is the largest phylogenomic data set explored so far in annelid phylogeny and has a mean data coverage of 41.7% per taxon.

Sensitivity analyses of our data (Supplementary Tables 4 and 6) showed that increasing the number of positions and mean leaf stability had a positive impact on BS, whereas increasing the data coverage by removing either genes or taxa with low coverage had no such impact (Supplementary Fig. 1). Therefore, the largest data set (47,953 positions), with either all taxa (denoted ALL) or excluding the five annelid taxa that showed leaf stabilities below 0.925 (denoted EX) was used in maximum likelihood and Bayesian inference analyses. These analyses retrieved a clade (called clade 1) comprising all annelids with the exception of Chaetopteridae, Sipuncula and Myzostomida. This clade received significant branch support: ALL, PP = 0.98 (Bayesian inference), BS = 88 (maximum likelihood); EX, PP = 0.99 (Bayesian inference), BS = 100 (maximum likelihood) (Fig. 1 and Supplementary Figs 2, 3, 6 and 8). Reconstructing ancestral morphological traits for clade 1 and Annelida revealed that they were similar, except for some larval characters (Fig. 2a and Supplementary Table 5).

On the basis of this reconstruction, the ancestral annelid had a pair of anterior appendages (that is, grooved palps), which functioned in food gathering and sensory perception. Other head or pygidial appendages were absent. Eyes and nuchal organs were present as sensory organs. Of the different chaetal types, only internalized supporting chaetae and simple chaetae were part of the ancestral pattern. Reconstructions of most other parapodial characters were uncertain, except for the possession of prominent notopodial lobes. Although the fossil record of early annelids from the Cambrian period is sparse, it nonetheless reveals that, congruent with our reconstructions, the early annelids had palps, simple chaetae and internalized supporting chaetae but did not have other chaetae or appendages such as tentacular, parapodial or pygidial cirri⁶,¹⁷.

In agreement with previous molecular studies¹–³,¹⁸, Chaetopteridae, which have three distinct body regions, are found in the basal part of the annelid tree. Thus, the evolution of segmentation—with predominantly homonomous segmentation in clade 1 and Myzostomida, heteronomous segmentation in Chaetopteridae and complete reduction in Sipuncula—is already highly variable at basal nodes in the
Figure 1 | Reconstruction of the Annelida phylogenetic tree. Majority rule consensus trees of the Bayesian inference analysis using the site-heterogeneous CAT model of the data set with 39 taxa and 47,953 amino acid positions. Only PP (top of branch or alone) and BS (bottom) values ≥ 0.70 or 70, respectively, are shown. The branch leading to Myzostomida is reduced by 75%. Annelida are highlighted in red, with Sedentaria in blue and Errantia in green. Grey bars indicate additional annelid groups. *, BS value for the monophyly of Annelida without Myzostomida in the maximum likelihood analysis is 99.

Figure 2 | Ancestral reconstructions of body and parapodial characters. a, Annelida and clade 1. b, Errantia. c, Sedentaria. Body characters (left) and parapodial characters (right) are depicted. The state of several parapodial characters in Annelida and clade 1 is uncertain, so we depict the two most extreme possibilities. Dashed lines or question marks indicate that the state of the character is uncertain. bie, bicellular eyes; doc, dorsal cirrus; grp, grooved palps; isc, internalized supporting chaetae; laa, lateral antenna; mue, multicellular eyes; nuc, nuchal organ; pyc, pygidial cirrus; sic, simple chaeta; sop, solid palps; un or unvh, uncini/hooks; vec, ventral cirrus.

annelid phylogeny. Moreover, we acknowledge that, in addition to Chaetopteridae, Myzostomida and Sipuncula, other taxa such as Oweniidae, Dinophilidae or Protodrilida, which were not covered here because of a lack of data, might also be placed in the basal part of the annelid tree.

The major difference between the maximum likelihood and Bayesian inference analyses is the placement of Myzostomida. Myzostomids are either ectocommensals or endoparasites of echinoderms, and the systematic placement of this aberrant taxon has proved to be problematic. Bayesian inference analysis places Myzostomida within Annelida (PP = 0.99 for both data sets (ALL and EX); Fig. 1 and Supplementary Fig. 6). By contrast, by maximum likelihood analyses, long-branched Myzostomida are grouped with Ectoprocta, the outgroup taxon with the longest branch (Supplementary Figs 2 and 3). There is conclusive support from mitochondrial gene order and morphological data that Myzostomida are part of the annelid radiation and it has been shown that their derived sequences can be affected by long-branch attraction (LBA). The CAT model of Bayesian inference analyses is known to be less affected by LBA than other models, and this model proved to be better suited for our data set than was the LG model of maximum likelihood analyses (Supplementary Information). Notwithstanding the different position of Myzostomida (possibly owing to LBA), both maximum likelihood analyses support the monophyly of Annelida: ALL, BS = 99; EX, BS = 100 (Supplementary Figs 2 and 3). Moreover, the exclusion of Myzostomida did not substantially
alter the phylogenetic reconstruction of annelid ingroup relationships and BS values (Supplementary Fig. 7). Finally, the different placement of Myzostomida in the Bayesian inference and maximum likelihood analyses did not affect the reconstructions of ancestral morphological traits (Supplementary Table 5).

Clade 1 split into two well-supported clades: Errantia, which comprised Phyllodocida, Eunicida, Amphionimida and Orbiniidae; and Sedentaria, which comprised Clitellata and Echiura, as well as most other Scolecia (Capitellidae, Opheliidae and Arenicolidae) and Canali palpata (Terebelliformia, Cirratuliformia, Siboglinidae, Serpulidae and Spionidae). Both clades were significantly supported: ALL, PP = 0.99 (Bayesian inference), BS = 79 (maximum likelihood); EX, PP = 0.99 (Bayesian inference), BS = 100 (maximum likelihood) (Fig. 1 and Supplementary Figs 2, 3, 6 and 8). The placement of Clitellata indicated a closer relationship to Terebelliformia/Arenicolidae, Opheliidae and Capitellidae/Echiura. Moreover, analyses of branch attachment frequencies based on the data set comprising all taxa showed that each of the five removed annelid taxa is nested in either Sedentaria (Ridgeia, Ophelia, Pomatoecos and Malacoceros) or Errantia (Eurythoe), and none is moving between clades (Supplementary Figs 4 and 5).

In an influential publication in the 1990s, the two main competing hypotheses of annelid evolution were discussed15; one, starting with a ground pattern that resembles an errant, epibenthic organism; and, the other, starting with an infanual burrowing form. Interestingly, we found both trends to be realized within annelids. The ground pattern of Errantia reveals some important changes with respect to sensory perception and motility. On the basis of our reconstructions, the last common ancestor of Errantia had lateral antennae, palps (which are solid and restricted to sensory perception), a pair of pygidial cirri, nuchal commissures and two pairs of multicellular eyes facing in different directions23 (Fig. 2b and Supplementary Table 5). The parapodia had prominent notopodial and neuropodial lobes supported by internalized chaetae, as well as ventral cirri. Overall, this pattern can be regarded as adaptations to a more active and mobile lifestyle, which requires increased perception of the environment, as well as motility by undulation. Prominent parapodial lobes are advantageous for rapid movements based on undulation, which is mainly achieved by the well-developed longitudinal musculature arranged in at least four separate bundles. For example, in sexually mature (that is, epigamous) nereids, adopting a temporary pelagic reproductive stage, parapodial lobes are even further enlarged and paddle-like than in immature stages17. Most taxa of Phyllodocida, Eunicida and Amphimimida show such an errant, often predatory, mode of life and hence were traditionally named Errantia11. The position of Orbiniidae, which were previously grouped with Scolecia15, might be surprising; however, placing them within or close to the errant forms had previously been debated on the basis of morphological and molecular evidence23,12,14. Therefore, we named this clade Errantia, as it is characterized by adaptation to a more errant life.

The evolution of parapodia in Sedentaria shows the opposite trend. Neuropodial and notopodial lobes are generally smaller than in Errantia and lack internalized supporting chaetae (Fig. 2c). In general, chaetae are in close proximity to the stiff body wall, an arrangement that facilitates a better anchorage in tubes and burrows. Moreover, antennae are absent, and palps have been lost independently in several taxa. The taxa of this clade are commonly characterized by a sedentary life, as more or less sessile organisms that live below stones, tube builders, or burrowers by means of peristalsis such as earthworms17. Sedentaria are generally microphagous. Taxa without appendages such as those formerly grouped as Scolecia15 are deposit feeders, often ingesting sediment. By contrast, taxa with sometimes elaborate head appendages such as terebellids or serpulids are surface deposit feeders or filter feeders, respectively21. The deposit feeding lifestyle also generally applies to most Clitellata. Therefore, we named this clade Sedentaria14 (now including Clitellata), and it is characterized by adaptations to a more sedentary lifestyle by, for example, the reduction of parapodia and loss of internalized supporting chaetae. A key feature is that the chaetae are in closer proximity to the stiff body wall, rather than being embedded in parapodial lobes (which are more flexible) as typical for errant annelids. Interestingly, errant polychaetes with sedentary life strategies such as Lumbrineriidae or Onuphidae have adapted to such a lifestyle by using different solutions17.

Hence, within Annelida, there are two major clades, Errantia and Sedentaria, whose evolution was driven by the adaptation to two different modes of life. Errantia show a more mobile and active life strategy than Sedentaria, and this is correlated to increased sensory perception and motility. Sedentaria are more sessile, with accompanying reductions of head and body appendages and the position of the chaetae being in closer proximity to the body wall than in Errantia. Annelids have been successfully established as models in evolutionary developmental studies to elucidate the characteristics of the last common bilaterian ancestor24. Of the recent model organisms, Platynereis, with its well-developed head and parapodial appendages, is a good representative of Errantia. By contrast, Capitella (as a burrower with reduced appendages), Helobdella (as a clitellate) and Hydroides (as a filter feeder using its radiolar crown) represent different microphagous feeder types in Sedentaria.

METHODS SUMMARY

EST libraries of 1,370 clones, on average, were prepared for 17 annelid species (Supplementary Table 1). All original sequence data have been deposited in the NCBI Expressed Sequence Tag database (dbEST). EST or genomic data from 17 additional annelid species and 5 outgroup species were obtained from public archives (Supplementary Table 1). These raw EST data were further processed as described previously23. Sets of orthologous genes were determined using the program HaMSR in combination with the InParanoid database (without ribosomal proteins)28, and were translated into amino acid sequences using the program ESTwise26. In parallel, we retrieved all ribosomal proteins from databases as described previously25 (Supplementary Table 2). Each orthologous gene set was aligned using MAFFT software27 and masked using the program REAP28. Only genes that had taxon coverage of at least 33.3% were included in the final super-matrix.

Phylogenetic trees were inferred from this data set of 39 taxa by using a Bayesian inference approach (using the site-heterogeneous CAT model) and a maximum likelihood approach (using the LG model). Stabilities of taxa were assessed using the leaf stability index as calculated by Phyutility software29 (Supplementary Table 3). The five annelid taxa with an index below 0.295 were removed in the second data set, and the Bayesian inference analysis was repeated. Branch attachment frequencies of these unstable annelid taxa were assessed using the lineup movement option in Phyutility30 based on the data set with all taxa.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 2 September 2010; accepted 18 January 2011.


3 MARCH 2011 | VOL 471 | NATURE | 97
Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to I. Ebersberger, S. Strauss and A. von Haeseler for the processing of our raw EST libraries. We also thank M. Aguado for species identification of Typossyllis pigmamenta, as well as K. M. Halanych and P. A. Rainey-Balci for suggestions. T.H.S. and C.B. acknowledge support from the marine biological stations in Barnfield, Helgoland, List and Roscoff for collection of annelids. This work was funded by the priority programme 'Deep Metazoan Phylogeny' of the Deutsche Forschungsgemeinschaft by grants DFG-STR 683/5-2 (T.H.S.), DFG-Pu-84/5-1 (G.P. and T.H.S.), DFG-Ti-349/4-1 (R.T.), DFG U 998/3-1 (B.L.), DFG-FA 9744/1-1 (S.H.) and DFG-BL 787/2-1 (C.B.).

Author Contributions T.H.S., G.P., R.T. and C.B. conceived this study. T.H.S. took the lead on data collection of sedentary polychaetes, and writing. T.H.S. and S.H. performed phylogenomic analyses. C.H. aided in the data collection of Sedentaria. C.B. and C.P. took the lead on data collection of errant polychaetes, and C.B., S.H. and N.H. on compilation of the data sets from the EST libraries. A.M. and B.L. generated the EST library of Spionulus nudus, and M.K. was responsible for the sequencing of the EST libraries. T.H.S., G.P., R.T. and C.B. were the main contributors to the writing of the manuscript.

Author Information Sequence data have been deposited in the NCBI Expressed Sequence Tag database (dbEST) under accession numbers FN424437–FN428571, FR754554–FR771822, HQ729923–HQ729975. The largest aligned data set has been deposited at http://www.treebase.org. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.H.S. (struck@biologie.uni-osnabrueck.de) or C.B. (bieford@uni-leipzig.de).