Hemichordate genomes and deuterostome origins

Oleg Simakov1,2,*, Takeshi Kawashima3,*, Ferdinand Marlétaz4, Jerry Jenkins5, Ryo Koyanagi6, Therese Mitros7, Kanako Hisata3, Jessen Bredeson3, Eiichi Shoguchi3, Fuki Gyoja3, Ita–Xing Yue4,†, Yi–Chih Chen7, Robert M. Freeman Jr10,†, Akane Sasaki11, Tomoe Hikosaka–Katayama12, Atsuko Sato13, Manabu Fujie6, Kenneth W. Baughman1, Judith Levine14, Paul Gonzalez14, Christopher Cameron14, Jens H. Fritzenwanker14, Ariel M. Pani15, Hiroki Goto6, Mlyuki Kanda6, Nana Arakaki6, Shinichi Yamasaki16, Jiaxin Qu17, Andrew Cree17, Yan Ding17, Huyen H. Dinh17, Shannon Dugan17, Michael Holder17, Shalini N. Jhangiani17, Christie L. Kovar17, Sandra L. Lee17, Lora R. Lewis17, Donna Morton17, Lynne V. Nazareth17, Geoffrey Okwuonu17, Jireh Santibanez17, Rui Chen7, Stephen Richards7, Donna M. Muzny17, Andrew Gillis18, Leonid Peshkin10, Michael Wu19, Tom Humphreys19, Yi–Hsien Su19, Nicholas H. Putnam19, Jeremy Schmutz17, Asao Fujitaya20, Jr–Kai Yu17, Kunifumi Tagawa11, Kim C. Worley17, Richard A. Gibbs17, Marc W. Kirschner10, Christopher J. Lowe14, Noriyuki Satoh11, Daniel S. Rokhsar1,2,† & John Gerhart2

*These authors contributed equally to this work.

The prominent pharyngeal gill slits, rigid stomochord, and midline nerve cords of acorn worms led 19th century zoologists to designate them as ‘hemichordates’ and group them with vertebrates and other chordates3–4, but their early embryos and larvae also linked them to echinoderms5,6. Current molecular phylogenies strongly support the affinities of hemichordates and echinoderms as sister phyla, together called ambulacrarians7, and unite ambulacrarians and chordates within the deuterostomes (see glossary in Supplementary Note 1). Of all the shared derived morphological characters proposed between hemichordates and chordates, the pharyngeal gill slits have emerged with unambiguous morphological and molecular support, notably the shared expression of the pax1/9 gene6–10. These structures were ancestral deuterostome characters elaborated upon the bilaterian ancestral body plan, but the gill slits were subsequently lost in extant echinoderms and annelids11. Since extant invertebrate deuterostomes use this apparatus for efficient suspension and/or deposit feeding, the early Cambrian or Precambrian deuterostome ancestor probably also shared this lifestyle. This perspective on the last common deuterostome ancestor informs our understanding of the subsequent evolution of hemichordates, echinoderms and chordates10,12–16.

Hemichordates share bilateral symmetry, gill slits, soft bodies and early axial patterning with chordates, making them key comparators for inferring the ancestral genomic features of deuterostomes. To this end, we sequenced and analysed the genomes of acorn worms belonging to the two main lineages of enteropneust hemichordates (Supplementary Note 1): Saccoglossus kowalevskii (Harrimaniidae; Atlantic, North America, Fig. 1a) and Ptychodera flava (Ptychoderidae; Pacific, pan-tropical, Fig. 1b). Both have characteristic three-part
bodies comprising proboscis, collar and trunk, the last with tens to hundreds of pairs of gill slits. While S. kowalevskii develops directly to a juvenile worm with these traits within days (Fig. 1c, e), P. flava develops indirectly through a feeding larva that metamorphoses to a juvenile worm after months in the plankton (Fig. 1d, e). Our analyses begin to integrate macroscopic information about morphology, organisational physiology, and descriptive embryology of these deuterostomes with genomic information about gene homologies, gene arrangements, gene novelties and non-coding elements.

Genomes

We sequenced the two acorn worm genomes by random shotgun methods with a variety of read types (Methods; Supplementary Note 2), each starting from sperm from a single outbred diploid individual. The haploid lengths of the two genomes are both about 1 Gbp (Extended Data Fig. 1), but differ in nucleotide heterozygosity. Both acorn worm genomes were annotated using extensive transcriptome data as well as standard homology-based and de novo methods (Supplementary Note 3). Counting gene models with at least one detectable orthologue as standard homology-based and de novo gene predictions include divergent and/or novel genes (Extended Data Fig. 1). Despite the ancient divergence of the Saccoglossus and Ptychodera lineages (more than 370 million years ago, see below) and their different modes of development, the two acorn worm genomes have similar bulk gene content, as discussed later (Extended Data Fig. 2 and Supplementary Note 4), and similar repetitive Landscapes (Supplementary Note 5).

Deuterostome phylogeny

Deuterostome relationships were originally inferred from developmental and morphological characters25,17 and these hypotheses were later tested and refined with molecular data8,9. Aspects of deuterostome phylogeny continue to be controversial, however, notably the position of the sessile pterobranchs among hemichordates, and the surprising association of Xenoturbella19 and acoelomorph flatworms with ambulacrarians19 proposed by some studies. We explored these issues using genome-wide analyses of the newly sequenced hemichordate genomes augmented with extensive new RNAseq from five echinoderms, three additional hemichordates (including a rhabdopleurid pterobranch) and two acoels (Fig. 2, Extended Data Fig. 3, Methods and Supplementary Note 6). We recovered the monophyly of hemichordates, echinoderms, ambulacrarians and deuterostomes, using not only amino acid characters but also presence–absence characters for introns and coding indels (Supplementary Note 4). Our analyses also placed pterobranch hemichordates as the sister-group to enteropneusts rather than within them12. These phylogenetic analyses imply that genomic traits shared by chordates and ambulacrarians can be attributed to the last common deuterostome ancestor (see below). Using a relaxed molecular clock, we estimate a Cambrian origin of hemichordates (Methods, Extended Data Fig. 3 and Supplementary Note 6).

We also performed several analyses to assess the controversial relationships between Xenoturbella, acoelomorphs and deuterostomes (Supplementary Note 6). With conventional site-homogeneous models, acoels remain outside deuterostomes20–23 (Fig. 2, Supplementary Figs 6.1 and 6.2). Alternative models, however, show equivocal branching of acoels depending on the inclusion of the current sparse data for Xenoturbella (Supplementary Note 6). Notably, without Xenoturbella, acoels are positioned as a bilaterian sister group (Supplementary Fig. 6.3)20–23. Although we cannot rule out a deuterostome placement for Xenoturbella, our analyses generally do not support a grouping of acoels with deuterostomes19.

The gene set of the deuterostome ancestor

By comparative analysis, we identified 8,716 families of homologous genes whose distributions in sequenced extant genomes imply their presence in the deuterostome ancestor (Methods; Supplementary Note 4). Owing to gene duplication and other processes the descendants of these ancestral genes account for ~14,000 genes in extant deuterostome genomes including human (Supplementary Table 4.1.2). The distributions of gene functions, domain compositions, and gene family sizes of hemichordates resemble those of amphioxus, sea urchin, and sequenced lophotrochozoans more than those of ecdysozoans; vertebrates also form a distinct group (Extended Data Fig. 2, Supplementary Note 4 and Supplementary Fig. 4.2).

Exon–intron structures of genes are generally well conserved among hemichordates, chordates, and many non-deuterostome metazoans, allowing us to infer 2,061 ancestral deuterostome splice sites (Supplementary Note 4). Among orthologous bilaterian genes we found 23 introns and 4 coding sequence indels present only in deuterostomes (shared between at least one ambulacarian and chordate), suggesting that these shared derived characters may be useful to diagnose clade membership of new candidate organisms (Supplementary Note 4).

Based on whole-genome alignments, we identified 6,533 conserved non-coding elements (CNE) longer than 50 bp that are found in all of the five deuterostomes Saccoglossus, Ptychodera, amphioxus,
genomic rearrangement or, more interestingly, as a result of selection to stomie genomes (Methods, Fig. 3b and Extended Data Figs 5 and 6). Saccoglossus acorn worms (Extended Data Fig. 4), as also found in echinoderms 32, 33. Supplementary Note 7) including Hox 30 and ParaHox 31 clusters in both implying that these chromosome-scale linkages were also present in chordate-defined linkage groups (Fig. 3a and Supplementary Note 7), organization, we find that the is too limited to determine whether it shares this chromosome-scale While the contiguity of the draft of the sea urchin genome assembly 29 Ciona linkage groups across chordates, including amphioxus and 27. Ancient gene linkages (‘macro-synteny’ 27) are often preserved in extant sox14/21 ortho - these conserved sequences is a previously identified vertebrate brain alignments usually do not exceed 250 bp (as has been reported among (3,611 CNE loci; 55%, Fisher’s exact test value P value < 2.2 × 10⁻¹⁶). Those alignments usually do not exceed 250 bp (as has been reported among vertebrates25) and occur in clusters (Supplementary Note 8). Among these conserved sequences is a previously identified vertebrate brain and neural tube specific enhancer, located close to the sox14/21 orthologue in all five species26.

Conserved gene linkage

Ancient gene linkages (‘macro-synteny’27) are often preserved in extant bilaterian genomes27,28. Comparative analysis revealed 17 ancestral linkage groups across chordates, including amphioxus and Ciona27. While the contiguity of the draft of the sea urchin genome assembly29 is too limited to determine whether it shares this chromosome-scale organization, we find that the Saccoglossus genome clearly shares these chordate-defined linkage groups (Fig. 3a and Supplementary Note 7), implying that these chromosome-scale linkages were also present in the ancestral deuterostome.

On a more local scale, we find hundreds of tightly linked conserved gene clusters of three or more genes (‘micro-synteny’; Methods; Supplementary Note 7) including Hox30 and ParaHox31 clusters in both acorn worms (Extended Data Fig. 4), as also found in echinoderms32,33 Saccoglossus and amphioxus share more micro-syntenic linkages with each other than either does with sea urchin, vertebrates, or available protostome genomes (Methods, Fig. 3b and Extended Data Figs 5 and 6). Conservation of micro-syntenic linkages can occur due to low rates of genomic rearrangement or, more interestingly, as a result of selection to retain linkages between genes and their regulatory elements located in neighbouring genes28.

A deuterostome pharyngeal gene cluster

One conserved deuterostome-specific micro-syntenic cluster with functional implications for deuterostome biology is a cluster of genes expressed in the pharyngeal slits and surrounding pharyngeal endo - (Fig. 4; Supplementary Note 9). This six-gene cluster contains four transcription factor genes in the order ntx2.1, ntx2.2, pax1/9 and foxA, along with two non-transcription-factor genes sulc25A21 and mipol1, whose introns harbour regulatory elements for pax1/9 and foxA, respectively34-36. The cluster was first found conserved across vertebrates including humans (see chromosome 14; 1.1 Mb length from ntx2.1 to foxA34,35). In S. kowalevskii, it is intact with the same gene order as in vertebrates (0.5 Mb length from ntx2.1 to foxA), implying that it was present in the deuterostome and ambulacrarian ancestors. The full ordered gene cluster also exists on a single scaffold in the crown-of-thorns sea star Acanthaster planci. Since these genes are not clustered in available protostome genomes, there is no evidence for deeper bilaterian ancestry. Two non-coding elements that are conserved across vertebrates and amphioxus34 are found in the hemichordate and A. planci clusters at similar locations (A2 and A4, in Fig. 4a).

The pax1/9 gene, at the centre of the cluster, is expressed in the pharyngeal endodermal primordium of the gill slit in hemichordates, tunicates, amphioxus, fish, and amphibians8,9 and in the branchial pouch endoderm of amniotes (which do not complete the last steps of gill slit formation), as well as other locations in vertebrates. The ntx2.1 (thryroid transcription factor 1) gene is also expressed in the hemichordate

---

Figure 2 | Phylogenetic placement of deuterostome taxa within the metazoan tree. Maximum-likelihood tree obtained with a super-matrix of 506,428 amino-acid residues gathered from 1,564 orthologous genes in 52 species (65.1% occupancy) and using a LG+Γ model partitioned for each sea urchin, and human (Methods; Supplementary Note 8). The identified CNEs overlap extensively with human long non-coding RNAs (3,611 CNE loci; 55%, Fisher’s exact test P value < 2.2 × 10⁻¹⁶). Those alignments usually do not exceed 250 bp (as has been reported among vertebrates25) and occur in clusters (Supplementary Note 8). Among these conserved sequences is a previously identified vertebrate brain and neural tube specific enhancer, located close to the sox14/21 orthologue in all five species26.

Conserved gene linkage

Ancient gene linkages (‘macro-synteny’27) are often preserved in extant bilaterian genomes27,28. Comparative analysis revealed 17 ancestral linkage groups across chordates, including amphioxus and Ciona27. While the contiguity of the draft of the sea urchin genome assembly29 is too limited to determine whether it shares this chromosome-scale organization, we find that the Saccoglossus genome clearly shares these chordate-defined linkage groups (Fig. 3a and Supplementary Note 7), implying that these chromosome-scale linkages were also present in the ancestral deuterostome.

On a more local scale, we find hundreds of tightly linked conserved gene clusters of three or more genes (‘micro-synteny’; Methods; Supplementary Note 7) including Hox30 and ParaHox31 clusters in both acorn worms (Extended Data Fig. 4), as also found in echinoderms32,33 Saccoglossus and amphioxus share more micro-syntenic linkages with each other than either does with sea urchin, vertebrates, or available protostome genomes (Methods, Fig. 3b and Extended Data Figs 5 and 6). Conservation of micro-syntenic linkages can occur due to low rates of genomic rearrangement or, more interestingly, as a result of selection to retain linkages between genes and their regulatory elements located in neighbouring genes28.

A deuterostome pharyngeal gene cluster

One conserved deuterostome-specific micro-syntenic cluster with functional implications for deuterostome biology is a cluster of genes expressed in the pharyngeal slits and surrounding pharyngeal endo - (Fig. 4; Supplementary Note 9). This six-gene cluster contains four transcription factor genes in the order ntx2.1, ntx2.2, pax1/9 and foxA, along with two non-transcription-factor genes sulc25A21 and mipol1, whose introns harbour regulatory elements for pax1/9 and foxA, respectively34-36. The cluster was first found conserved across vertebrates including humans (see chromosome 14; 1.1 Mb length from ntx2.1 to foxA34,35). In S. kowalevskii, it is intact with the same gene order as in vertebrates (0.5 Mb length from ntx2.1 to foxA), implying that it was present in the deuterostome and ambulacrarian ancestors. The full ordered gene cluster also exists on a single scaffold in the crown-of-thorns sea star Acanthaster planci. Since these genes are not clustered in available protostome genomes, there is no evidence for deeper bilaterian ancestry. Two non-coding elements that are conserved across vertebrates and amphioxus34 are found in the hemichordate and A. planci clusters at similar locations (A2 and A4, in Fig. 4a).

The pax1/9 gene, at the centre of the cluster, is expressed in the pharyngeal endodermal primordium of the gill slit in hemichordates, tunicates, amphioxus, fish, and amphibians8,9 and in the branchial pouch endoderm of amniotes (which do not complete the last steps of gill slit formation), as well as other locations in vertebrates. The ntx2.1 (thyroid transcription factor 1) gene is also expressed in the hemichordate

---

Figure 2 | Phylogenetic placement of deuterostome taxa within the metazoan tree. Maximum-likelihood tree obtained with a super-matrix of 506,428 amino-acid residues gathered from 1,564 orthologous genes in 52 species (65.1% occupancy) and using a LG+Γ model partitioned for each sea urchin, and human (Methods; Supplementary Note 8). The identified CNEs overlap extensively with human long non-coding RNAs (3,611 CNE loci; 55%, Fisher’s exact test P value < 2.2 × 10⁻¹⁶). Those alignments usually do not exceed 250 bp (as has been reported among vertebrates25) and occur in clusters (Supplementary Note 8). Among these conserved sequences is a previously identified vertebrate brain and neural tube specific enhancer, located close to the sox14/21 orthologue in all five species26.© 2015 Macmillan Publishers Limited. All rights reserved
pharyngeal endoderm in a band passing through the gill slit, but not localized to a thyroid-like organ. Here we also examined the expression of *nkx2.2* and *foxA* in *S. kowalevskii*. We find that *nkx2.2*, which is expressed in the ventral hindbrain in vertebrates, is expressed in pharyngeal ventral endoderm in *S. kowalevskii*, close to the gill slit (Fig. 4b), and that *foxA* is expressed throughout endoderm but repressed in the gill slit region (Fig. 4b). The co-expression of this ordered cluster of the four transcription factors during pharyngeal development strongly supports the functional importance of their genomic clustering.

The presence of this cluster in the crown-of-thorns sea star, an echinoderm that lacks gill pores, and in amniote vertebrates that lack gill slits, suggests that the cluster’s ancestral role was in pharyngeal apparatus patterning as a whole, of which overt slits (perforations of adult gill pores) as a whole, of which overt slits (perforations of gill slits, suggests that the cluster’s ancestral role was in pharyngeal echinoderm that lacks gill pores, and in amniote vertebrates that lack genomic clustering.

The presence of this cluster in the crown-of-thorns sea star, an echinoderm that lacks gill pores, and in amniote vertebrates that lack gill slits, suggests that the cluster’s ancestral role was in pharyngeal apparatus patterning as a whole, of which overt slits (perforations of gill slits, suggests that the cluster’s ancestral role was in pharyngeal echinoderm that lacks gill pores, and in amniote vertebrates that lacks gill pores, and in amniote vertebrates that lack genomic clustering.
Figure 5 | Examples of deuterostome gene novelties. a. Steps of biosynthesis of sialic acid and its addition to and removal from glycoproteins. b–d. Novel genes in TGFβ signalling pathways. The encoded proteins are shown and include Lefty (b), an antagonist of Nodal signalling, which activates Smad2/3-dependent transcription when not antagonized; Univin (c), an agonist of Nodal signalling, also called Vg1, DVR1, and GDF1; and TGFβ3 (d), a ligand that activates Smad2/3-dependent transcription by binding to a deuterostome-specific TGFβ receptor type II, which contains a novel ectodomain (not shown). Also shown in d is the novel protein thrombospondin 1 that activates TGFβ signalling, including Smad2/3 and Smad4. Red boxes around protein names indicate their deuterostome novelty. Green boxes around the names indicate genes with pan-metazoan/bilaterian ancestry and without accelerated sequence change in the deuterostome lineage.

Deuterostome novelties

We found >30 deuterostome genes with sequences that differ markedly from those of other metazoans, related to functional innovation in deuterostomes. Some plausibly arose from accelerated sequence change on the deuterostome stem from distant but identifiable bilaterian homologues, others represent new protein domain combinations in deuterostomes, while others lack identifiable sequence and domain homologues in other animals. In the latter group, we found over a dozen deuterostome genes that have readily identified relatives in marine microbes, often cyanobacteria or eukaryotic micro-algae, but are not known in other metazoans (Extended Data Table 1 and Extended Data Fig. 7; Supplementary Notes 10.4 and 10.5). Such genes include two of the novel deuterostome sequences associated with sialic acid metabolism (found in many microbes, see below), enzymes that modify proteins (for example, protein arginine deiminase) and RNA (for example, FATSO methyladenosine demethylase) as well as others that provide specialized reactions of secondary metabolism (Extended Data Table 1 and Extended Data Fig. 7; Supplementary Note 10.5). Possible explanations for the unusual phylogenetic distribution of these genes include horizontal transfer on the deuterostome stem from early marine microbes (which were plausibly commensals, pathogens, or food sources of stem deuterostomes), or convergent gene loss and/or extensive sequence divergence along five or more opisthokont lineages (Supplementary Note 10.2).

Regardless of their mechanism of origination, the various deuterostome novelties and gene family expansions of sialic acid metabolism are noteworthy. Deuterostomes are unique among metazoans in their high level and diverse linkage of addition of sialic acid (also known as neuraminic acid), a nine carbon negatively charged sugar, to the terminal sugars of glycoproteins, mucins and glycolipids. We find expanded families of enzymes for several of these reactions in hemichordates (Fig. 5a and Extended Data Table 1). Based on the presence/absence of relevant enzymes we infer that 5 of the 11 steps of the pathways of sialic acid formation, addition to termini, and removal are not found in protostomes or other metazoans, and are deuterostome novelties (Fig. 5a and Supplementary Note 10), whereas the other steps use
enzymes similar to those of the more limited pathway of some protostomes (for example, insects such as Drosophila). The importance of glycoproteins for muco-ciliary feeding and other hemichordate activities is further supported by novel and expanded families of genes encoding the polypeptide backbones of glycoproteins, those with von Willebrand type-D and/or cysteine-rich domains (PTHR11339 classifier), including mucins, present in hemichordates and amphioxus as large tandemly duplicated clusters (with varied expression patterns as shown in Extended Data Fig. 8), but not in sea urchin, which has a different mode of feeding (Supplementary Note 10). As in amphioxus, the pharynx of Saccoglossus is heavily ciliated, and cells of the pharyngeal walls in hemichordates and the ventral endostyle in amphioxus secrete abundant mucins and glycoproteins. Similarly, in the deuterostome ancestor these glycoproteins probably enhanced the muco-ciliary filter-feeding capture of food particles from the microbe-rich marine environment and protected its inner and outer tissue surfaces.

Novelty in the TGFβ3 signalling pathway

The signalling ligands Lefty (a Nodal antagonist) and Uninv/Vg1/GDF1 (a Nodal agonist) are deuterostome innovations that modulate Nodal signalling during the major developmental events of endomesoderm induction and axial patterning in vertebrates, axial patterning in hemichordates and echinoderms, and left–right patterning in all deuterostomes (see Fig. 5b–d and Extended Data Fig. 9a, b). Uninv is tightly linked to the related bilaterianbmp2/4 in the sea urchin genome and also, we now report, in hemichordates and amphioxus, supporting its origin by tandem duplication and divergence from an ancestralbmp2/4-type gene, as suggested previously.

TGFβ3 signalling (TGFβ1, 2 and 3 in vertebrates) is a deuterostome innovation that controls cell growth, proliferation, differentiation and apoptosis at later developmental stages. Accompanying the novel TGFβ3 ligand, the type II receptor has a novel ectodomain. The extra-cellular matrix protein thrombospondin 1, which activates TGFβ3 in vertebrates, contains a deuterostome-unique combination of domains including three thrombospondin type 1 (TSP1) domains that bind the TGFβ3 pro-domain region. While these signalling novelties have clear sequence similarity to pan-bilaterian components, they form long stem branch clades on the phylogenetic trees, indicating extensive sequence divergence on the deuterostome stem (Supplementary Note 10). Together, these innovations appear to contribute to the increased amount and complex patterning of Smad2/3-mediated signalling in deuterostomes compared with protostomes and other metazoans.

Conclusion

The two acorn worms whose genomes are described here represent the two main enteropneust lineages, separated by at least 370 million years and differing in their developmental modes. These analyses reveal (1) extensive conserved macro-synteny among deuterostomes; (2) a widely conserved deuterostome-specific cluster of six ordered genes, including four transcription factor genes that are expressed during the development of pharyngeal gill slits and the branchial apparatus, the most prominent morphological innovation of the deuterostome ancestor; and (3) numerous gene novelties shared among deuterostomes, many expanded into large families, with putative protein functions that imply physiological, metabolic and developmental specializations of the filter-feeding deuterostome ancestor. Some of these genes lack identifiable orthologues in other metazoans but do resemble microbial sequences and domain types. In addition to their contributions towards defining the deuterostome ancestor and illuminating chordate origins, the two genomes should inform hypotheses of larval evolution by providing a basis for future comparisons of direct-developing and indirect-developing acorn worms, which achieve remarkably similar adult forms by distinct embryological routes (Fig. 1).
31. Ikuta, T. et al. Identification of an intact ParaHox cluster with temporal colinearity but altered spatial colinearity in the hemichordate Ptychodera flava. BMC Evol. Biol. 13, 129 (2013).
32. Cameron, R. A. et al. Unusual gene order and organization of the sea urchin hox cluster. J. Exp. Zool. B Mol. Dev. Evol. 306, 45–58 (2006).
33. Baughman, K. W. et al. Genomic organization of Hox and ParaHox clusters in the echinoderm, Acanthaster plancis. Genesis 52, 952–958 (2014).
34. Santagati, F. et al. Identification of cis-regulatory elements in the mouse Pax9/Nkx2–5 genomic region: implication for evolutionary conserved synteny. Genetics 165, 235–242 (2003).
35. Lowe, C. J. Antero-posterior patterning in hemichordates: insights into early chordate evolution. PLoS Biol. 4, e291 (2006).
36. Wang, W., Zhong, J., Su, B., Zhou, Y. & Wang, Y. Q. Comparison of Pax1/9 locus reveals 500-Myr-old syntenic block and evolutionary conserved noncoding regions. Mol. Biol. Evol. 24, 784–791 (2007).
37. Santagati, F. et al. Comparative analysis of the genomic organization of Pax9 and its conserved physical association with Nkx2–9 in the human, mouse, and pufferfish genomes. Mamm. Genome 12, 232–237 (2001).
38. Wang, S., Zhang, S., Zhao, B. & Lun, L. Up-regulation of C/EBP by thyroid hormones: a case demonstrating the vertebrate-like thyroid hormone signaling pathway in amphioxus. Mol. Cell. Endocrinol. 313, 57–63 (2009).
39. Lowe, C. J. et al. Antero-posterior patterning in hemichordates and the origins of the chordate nervous system. Cell 113, 853–865 (2003).
40. Kokubu, C. et al. Comparison of genomic organization of ST8Sia and sialyltransferase-related genes: a phylogenetic approach. Glycobiology 15, 805–817 (2005).
41. Giacopuzzi, E., Bresciani, R., Schauer, R., Monti, E. & Borsani, G. New insights into the chordate nervous system. Cell 134, 3649–3664 (2008).
42. Pardos, F. Fine structure and function of pharynx cilia in Saccoglossus kowalevskii (Enteropneusta). Acta Zoológica 68, 1–12 (1988).
43. Kauf-Strehlow, S. & Stach, T. A detailed description of the development of the hemichordate Saccoglossus kowalevskii using SEM, TEM, Histology and 3D-reconstructions. Front. Zool. 10, 53 (2013).
44. Rudert, F. E., Gomaa, M. O. & Frick, J. E. Endostyle-like features of the dorsal epibranchial ridge of an enteropneust and the hypothesis of dorsal-ventral axis inversion in chordates. Invertebr. Biol. 118, 202–212 (1999).
45. Range, R. & Lepage, T. Maternal Oct1/2 is required for Nodal and Vg1/Univin expression during dorsal-ventral axis specification in the sea urchin embryo. Dev. Biol. 357, 440–449 (2011).
46. Massagué, J. TGFβ signalling in context. Nature Rev. Mol. Cell Biol. 13, 616–630 (2012).
47. Range, R. et al. Cis-regulatory analysis of nodal and maternal control of dorsal-ventral axis formation by Univin, a TGF-β1 related to Vg1. Development 134, 3649–3664 (2007).

Supplementary Information is available in the online version of the paper.

Author Contributions J.Q., K.C.W. assembled the initial S. kowalevskii genomic assembly and performed quality assessments of the genome assemblies. Ptychodera collection, genome sequencing, and assembly: T.K., K.T., A.S., R.K., H.G., M.F., M.I.K., N.A., S.Y., A.F., T.H. Rhabdopleura collection: A.S. Saccoglossus RNA sequencing and analysis: R.M.F., M.W.K., R.C., C.L.K., S.L.L., M.H., S.R., D.M.M., K.C.W. Genome sequence production: A.C., X.D., H.H.D., S.D., M.H., S.N.J., C.L.K., S.L.L., L.R.L., D.M., L.V.N., G.O., J.Sa., S.R., K.C.W., D.M.M., L.P., B.F., M.W.K. Saccoglossus sequence finishing: S.D., Y.D., D.M.M. Final Saccoglossus assembly: J.J., J.Sc. Saccoglossus gene model and validation: T.M., J.B., J.H.F., A.M.P., M.W. Ptychodera gene modelling and analyses: T.K., R.K., K.H., E.S., F.G., K.W.B., K.T., O.S., J.G., N.S. Gene family analyses: O.S., T.K., L.P., R.M.F., C.L., J.G. Synteny: N.H.P., O.S., J.–Y.Y. Repeats: O.S. Saccoglossus sequencing and assembly project management: S.R., D.M.M., K.C.W., R.A.G. Ptychodera expression analysis: Y.C.C., Y.H.S., J.–Y.Y. Phylogenetic analyses: F.M. Additional EST collections: T.H.–K., K.T., A.S., A.T.S., J.P., P.G., C.C., C.L. HGT and novelties: J.G., O.S. Pharyngeal cluster analysis and expression: J.G., O.S., N.S., K.B., A.G. Project coordination, manuscript writing: O.S., T.K., F.M., K.T., N.S., J.G., C.L., D.S.R.

Author Information Sequencing data have been deposited in NCBI Bioproject under accession number PRJNA12887 (Saccoglossus kowalevskii) and DDBJ under accession number PRJDB182 (Ptychodera flava). 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 the paper. Correspondence and requests for materials should be addressed to O.S. (oleg.simakov@oist.jp), J.G. (jgerhart@berkeley.edu), N.S. (norsisky@oist.jp) and D.S.R. (dsrokhsar@gmail.com).

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported licence. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/3.0/
METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Sequencing. Sperm DNA from adult males was extracted for sequencing as described in Supplementary Note 2. A single male was used for each species to minimize the impact of heterozygosity on assembly. For Saccoglossus, approximately eightfold redundant random shotgun coverage (totaling 8.1 Gb) was obtained with Sanger dideoxy sequencing at the Baylor College of Medicine Genome Center, including 34,279 BAC ends and 459,052 fosmid ends. For Ptychodera, 1.3 Gb in Sanger shotgun sequences, 15.3 Gb in Roche 454 pyrosequence reads, and 52-Gb paired-end sequences with Illumina MiSeq, along with mate-pairs, were generated at the Okinawa Institute of Science and Technology Graduate University. More sequencing details are available in Supplementary Note 2.

Genome assemblies. We assembled the Saccoglossus genome with Arachne2, combined with BAC/fosmid pair information to produce the final assembly. This Saccoglossus assembly includes 7,282 total scaffold sequences spanning a total length of 758 Mb. The relatively modest nucleotide heterozygosity (0.5%) of S. kowalevskii, coupled with longer read lengths, enabled assembly of a single composite reference sequence. Half of the assembly is in scaffolds longer than 552 kb (the N50 scaffold length), and 82% of the assembled sequence is found in 1,602 scaffolds longer than 100 kb. For Ptychodera we used the Platanus51 assembler. The resulting total scaffold length was 1,229 Mb, with half the assembly in scaffolds longer than 196 kb (N50 scaffold length). P. flavu exhibited a notably higher heterozygosity (1.3% single nucleotide heterozygosity with frequent indels) than S. kowalevskii, presumably related to its pelagic dispersal and larger effective population size25. We therefore initially produced stringent separate assemblies of the two divergent haplotypes, and found that many scaffolds had a closely related second scaffold with ~94% BLASTN identity (over longer stretches, including indels). To avoid reporting both haplotypes at these loci, scaffolds with less than 6% divergence over at least 75% of their length were merged into a single haplotype reference for comparative analysis. To further classify regions with ‘double’ depth and single haplotype regions we implemented a Hidden Markov Model classifier. We find that at least 63% of the initial Platanus assembly constitutes merged haplotypes. The inferred SNP rate for these regions is 1.3%, while for the remaining haplotype regions it is below 0.1%. Further details of assemblies are described in Supplementary Note 2.

Gene predictions. Transcriptome data for both species were used, along with homology-guided and ab initio methods, to predict protein-coding genes (Supplementary Note 3). For Saccoglossus, 8.6 million RNAseq reads were generated from 7 adult tissues and 15 developmental stages using Roche 454 sequencing, along with previously deposited ESTs in GenBank. For Ptychodera, extensive EST data from egg, blastulae, gastrulae, larvae, juveniles, adult proboscis, stomochord, and gills defining 34,159 cDNA clones24, and 879,000 Roche/454 RNAseq reads from a mixed library of developmental stages48 were used. The Saccoglossus genome was annotated using IGI gene prediction pipeline52, while Augustus56 was used to produce gene models for Ptychodera. We find a total of 34,239 gene predictions for Saccoglossus (68% with transcript evidence) and 34,687 for Ptychodera (43% with transcript evidence), although these are overestimates of the true gene number due to fragmented gene predictions, mis-annotated repetitive sequences, and spurious predictions. As described in the main text, 18–19,000 gene models in each species have known annotations and/or orthologues in other species.

Gene family analysis. Gene family clustering was done using a progressive (leaf to root) BLASTP-based clustering algorithm, where at a given phylogenetic node the gene families are constructed taking into account protein similarities among ingroups and outgroups57. For the inference of deuterostome gene families we use the bilaterian node of the clustering. To call gene families present in the deuterostome clade we required (1) at least two ambulacrarian orthologues out of the three available ambulacrarian genomes and at least two chordate orthologues, or (2) at least two deuterostomes (chordates and/or ambulacrarians) and two outgroups in the bilaterian level clusters.

Transposable elements. Repetitive sequences were identified using RepeatScout48, followed by manual curation and annotation using both a Repbase release (version 20140131)59 and BLASTX-based search against a custom collection of transposons, using a previously described repeat identification and annotation pipeline60 (Supplementary Note 5). The assemblies were then masked with RepeatMasker version open-4.0.56. The repetitive complements of the two hemichordate genomes are summarized in Supplementary Table 5.1.

Phylogenetic analysis. Phylogenetic analyses were done using metazoan-level gene family clusters based on whole-genome sequences (Supplementary Note 4), selecting a single orthologue per genome with the best cumulative BLASTP to other species, and best reciprocal BLASTP hits to species with transcriptome-only information (Supplementary Note 6). Single gene alignments were built using Muscle61 and filtered using Trimal62 for each orthologue, and were concatenated, yielding a supermatrix of 506,428 positions with 34.9% missing data. This supermatrix was analysed with ExaML assuming a site-homogeneous LG+Γ1 model partitioned for each gene63. A slow-fast analysis was conducted to stratify marker genes based on the length of the branch leading to acoids in individual trees. A subset of the slowest 10% of genes was analysed with the site-heterogeneous CAT+GTR+Γ1 model using Phylogeny64. Molecular dating was carried out using Phylogeny64 using the log-normal relaxed clock model and the calibrations described in Supplementary Table 6.2.

Synteny analysis. Macro- and micro-syntenic linkages were calculated as described in Supplementary Note 7. For Fig. 3a, we merged the amphioxus scaffolds into 17 pre-defined scaffold groups as suggested in ref. 27. These 17 merged scaffold groups represent the 17 ancestral linkage groups (ALGs) shared in chordates. Then we calculated the orthologous gene groups shared by each amphioxus ALG–Saccoglossus scaffold pair and generated the dot plot as described in Supplementary Note 7. For micro-synteny we required at least three genes (separated by a maximum of ten genes) to be present in pairwise comparisons. Under random reshuffling of the genome, this yields 10% false positives in pairwise genome comparisons, that is, we observe approximately one-tenth as many micro-syntenic blocks between the two genomes when gene orders are shuffled. This false-positive rate, however, falls to 1% when considering more than two species. For our inference of deuterostome ancestral and novel synteny we therefore focus on blocks present in at least three species (and both ingroup representatives, that is, ambulacrarians and chordates). This yields 698 blocks that can be traced back to the deuterostome ancestor, including 71 blocks found exclusively in deuterostome species (shared among ambulacrarians and chordates), including the pharyngeal cluster discussed in Fig. 4.

Whole-genome alignment. Whole-genome alignments were conducted with MEGALIST65 using parameters previously reported66. We assessed the distribution of the resulting 12,722 aligned loci across known gene annotations in ENSEMBL67, previously identified conserved pan-vertebrate elements68, as well as known enhancers in human according to LIR database67.

Gene novelties. Deuterostome gene novelties were assessed initially through bilaterian gene clusters (Supplementary Note 10) by requiring at least two species on both ambulacrarian and chordate side to be present. The novelties were further automatically subdivided into four categories: G1 (gain type I), with no BLASTP hit outside of deuterostomes; G2 (gain type II), with a novel PFAM domain present only in deuterostomes; G3 (gain type III) having a novel PFAM combination unique to deuterostomes; and G4 (gain type IV), those that do not fall under any of the G1–3 categories and define novelties due to acceleration in the substitution rate on the deuterostome stem. To confirm the novel nature, especially for G4 novelties, we have constructed phylogenies for the members and non-deuterostome BLASTP hits (up to an e-value of 1 × 10−20) using MAFFT-alignment-based FastTree calculations. The trees were assessed for the accelerated rate of evolution at the deuterostome stem (Supplementary Fig. 9.1.1). The final result is provided in the Supplementary Information.

Curation of candidates for horizontal gene transfer on the deuterostome stem. We examined in detail gene families found broadly in deuterostomes whose encoded peptides were readily alignable to microbial sequences but had no detectable similarity in non-deuterostome animals. Criteria for evaluation included: (1) the hemichordate gene matches microbial genes at least ten orders of magnitude in the e-value better than it matches sequences of non-deuterostome metazoans (most of the putative HGTs we describe have no non-deuterostome metazoan hit at all); (2) it has a defined genomic locus among bona fide metazoan genes; (3) it shares an exon–intron structure with genes of chordates and other ambulacraria; and (4) when a low bitscore match is found to a non-deuterostome metazoan sequence, that sequence is identified as containing different domains (domain switching) and having different CDD69 and Pfam70 structure, implying a dubious relatedness. When phylogenetic trees are constructed for these HGT-candidate proteins, the trees contain numerous branches for microbial sequences and none for non-deuterostome metazoan sequences, or only very long branches for dubious relatives, and hence the trees differ greatly from the metazoan species tree, except within the deuterostome clade.

Code availability. Original data and code can be accessed at https://groups.ouist.jpl/moglenu.

50. Jaffe, D. B. et al. Whole-genome sequence assembly for mammalian genomes: Arachne 2. Genome Res. 13, 91–96 (2003).
51. Kajitani, R. et al. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome Res. 24, 1384–1395 (2014).
52. Romiguier, J. et al. Comparative population genomics in animals uncovers the determinants of genetic diversity. Nature 515, 261–263 (2014).
53. Tagawa, K. et al. A cDNA resource for gene expression studies of a hemichordate, Ptychodera flava. Zoolog. Sci. 31, 414–420 (2014).

54. Chen, S. H. et al. Sequencing and analysis of the transcriptome of the acorn worm Ptychodera flava; an indirect developing hemichordate. Mar. Genomics 15, 35–43 (2014).

55. Salamov, A. A. & Solovyev, V. V. Ab initio gene finding in Drosophila genomic DNA. Genome Res. 10, 516–522 (2000).

56. Stanke, M. & Waack, S. Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics 19 (Suppl. 2), ii215–ii225 (2003).

57. Simakov, O. et al. Insights into bilaterian evolution from three spiralian genomes. Nature (2013).

58. Price, A. L., Jones, N. C. & Pevzner, P. A. De novo identification of repeat families in large genomes. Bioinformatics 21 (Suppl. 1), i351–i358 (2005).

59. Jurka, J. et al. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet. Genome Res. 110, 462–467 (2005).

60. Smit, A., Hubley, R. & Green, P. RepeatMasker http://www.repeatmasker.org. (2007).

61. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5, 113 (2004).

62. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldon, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973 (2009).

63. Aberer, A. & Stamatakis, A. ExaML: Exascale maximum likelihood: program and documentation. See http://sco.h-its.org/exelixis/web/software/examl/index.html (2013).

64. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402 (1997).

65. Lee, A. P., Kerk, S. Y., Tan, Y. Y., Brenner, S. & Venkatesh, B. Ancient vertebrate conserved noncoding elements have been evolving rapidly in teleost fishes. Mol. Biol. Evol. 28, 1205–1215 (2011).

66. Cunningham, F. et al. Ensembl 2015. Nucleic Acids Res. 43, D662–D669 (2015).

67. Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L. A. VISTA Enhancer Browser–a database of tissue-specific human enhancers. Nucleic Acids Res. 35, D88–D92 (2007).

68. Marchler-Bauer, A. et al. CDD: NCBI’s conserved domain database. Nucleic Acids Res. 43, D222–D226 (2015).

69. Marinic, M., Aktas, T., Ruf, S. & Spitz, F. An integrated holo-enhancer unit defines tissue and gene specificity of the Fgf8 regulatory landscape. Dev. Cell 24, 530–542 (2013).
Extended Data Figure 1 | Summary of genome assemblies and heterozygosity distributions for *Saccoglossus* and *Ptychodera*.

a, Genome statistics summary. b, c, The single nucleotide polymorphism distribution across 100-bp windows for *Saccoglossus* (b) and the corresponding distribution for *Ptychodera* (c). The distributions in b and c are fitted with a geometric (expected when high recombination rate is present) and a Poisson distribution (expected with low recombination rate). The distribution for *Saccoglossus* is fitted to windows with one or more SNPs only, as there is an excess of zero SNP windows (approximately 84% of total 94,324 selected windows). For methods refer to Supplementary Note 2.
Extended Data Figure 2 | Ambulacrarians approximate the ancestral metazoan gene repertoire. a, Principal component analysis of Panther gene family sizes. Variances of the first two components are plotted in parentheses. Blue indicates deuterostomes; green indicates lophotrochozoans; red, ecdysozoans; yellow/orange, non-bilaterian metazoans. Note the clustering of the ambulacrarians Sko, Pfl and Spu with the non-vertebrate deuterostomes Bfl and Cin in the lower right corner, also with the lophotrochozoans Cgi, Lgi, Hro, Cte and the non-bilaterians Hma, Nve and Adi. b, Heat map of gene family counts showing significant (Fisher’s exact test P value < 0.01 after Bonferroni multiple testing correction) expansion in ambulacrarians as well as in Saccoglossus/Ptychodera/amphioxus. The cases discussed in the main text are highlighted in red. See Supplementary Note 4 for details. Species abbreviations are defined in Supplementary Note 4.1.
Extended Data Figure 3 | Molecular dating of deuterostome and metazoan radiations using PhyloBayes assuming a log-normal relaxed clock model. Yellow circles on particular nodes indicate the calibration dates applied from the fossil record, as indicated in Supplementary Note 6.2. Bars are 95% credibility intervals derived from posterior distributions. Note the estimated times of divergence of chordates and ambulacraria (the deuterostome ancestor) at 570 million years ago (Ma; mid-Ediacaran), hemichordates and echinoderms at 559 Ma, enteropneusts and pterobranchs at 547 Ma, and Harrimaniid and Ptychoderm enteropneusts at 373 Ma.
Extended Data Figure 4 | Homeobox gene complement of the two hemichordates in comparison to that of amphioxus. The numbers of homeobox-containing gene models are 170 in Saccoglossus and 139 in Ptychodera. These homeobox domains were aligned with 128 homeobox genes of Branchiostoma floridæ using ClustalW2, then gaps and unaligned regions were manually removed. Since some genes have more than one homeobox domain, we kept all domains or chose the longest one according to the state of domain conservation. In total, 448 homeobox sequences were aligned. See Supplementary Information for details. The clusters of homeobox genes on scaffolds in Saccoglossus and Ptychodera were identified and drawn at positions around the tree. Conserved clusters between the two species were aligned. In addition to the well-known Hox and ParaHox cluster, 17 clusters were found in at least one of the hemichordates or some in both. Sixteen genes of the Nkx class are distributed over four clusters: (i) nks1a-vent1-vent2.1-vent2.2; (ii) nks2.1-nks2.2-msx1s; (iii) nks3-msx-nks3.2-nkx4-lbx-hex; and (iv) voxvent-nk7like-nk7like2. The second cluster (ii) of these is part of the pharyngeal cluster (Fig. 4). Another five-gene cluster consists of one Lim class homeobox gene and four PRD class homeobox genes; isl-otp-rax-gsc. A cluster of six3/6-six1/2-six4/5 was found in both species, and a cluster of three unx genes was found only in F. flava. Ten more clusters were found containing two homeobox genes each. Notably, we found species-specific homeobox clusters in both species. Three remarkable clusters were found in S. kowalevskii in which 10, 12 and 5 homeobox-containing genes are tandem duplicated in scaffold_1710, _52 and _4796, respectively. We also found such clusters in P. flava in which 7, 4, 8 and 10 genes are aligned on scaffold 19451, scaffold 1398, scaffold 12422 and scaffold 154657, respectively. All homeobox genes identified in the genomes of the two hemichordates and amphioxus are listed in the Supplementary Table for Extended Data Fig. 4. This list includes some genes not containing a homeobox (for example, pax1/9) in cases where other family members do (for example, pax2).
Extended Data Figure 5 | High retention rate of micro-synteny in *Saccoglossus*. Circos plot showing micro-syntenic conservation in blocks of genes ($n_{\text{max}} = 10$ and $n_{\text{min}} = 2$) for six metazoan species for observed (left) and simulated (right) linkages. The width of connecting segments is proportional to the number of genes participating in the syntenic linkages (normalized by the total gene count). In this representation scaffolds are placed end-to-end, and adjacent scaffolds need not be from the same chromosome. While simulated data yields some blocks shared between pairs of species, few or no synteny blocks can be recovered among three or more species (Methods). *Saccoglossus* shows one of the highest retentions among the selected species (and the highest among the sequenced ambulacrarians). *Xenopus* (and vertebrates in general) have lost some micro-synteny due to whole-genome duplications and differential loss of paralogues. The matching between the hemichordate *S. kowalevskii* and the chordate amphioxus is highest, consistent with the fact that neither genome has undergone extensive gene loss (as have tunicates) or pseudotetraploidization with extensive loss of paralogues (as have vertebrates).
Extended Data Figure 6 | Deuterostome specific micro-syntenic linkages. **a, b,** Very tight linkages with no intervening genes. **a,** ParaHox cluster shown in *S. kowalevskii, P. flava,* and human. **b,** bmp2/4 and univin cluster in the hemichordates *S. kowalevskii* and *P. flava,* the sea urchin *S. purpuratus,* and the cephalochordate *B. floridae.*

c–e, Loose micro-syntenic linkages with a maximum of five intervening genes: *lefty* (c), *six1–six4* (d), and *fgf8–fbxw* (e) clusters. For c to e all species with micro-synteny are shown. Numbers above the genes indicate the copy number in the locus.
Extended Data Figure 7 | Three examples showing the domain structures of some proteins encoded by genes found in deuterostomes and marine microbes but not non-deuterostome animals. Best BLASTP hits of the Saccoglossus sequence in human/mouse, as well as in non-deuterostome metazoans and in non-metazoans (such as the cyanobacterium Staniera cyanosphaera, or the eukaryotic micro-alga Ostreococcus tauri) are shown. a, Cytidine monophosphate-N-acetylneuraminic acid hydroxylase (CMAH), an enzyme of sialic acid modification; b, peptidyl arginyl deiminase (PAD), an enzyme of post-translational modification of proteins; c, FATSO-like, also called α-ketoglutarate-dependent dioxygenase FTO, an enzyme that de-methylates \( N^6 \)-methyladenosine in nuclear RNA. Other analyses of these and other genes with the unusual phylogenetic distributions can be found in Supplementary Note 10.
Extended Data Figure 8 | *In situ* hybridization demonstration of the expression of von Willebrand type D (vWD) domain-encoding genes (putative glycoproteins/mucins) in *Saccoglossus* and *Ptychodera*. 

a, In *Saccoglossus* the genes are specifically expressed in different subregions of the ectoderm of the proboscis or collar at these pre-feeding stages. 
b, In *Ptychodera*, several of the genes are expressed in endoderm as well as ectoderm of the developing tornaria larva. The sequence IDs for the genes are provided in Supplementary Note S10.4.
Extended Data Figure 9 | Gene innovation in deuterostomes. a, FastTree phylogenetic tree of the TGFβ family members Lefty, TGFβ32, GDF8/11 and Nodal ligands (using GTR model). Bootstrap support is plotted as filled circles (size proportional to the support value) on each node. While Lefty shows deuterostome unique sequence composition, TGFβ32 has an acceleration of sequence change at the deuterostome stem branch, compared to the GDF8/11 or Nodal groups. b, Temporal co-expression of Lefty and TGFβ receptor type III in Saccoglossus at pre-gastrulation developmental stages and of TGFβ2 and TGFβ receptor type II at post-gastrulation stages. c, In situ hybridization demonstration of the expression in S. kowalevskii of one of the putative type I novelty genes (c9orf9, also known as rsb66) and of two of AAADC genes (aromatic amino acid decarboxylases of the microbial type) of S. kowalevskii (also in P. flava and B. floridae), which closely resemble sequences from bacteria rather than from non-deuterostome metazoans. gs, gill slits. d, The temporal expression profile for c9orf9 during S. kowalevskii development, taken from transcriptome data.
| ID | gene family name                                                                 | BLAST e-value in raw deuterostome dataset | BLAST e-value in raw echinodermata dataset with | Phylogeny support for novelty/characteristics in | Origin of novelty/characteristics | Presence of functions in deuterostome? |
|----|----------------------------------------------------------------------------------|-----------------------------------------|------------------------------------------------|------------------------------------------------|----------------------------------|----------------------------------|
| 1  |                                                                  |                                          |                                                |                                                |                                  |                                  |
| 2  |                                                                  |                                          |                                                |                                                |                                  |                                  |
| 3  |                                                                  |                                          |                                                |                                                |                                  |                                  |
| 4  |                                                                  |                                          |                                                |                                                |                                  |                                  |
| 5  |                                                                  |                                          |                                                |                                                |                                  |                                  |
| 6  |                                                                  |                                          |                                                |                                                |                                  |                                  |
| 7  |                                                                  |                                          |                                                |                                                |                                  |                                  |
| 8  |                                                                  |                                          |                                                |                                                |                                  |                                  |
| 9  |                                                                  |                                          |                                                |                                                |                                  |                                  |
| 10 |                                                                  |                                          |                                                |                                                |                                  |                                  |
| 11 |                                                                  |                                          |                                                |                                                |                                  |                                  |
| 12 |                                                                  |                                          |                                                |                                                |                                  |                                  |
| 13 |                                                                  |                                          |                                                |                                                |                                  |                                  |
| 14 |                                                                  |                                          |                                                |                                                |                                  |                                  |
| 15 |                                                                  |                                          |                                                |                                                |                                  |                                  |
| 16 |                                                                  |                                          |                                                |                                                |                                  |                                  |
| 17 |                                                                  |                                          |                                                |                                                |                                  |                                  |
| 18 |                                                                  |                                          |                                                |                                                |                                  |                                  |
| 19 |                                                                  |                                          |                                                |                                                |                                  |                                  |
| 20 |                                                                  |                                          |                                                |                                                |                                  |                                  |
| 21 |                                                                  |                                          |                                                |                                                |                                  |                                  |
| 22 |                                                                  |                                          |                                                |                                                |                                  |                                  |
| 23 |                                                                  |                                          |                                                |                                                |                                  |                                  |
| 24 |                                                                  |                                          |                                                |                                                |                                  |                                  |
| 25 |                                                                  |                                          |                                                |                                                |                                  |                                  |
| 26 |                                                                  |                                          |                                                |                                                |                                  |                                  |
| 27 |                                                                  |                                          |                                                |                                                |                                  |                                  |
| 28 |                                                                  |                                          |                                                |                                                |                                  |                                  |
| 29 |                                                                  |                                          |                                                |                                                |                                  |                                  |
| 30 |                                                                  |                                          |                                                |                                                |                                  |                                  |

This table summarizes the description of the novelties in Supplementary Note 10.