

# Tunicates and not cephalochordates are the closest living relatives of vertebrates

Frédéric Delsuc<sup>1†</sup>, Henner Brinkmann<sup>1</sup>, Daniel Chourrout<sup>2</sup> & Hervé Philippe<sup>1</sup>

Tunicates or urochordates (appendicularians, salps and sea squirts), cephalochordates (lancelets) and vertebrates (including lamprey and hagfish) constitute the three extant groups of chordate animals. Traditionally, cephalochordates are considered as the closest living relatives of vertebrates, with tunicates representing the earliest chordate lineage<sup>1,2</sup>. This view is mainly justified by overall morphological similarities and an apparently increased complexity in cephalochordates and vertebrates relative to tunicates<sup>2</sup>. Despite their critical importance for understanding the origins of vertebrates<sup>3</sup>, phylogenetic studies of chordate relationships have provided equivocal results<sup>4–7</sup>. Taking advantage of the genome sequencing of the appendicularian *Oikopleura dioica*, we assembled a phylogenomic data set of 146 nuclear genes (33,800 unambiguously aligned amino acids) from 14 deuterostomes and 24 other slowly evolving species as an outgroup. Here we show that phylogenetic analyses of this data set provide compelling evidence that tunicates, and not cephalochordates, represent the closest living relatives of vertebrates. Chordate monophyly remains uncertain because cephalochordates, albeit with a non-significant statistical support, surprisingly grouped with echinoderms, a hypothesis that needs to be tested with additional data. This new phylogenetic scheme prompts a reappraisal of both morphological and palaeontological data and has important implications for the interpretation of developmental and genomic studies in which tunicates and cephalochordates are used as model animals.

The introduction of molecular data into classical systematics has already tested several evolutionary hypotheses through the analysis of individual genes such as ribosomal RNA (rRNA). However, phylogenies reconstructed from a single gene or a small number of genes are hampered by stochastic effects limiting the statistical significance of the results. The genomic era is now providing the opportunity for phylogenetics to resolve a number of outstanding evolutionary questions through an increase of resolving power<sup>8</sup>. This applies to the origin and early evolution of vertebrates, a fundamental evolutionary question that has been revived by recent advances in molecular and developmental biology as well as new fossil discoveries<sup>3</sup>.

Our understanding of these events has to be considered in the context of chordate phylogeny, for which the traditional textbook view considers cephalochordates as the closest living relatives of vertebrates (a group named Euchordata), to the exclusion of the morphologically more distinct tunicates<sup>2</sup>. Although almost universally accepted, this classical picture is supported by only a limited number of morphological features that are far from being unambiguous. For example, the presence of metameric segmentation<sup>1</sup> used to link cephalochordates and vertebrates might in fact be considered as an ancestral feature of deuterostomes<sup>9</sup>. The classical view (Euchordata) has also found some support in molecular studies

of rRNA genes<sup>5</sup>. However, a competing hypothesis grouping tunicates and vertebrates into a clade named Olfactores<sup>10</sup> was recovered in cladistic analyses of combined rRNA and morphology<sup>4</sup> and suggested by the structure of cadherin genes<sup>11</sup>. However, the statistical significance of these apparently conflicting results was limited by the relatively few characters considered.

Recently, two multigene studies based on nuclear proteins have provided some support for Olfactores<sup>6,7</sup>. However, the extremely limited chordate species sampling considered in these studies meant that no firm conclusions were possible, given the potentially deleterious effects of poor taxon sampling on phylogenetic inference<sup>8</sup>. We have therefore extended the 146-gene data set of ref. 7 from 4 to 13 chordates, including one cephalochordate, four tunicates, and eight vertebrates, with the notable inclusion of the early-branching agnaths (hagfish and lamprey). Within tunicates, the incorporation of *O. dioica* is particularly important because it belongs to appendicularians (or larvaceans), which are morphologically and molecularly very divergent from the ascidians previously included.

Phylogenetic analyses of this multigene data set using maximum parsimony (MP), maximum likelihood (ML) and bayesian inference all converged to the same topology (Fig. 1). The statistical support was maximal in bayesian analyses, where all nodes received a posterior probability of 1.0. All non-controversial groups (choanoflagellates, cnidarians, molluscs, arthropods, tunicates and vertebrates) were recovered with strong bootstrap support (bootstrap percentage BP<sub>MP-ML</sub> > 95%), as were also metazoans, bilaterians, protostomes and lophotrochozoans (Fig. 1). Weak statistical support (BP<sub>MP-ML</sub> < 72%) was only observed for some relationships within insects and bivalves. Within vertebrates, our results strongly support the controversial monophyly of cyclostomes (lamprey and hagfish)<sup>2</sup>. Also, both MP and ML provided reasonable support for the monophyly of deuterostomes (BP<sub>MP-ML</sub> = 87–93%). However, within deuterostomes, chordates appeared not to be monophyletic, because cephalochordates grouped with echinoderms, albeit with moderate ML bootstrap support (BP<sub>MP-ML</sub> = 97–89%). By contrast, whereas MP moderately supported the grouping of tunicates and vertebrates (90%), the more accurate ML method<sup>12</sup> provided unambiguous bootstrap support (100%) for Olfactores (Fig. 1).

To test further the stability of phylogenetic relationships within deuterostomes, we evaluated in a likelihood framework the 15 rooted topologies corresponding to all possibilities of connecting the four major groups under study (echinoderms, cephalochordates, tunicates and vertebrates). Thirteen alternatives to the ML topology were significantly rejected at the 5% confidence level by all statistical tests (Table 1). Only the topology where chordates are monophyletic, with cephalochordates joining the tunicates plus vertebrate clade, was not rejected (Table 1). The traditional hypothesis of euchordate monophyly was ranked only fourth in terms of log-likelihood, after the

<sup>1</sup>Département de Biochimie, Centre Robert-Cedergren, Université de Montréal, Succursale Centre-Ville, Montréal, Québec H3C3J7, Canada. <sup>2</sup>Sars Centre for Marine Molecular Biology, Bergen High Technology Centre, University of Bergen, Thormøhlensgaten 55, 5008 Bergen, Norway. <sup>†</sup>Present address: Laboratoire de Paléontologie, Phylogénie et Paléobiologie, Institut des Sciences de l'Évolution, UMR 5554-CNRS, Université Montpellier II, France.

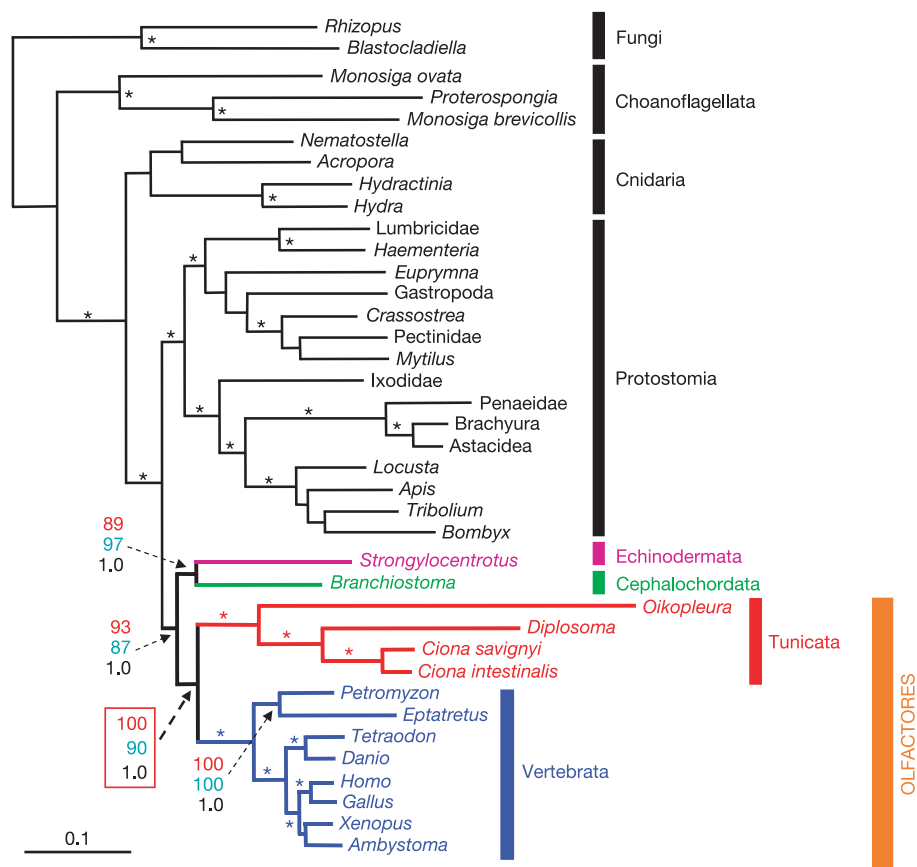
alternative in which cephalochordates emerge before echinoderms. These two topologies appeared significantly worse than the ML tree of Fig. 1, even for the conservative Shimodaira–Hasegawa test<sup>13</sup>.

Our results therefore indicate a strong phylogenetic affinity between tunicates and vertebrates to the exclusion of cephalochordates. However, obtaining high statistical support for a given topology does not necessarily indicate that the phylogenetic inference is correct. Indeed, the phylogenetic analysis of large-scale data sets requires particular attention to potential systematic biases associated, for instance, with differences in evolutionary rates among species, compositional biases and heterotachy<sup>8</sup>. In particular, a long-branch attraction (LBA) artefact<sup>14</sup> may potentially occur since tunicates include fast (*Ciona* spp.) and very fast (*O. dioica*) evolving species (Fig. 1). A high evolutionary rate of tunicate genes was already noticed in rRNA genes<sup>5</sup> and in complete mitochondrial genomes<sup>15</sup>. Our results confirm these observations for a large number of nuclear genes. As fast evolutionary rates are also often associated with compositional bias or with heterotachy, it is a necessary first step to exclude the possibility that the observed grouping of tunicates with vertebrates results from a tree reconstruction artefact.

The most obvious potential artefact, an LBA, predicts that the fast-evolving tunicates would be attracted towards the outgroup, and not by the slowly evolving vertebrates. This would produce a topology compatible with the classical hypothesis of chordate evolution where the slow-evolving cephalochordates and vertebrates group together. This prediction is perfectly congruent with the lower support for Olfactores observed with MP (90%), a method known to be more sensitive to LBA than probabilistic methods<sup>14</sup>. Indeed, when *O. dioica*

was used as the single representative of tunicates, MP unambiguously supported (BP = 100) an aberrant position for this group which emerged before cnidarians, disrupting the monophyly of bilaterians (Supplementary Fig. S1). By contrast, the less sensitive ML method recovered Olfactores, albeit with decreased bootstrap support (BP = 84) (Supplementary Fig. S2). Therefore, despite its extreme evolutionary rate, *O. dioica* retained enough phylogenetic signal for its position to be recovered with ML. This demonstrates that LBA is not responsible for the inferred grouping of tunicates and vertebrates, and represents a strong argument in favour of the authenticity of Olfactores. In addition, neither compositional bias nor heterotachy significantly influenced phylogenomic inference with our data set (see Supplementary Information). In fact, the compositional effect would act against Olfactores, because vertebrates and the amphioxus share similar amino acid compositions (Supplementary Fig. S3). In conclusion, the strongly supported monophyly of Olfactores cannot be explained by any kind of identifiable systematic bias (LBA, compositional bias, and heterotachy) and therefore constitutes the best current hypothesis for chordate phylogeny.

The monophyly of deuterostomes remained moderately supported in our phylogenomic analyses (Fig. 1). Also, the monophyly of chordates is not found in the ML tree, but is the only alternative not significantly rejected by likelihood-based statistical tests (Table 1). A unique origin of chordates and their distinctive features such as notochord and hollow nerve cord cannot be excluded. Our results nevertheless favoured the intriguing possibility of a sister-group relationship between cephalochordates and echinoderms that seems robust to analyses aimed at avoiding compositional bias and



**Figure 1 | Phylogenetic analyses of genomic data strongly support the grouping of tunicates and vertebrates into Olfactores.** ML tree obtained from the analysis of 33,800 aligned amino acid positions under a WAG substitution matrix plus a four-category gamma rate correction ( $\alpha = 0.5$ ) using two independent reconstruction algorithms (see Supplementary Information). Weighted maximum parsimony and bayesian inference using

the same WAG + F +  $\Gamma_4$  model and WAG + F +  $\Gamma_4$  plus covarion model also retrieved the same topology (see Supplementary Information). Bootstrap proportions obtained after 100 ML (red) and 1,000 MP replicates (blue), as well as bayesian posterior probabilities (black) are shown for selected branches. A star indicates that all three values are maximal (100%, 100% and 1.0). Scale bar indicates number of changes per site.

heterotachy (see Supplementary Information). Such a relationship has also been inferred from mitochondrial genomes<sup>16</sup>, but it lacks significant statistical support in both nuclear and mitochondrial analyses.

Although seemingly heretic, the grouping of echinoderms and cephalochordates constitutes an interesting working hypothesis. A similar situation was encountered a few years ago for the recently established sister-group relationship of echinoderms and hemichordates (Ambulacraria). The Ambulacraria hypothesis led to a re-evaluation of morphological character evolution, with the presence of pharyngeal slits being interpreted as an ancestral feature of deuterostomes<sup>9</sup>. Similarly, a close relationship between echinoderms and cephalochordates would imply that a dorsal nerve chord was already present in the last common deuterostome ancestor and subsequently evolved into derived nervous systems in both hemichordates and echinoderms. Such a scenario seems a priori possible given the difficulties encountered in polarizing morpho-anatomical characters in both extant<sup>9</sup> and fossil<sup>17</sup> deuterostomes. However, a definitive conclusion will only be achieved through the phylogenetic analysis of more genes combined with an increased taxon sampling including the enigmatic xenoturbellidans, hemichordates, and a greater diversity of echinoderms. Nonetheless, the strong support obtained for Olfactores will probably not be affected, as these additional taxa are considered to be on the echinoderm side of the deuterostome tree<sup>5,18</sup>. This prediction is supported by the observation that removing the sea-urchin from our data set has almost no effect (Supplementary Fig. S4).

Despite this remaining uncertainty, our new phylogenetic hypothesis implies a serious re-evaluation of fundamental aspects of deuterostome evolution. The nature of their last common ancestor has been most extensively addressed from the paleontological point of view<sup>3</sup>. However, extant deuterostome lineages are morphologically so distinct that possible stem-group representatives found in the fossil record are difficult to recognize<sup>17</sup>. A sister-group relationship of tunicates and vertebrates to the exclusion of cephalochordates is compatible with the controversial calcichordate theory of chordate origins proposed by Jefferies<sup>19</sup>. However, it does not mean that this evolutionary scenario based on the functional reconstruction of

unusual fossils with calcite skeletons (cornutes and mitrates) and their interpretation as stem-group chordates<sup>19</sup> is necessarily true. In fact, Jefferies<sup>10</sup> coined the name Olfactores on the basis of the presence of a putatively homologous olfactory apparatus in fossils that were proposed to be precursors of tunicates and vertebrates. However, the phylogenetic position of cornutes and mitrates is still much debated, with the majority advocating echinoderm affinities for these controversial fossils<sup>20,21</sup>. At any rate, the present molecular evidence for a monophyletic group of tunicates and vertebrates might help to polarize morphological characters of basal deuterostome fossils, thereby leading to a better understanding of early deuterostome evolution.

Our results also prompt a reinterpretation of morphological data in deuterostome phylogeny. In particular, a close proximity between tunicates and vertebrates suggests that the presence of metameric segmentation classically used to unify cephalochordates and vertebrates might be considered as an ancestral feature that underwent a secondary reduction in tunicates<sup>9</sup>. More generally, this new phylogenetic picture is in agreement with an alternative hypothesis for chordate evolution based on a recent homology analysis of morphological structures in hemichordates and chordates<sup>22</sup>. This unorthodox view proposes that cephalochordates have retained many ancestral characters that have been secondarily lost in the morphologically more derived tunicates, and reveals 13 putative synapomorphies uniting tunicates and vertebrates to the exclusion of cephalochordates<sup>22</sup>. The monophyly of Olfactores invalidates the traditional textbook representation of chordate, and even deuterostome, evolution as a steady increase towards complexity culminating in the highly specialized brain of vertebrates<sup>9</sup>. This anthropocentric interpretation is perhaps best reflected by the terms 'Euchordata' (that is, 'true chordates') or 'chordates with a brain', which are used to designate the grouping of cephalochordates and vertebrates<sup>2</sup>. Tunicates should therefore no longer be considered as 'primitive' but rather as derived chordates with highly specialized lifestyles and developmental modes.

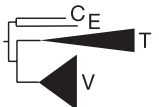
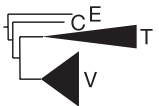
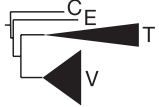
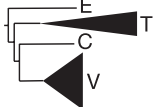
From the developmental point of view, our phylogenetic results help to understand the origin of the major evolutionary novelty constituted by the neural crest. This vertebrate innovation can be traced back to the origins of the chordate lineage, because 'latent homologues' of neural crest cells have been identified in both cephalochordates and tunicates<sup>23</sup>. However, evidence for migratory neural crest cells has so far been reported only in tunicates<sup>24</sup>, whereas their existence is still unproved in amphioxus. In light of the Olfactores hypothesis, these migratory cells may well have evolved in the last common ancestor of tunicates and vertebrates, after the divergence from cephalochordates, with these evolutionary precursor cells later giving birth to the neural crest along the vertebrate lineage<sup>24</sup>.

The newly proposed deuterostome phylogeny strengthens the view that tunicates and cephalochordates represent complementary models for studying the origin of the vertebrate developmental program. Indeed, tunicates are phylogenetically closer to vertebrates but are morphologically and molecularly highly derived with a trend towards genomic simplification<sup>25,26</sup>, whereas the more distantly related cephalochordates might have retained more ancestral characters<sup>27</sup>. The comparative analysis of available tunicate and vertebrate genomes with the upcoming amphioxus and sea-urchin genome sequences will be particularly valuable for understanding the evolution of new gene systems and structures involved in early vertebrate development.

## METHODS

**Data assembly.** We built upon a phylogenomic data set consisting of 146 nuclear genes previously assembled to study animal phylogeny<sup>7</sup>. This data set was updated using the same protocol (see Supplementary Information) with new sequences publicly available from the Trace Archive (<http://www.ncbi.nlm.nih.gov/Traces/>) and the EST Database (<http://www.ncbi.nlm.nih.gov/dbEST/>) of

**Table 1 | Results of likelihood-based tests of alternative topologies within deuterostomes**

Trees	−ln[L]	Δln[L]	AU	SH	RELL BP
	554,914.8	Best	0.947	1.000	0.938
	554,967.2	52.4	0.071	0.415	0.061
	555,051.5	136.7	0.000*	0.019*	0.000*
	555,066.4	151.6	0.004*	0.007*	0.002*

Results computed with a concatenated WAG + F + T<sub>4</sub> model are given for the top four ranking topologies; the other 11 alternative topologies were rejected by all tests with  $P < 0.001$ . L, likelihood value; AU, Approximately Unbiased test; SH, Shimodaira–Hasegawa test; RELL BP, Resampling of Estimated Log-Likelihood Bootstrap Percentage. \*Statistically significant at the 5% level. E, echinoderms; C, cephalochordates; T, tunicates; V, vertebrates.

GenBank at the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). Sequences from the appendicularian were generated by the *Oikopleura dioica* genome project ([http://www.genoscope.cns.fr/externe/English/Projets/Projet\\_HG/HG.html](http://www.genoscope.cns.fr/externe/English/Projets/Projet_HG/HG.html)).

As previously demonstrated<sup>7,8</sup>, taxon sampling has a major impact on phylogenomic studies. As an outgroup to the 14 available deuterostomes, we therefore selected the slowest-evolving taxa among available protostomes and fungi to reduce the potential impact of long-branch attraction<sup>14</sup>. Furthermore, we also incorporated all available cnidarians and choanoflagellates, allowing us efficiently to break the long branch leading to the distantly related fungal outgroup.

**Phylogenetic analyses.** Multiple methods using different optimality criteria and algorithms were used to analyse our phylogenomic data set. Weighted MP heuristic searches were conducted using PAUP<sup>28</sup> with ten random additions of species and Tree Bisection and Reconnection branch swapping. MP bootstrap percentages were obtained after 1,000 replications using the same heuristic search strategy. Because of the computational difficulties involved in conducting ML searches for such a large data set<sup>7</sup>, ML analyses were conducted with different algorithms (see Supplementary Information for details). ML bootstrap percentages were obtained after 100 replications. Bayesian phylogenetic inferences were also conducted using parallel computing (see Supplementary Information for details).

Likelihood-based tests of alternative topologies were calculated using CONSEL<sup>13</sup>. ML branch lengths of alternative topologies were first inferred assuming a concatenated WAG + F +  $\Gamma_4$  model using TREE-PUZZLE<sup>29</sup>, site-wise log-likelihood values were then computed with CODEML<sup>30</sup> and the *P* values of the different likelihood-based tests were finally calculated with CONSEL.

Received 9 August; accepted 19 October 2005.

- Schaeffer, B. Deuterostome monophyly and phylogeny. *Evol. Biol.* **21**, 179–235 (1987).
- Rowe, T. in *Assembling the Tree of Life* (eds Cracraft, J. & Donoghue, M. J.) 384–409 (Oxford Univ. Press, Oxford, 2004).
- Gee, H. *Before the Backbone: Views on the Origin of the Vertebrates* (Chapman & Hall, London, 1996).
- Zrzavy, J., Mihulka, S., Kepka, P., Bezdek, A. & Tietz, D. Phylogeny of the Metazoa based on morphological and 18S ribosomal DNA evidence. *Cladistics* **14**, 249–285 (1998).
- Winchell, C. J., Sullivan, J., Cameron, C. B., Swalla, B. J. & Mallatt, J. Evaluating hypotheses of deuterostome phylogeny and chordate evolution with new LSU and SSU ribosomal DNA data. *Mol. Biol. Evol.* **19**, 762–776 (2002).
- Blair, J. E. & Hedges, S. B. Molecular phylogeny and divergence times of deuterostome animals. *Mol. Biol. Evol.* **22**, 2275–2284 (2005).
- 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).
- Delsuc, F., Brinkmann, H. & Philippe, H. Phylogenomics and the reconstruction of the tree of life. *Nature Rev. Genet.* **6**, 361–375 (2005).
- Gee, H. in *Major Events in Early Vertebrate Evolution: Palaeontology, Phylogeny, Genetics, and Development* (ed. Ahlberg, P. E.) 1–14 (Taylor and Francis, London, 2001).
- Jefferies, R. P. S. in *Biological Asymmetry and Handedness* (eds Bock, G. R. & Marsh, J.) 94–127 (Wiley, Chichester, 1991).
- Oda, H., Akiyama-Oda, Y. & Zhang, S. Two classic cadherin-related molecules with no cadherin extracellular repeats in the cephalochordate amphioxus: distinct adhesive specificities and possible involvement in the development of multicell-layered structures. *J. Cell Sci.* **117**, 2757–2767 (2004).
- Felsenstein, J. *Inferring Phylogenies* (Sinauer, Sunderland, Massachusetts, 2004).
- Shimodaira, H. & Hasegawa, M. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* **17**, 1246–1247 (2001).
- Felsenstein, J. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* **27**, 401–410 (1978).
- Yokobori, S., Oshima, T. & Wada, H. Complete nucleotide sequence of the mitochondrial genome of *Doliolum nationalis* with implications for evolution of urochordates. *Mol. Phylogenet. Evol.* **34**, 273–283 (2005).
- Ruiz-Trillo, I., Riutort, M., Fourcade, H. M., Baguna, J. & Boore, J. L. Mitochondrial genome data support the basal position of Acoelomorpha and the polyphyly of the Platyhelminthes. *Mol. Phylogenet. Evol.* **33**, 321–332 (2004).
- Conway Morris, S. The Cambrian “explosion”: slow-fuse or megatonnage? *Proc. Natl Acad. Sci. USA* **97**, 4426–4429 (2000).
- Bourlat, S. J., Nielsen, C., Lockyer, A. E., Littlewood, D. T. & Telford, M. J. *Xenoturbella* is a deuterostome that eats molluscs. *Nature* **424**, 925–928 (2003).
- Jefferies, R. P. S. *The Ancestry of the Vertebrates* (Cambridge Univ. Press, London, 1986).
- Peterson, K. J. A phylogenetic test of the calcichordate scenario. *Lethaia* **28**, 25–38 (1995).
- Jefferies, R. P. S. A defence of the calcichordates. *Lethaia* **30**, 1–10 (1997).
- Ruppert, E. E. Key characters uniting hemichordates and chordates: homologies or homoplasies? *Can. J. Zool.* **83**, 8–23 (2005).
- Stone, J. R. & Hall, B. K. Latent homologues for the neural crest as an evolutionary novelty. *Evol. Dev.* **6**, 123–129 (2004).
- 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).
- Seo, H. C. et al. Hox cluster disintegration with persistent anteroposterior order of expression in *Oikopleura dioica*. *Nature* **431**, 67–71 (2004).
- Edwards, R. B. et al. Remodelling of the homeobox gene complement in the tunicate *Oikopleura dioica*. *Curr. Biol.* **15**, R12–R13 (2005).
- Holland, L. Z., Laudet, V. & Schubert, M. The chordate amphioxus: an emerging model organism for developmental biology. *Cell. Mol. Life Sci.* **61**, 2290–2308 (2004).
- Swofford, D. L. *PAUP\*: Phylogenetic Analyses Using Parsimony and Other Methods* (Sinauer, Sunderland, Massachusetts, 2000).
- Schmidt, H. A., Strimmer, K., Vingron, M. & von Haeseler, A. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**, 502–504 (2002).
- Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556 (1997).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank S. Conway Morris, R. P. S. Jefferies, W. R. Jeffery and J. Mallatt for suggestions, and N. Lartillot and N. Rodrigue for critical readings of early versions of the manuscript. *Oikopleura* genome data were generated at Génoscope Evry (France) with material and co-funding from the Sars International Centre. We are grateful to P. Wincker and the Génoscope team. We gratefully acknowledge the financial support provided by Génome Québec, the Canadian Research Chair and the Université de Montréal, and the Réseau Québécois de Calcul de Haute Performance for computational resources.

**Author Contributions** H.P. conceived the study. D.C. contributed sequence data from the *Oikopleura* genome project. F.D., H.B. and H.P. assembled the data set and performed phylogenetic analyses. F.D. wrote the first draft of the manuscript and all authors contributed to the writing of its final version.

**Author Information** Reprints and permissions information is available at [npg.nature.com/reprintsandpermissions](http://npg.nature.com/reprintsandpermissions). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.P. ([herve.philippe@umontreal.ca](mailto:herve.philippe@umontreal.ca)).