Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida

Sarah J. Bourlat[1], Thorhildur Juliusdottir[2], Christopher J. Lowe[3], Robert Freeman[4], Jochanan Aronowicz[3], Mark Kirschner[5], Eric S. Lander[4,6], Michael Thorndyke[7], Hiroaki Nakano[7], Andrea B. Kohn[8], Andreas Heyland[8], Leonid L. Moroz[8], Richard R. Copley[2] & Maximilian J. Telford[1]

Deuterostomes comprise vertebrates, the related invertebrate chordates (tunicates and cephalochordates) and three other invertebrate taxa: hemichordates, echinoderms and *Xenoturbella*[1]. The relationships between invertebrate and vertebrate deuterostomes are clearly important for understanding our own distant origins. Recent phylogenetic studies of chordate classes and a sea urchin have indicated that urochordates might be the closest invertebrate sister group of vertebrates, rather than cephalochordates, as traditionally believed[2–5]. More remarkable is the suggestion that cephalochordates are closer to echinoderms than to vertebrates and urochordates, meaning that chordates are paraphyletic[2]. To study the relationships among all deuterostome groups, we have assembled an alignment of more than 35,000 homologous amino acids, including new data from a hemichordate, starfish and *Xenoturbella*. We have also sequenced the mitochondrial genome of *Xenoturbella*. We support the clades Olfactores (urochordates and vertebrates) and Ambulacraria (hemichordates and echinoderms[6]). Analyses using our new data, however, do not support a cephalochordate and echinoderm grouping and we conclude that chordates are monophyletic. Finally, nuclear and mitochondrial data place *Xenoturbella* as the sister group of the two ambulacrarian phyla[1]. As such, *Xenoturbella* is shown to be an independent phylum, Xenoturbellida, bringing the number of living deuterostome phyla to four.

Although molecular phylogenetic studies have prompted changes in the membership of the deuterostome superphylum (two groups—chaetognaths[7] and lophophorates[8]—have been reclassified as protostomes and one species, *Xenoturbella bocki*, added[1]) there remains an unchanging core of three phyla: echinoderms, hemichordates and chordates. Three main groups are recognized within the chordates themselves—urochordates, including the ascidians and larvaceans; cephalochordates or lancelets; and vertebrates, including fish and tetrapods.

Molecular studies have also changed our view of the relationships between deuterostome taxa. Hemichordates, long thought of as the sister group of the chordates, are now considered to be closest to echinoderms in a supraphyletic clade named the Ambulacraria[6]. Most recently, analyses of hundreds of genes from expressed sequence tag (EST) and whole-genome sequencing projects have indicated that, contrary to previous studies, urochordates are the invertebrate group that is most closely related to vertebrates[2]. Independent molecular[3,4] and morphological[9] support for this unexpected grouping also exists. Although Ambulacraria and Olfactores (the urochordate plus vertebrate clade[10]) seem to be credible components of the deuterostomes, two further aspects of the phylogeny of the group remain contentious.

The first stems from a remarkable result in a phylogenomic analysis in which cephalochordates were associated with echinoderms rather than with other chordates[2]. This result, if correct, has profound implications for our understanding of early deuterostome evolution and the origins of the chordate body plan, because it suggests that a number of conserved chordate features were present in the common ancestor of all deuterostomes and were secondarily lost in echinoderms and hemichordates. Although this is clearly a controversial finding, it deserves serious consideration. The initially surprising relocation of the hemichordates from chordate sister group to the Ambulacraria has already shown that certain characteristics that were thought to be specific to the chordate linage are likely to derive from the deuterostome common ancestor; the homology of gill slits and endostyle in hemichordates and chordates is now supported by studies both of morphology and of shared gene-expression patterns[9,11,12].

The second open question concerns the position of the worm *Xenoturbella*. Consideration of its small subunit ribosomal RNA gene (SSU) indicated that *Xenoturbella* might be a deuterostome, related to the Ambulacraria but not included in either the echinoderms or the hemichordates[1]. This result is supported by the lack in *Xenoturbella* of a complex molecular character that unites the two ambulacrarian phyla—a change in their mitochondrial genetic code whereby the codon ATA codes for the amino acid isoleucine rather than for methionine as in most other animals, including *Xenoturbella*[1]. At the same time, there are several reasons to question the position of *Xenoturbella*. There is low statistical support for its exclusion from the hemichordates in the SSU study[1], the change in genetic code is known to be homoplastic because it has occurred independently in the rhabditophoran playthelminths[13]; separate analyses of mitochondrial *Cytochrome oxidase I* sequences placed *Xenoturbella* with the hemichordates[1], and there are detailed similarities between the structure of the epidermis of hemichordates and *Xenoturbella*[14] that have led to suggestions that *Xenoturbella* is a neotenous hemichordate[15].

To test the robustness of previous work and to extend the phylogenomic approach to all of the main deuterostome groups, we have gathered data from the two groups previously missing: a hemichordate, *Saccoglossus kowalevski* (206,715 ESTs including those from Lowe *et al.*[16]) and *Xenoturbella bocki* (1,372 ESTs). We added to

[1]Department of Biology, University College London, Darwin Building, Gower Street, London WC1E 6BT, UK. [2]Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK. [3]Department of Organismal Biology and Anatomy, 1027 E. 57th Street, University of Chicago, Chicago, Illinois 60637, USA. [4]Department of Systems Biology, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115, USA. [5]University of California, Berkeley, Department of Molecular and Cell Biology, 142 Life Sciences Addition #3200, Berkeley, California 94720-3200, USA. [6]Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. [7]Royal Swedish Academy of Sciences, Kristineberg Marine Research Station, Fiskebäckskil, S 450 34, Sweden. [8]The Whitney Laboratory for Marine Bioscience, University of Florida, 9505 Ocean Shore Boulevard, St Augustine, Florida 32080, USA.
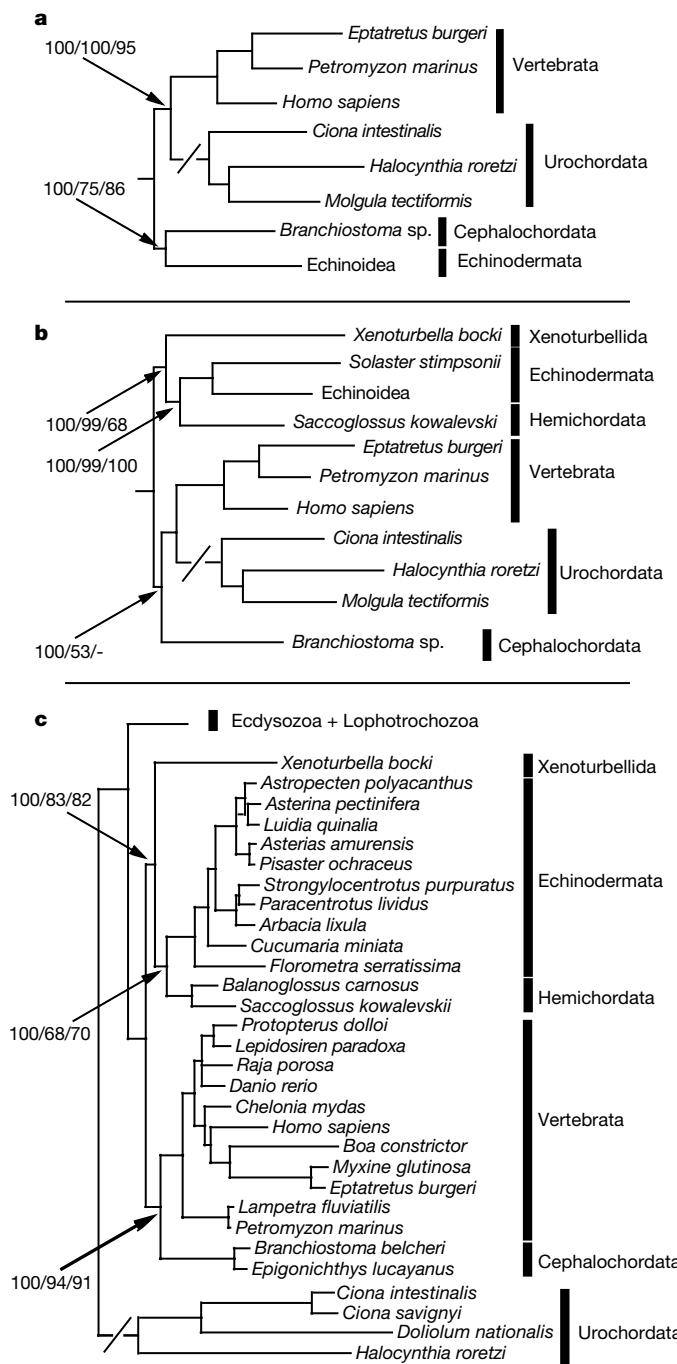
Figure 1 | **Phylogenetic analyses of 170 nuclear proteins and 13 mitochondrial proteins support a monophyletic chordate clade and an independent deuterostome phylum of Xenoturbellida. a,** Bayesian analysis of nuclear data before the addition of the new asteroid, hemichordate and xenoturbellid data. The cephalochordate groups with the echinoderm, implying that the chordates are paraphyletic. **b,** Bayesian analysis of nuclear data after the addition of asteroid, hemichordate and xenoturbellid data. The new sequences join the branch to the echinoderm, and the cephalochordates now join the chordate branch. This indicates that the previous result is due to systematic error. *Xenoturbella* is the sister group of the Ambulacraria (echinoderms plus hemichordates). **c,** Bayesian analysis of mitochondrial data with the amino acids M, I, N and K recoded as missing data places the cephalochordates with vertebrates; *Xenoturbella* is the sister group of Ambulacraria. Support for critical nodes is shown in the order BPP (%), TF (%) and MLBP (%). Long branches leading to the urochordates have been shortened as indicated by a break (/) in the branch (see Supplementary Information for unaltered tree and scale).

the sampling of the echinoderm clade with data from the sun star *Solaster stimpsoni* (5,614 ESTs). We searched these ESTs for orthologues of the 146 genes used by Philippe *et al.*[17]. We then supplemented these with 35 genes having a representative orthologue in our *Xenoturbella* ESTs, as this taxon had the fewest matches to the data set from ref. 17. We culled 11 alignments where orthology was uncertain, leaving 170 concatenated, aligned orthologous genes (35,588 reliably aligned amino acids). *Saccoglossus* had data from 153 of 170 genes and 29,777 amino acids, *Xenoturbella* had 63 genes and 8,370 amino acids and *Solaster* had 67 genes and 11,033 amino acids.

We compared results from our extended data set of 170 genes to those of the 146-gene data set of Delsuc *et al.*[2] by performing phylogenetic analyses without the new hemichordate, *Xenoturbella* and sun star sequences. Phylogenetic analyses using maximum-likelihood and bayesian approaches recapitulated previous results (Fig. 1a). All nodes received maximum support from bayesian analysis (bayesian posterior probability (BPP) = 100%). Maximum-likelihood non-parametric bootstrap support (MLBP) was lower for some nodes (see Supplementary Information). The overall topology of this tree conforms to the 'new animal phylogeny'[18] with monophyletic protostomes, deuterostomes, lophotrochozoans (represented here by molluscs and an annelid), and ecdysozoans (represented here by arthropods and a nematode; see Supplementary Information). Within the deuterostomes, we find support for monophyletic groups of vertebrates, cyclostomes (hagfish and lamprey) and Olfactores (vertebrates plus urochordates). Importantly, we recovered the relationship linking cephalochordates not to other chordates but to echinoderms: BPP = 100%, Treefinder[19] LRSH support (TF) = 75% and MLBP = 81%. This relationship was unexpected and we wanted to test the possibility that it was caused by systematic error. Two powerful ways of addressing systematic errors are to include additional taxa, thereby enabling better detection of misleading multiple substitutions, and to remove problematic taxa (for example, long branches); we have used both approaches[20].

By adding sequences from three taxa that are likely to be related to the echinoids (asteroid, hemichordate and *Xenoturbella*) we could test the possibility that the cephalochordate/echinoderm relationship might result from a systematic error such as long-branch attraction[21]. The effect of including these three taxa in our analysis is clear; support for a cephalochordate plus echinoderm clade disappears and support for the canonical monophyletic chordates increases (Fig. 1b). The monophyletic chordate clade in our new analysis has BPP = 100% and TF = 53%, but no MLBP support. This change in topology on the addition of taxa that divide the echinoid branch is consistent with the idea that the attraction between the cephalochordate and the echinoid results from a systematic error; the addition of non-problematic taxa tends to improve the reliability of a phylogeny[21]. We compared the support for the optimal topology to that for the alternative topology, which groups the cephalochordates with Ambulacraria and *Xenoturbella* using the Shimodaira–Hasegawa (SH) tests and Bayes factors. The SH test did not reject the alternative topology as significantly worse than the optimal topology (SH $P = 0.476$), but the alternative topology was rejected as significantly worse using the Bayes factors test ($2\log e (B10) = 24$), although this test might not be conservative. The inclusion of data from the other major taxon of cephalochordates—the Epigonichthyidae—will divide the cephalochordate branch and, we predict, will further increase support for the monophyletic chordates.

A parallel possibility is that rapidly evolving urochordates, along with their sister group the vertebrates, have been attracted towards the base of the tree by long-branch effects[22]. Two lines of evidence support this possibility. First, in the analyses described above, we used diploblast metazoans (ctenophore and cnidarian) as close outgroups; the inclusion of more distant outgroups (yeast and choanoflagellates), as in ref. 2, results in a more basal position for Olfactores with the corollary that cephalochordates group with echinoderms. This result is observed even when the asteroid, hemichordate

Table 1 | Proportions of amino acids affected by mitochondrial genetic code change in echinoderms and other metazoans

	M	I	N	K
Echinoderm	2.32 (0.09)	10.28 (1.03)	7.05 (0.64)	1.43 (0.02)
Metazoan	5.51 (2.31) **	8.1 (2.6) **	3.52 (0.66) n.s.	4.14 (0.20) **
Deuterostome	5.31 (1.34) **	7.79 (1.25) n.s.	3.58 (0.24) **	3.96 (0.28) **

Mean and variance (in parentheses) of percentage proportions of the amino acids methionine (M), isoleucine (I), asparagine (N) and lysine (K) in the mitochondrial genes of 12 echinoderms, 70 metazoans and 23 non-echinoderm deuterostomes. *F*-test two-sample for variance test was used to compare metazoan and deuterostome values to those of the echinoderms. ** indicates $P < 0.05$. n.s. indicates non-significant difference.

and *Xenoturbella* data are included. Second, exclusion of the long-branched urochordates always results in cephalochordates grouping with vertebrates with or without our additional taxa and even in the presence of distant outgroups (see Supplementary Information).

We also considered the phylogeny of the other deuterostome groups. The hemichordate is the sister group of the echinoderms (asteroid and echinoid), which gives weight to a monophyletic Ambulacraria. Of greater interest is our finding that *Xenoturbella* is related to but outside the monophyletic Ambulacraria. The clade (*Xenoturbella* plus Ambulacraria) has BPP = 100%, TF = 99% and MLBP = 68% whereas ambulacrarian monophyly excluding *Xenoturbella* has BPP = 100%, TF = 99.3% and MLBP = 100%. We compared the support for the optimal topology to that for alternative possibilities in which *Xenoturbella* is grouped with hemichordates or echinoderms using the SH tests and Bayes factors. Both tests rejected as significantly less well supported the grouping of *Xenoturbella* with hemichordates (SH $P = 0.012$, Bayes factors $2\log e(B10) = 96$) or *Xenoturbella* with echinoderms (SH $P = 0.037$, Bayes factors $2\log e(B10) = 118$).

To test our findings further, we sequenced the entire mitochondrial genome (15,234 nucleotides) of *Xenoturbella bocki*. We have used the amino-acid sequences of all 13 protein-coding loci to examine the position of *Xenoturbella* within the deuterostomes. Our initial results placed *Xenoturbella* at the base of the deuterostomes rather than as the sister group of Ambulacraria. We were aware, however, that the mitochondrial genetic code changes in the lineage leading to Ambulacraria and an additional change in the echinoderms[13] might have led to compositional bias that could artefactually distance them from *Xenoturbella*. Indeed, the proportions of amino acids that were affected by the changes of genetic code (methionine and isoleucine, asparagine and lysine (M, I, N, K)) in the echinoderms differ from the average seen in deuterostomes and metazoans (Table 1). We tested the effects of this imbalance on the position of *Xenoturbella* in two ways. First we excluded the most biased taxa—the echinoderms—leaving the hemichordates as representatives of Ambulacraria. The result of excluding the echinoderms is to move *Xenoturbella* next to the hemichordates, as predicted if its basal position were due to compositional bias in the echinoderms (see Supplementary Information). Second, we avoided the influence of the problematic amino acids (M, I, N, K)) by recoding each instance of them as 'missing data', which affected 33.9% of our reliably aligned amino acids. The effect of this was again to move *Xenoturbella* to become the sister group of Ambulacraria. *Xenoturbella* plus Ambulacraria monophyly has BPP = 100%, TF = 83% and MLBP = 82% whereas ambulacrarian monophyly excluding *Xenoturbella* has BPP = 100%, TF = 68% and MLBP = 70% (Fig. 1c).

We note that according to our mitochondrial data set, the cephalochordates and vertebrates form a highly supported monophyletic group (BPP = 100%, TF = 96.5% and MLBP = 91%), which further diminishes the credibility of the cephalochordate–echinoderm link. In common with previous analyses[23], our tree reconstructions place the long-branched urochordates as a sister group to all other Bilateria.

What conclusions can we draw from this knowledge of the relationships of all the main deuterostome clades? Our analyses indicate that the relationship between cephalochordates and echinoderms

probably results from systematic error. This makes excellent sense from a morphological standpoint because chordates share several characteristics (somites, dorsal nerve chord, notochord and hypophysis) that are absent from Ambulacraria, *Xenoturbella* and non-deuterostome phyla[24]. Although hemichordates have in the past been considered to possess homologues of the chordate notochord and dorsal nerve cord, these similarities do not stand up to close scrutiny on the basis of ultrastructure or shared expression of orthologous patterning genes[9,16].

We have also shown compelling evidence that *Xenoturbella* should be positioned within the deuterostomes as a sister group to Ambulacraria. This result has important consequences for our understanding of deuterostome evolution. First, it means that *Xenoturbella* is not a neotenous member of the hemichordates[15] but an independent lineage—Xenoturbellida[25]—which constitutes a newly recognized deuterostome phylum. We propose the name Xenambulacraria for the highly supported monophyletic grouping of Xenoturbellida plus Ambulacaria.

The second implication of this result is that characteristics shared by the hemichordates and *Xenoturbella* can be parsimoniously reconstructed as having been present in the last common ancestor of Xenambulacraria. Characteristics of this ancestor must have included those characteristics shared by all deuterostomes (deuterostomy, radial cleavage, enterocoely); those shared by hemichordates and chordates (gill slits and endostyle); and the basiepithelial nervous system[15,26,27] and unusual ciliary structure shared by the hemichordates and *Xenoturbella*[28,29]. Evidence for most of these characteristics remains to be found in the little-studied *Xenoturbella*.

Although our results, which support a monophyletic chordate clade, conform to a reasonable view of unique chordate characters, we are, nevertheless, returned to a state of relative ignorance concerning the characteristics and evolution of the earliest deuterostomes. Perhaps the biggest question is whether the ancestral deuterostome nervous system was centralized, as in chordates, or diffuse, as in Xenambulacraria. The central nervous system that is prevalent in the protostomes—the closest sister group of the deuterostomes—suggests the former, but a diffuse nervous system is found in the cnidarians and acoelomorph flatworms, which are widely thought to be the sister group of the protostomes plus deuterostomes[30]. Further progress in understanding the origins and evolution of deuterostome and chordate characteristics is likely to come from the integration of palaeontological, comparative embryological and genomic approaches.

METHODS

Data set assembly. Representative sequences from the data set of Philippe *et al.*[17] were searched the est_others subset of GenBank, supplemented with EST sequences from *Saccoglossus* (206,715 sequences), *Solaster* (5,614 sequences) and *Xenoturbella* (1,372 sequences) using tblastn. Hits from our EST data sets plus those from the lamprey *Petromyzon marinus*, hagfish *Eptatretus burgeri* and the tunicates *Molgula tectiformis* and *Oikopleura diocia* were then assembled and added to these alignments. Additional data sets were assembled using a starting set of *Xenoturbella* proteins that had an orthologue in *Homo*, *Drosophila* and *Caenorhabditis* (see Supplementary Information). The *Xenoturbella* mitochondrial genome was sequenced as described in the Supplementary Information. Additional mitochondrial genomes came from the OGRE website (drake. physics.mcmaster.ca/ogre/index.shtml).

Phylogenetic analyses. The concatenated amino-acid alignments were analysed using bayesian and maximum-likelihood approaches and maximum-likelihood

non-parametric bootstrapping. Likelihood-based tests of alternative topologies used SH tests and Bayes factors (see Supplementary Information).

1. Bourlat, S. J., Nielsen, C., Lockyer, A. E., Littlewood, D. T. J. & Telford, M. J. *Xenoturbella* is a deuterostome that eats molluscs. *Nature* **424**, 925–928 (2003).
2. Delsuc, F., Brinkmann, H., Chourrout, D. & Philippe, H. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* **439**, 965–968 (2006).
3. Oda, H. *et al.* A novel amphioxus cadherin that localises to epithelial adherens junctions has an unusual domain organisation with implications for chordate phylogeny. *Evol. Dev.* **4**, 426–434 (2002).
4. Wada, H., Okuyama, M., Satoh, N. & Zhang, S. Molecular evolution of fibrillar collagen in chordates, with implications for the evolution of vertebrate skeletons and chordate phylogeny. *Evol. Dev.* **8**, 370–377 (2006).
5. Jeffery, W. R., Strickler, A. G. & Yamamoto, Y. Migratory neural crest-like cells form body pigmentation in a urochordate embryo. *Nature* **431**, 696–699 (2004).
6. Turbeville, J. M., Schulz, J. R. & Raff, R. A. Deuterostome phylogeny and the sister group of the chordates: evidence from molecules and morphology. *Mol. Biol. Evol.* **11**, 648–655 (1994).
7. Telford, M. J. & Holland, P. W. H. The phylogenetic affinities of the chaetognaths: a molecular analysis. *Mol. Biol. Evol.* **10**, 660–676 (1993).
8. Halanych, K. M. *et al.* Evidence from 18S ribosomal DNA that the lophophorates are protostome animals. *Science* **267**, 1641–1643 (1995).
9. Ruppert, E. E. Key characters uniting hemichordates and chordates: homologies or homoplasies? *Can. J. Zool.* **83**, 8–23 (2005).
10. Jefferies, R. P. S. in *Biological Asymmetry and Handedness* (eds Bock, G. R. & Marsh, J.) 94–127 (John Wiley, Chichester, 1991).
11. Ogasawara, M., Wada, H., Peers, H. & Satoh, N. Developmental expression of Pax 1/9 genes in urochordate and hemichordate gills: insight into function and evolution of the pharyngeal epithelium. *Development* **126**, 2539–2550 (1999).
12. Rychel, A. L., Smith, S.E., Shimamoto, H. T & Swalla, B. J. Evolution and development of the chordates: collagen and pharyngeal cartilage. *Mol. Biol. Evol.* **23**, 541–549 (2006).
13. Telford, M. J., Herniou, E. A., Russell, R. B. & Littlewood, D. T. J. Changes in mitochondrial genetic codes as phylogenetic characters: two examples from the flatworms. *Proc. Natl Acad. Sci. USA* **97**, 11359–11364 (2000).
14. Stach, T. *et al.* Nerve cells of *Xenoturbella* (phylum uncertain) and *Harrimania kupfferi* (Enteropneusta) are positively immunoreactive to antobodies rasied against echinoderm neuropeptides. *J. Mar. Biol. Assoc. UK* **85**, 1519–1524 (2005).
15. Reisinger, E. Was ist *Xenoturbella*? *Z. Wissenschaftlische Zool.* **164**, 188–198 (1960).
16. Lowe, C. J. *et al.* Anteroposterior patterning in hemichordates and the origins of the chordate nervous system. *Cell* **113**, 853–865 (2003).
17. Philippe, H., Lartillot, N. & Brinkmann, H. Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. *Mol. Biol. Evol.* **22**, 1246–1253 (2005).
18. Aguinaldo, A. M. A. & Lake, J. A. Evolution of the multicellular animals. *Am. Zool.* **38**, 878–887 (1998).
19. Jobb, G. TREEFINDER version of May 2006. ⟨http://www.treefinder.de⟩ (2006).
20. Phillipe, H. & Telford, M. J. Large-scale sequencing and the new animal phylogeny. *Trends Ecol. Evol.* (in the press).
21. Wiens, J. J. Can incomplete taxa rescue phylogenetic analyses from long-branch attraction? *Syst. Biol.* **54**, 731–742 (2005).
22. Felsenstein, J. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* **27**, 401–410 (1978).
23. Yokobori, S., Watanabe, Y. & Oshima, T. Mitochondrial genome of *Ciona savigny* (Urochordata, Ascidiacea, Enterogona): comparison of gene arrangement and tRNA genes with *Halocynthia roretzi* mitochondrial genome. *J. Mol. Evol.* **57**, 574–587 (2003).
24. Weichert, C. K. *Anatomy of the Chordates* (McGraw-Hill, New York, 1965).
25. Westblad, E. *Xenoturbella bocki* n.g, n.sp, a peculiar, primitive turbellarian type. *Ark. Zool.* **1**, 3–29 (1949).
26. Pedersen, K. J. & Pedersen, L. R. Ultrastructural observations on the epidermis of *Xenoturbella bocki* Westblad, 1949—with a discussion of epidermal cytoplasmic filament systems of invertebrates. *Acta Zool.* **69**, 231–246 (1988).
27. Raikova, O. I., Reuter, M., Jondelius, U. & Gustafsson, M. K. S. An immunocytochemical and ultrastructural study of the nervous and muscular systems of *Xenoturbella westbladi* (Bilateria inc. sed.). *Zoomorphology* **120**, 107–118 (2000).
28. Franzen, A. & Afzelius, B. A. The ciliated epidermis of *Xenoturbella bocki* (Platyhelminthes, Xenoturbellida) with some phylogenetic considerations. *Zool. Scr.* **16**, 9–17 (1987).
29. Pardos, F. Fine structure and function of pharynx cilia in *Glossobalanus minutus* Kowalewsky (Entropneusta). *Acta Zool.* **69**, 1–12 (1988).
30. Ruiz Trillo, I., Riutort, M., Littlewood, D. T. J., Herniou, E. A. & Baguñà, J. Acoel flatworms: earliest extant bilaterian metazoans, not members of Platyhelminthes. *Science* **283**, 1919–1923 (1999).