

## ARTICLE

<https://doi.org/10.1038/s42003-020-1091-1>

OPEN

## Genomic insights of body plan transitions from bilateral to pentameral symmetry in Echinoderms

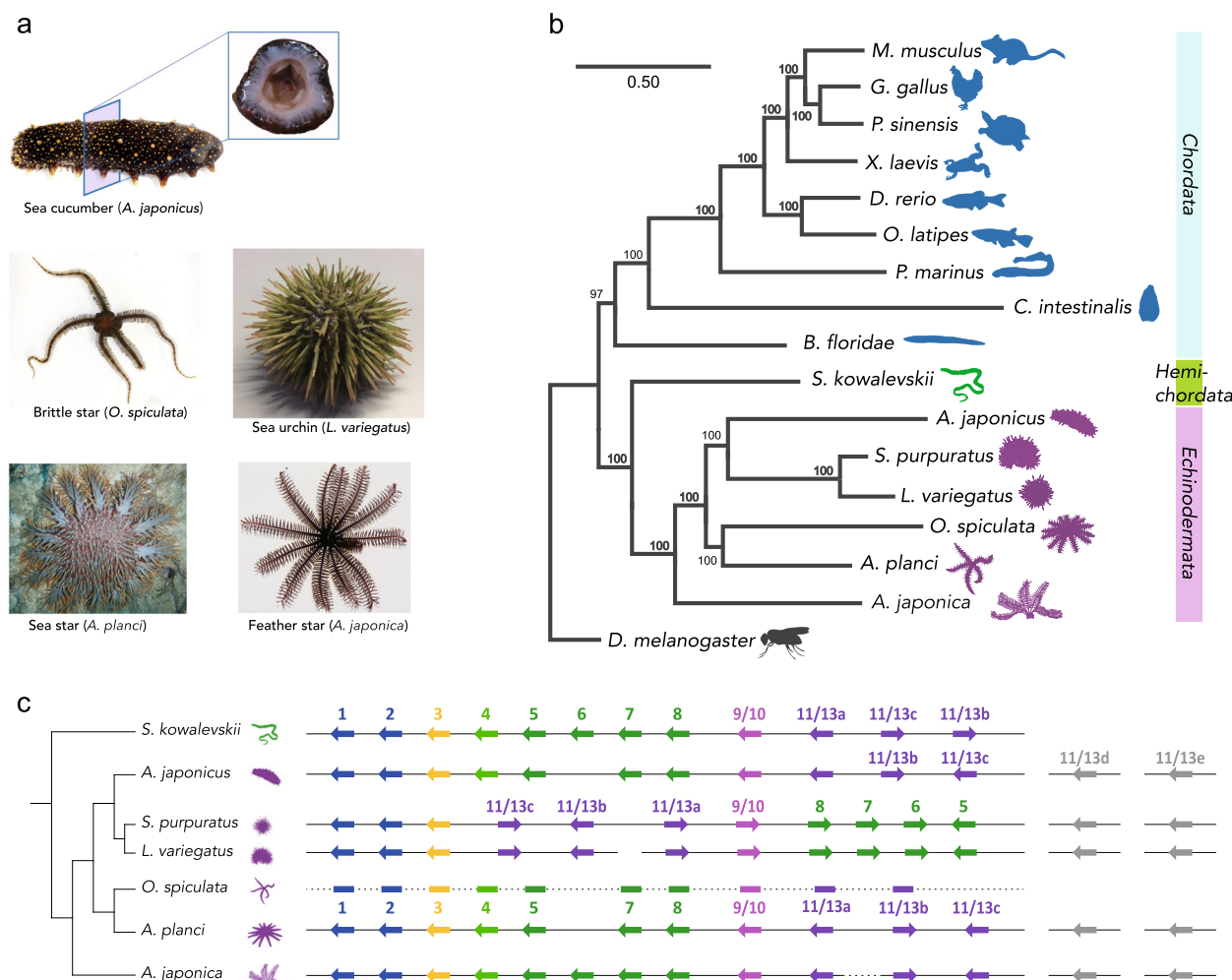
Yongxin Li<sup>1</sup>, Akihito Omori<sup>2</sup>, Rachel L. Flores<sup>3</sup>, Sheri Satterfield<sup>3</sup>, Christine Nguyen<sup>3</sup>, Tatsuya Ota<sup>4</sup>, Toko Tsurugaya<sup>5</sup>, Tetsuro Ikuta<sup>6,7</sup>, Kazuho Ikeo<sup>4</sup>, Mani Kikuchi<sup>8</sup>, Jason C. K. Leong<sup>9</sup>, Adrian Reich<sup>10</sup>, Meng Hao<sup>1</sup>, Wenting Wan<sup>1</sup>, Yang Dong<sup>11</sup>, Yaondong Ren<sup>1</sup>, Si Zhang<sup>12</sup>, Tao Zeng<sup>12</sup>, Masahiro Uesaka<sup>13</sup>, Yui Uchida<sup>9,14</sup>, Xueyan Li<sup>1</sup>, Tomoko F. Shibata<sup>9</sup>, Takahiro Bino<sup>15</sup>, Kota Ogawa<sup>16</sup>, Shuji Shigenobu<sup>15</sup>, Mariko Kondo<sup>9</sup>, Fayou Wang<sup>12</sup>, Luonan Chen<sup>12,17</sup>, Gary Wessel<sup>10</sup>, Hidetoshi Saiga<sup>7,9,18</sup>, R. Andrew Cameron<sup>19</sup>, Brian Livingston<sup>3</sup>, Cynthia Bradham<sup>20</sup>, Wen Wang<sup>1,21,22</sup>✉ & Naoki Irie<sup>9,14,22</sup>✉

Echinoderms are an exceptional group of bilaterians that develop pentameral adult symmetry from a bilaterally symmetric larva. However, the genetic basis in evolution and development of this unique transformation remains to be clarified. Here we report newly sequenced genomes, developmental transcriptomes, and proteomes of diverse echinoderms including the green sea urchin (*L. variegatus*), a sea cucumber (*A. japonicus*), and with particular emphasis on a sister group of the earliest-diverged echinoderms, the feather star (*A. japonica*). We learned that the last common ancestor of echinoderms retained a well-organized Hox cluster reminiscent of the hemichordate, and had gene sets involved in endoskeleton development. Further, unlike in other animal groups, the most conserved developmental stages were not at the body plan establishing phase, and genes normally involved in bilaterality appear to function in pentameric axis development. These results enhance our understanding of the divergence of protostomes and deuterostomes almost 500 Mya.

<sup>1</sup>State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China. <sup>2</sup>Faculty of Agriculture, Niigata University, Niigata, Japan. <sup>3</sup>Dept. of Biological Sciences, California State University, Long Beach, CA, USA. <sup>4</sup>SOKENDAI, Kanagawa, Japan. <sup>5</sup>Urawa University, Saitama, Japan. <sup>6</sup>Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Kanagawa, Japan. <sup>7</sup>Tokyo Metropolitan University, Yokosuka, Tokyo, Japan. <sup>8</sup>Hokkaido University, Sapporo, Hokkaido, Japan. <sup>9</sup>Dept. of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo, Japan. <sup>10</sup>Providence Institute of Molecular Oogenesis, Brown University, Providence, RI, USA. <sup>11</sup>Yunnan Agricultural University, Kunming, China. <sup>12</sup>Shanghai Institute of Biochemistry and Cell Biology, Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences, Shanghai, China. <sup>13</sup>RIKEN Center for Biosystems Dynamics Research (BDR), Kobe, Hyogo, Japan. <sup>14</sup>Universal Biology Institute, University of Tokyo, Tokyo, Japan. <sup>15</sup>NIBB Core Research Facilities, National Institute of Basic Biology, Okazaki, Aichi, Japan. <sup>16</sup>Faculty of Social and Cultural Studies, Kyushu University, Fukuoka, Japan. <sup>17</sup>Key Laboratory of Systems Biology, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Hangzhou, China. <sup>18</sup>Chuo University, Tokyo, Japan. <sup>19</sup>Beckman Institute, Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA. <sup>20</sup>Department of Biology, Boston University, Boston, MA, USA. <sup>21</sup>School of Ecology and Environment, Northwestern Polytechnical University, Xi'an, China. <sup>22</sup>These authors jointly supervised this work: Wen Wang, Naoki Irie. ✉email: [wwang@mail.kiz.ac.cn](mailto:wwang@mail.kiz.ac.cn); [irie@bs.u-tokyo.ac.jp](mailto:irie@bs.u-tokyo.ac.jp)

## Results

**Genetic changes behind echinoderm evolution.** The estimated genome sizes were 952 Mb for green sea urchin and 553 Mb for the feather star (Supplementary Fig. 1), with 30,238 and 26,838 protein-coding genes, respectively (Supplementary Tables 18–20 and 23). Basic features of these genomes such as GC content, gene length, and exon number were comparable to those of chordate and hemichordate species (Supplementary Fig. 2–7). Genome-wide analysis with 1196 one-to-one orthologs (Fig. 1b, Supplementary Fig. 7, and Supplementary Table 23–24) robustly corroborated recent reports<sup>4,5</sup> that Echinodermata consists of the early diverged Crinoidea (including feather star), Asterozoa (including brittle star and sea star), and Echinozoa (including sea urchin and sea cucumber). This is consistent with paleontological evidence that suggests echinoderms first evolved with a stemmed, or imperforate extra-axial morphology<sup>3</sup>. We also found that the overall protein sequences of an early diverged echinoderm species, the feather star, showed a relatively low evolutionary rate. Meanwhile, echinoderms showed only slightly diverged protein sequences from vertebrates, as opposed to an ascidian, a species with highly derived morphological features, and with significantly



**Fig. 1 Echinoderms and their evolutionary diversity.** **a** Echinoderm species of five living classes were analyzed in this study. Pentameral symmetry can also be observed in the transverse section of the sea cucumber (top), which otherwise shows apparent bilaterality. **b** Evolutionary rate and the phylogenetic tree constructed by RAxML software using the 1196 orthologous protein sequences identified by reciprocal best blast hit (RBBH). The values on branches represent bootstrap values. **c** Schematic representation of genomic organization of ambulacrarian Hox clusters. Arrows and horizontal lines represent Hox genes and chromosomal DNAs, respectively. Dashed lines indicate the presence of unconnected scaffolds. See Supplementary Fig. 13 for more detailed Hox cluster structures. Hox cluster structures of *S. kowalevskii*<sup>9</sup>, *A. japonicus*<sup>56</sup>, *S. purpuratus*<sup>7</sup>, *O. spiculata*<sup>56</sup> and *A. planci*<sup>8</sup> are according to the previous studies.

diverged genomic sequences from vertebrates (Supplementary Fig. 8 and Supplementary Tables 26–28).

Given that echinoderms evolved unique features without significant genome-wide changes detected, we tested abundance in sets of gene families that may have played critical roles in the evolution of the echinoderm features. We first tested if numbers of genes potentially involved in development (such as genes involved in cell-cell communications) have expanded in the common ancestor of echinoderms (Supplementary Fig. 10). In contrast to our expectation, GO term enrichment analysis suggested that no such terms were enriched in the echinoderm-expanded gene set (Supplementary Fig. 11). Rather, GO terms such as “cell communication”, or “signal transduction”, were present in echinoderm-contracted genes (Supplementary Fig. 11a). Further, genes potentially involved in cytoskeletal regulation appear to have experienced extensive modifications during echinoderm evolution. For example, the “plectin repeat domain”, one of the important domains of cytolinkers that connect cytoskeletal elements with each other and to junctional complexes<sup>6</sup>, was not found in any of the five echinoderm species (Supplementary Fig. 12). Similarly, genes potentially involved in actin cytoskeleton regulation such as BCAR1/CAS and PIP5K were found to be positively selected during echinoderm evolution (Supplementary Table 29), suggesting substantial modifications of cytoskeletal function during echinoderm evolution.

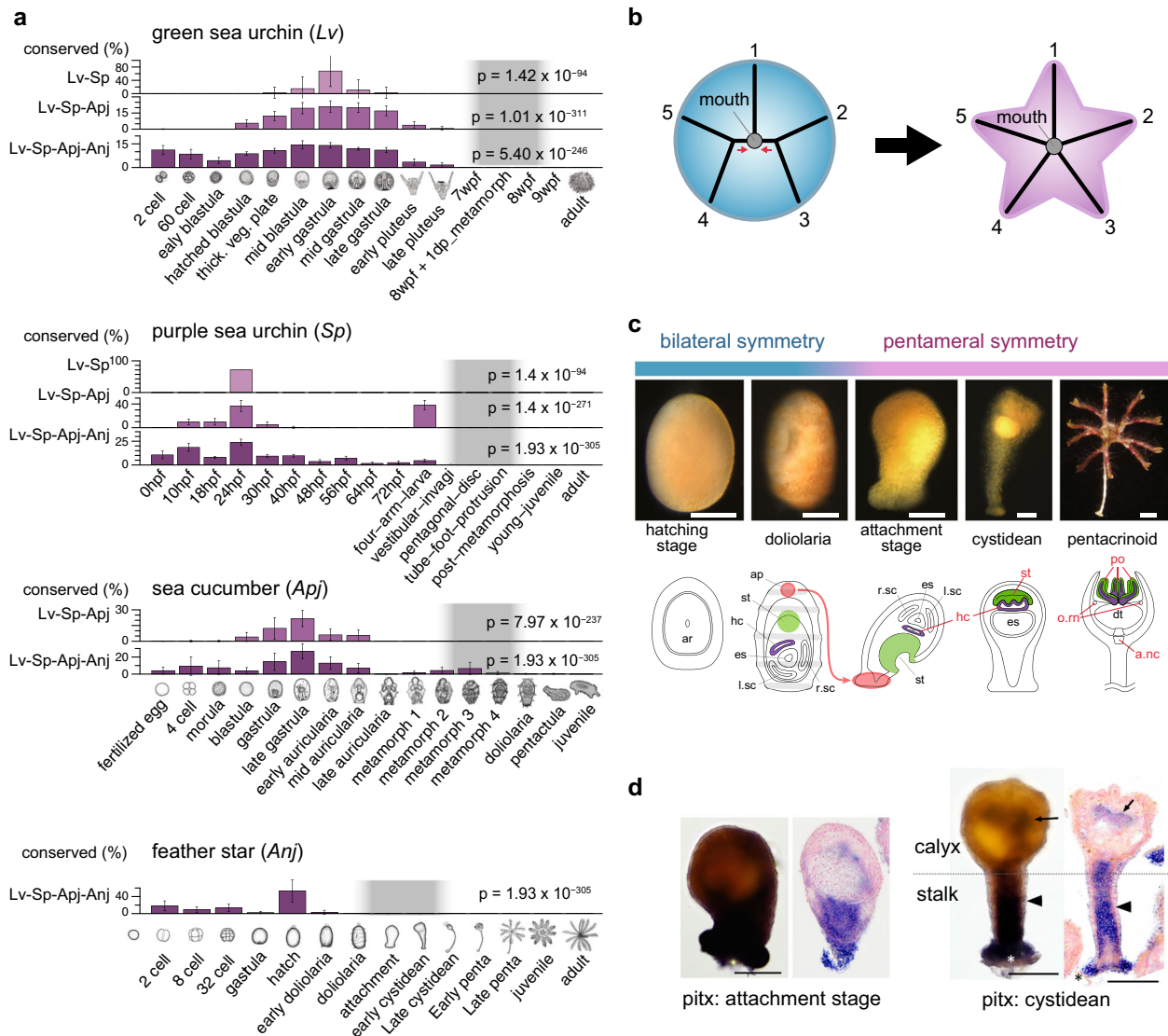
**Hox clusters in ambulacrarians.** We next analyzed genes in the Hox cluster, since previous studies implied that echinoderms may have undergone extensive changes to the genomic-arrangement of these genes early in their evolution<sup>7</sup>, but this is controversial<sup>8</sup>. By analyzing the echinoderm genomes, together with BAC sequencing and fluorescent in situ hybridization (FISH) experiments on feather star (see Supplementary Fig. 13), we found that *A. japonica* have 10 clustered Hox genes (Hox1 through Hox11/13a), together with two posterior genes (Hox11/13b and Hox11/13c) located with inverse directions more than 360 kb apart from the cluster of 10 Hox genes (Supplementary Fig. 14). This situation is reminiscent of hemichordate Hox gene clusters, which consist of 12 genes with the inversion/translocation of two posterior genes<sup>9</sup>. The consistent differences of Hox genes in feather star from those of hemichordates (*Saccoglossus kowalevskii* and *Ptychodera flava*<sup>9</sup>) were that the two posterior genes have face-to-face orientations, and two additional posterior genes (Hox11/13d and Hox11/13e) are present as reported in other echinoderm genomes<sup>10</sup>. These characteristics in turn suggest that the last common ancestor of echinoderms retained a canonical arrangement of 10 Hox genes, while its posterior genes had increased in number and changed their locations. This view accordingly indicates that the changes previously noted with the Hox gene clusters of echinoderms<sup>7,8</sup>, including the loss of Hox4 or Hox6 and the inversion/translocation of anterior genes, are lineage-specific events, and therefore are unlikely to be involved in the establishment of pentameral body plan. On the other hand, the involvement of the posterior Hox genes, are important candidates in the establishment of the pentameral body plan (Fig. 1c).

**Echinoderm embryogenesis show hourglass-like conservation.** We next tested if evolution of echinoderm embryogenesis follows the developmental hourglass model<sup>11</sup> as supported in several animal groups<sup>12–16</sup>. The phylotype hypothesis of the hourglass model predicts that anatomical features of most conserved-embryonic phase represent the body plan of their animal phylum<sup>11,17,18</sup>. We tested if the developmental stages most critical for the pentameral body plan show the highest transcriptomic conservation. Recent studies reported that the transcriptomic

conservation is instead around blastula to gastrula in sea urchins<sup>19,20</sup>, however, these studies only examined embryonic sea urchins. We thus analyzed gene expression profiles by covering early-to-late embryos of four diverse echinoderm species, including the publicly available data of purple sea urchins<sup>21,22</sup> and a sea cucumber<sup>23</sup> (Supplementary Figs. 15 and 16). Unexpectedly, while the hourglass-like conservation was observed, the most conserved phase was not bracketing the pentameral establishing stages (defined as stages when first pentameral symmetric structures appear), but it was instead during gastrulation (Fig. 2a and Supplementary Figs. 17–18). This mismatch between the most conserved phase and the phase for establishing the body plan was also supported by pair-wise comparisons of one-to-one orthologs (Supplementary Fig. 19). These results suggested that, unlike in other animal groups such as vertebrates, the phase for establishing the body plan in echinoderms has experienced substantial diversification during their evolution, further implying that the phylotype hypothesis may not fit within echinoderm embryogenesis. However, a potential caveat of this conclusion would be that the conservation signals from adult rudiments could have been obscured by larval tissues, as adult rudiments share only a small proportion of embryo in early metamorphic stages. Similar analyses with dissected adult rudiment, or single-cell RNAseq technology could clarify this point. While possible contribution of pleiotropic constraints were supported between the closely related species as reported previously<sup>24,25</sup> (Supplementary Figs. 20 and 21), further studies are needed to clarify the evolutionary mechanism behind the unique evolution of echinoderm embryos.

**Partial co-option in pentameral body plan establishment.** Paleontological studies suggested the possible evolution of pentameral body axes through changes in the mechanisms of bilateral symmetry<sup>26</sup> (Fig. 2b). However, developmental genes that control the pentameral symmetry remains largely unknown. We thus focused on genes that are involved in Left/Right-patterning and other axis-forming in bilaterians<sup>27–31</sup> and examined expression patterns of their homologs in the feather star, particularly at the attachment and cystidean stages when the pentameral body plan forms (Supplementary Figs. 22 and 23). Among the genes examined, *pitx* exhibited relatively strong expression in the calyx, where the pentameral structure first becomes evident (Fig. 2c–d). Weak expression of *chordin* was also detected in the calyx (Supplementary Fig. 23). Meanwhile, expression of *bmp2/4*, *nodal*, *lefty* and *not* were detected most during gastrula to doliolaria stages (Supplementary Fig. 22), but not in the calyx (Supplementary Fig. 23), implying that these genes are possibly involved in the body patterning during bilateral planktonic development rather than pentameral body plan development. These results suggest that evolution of the pentameral body plan may have associated partial co-option of genes involved in existing body axes, which in part, coincides with paleontological predictions that modification of bilateral patterning system contributed to the pentameral body plan establishment<sup>26</sup>.

**Proteome analyses of echinoderm skeleton.** Mineralized endoskeleton is another notable feature of echinoderms. Despite the widely conserved endoskeleton structures in echinoderms, some of the proteins first identified in sea urchin biomineralization, such as MSP130, have not been identified in the skeleton of other echinoderm species. The MSP130 gene was suggested to have originated in prokaryotes and was introduced into metazoan genomes, including echinoderms, by horizontal gene transfer<sup>32</sup>. An MSP130-like gene involved in biomineralization has also been identified in a polychaete<sup>33</sup>. The authors suggest that the MSP130

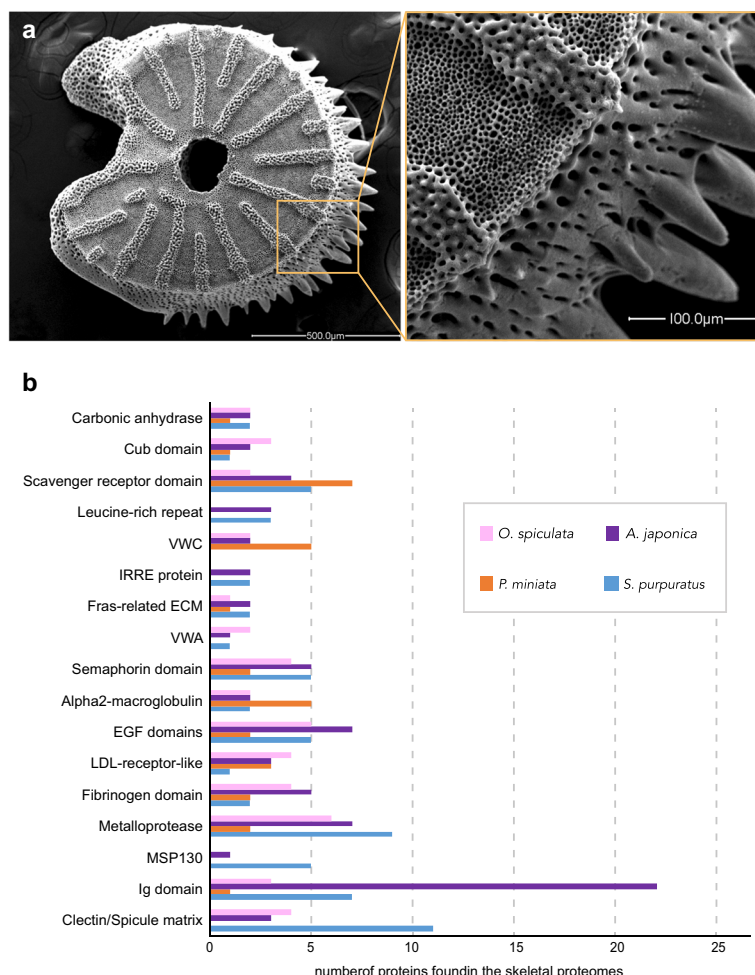


**Fig. 2 Evolutionarily conserved echinoderm stages and potential involvement of *pitx* signal in pentameral body plan development.** **a** Mid-embryonic conservation found in echinoderm species. Based on expression distance (expDists, see also Supplementary Fig. 17) of orthologous groups (defined by orthomc<sup>57</sup>), an evolutionary conservation of developmental stages were estimated for three taxonomic levels (Lv-Sp, Lv-Sp-Apj, Lv-Sp-Apj-Anj, see also “methods”). The vertical axis represents percentages of the stage being included in the most (top 1%) conserved stage-combinations<sup>13</sup> (Ptop). Changes of the Ptop scores were significant among stages (Friedman test). Error bars represent S.D. of Ptop values. In each species, the developmental phase in which pentameral body plan establishment begins is colored in gray. **b** Possible evolutionary transition from bilateral symmetry to pentameral symmetric body plan suggested by paleontological studies<sup>61</sup>. The basal echinoderms had a bilaterally symmetric ambulacral system that is arranged in a 2-1-2 pattern (left); consisting of one unpaired ambulacrum (1) and two ambulacra with a distal bifurcation (2&3, 4&5) and a single unpaired ambulacrum (1). **c** Feather star development from bilateral symmetry to pentameral symmetry. a.nc aboral nerve center, ar archenteron, ap adhesive pit, dt digestive tract, es enteric sac, hc hydrocoel, l.sc left somatocoel, o.rn oral ring nerve, po podia, r.sc right somatocoel, st stomodeum. **d** *pitx* gene expression detected in embryos of attachment stage and cystidean stage. In cystidean embryos, *pitx* was expressed in the tissues around the gut (arrows) and the inner tissue of the whole stalk (arrowheads). Scale bars: 100  $\mu$ m. The expression was detected by in situ hybridization with whole mount (left) and sectioned specimens (right).

gene was present in the common ancestor to bilaterians, rather than being introduced into protostomes and deuterostomes in separate lateral transfer events. The MSP130 protein was then co-opted into skeleton formation at some point in echinoderm evolution. The gene was duplicated in sea urchins and the resulting paralogues acquired repetitive regions<sup>32</sup>. The MSP130 gene is present in other echinoderm genomes, but is not utilized in the proteome of brittle stars<sup>34,35</sup> or sea stars<sup>36</sup>. Here, we performed a proteome analysis and identified 280 proteins that are included within the mineral of the adult feather star skeleton (see “Methods”). These skeletal proteins included a protein similar to the urchin MSP130 proteins, suggesting that the ancestral

echinoderm had co-opted this single protein into biomineralization (Supplementary Fig. 24). A number of other genes encoding skeletal proteins and domains conserved between the purple sea urchin (*S. purpuratus*) and the feather star were also identified (Fig. 3). Among these, we found two proteins with C-type lectin domains, which are also found on the urchin spicule matrix proteins. C-type lectin proteins are absent in sea star skeletons<sup>36</sup>, and present in only a few copies in brittle star skeletons<sup>34,35</sup>. Sea urchin skeletons utilize a large number of C-type lectins, mostly with repetitive stretches of acidic amino acids<sup>37</sup>, while the feather star and brittle star proteins lack these repetitive domains (Supplementary Fig. 25). A possible evolutionary scenario to explain





**Fig. 3 Skeletal element related proteins/domains identified in echinoderms.** **a** SEM image of a skeletal element isolated from the feather star (*A. japonica*). **b** Proteins present in the feather star skeletal proteome were isolated from adult skeleton and identified by comparison of LC/MS/MS data to the genes computationally identified in the feather star genome. These proteins were compared to those found in skeletal proteomes of the sea urchin *S. purpuratus*<sup>62–64</sup>, the sea star *P. miniata*<sup>36</sup> and the brittle star *O. spiculata*<sup>35</sup>. The most prevalent proteins are shown in the figure, along with the number of different proteins from the listed groups present in each species' proteome. The feather star skeletal proteome contains members of each of the protein families shown. The other echinoderm species are missing some of these proteins in their skeletal proteome and the number of members in each protein family varies between groups.

these differences is that the ancestral skeletal C-type lectin genes experienced extensive duplication and acquisition of repetitive domains in the sea urchin lineage. The use of C-type lectins in the sea star skeleton was lost, while the C-type lectins in the feather star and brittle star skeletal proteomes remain largely unchanged. Together, these results suggest that the precursors to all of the genes and domains used in echinoderm skeleton were already present in the common ancestor to echinoderms, which emerged 589.7 Mya (Supplementary Fig. 7). Additionally, these skeleton-related genes may have undergone frequent duplication and loss in specific lineages, together with frequent changes in gene expression, since expression of MSP130-like genes and C-type lectin genes in the skeleton forming cells were lost in some lineages, even though these genes exist in their genomes. In summary, our study highlights the genomic, transcriptomic and proteomic changes behind the evolution of unique features in echinoderms, and offers an exceptional case in understanding the general tendency for the evolution of body plans.

## Methods

**Animal care and use.** Animal care and experimental procedures were conducted in strict accordance with guidelines approved by the Animal Experiments

Committee of University of Tokyo (approval ID: 14–03, 16–2). All efforts were made to minimize suffering. Individual animals and embryos were selected blindly from wild types.

**DNA extraction, library construction, and genome sequencing.** *Lytechinus variegatus*: Genomic DNA was extracted from sperms from a single male. We first constructed five different short-insert libraries (394, 424, 479, 496, and 522 bp. See also Supplementary Table 1) from the genomic DNA samples and sequenced them using the Illumina HiSeq 4000 system to survey the genome complexity. After obtaining the genome size, we further constructed four mate-pair libraries (2–18 Kb. See also Supplementary Table 1) from the same DNA sample, and sequenced them for further assembly. The DNA and genome we obtained in this project is independent from those available through EchinoBase (*Lytechinus variegatus* genome v.2.2).

*Anneissia japonica*: After collecting adult feather stars in the cove of Koajiro, Sagami Bay (Misaki, Japan) by scuba diving, sperms from a single male was collected during the breeding season when the gonads were ripe with mature gametes. Sperms were embedded in ~0.5% low-melting agarose plugs (SeaPlaque GTG Agarose, Lonza), and in-gel digestion of proteins was performed by immersing the plugs in digestion buffer (10 mM Tris-Cl pH 7.5, 50 mM NaCl, 10 mM EDTA, 0.5% SDS, 200 mg/mL Proteinase K) at 55 °C, overnight. The gel-plugs were washed repeatedly with TE buffer and stored in TE at 4 °C until use. DNA was released from the gel-plugs using GELase (Epicenter). The DNA was further purified using QIAGEN Genomic-tip 20/G (QIAGEN) and dissolved in TE. Five different short-insert libraries (277, 324, 381, 450, and 477 bp) were constructed and sequenced with Illumina HiSeq 4000 system to survey the genome complexity. After obtaining the genome size, we

further constructed six mate-pair libraries (2–18 Kb) and sequenced them for further assembly (Supplementary Table 1).

**K-mer-based estimation of genome sizes.** We first compared the performances in genome size estimation by K-mer frequency method and GenomeScope. The genome size of green sea urchin estimated by GenomeScope was about 650 Mbp, while that of kmerfreq method was about 952 Mbp. Considering that the genome size estimated by kmerfreq was closer to the genome size estimated from *C*-value (0.92, [www.genomesize.com](http://www.genomesize.com)) than GenomeScope, we decided to apply kmerfreq method for the genome size estimation. Following formula was used for estimating genome size: *Genome size (bp) = K-mer number/average depth of K-mer*. Based on the rate of occurrence of K-mers in each genome, the read depths for feather star and green sea urchin were estimated as 147 and 124, respectively, leading to genome size estimations of approximately 553 Mb for feather star and 952 Mb for green sea urchin (Supplementary Fig. 1).

**Raw read filtering and error correction of short-read libraries.** HiSeq raw reads with the following features were regarded as low-quality reads and were filtered out: [1] Reads containing >10 bp adapter sequences; [2] Reads in the small insert libraries (refer to Supplementary Tables 1 and 2) having >10 bp overlap; [3] Reads having N's >10% of their length; [4] PCR duplicates (paired-end reads completely identical); [5] Reads containing >40 bp low-quality (phred quality score ≤ 5) bases. After the filtering process, we further corrected the qualified K-mers. In brief, K-mers with sequencing errors are usually low in frequency, and we thus corrected these K-mer sequences by referring to high-frequency reads. If the erroneous sites could not be corrected, the low-frequency K-mers from the reads were trimmed. No error correction was made for the long-insert libraries (refer to Supplementary Tables 1 and 2), as these were only used for scaffolding. The SOA-Pec\_bin\_v2.03 software was used to correct the error within reads. Command line: SOA-Pec\_bin\_v2.03/bin/KmerFreq\_AR -q 33 -b 100000000000 -k 17 -p output\_reads\_files\_list; SOA-Pec\_bin\_v2.03/bin/Corrector\_AR -Q 33 -k 17 output.freq.cz output.freq.cz.len reads\_files\_list.

**Gene set and genomes obtained from public database.** Refer to Supplementary Table 24 for the publicly available gene set and genomes used in this project.

**Genome assembly.** Genome sequences with the filtered and/or corrected data were assembled by Platanus software. The assembly was carried out using the following steps: (a) Contig construction: Reads from short-insert (<1 Kb) libraries were split into K-mers and used to construct a de Bruijn graph. Short branches caused by errors were removed by “tip removal” step and short repeats were resolved by K-mer extension. Bubble structures caused by heterozygosity or errors were removed. At last, subgraphs without any junctions represent the contigs. (b) Scaffold construction: All the filtered clean reads were re-aligned onto the contig sequences, and the scaffolds were constructed by weighting the consistent rate and paired-end reads relationships on the contigs. Heterozygous regions were removed as bubble or branch structures on the graph by the “bubble removal” or “branch cut” step. These simplification steps are characteristic of Platanus and especially effective for assembling complex heterozygous regions. (c) Gap filling: Paired-end reads have one end mapped on the contig with the other end located in the gap region were used to fill the gaps in the genome assembly by GapCloser1.10 software. Then the very short assembly sequences (contig shorter than 500 bp) were removed in the genome assembly. The detailed command lines of the Platanus assembly were shown as below: Feather star: `platanus assemble -o contig.fa -f short_clean_reads.fq -k 69 -u 0.2 -m 200; platanus scaffold -c contig.fa -b contigBubble.fa -o scaffold.fa -IP R1.fq R2.fq -OP R1.fq R2.fq -u 0.2`. Green sea urchin: `platanus assemble -o contig.fa -f short_clean_reads.fq -k 29 -u 0.3 -m 200; platanus scaffold -c contig.fa -b contigBubble.fa -o scaffold.fa -IP R1.fq R2.fq -OP R1.fq R2.fq -u 0.3`. Results by K-mer analysis Supplementary Fig. 1, and statistics of feather star and green sea urchin genomes are shown in Supplementary Tables 3 and 4.

**Assessment of assembled genomes.** The completeness of the feather star and green sea urchin assemblies was assessed by the BUSCO program (version 2.0), using the eukaryotic and metazoan libraries (Supplementary Tables 6–8). Reads from the short-insert libraries were also mapped to these assembled genomes by BWA and SAMtools software (`bwa index -a bwtsw genome.fa; bwa aln -t 6 genome.fa reads.fq; samtools view -b -S out.sam > out.bam; samtools flagstat out.bam`) to assess the genomic quality (Supplementary Table 9). In addition, coverage ratio of de novo assembled transcripts obtained by Trinity <ver. 2.2.0> (`perl Trinity --JM 200 G --seqType fq --left reads_R1.fq --right reads_R2.fq --SS lib_type FR --output out`) and TGICL software (`tgicl -F transcripts.fasta`) over the sequenced genomes using BLAT software (`blat genome.fa transcript.fa -t dna -q rna out.psl`) (Supplementary Tables 10–13. De novo assembled the transcripts (made by the RNAseq data we obtained for each species) were also aligned to the filtered genome and confirmed that 98.64% transcripts in feather star and 99.53% transcripts in green sea urchin were aligned.

**GC content of genome.** GC content of the feather star and green sea urchin genomes were estimated using a sliding window approach. Briefly, a 500 bp sliding window (250 bp stepwise) was employed to scan along the genome and calculate the GC content, and found that the average GC content of feather star and green sea urchin is about 33.22% and 33.71%, respectively. Both of these values were found to be similar with those of hemichordate and most chordate species except lamprey (Supplementary Fig. 2).

**Repeat annotation.** Tandem repeats in the genomes were identified using Tandem Repeat Finder<sup>38</sup> (v4.04 <http://tandem.bu.edu/trf/trf.html>) with default parameters (trf sequence.txt 2 7 7 80 10 50 2000 -d -h, these number means: Match, Mismatch, Delta, PM, PI, Minscore, and MaxPeriod), and non-interspersed repeats in the genome using RepeatMasker<sup>38</sup> (open-4–0–5) with default parameters (-nolow (Not mask low-complexity DNA or simple repeats) -no\_is (Skips bacterial insertion element check) -norna (Does not mask small RNA (pseudo) genes) -parallel 1 (The number of processors to use in parallel)). Transposable elements (TEs) were identified on both the DNA and protein levels. On the DNA level, RepeatModeler (v1.0.4) and RepeatScout<sup>39</sup> (version 1.0.5) was used to build repeat libraries. In feather star, LTR\_FINDER<sup>40</sup> (v1.0.5) software were additionally used to build de novo repeat libraries. RepeatMasker was performed on both de novo libraries and repbase (RepBase1.6.02) separately to identify homologous repeats with default parameters with format set with 2 (-w 2-table). On the protein level, RM-BLASTX within RepeatProteinMask was used to query the TE protein database with -noLowSimple and *P*-value 0.0001 (Supplementary Fig. 3 and Supplementary Tables 14–17).

**Prediction of protein-coding genes.** Prediction of protein-coding genes was based on integration of three different methods, namely, ab initio prediction, homology-based annotation and RNAseq-based annotation. For ab initio prediction, Augustus<sup>41</sup> (v2.5.5, --uniqueGeneId=true [output gene identifiers] --noIn-FrameStop=true [Do not report transcripts with in-frame stop codons] --gff3=on [output in gff3 format] --strand=both [-strand=forward and --strand=back-ward]) and GENSCAN<sup>42,43</sup> (v1.0, -mini\_cds 150 -cds\_ns 10) software were used to predict genes. In feather star, SNAP<sup>44</sup> (using \*\* species for gene prediction) and GlimmerHMM<sup>45</sup> (v3.02, using \*\* species for gene prediction, and -f [Do not make partial gene predictions] -g [Print output in gff format]) softwares were also used in this analysis. These four software programs were trained by using lamprey, human, ciona, and zebrafish, respectively. Short genes (CDS length < 150 bp) and low-quality genes (gaps covered more than 10% of the coding region) were discarded. Proteins from human (Ensembl:GRCh38), mouse (Ensembl: GRCm38), chicken (Ensembl: Gallus\_gallus-5.0), green anole lizard (Ensembl: AnoCar2.0), Xenopus tropicalis (Ensembl: JGI\_4.2), zebrafish (Ensembl: GRCz10), sea lamprey (Ensembl: Pmarinus\_7.0), lancelet (LanceletDB: v18h27.r3\_ref), *Ciona intestinalis* (NCBI: GCA\_000224145.1), acorn worm (NCBI: GCF\_000003605.2) and purple sea urchin (NCBI: GCF\_000002235.4) were used in the homology-based annotation using tblastn with *e*-value 1e-5. Blast hits that correspond to reference proteins were concatenated by Solar software and low-quality records were filtered out. Sequence of each reference protein was extended to upstream and downstream by 2 Kb to represent the protein-coding region with default parameters. GeneWise software was used to predict gene structure contained in each protein-coding region. For each gene locus, the longest coding region and/or highest genewise score was retained. In RNAseq-based method, the coding sequences defined by transcripts was aligned against the genome by BLAT<sup>46</sup> (v34, identity > 90%, coverage > 90%), thereby defining the splicing orientation of coding region. Then, PASA software was used to link the spliced alignments with default parameters. The Evidence-Modeler<sup>47</sup> (EVM, ver. 1.1) software was further used to integrate data derived from the three methods into an EVM-derived gene set with default parameters, the weight of de novo, homolog and complementary DNA (cDNA) are 1, 5, and 10. Finally, 26,838 and 30,238 protein-coding gene models were annotated in feather star and green sea urchin genome, respectively (Supplementary Table 18).

**Annotation of gene function, non-protein-coding genes.** InterProScan (v4.8) was used to screen these genes' protein sequences against five databases (including: Pfam, release 27.0, prints, release 42.0, prosite, release 20.97, ProDom, 2006.1, and smart, release 6.2) to determine the InterPro and GO number of those predicted protein-coding genes. In addition, KEGG, COG, NR, Uniprot/SwissProt, and UniProt/TrEMBL databases were searched for homology-based functions (Supplementary Tables 19 and 20) using blastp (v2.2.26) with *e*-value (1e-5). For non-coding genes, tRNAscan-SE 53 (v1.3) software for eukaryotes was used for tRNA annotation in the genomic assembly with default parameters. Ribosomal RNA (rRNA) annotation was based on homology information of invertebrate rRNA collections using BLASTN (v2.2.26) with *e*-value set as 1e-5. The small nuclear RNA (snRNA) and microRNA (miRNA) were predicted by INFERNAL software (v0.81) against the Rfam database (Release 9.1) with default parameters. The statistical results are shown in Supplementary Tables 21 and 22.

Potential functions of protein-coding genes were predicted using InterProScan<sup>48</sup> (v4.5), against five databases (Pfam, release 27.0, PRINTS, release 42.0, PROSITE, release 20.97, ProDom, 2006.1, and SMART, release 6.2). In addition, KEGG, COG, NR, Uniprot/SwissProt and UniProt/TrEMBL databases

were searched for homology-based functions (Supplementary Tables 19 and 20). For non-coding genes, the tRNAscan-SE<sup>49</sup> (v1.3) software for eukaryote was used for tRNA annotation in the genomic assembly. rRNA annotation was based on homology information of invertebrate rRNA collections using BLASTN (v2.2.26) with *e*-value (1e-5). The snRNA and miRNA were predicted by INFERNAL software (v0.81) against the Rfam database (Release 9.1). The statistical results are shown in Supplementary Tables 21 and 22.

**Gene family analysis.** *orthoMCL*: *orthoMCL*<sup>50</sup> was used to find ortholog genes and/or gene families (ortholog groups) among different species. Amphioxus (*Branchiostoma floridae*), zebrafish (*Danio rerio*), ciona (*Ciona intestinalis*), *Drosophila* (*Drosophila melanogaster*), chicken (*Gallus gallus*), acorn worm (*Saccoglossus kowalevskii*), green sea urchin (*Lytechinus variegatus* [Lv]), purple sea urchin (*Strongylocentrotus purpuratus* [Sp]), medaka (*Oryzias latipes*), mouse (*Mus musculus*), brittle star (*Ophiothrix spiculata*), feather star (*Anneissia japonica* [Anj]), lamprey (*Petromyzon marinus*), sea cucumber (*Apostichopus japonicus* [Apj]), sea star (*Acanthaster planci*), frog (*Xenopus laevis*), turtle (*Pelodiscus sinensis*) gene set were prepared and used here. Gene families and ortholog genes identified by this *OrthoMCL* is shown in Supplementary Figs. 5 and 6. For ortholog groups identified among echinoderms were 15618 for Lv-Sp-Apj-Anj, 14758 for Lv-Apj-Anj, 15035 for Lv-Sp-Apj, 14231 for Lv-Sp and 13649 for Lv-Anj.

**Reciprocal best blast hit (RBBH), 1:1 orthologs:** We also analyzed the ortholog genes by RBBH method. We first selected feather star as the reference species, and aligned the protein sequences in all other 16 species to feather star gene set and vice versa by blast. Second, the aligned results were filtered by *e*-value (1e-5) and retained only the reciprocal best blast hit for each gene-gene pairs. Third, orthologous gene pairs in all of the 16 species were extracted. Finally, we identified 1196 ortholog genes among these 17 species.

**Phylogenetic tree construction and divergence time.** *Molecular phylogenetic analysis:* 1196 RBBH ortholog genes (1,447,456 aa) in each species were combined into a super-gene in the same gene order, followed by phylogenetic analysis using RaxML<sup>51</sup> (with PROTGAMMAAUTO model, *Drosophila melanogaster* was used as the outgroup species) through these super-genes (Fig. 1b and Supplementary Table 25). Both of the reconstructed phylogenetic trees robustly showed three clusters, including echinoderms, acorn worm, and chordates. Among them, feather star was the earliest diverging species in echinoderms, brittle star and sea star form one branch, sea urchin and sea cucumber form another branch.

**Divergence time estimation:** To estimate the divergence time, the super-genes prepared above were analyzed by MCMCTree software, together with several calibration points downloaded from TimeTree website (<http://www.timetree.orgs>) (Supplementary Fig. 7).

**Relative evolutionary rates of species.** To determine the relatively evolutionary rates of echinoderm species, the super-genes, which we produced from 1196 RBBH orthologs were used. LINTRE software and RRT (Tajima's relative rate test) analysis were employed, and *Drosophila* was used as an outgroup to determine the root of the whole tree (Supplementary Fig. 8 and Supplementary Tables 26–28. In addition to the LINTRE analysis, R-package “APE” was also used to deduce robust conclusion (Supplementary Fig. 9).

**Expansion and contraction of gene families.** To identify expanded and contracted gene families in the common ancestor of echinoderms, the gene family result generated from *OrthoMCL* were used and analyzed by CAFE software (Supplementary Fig. 11). Profiles of GO terms, protein domains and KEGG pathways of these expanded and contracted genes are shown in Extended data 1 (Extended\_data1.xlsx).

**Domains lost in echinoderms.** Domains found in any of the chordate species, but not found in any of the echinoderm species were defined as domains lost in echinoderm lineage. Six echinoderm species (*Apostichopus japonicus*, *Lytechinus variegatus*, *Acanthaster planci*, *Ophiothrix spiculata*, *Anneissia japonica*, and *Strongylocentrotus purpuratus*) and nine chordate species (*Mus musculus*, *Branchiostoma floridae*, *Ciona intestinalis*, *Petromyzon marinus*, *Oryzias latipes*, *Gallus gallus*, *Xenopus laevis*, *Pelodiscus sinensis*, and *Danio rerio*) were blasted (>50% identity and >30% align ratio) to the acorn worm (*Saccoglossus kowalevskii*) protein gene set and searched for potential domains lost in echinoderms. Seven-hundred forty-seven genes were identified to be the lost genes in echinoderms. Among these genes, six genes were not found in any of the nine chordate species, but found in acorn worm. These genes were enriched with GO terms of biosynthetic process, metabolism process, and the establishment of localization.

**Hox cluster analysis.** *Cloning of Hox genes:* To further confirm sequences of Hox genes in the feather star, a total of nine hox genes had been cloned from *Anneissia japonica* using RT-PCR (Tsurugaya et al., in preparation).

**Preparation of A. japonica BAC library:** We constructed BAC library, using genomic DNA prepared from the male gonadal pinnules that contained testes. The DNA was partially digested with the restriction enzyme MboI, size-fractionated,

and cloned into the vector pCCBAC1(EPICENTER). The bacterial strain DH10B T1 phage resistant (Invitrogen) was used for transfecting the BACs for constructing the library. Single clones were picked into 384-well plates and preserved. Two batches of libraries were produced, which were named Oj1 (average insert size ~100 kb, 35,712 clones) and Oj2 (average insert size ~78 kb, 45,977 clones).

**Screening and cloning of BAC clones containing hox genes:** Using the cDNA fragments of nine hox genes (*hox1*, *hox2*, *hox4*, *hox5*, *hox7*, *hox8*, *hox9/10*, *hox11/13a*, and *hox11/13c*. Tsurugaya et al., in preparation), we screened the BAC library of *A. japonica* for the clones that contained Hox genes and their neighboring regions. This screening yielded 23 clones in total, which, however, were not contiguous but separated into four groups.

**FISH analysis:** Probes for FISH were derived from clones out of the *A. japonica* BAC library. BAC clones used for FISH were Oj1–26E10 (containing *hox1*), Oj2–17D15 (*hox2*), Oj2–75D03 (*hox4* and *hox5*), Oj2–78N14 (*hox7* and *hox8*), Oj1–50I03 (*hox8* and *hox9/10*), and Oj2–102A05 (*hox11/13c*). BAC clone DNAs were isolated using Qiagen Plasmid Midi Kit (Qiagen) and labeled with biotin or digoxigenin by using Nick Translation Kit (Roche). Hybridization mix was prepared as described previously (3). Two color-chromosomal FISH was carried out as described previously (1, 2) with the following modifications. Blastula or early gastrula stage embryos were treated with 0.08% colchicine (Sigma) in sea water for 30 min. Embryos were fixed in methanol glacial acetic acid (3:1) fixative at 4 °C overnight, then transferred to 100% ethanol, and stored at –20 °C. To prepare metaphase spreads, 80 µL of 60% acetic acid was added to a microfuge tube containing 50–100 embryos. Three minutes later, embryos were dropped onto a prewarmed (48 °C) slide glass, and left until dry (about 30 min). Before hybridization, the slides were treated with 0.5% pepsin (1:100, Wako) in 0.01 N HCl for 3 min, and washed in phosphate-buffered saline (PBS) three times. Then the slides were post-fixed in 1% paraformaldehyde in PBS at r.t. for 30 min, and washed in PBS twice. After dehydration, the air-dried slides were treated with acetone at r.t. for 10 min, and dried again. Following the denaturation of chromosomal DNA and dehydration, hybridization was carried out at 43 °C for 16 h. FISH images were taken using an Olympus BX60 microscope equipped with an Olympus DP70 camera.

**Identification of clustered Hox genes:** In scaffold 288292 (about 1.86 Mbp in length), *hox1*, *hox2*, *hox3*, *hox4*, *hox5*, *hox6*, *hox7*, *hox8*, *hox9/10*, and *hox11/13a* were identified. The ten Hox genes were aligned in the order, spanning about 480 kb in length, with the 3' end of *hox1* about 392 kb away from the end of the scaffold. In scaffold 287987 (about 96 kb), *hox11/13b* and *hox11/13c* were identified. To see whether the 12 Hox genes form a single cluster, we carried out two color-chromosomal FISH, using the BAC clones (described above) as probes. The FISH analysis revealed that the eight genes contained in the BAC clones were in close vicinity to one another on a single chromosome (comprising of two sister chromatids). However, the gene order or relative positions of the two scaffolds on the chromosome could not be clarified, leaving four possible gene orders of 12 Hox genes undetermined. Thus, it is suggested that 12 Hox genes are present on a single chromosome, forming two subclusters separated by at least 400 kb in the genome of *Anneissia japonica*<sup>10</sup>, *hox11/13d* and *hox11/13e*, in another scaffolds 2266 and 6788, respectively. This suggests that the two genes are localized apart from the subcluster of ten Hox genes, which situation is consistent with the previous report showing that *Hox11/13d* and *Hox11/13e* do not reside in the Hox gene cluster in echinoderm genomes<sup>10</sup>.

**Embryo collection and RNA extraction.** *Lytechinus variegatus:* Adult green sea urchins were originally obtained from Reeftopia in Florida (FL) or from the Duke Marine lab in Beaufort NC. *L. variegatus* total RNA was prepared from wild type embryos per timepoint using TRIzol (Invitrogen) and DNase treatment. RNA quantitation and integrity were determined using a Qubit® 2.0 Fluorometer (Life Technologies) and a 2100 Bioanalyzer (Agilent Technologies). Total RNA was subjected to three iterations of polyA selection using Dynabeads (Life Technologies) prior to cDNA synthesis. Following stages were collected for RNA extraction and fixation; 2 cell (1 h post fertilization), 60 cell (2.5 hpf), EB (Early Blastula, 4 hpf), HB (Hatched Blastula, 7 hpf), TVP (Thickened Vegetal Plate, 10 hpf), MB (Mesenchyme Blastula, 12 hpf), EG (Early Gastrula, 13 hpf), MG (Mid Gastrula, 15 hpf), LG (Late Gastrula, 18 hpf), EP (Early Pluteus, 36 hpf), LP (Late Pluteus, 48 hpf), 7 wpf (7 weeks post fertilization), 8 wpf (8 weeks post fertilization), 1 day post metamorphosis, 1 week post metamorphosis, and adult. In addition, RNA from larval region of 8 weeks post fertilization (8 wpf Larva), and rudiment region of 8 weeks post fertilization (8 wpf Rudiment) were also extracted by dissecting the 8 wpf embryo. Results based on analyses with RNAseq data from two cell to Late pluteus were published<sup>52</sup>. Two independent biological samples were prepared for all the sampled stages.

*Apostichopus japonicus:* Embryos of fertilized eggs, 4 cell (2 h post fertilization (hpf)), morula (6 hpf), blastula (14 hpf), gastrula (29 hpf), late gastrula (34 hpf), early auricularia larva (48 hpf), mid-auricularia larva (69 hpf), late auricularia larva (15 days post fertilization (dpf)), metamorphosis 1–4 (17–19 dpf), doliolaria larva (19 dpf), pentactula larva (27 dpf), and juvenile (51 dpf) stage were collected and used for this study, as previously described<sup>23</sup>. Three independent biological samples were prepared for all the sampled stages.

*Anneissia japonica:* Adult *Anneissia japonica* (previously called as *Oxycomanthus japonicus*, see Summers et al.<sup>53</sup> for the nomenclature) were



collected from rocky substrate of about 10 m depth at Koajiro, Sagami Bay, and kept in the sea until the day of spawning. Spawning check was carried out at every neap tide days during October and November, 2015. Spawning was observed in the evening of 20th and 21st of October, 2015, and seven females spawned in total. The obtained eggs were very fragile and surrounded by mucus. Small amounts of the spawned unfertilized eggs were separated in 1.5 mL tubes (100 µL each) for RNA extraction and fixation. The rest of the eggs were inseminated immediately by diluting concentrated sperm, which were directly collected from genital pinnules. The fertilized eggs were washed with filtered sea water several times to remove the mucus, and separated in the filtered sea water in plastic vessels for culture. The culture was done at room temperature (about 10–20 °C). Following stages were collected for RNA extraction and fixation (Supplementary Fig. 15); 2 cells (1.5 h post fertilization), 8 cells (2.5 hpf), 32 cells (3.5 hpf), gastrula (8 hpf), hatching stage (17 hpf), early doliolaria (24 hpf), mid-late doliolaria (36 hpf), attachment stage (3–4 days pf), early cystidean (4–7 days pf), late cystidean (7–9 days pf), early pentacrinoid (3 weeks pf), late pentacrinoid (1.5 months pf), juvenile (2.5 months pf), arm branching stage (6–7 months pf), and adult (9 months pf). For the RNA extraction, more than 50 µL of specimens were diluted in the 10x volume of TRIzol reagent (Invitrogen). The tissue of the specimens were destructed by pipetting with a micro syringe or grinding with a pestle and mortar in the TRIzol reagent, and immediately stored in –80 °C. For the fixation, specimens were fixed with 4% paraformaldehyde in 0.5 M NaCl and 0.1 M 3-(N-morpholino) propanesulfonic acid (MOPS), pH 7.0 for several days at room temperature (about 22 °C). Fixed specimens were washed with 70% ethanol three times, and stored in 70% ethanol at –20 °C. Two independent biological samples were prepared for all the sampled stages.

**RNA sequencing and gene expression data.** After adjusting total RNA amounts between samples, non-stranded sequencing libraries (with the TruSeq protocol) were constructed and sequenced using the Illumina HiSeq 4000 platform. For the sea cucumber (*A. japonicus*) samples, Quartz-seq amplified libraries were made as previously described<sup>13</sup>. Qualities of raw reads were evaluated using FastQC program (<http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/>). Read length and single/paired information are as follows; *L. variegatus* (100 bp, paired-end), *A. japonica* (150 bp, paired-end), *A. japonicus* (100 bp, single-end). Adapter sequences of Quartz-Seq samples (Mm early stages 2-cell-blastocyst) were removed using the fastq-mcf program (<https://code.google.com/p/ea-utils/wiki/FastqMcf>) as previously described<sup>13</sup>. RNAseq data were then mapped to genomes of each species using HISAT2 program<sup>54</sup> (ver. 2.05), and calculated relative expression levels by StringTie<sup>55</sup> (ver. 1.3.5) with species-specific GTF files.

*Apostichopus japonicus*: For the gene expression levels of Japanese sea cucumber, genome and GTF files reported by Zhang et al.<sup>56</sup> were used.

*Strongylocentrotus purpuratus*: Developmental transcriptomes of the purple sea urchin was obtained from SRA (Accession: PRJNA81157), sequenced by Tu et al.<sup>21,22</sup>. The dataset contained developmental stages of 0 hpf, 10 hpf, 18 hpf, 24 hpf, 30 hpf, 40 hpf, 48 hpf, 56 hpf, 64 hpf, 72 hpf, four arm larva, vestibular invagi, pentagonal disc, tube foot protrusion, post metamorphosis, young juvenile, adult.

**Identification of conserved stages.** Whole embryonic, comparative transcriptomic analysis was performed as previously described to find evolutionarily conserved developmental stages<sup>13</sup>. Relative expression levels (TPM) of ortholog groups (defined by orthomcl<sup>37</sup>) were calculated from the RNAseq data, and then compared their dissimilarities (expDists) among developmental stages of different species. In calculating dissimilarity (1 – Spearman) of ortholog-group-based whole-embryonic transcriptomes (expDists), phylogenetic relationship [Anj(Apj(Sp,Lv))] were taken into consideration to avoid unwanted bias arising from simple pair-wise comparisons<sup>58</sup>. By randomly picking-up one biological replicate sample for each developmental stage (in each species), 100 expression tables (100 biological replicate included expression table, or BRI-exp data) were created (method reported in Hu et al.<sup>13</sup>). We used this 100 BRI table to test statistical significance of changes in Ptop scores (Friedman test).

**Whole-mount in situ hybridization.** Digoxigenin (Dig) labeled riboprobe for *pitx* gene of *A. japonica* was prepared from PCR-amplified fragments (911 bp) using following primers: 5'-GAACGATTCGCTTCCGATGC-3' (forward primer), 5'-TGAGACCGCGTATTGACAC-3' (reverse primer). Whole-mount in situ hybridization (WISH) was conducted following the protocol for the planktonic larvae of a stalked crinoid *Metacrinus rotundus*<sup>59</sup> with some modifications. Specimens were fixed with 4% paraformaldehyde (PFA) in 0.5 M NaCl and 0.1 M 3-(N-morpholino) propanesulfonic acid (MOPS), pH 7.0 for over 1 day, and stored at –20 °C in 70% ethanol. The fixed specimens were washed three times with PBST (1× phosphate-buffered saline with 0.1% Tween 20), treated with 0.2 µg/mL proteinase K in PBST at 37 °C for 20 min, re-fixed with 4% PFA at 4 °C for 30 min, washed three times with PBST, and then incubated in hybridization buffer (50% formamide, 5× SSC, 100 µg/mL yeast RNA, 50 µg/mL heparin, 1% Tween 20) at 55 °C for 4–6 h. Hybridization was carried out with 0.2 µg/mL probes in the hybridization buffer at 55 °C for 5 days. The protocol after hybridization was as previously described<sup>59</sup>. WISH finished samples were observed and photographed under the BX-51 optical microscope (Olympus).

**Proteomic analysis.** Proteins were isolated from adult feather skeleton and analyzed as previously described<sup>36</sup>. All organic material was removed from the skeleton by extensive washing with sodium hypochlorite followed by guanidine isothiocyanate. The skeleton was demineralized with acetic acid followed by dialysis. Both soluble and insoluble protein fractions were analyzed. Proteins were separated by sodium dodecyl sulfate–polyacrylamide gel electrophoresis. Each lane was excised into 20 equal sized segments and processed. In-gel digestion with trypsin was performed on each fraction, followed by analysis by nano Liquid chromatography–mass spectrometry (LC-MS/MS) with a Waters nanoAcquity high-performance liquid chromatography system interfaced to a ThermoFisher Q Exactive hybrid quadrupole-orbitrap mass spectrometer. The mass spectrometer was operated in a data-dependent mode. Data were used to search predicted peptides from the *A. japonica* genome using Mascot and then parsed into the Scaffold algorithm for validation and filtering, using a 95% protein identification score with at least two peptides per protein.

**Statistics and reproducibility.** Alpha levels of 0.05 were regarded as statistically significant throughout the study, unless otherwise specified. Experiments were repeated multiple times to confirm the reproducibility of the data. See details for individual experiments in the “Methods” sections above.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Genomic sequence data and assembled genomes for the following species are available through the NCBI database at the indicated BioProject accession IDs: Brittle star genome (PRJNA182997), feather star genome (PRJNA553656), and green sea urchin genome (PRJNA553643). RNAseq data are available for the following species at the indicated BioProject accession IDs: green sea urchin RNAseq data (PRJNA554218), feather star RNAseq data (PRJNA553591), and Japanese sea cucumber RNAseq data (PRJNA553613). Cloned sequences of Hox genes of the feather star (hox1 LC462021, hox2 LC462022, hox4 LC462023, hox5 LC462024, hox7 LC462025, hox8 LC462026, hox9/10 LC462027, hox11/13a LC462028, hox11/13c LC462029) are also available through the NCBI database at the indicated nucleotide accession IDs. Assembled genomes and gene sets can also be accessed through the DRYAD database<sup>60</sup> at <https://doi.org/10.5061/dryad.rbnzs7h7n>. Proteomic data are available via the ProteomeXchange with identifier PXD019526. Source data for Figs. 2a and 3b can be found in Supplementary Data 1 and Supplementary Data 2.

## Code availability

No custom or proprietary software was used in the analysis. Versions and parameters for each software package used are described in the reporting summary and elsewhere in the “Methods.”

Received: 18 October 2019; Accepted: 19 June 2020;

Published online: 10 July 2020

## References

- Hyman, L. H. *The Invertebrates: Echinodermata* Vol. IV (McGraw-Hill, 1955).
- Mooi, R., David, B. & Wray, G. A. Arrays in rays: terminal addition in echinoderms and its correlation with gene expression. *Evol. Dev.* **7**, 542–555 (2005).
- Rozhnov, S. V. Symmetry of echinoderms: From initial bilaterally-asymmetric metamerism to pentaradiality. *Nat. Sci.* **06**, 171–183 (2014).
- Reich, A., Dunn, C., Akasaka, K. & Wessel, G. Phylogenomic analyses of Echinodermata support the sister groups of Asterozoa and Echinozoa. *PLoS ONE* **10**, e0119627 (2015).
- Telford, M. J. et al. Phylogenomic analysis of echinoderm class relationships supports Asterozoa. *Proc. Biol. Sci.* **281**, 20140479 (2014).
- Janda, L., Damborsky, J., Reznicek, G. A. & Wiche, G. Plectin repeats and modules: strategic cysteines and their presumed impact on cytolinker functions. *Bioessays* **23**, 1064–1069 (2001).
- 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).
- Baughman, K. W. et al. Genomic organization of Hox and ParaHox clusters in the echinoderm, *Acanthaster planci*. *Genesis* **52**, 952–958 (2014).
- Freeman, R. et al. Identical genomic organization of two hemichordate hox clusters. *Curr. Biol.* **22**, 2053–2058 (2012).
- Szabo, R. & Ferrier, D. E. K. Two more Posterior Hox genes and Hox cluster dispersal in echinoderms. *BMC Evol. Biol.* **18**, 203 (2018).
- Duboule, D. Temporal colinearity and the phylogenetic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies



- through heterochrony. *Development* 135–142 <https://dev.biologists.org/content/development/1994/Supplement/135.full.pdf> (1994).
12. Irie, N. & Kuratani, S. Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. *Nat. Commun.* 2, 248 (2011).
  13. Hu, H. et al. Constrained vertebrate evolution by pleiotropic genes. *Nat. Ecol. Evol.* 1, 1722–1730 (2017).
  14. Zalts, H. & Yanai, I. Developmental constraints shape the evolution of the nematode mid-developmental transition. *Nat. Ecol. Evol.* 1, 113 (2017).
  15. Kalinka, A. T. et al. Gene expression divergence recapitulates the developmental hourglass model. *Nature* 468, 811–814 (2010).
  16. Xu, F. et al. High expression of new genes in trochophore enlightening the ontogeny and evolution of trochozoans. *Sci. Rep.* 6, 34664 (2016).
  17. Raff, A. *The Shape of Life: Genes, Development, and the Evolution of Animal Form* (University of Chicago Press, 1996).
  18. Irie, N., Satoh, N. & Kuratani, S. The phylum Vertebrata: a case for zoological recognition. *Zool. Lett.* 4, 32 (2018).
  19. Israel, J. W. et al. Comparative developmental transcriptomics reveals rewiring of a highly conserved gene regulatory network during a major life history switch in the sea urchin genus *Helicidaris*. *PLoS Biol.* 14, e1002391 (2016).
  20. Malik, A., Gildor, T., Sher, N., Layous, M. & Ben-Tabou de-Leon, S. Parallel embryonic morphological programs evolve under distinct constraints and may enable morphological conservation amidst adaptation. *Dev. Biol.* 430, 202–213 (2017).
  21. Tu, Q., Cameron, R. A. & Davidson, E. H. Quantitative developmental transcriptomes of the sea urchin *Strongylocentrotus purpuratus*. *Dev. Biol.* 385, 160–167 (2014).
  22. Tu, Q., Cameron, R. A., Worley, K. C., Gibbs, R. A. & Davidson, E. H. Gene structure in the sea urchin *Strongylocentrotus purpuratus* based on transcriptome analysis. *Genome Res.* 22, 2079–2087 (2012).
  23. Li, Y. et al. Weighted gene co-expression network analysis reveals potential genes involved in early metamorphosis process in sea cucumber *Apostichopus japonicus*. *Biochem. Biophys. Res. Commun.* 495, 1395–1402 (2018).
  24. Uchida, Y., Uesaka, M., Yamamoto, T., Takeda, H. & Irie, N. Embryonic lethality is not sufficient to explain hourglass-like conservation of vertebrate embryos. *EvoDevo* 9, 7 (2018).
  25. Irie, N. Remaining questions related to the hourglass model in vertebrate evolution. *Curr. Opin. Genet. Dev.* 45, 103–107 (2017).
  26. Sumrall, C. D. & Wray, G. A. Ontogeny in the fossil record: diversification of body plans and the evolution of “aberrant” symmetry in Paleozoic echinoderms. *Paleobiology* 33, 149–163 (2007).
  27. Molina, M. D., de Croze, N., Haillet, E. & Lepage, T. Nodal: master and commander of the dorsal-ventral and left-right axes in the sea urchin embryo. *Curr. Opin. Genet. Dev.* 23, 445–453 (2013).
  28. Su, Y. H. Telling left from right: left-right asymmetric controls in sea urchins. *Genesis* 52, 269–278 (2014).
  29. Bessodes, N. et al. Reciprocal signaling between the ectoderm and a mesendodermal left-right organizer directs left-right determination in the sea urchin embryo. *PLoS Genet.* 8, e1003121 (2012).
  30. Duboc, V., Rottinger, E., Lapraz, F., Besnardeau, L. & Lepage, T. Left-right asymmetry in the sea urchin embryo is regulated by nodal signaling on the right side. *Dev. Cell* 9, 147–158 (2005).
  31. Luo, Y. J. & Su, Y. H. Opposing nodal and BMP signals regulate left-right asymmetry in the sea urchin larva. *PLoS Biol.* 10, e1001402 (2012).
  32. Eitensohn, C. A. Horizontal transfer of the msp130 gene supported the evolution of metazoan biomineralization. *Evol. Dev.* 16, 139–148 (2014).
  33. Szabo, R. & Ferrier, D. E. K. Another biomineralising protostome with an msp130 gene and conservation of msp130 gene structure across Bilateria. *Evol. Dev.* 17, 195–197 (2015).
  34. Seaver, R. W. & Livingston, B. T. Examination of the skeletal proteome of the brittle star *Ophiocoma wendtii* reveals overall conservation of proteins but variation in spicule matrix proteins. *Proteome Sci.* 13, 7 (2015).
  35. Flores, R. L., Gonzales, K., Seaver, R. W. & Livingston, B. T. The skeletal proteome of the brittle star *Ophiotrix spiculata* identifies C-type lectins and other proteins conserved in echinoderm skeleton formation. *AIMS Mol. Sci.* 3, 357–367 (2016).
  36. Flores, R. L. & Livingston, B. T. The skeletal proteome of the sea star *Patiria miniata* and evolution of biomineralization in echinoderms. *BMC Evol. Biol.* 17, 125 (2017).
  37. Livingston, B. T. et al. A genome-wide analysis of biomineralization-related proteins in the sea urchin *Strongylocentrotus purpuratus*. *Dev. Biol.* 300, 335–348 (2006).
  38. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573–580 (1999).
  39. Price, A. L., Jones, N. C. & Pevzner, P. A. De novo identification of repeat families in large genomes. *Bioinformatics* 21, i351–i358 (2005). Suppl 1.
  40. Xu, Z. & Wang, H. LTR\_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* 35, W265–W268 (2007).
  41. Stanke, M. & Waack, S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* 19, ii215–ii225 (2003). Suppl 2.
  42. Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* 268, 78–94 (1997).
  43. Burge, C. B. & Karlin, S. Finding the genes in genomic DNA. *Curr. Opin. Struct. Biol.* 8, 346–354 (1998).
  44. Korf, I. Gene finding in novel genomes. *BMC Bioinforma.* 5, 59 (2004).
  45. Majoros, W. H., Pertea, M. & Salzberg, S. L. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* 20, 2878–2879 (2004).
  46. Kent, W. J. BLAT-the BLAST-like alignment tool. *Genome Res.* 12, 656–664 (2002).
  47. Haas, B. J. et al. Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* 9, R7 (2008).
  48. Zdobnov, E. M. & Apweiler, R. InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17, 847–848 (2001).
  49. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964 (1997).
  50. Li, L., Stoeckert, C. J. Jr. & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189 (2003).
  51. Stamatakis, A. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313 (2014).
  52. Hogan, J. D. K. et al. The developmental transcriptome for *Lytechinus variegatus* exhibits temporally punctuated gene expression. *Chang. Dev.* 460, 139–154 (2020). Issue 2.
  53. Summers, M. M., Messing, C. G. & Rouse, G. W. Phylogeny of Comatulidae (Echinodermata: Crinoidea: Comatulida): a new classification and an assessment of morphological characters for crinoid taxonomy. *Mol. Phylogenet. Evol.* 80, 319–339 (2014).
  54. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360 (2015).
  55. Pertea, M., Kim, D., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* 11, 1650–1667 (2016).
  56. Zhang, X. et al. The sea cucumber genome provides insights into morphological evolution and visceral regeneration. *PLoS Biol.* 15, e2003790 (2017).
  57. Tabari, E. & Su, Z. PorthoMCL: Parallel orthology prediction using MCL for the realm of massive genome availability. *Big Data Analytics* 2, 4 (2017).
  58. Dunn, C. W., Zapata, F., Munro, C., Siebert, S. & Hejnol, A. Pairwise comparisons across species are problematic when analyzing functional genomic data. *Proc. Natl Acad. Sci. USA* 115, E409–E417 (2018).
  59. Omori, A., Akasaka, K., Kurokawa, D. & Amemiya, S. Gene expression analysis of Six3, Pax6, and Otx in the early development of the stalked crinoid *Metacrinus rotundus*. *Gene Expr. Patterns* 11, 48–56 (2011).
  60. Li, Y. Genomic tests of body plan transitions from bilateral to pentameric symmetry in Echinoderms, v3. *Dryad*, <https://doi.org/10.5061/dryad.rbnz57h7n> (2020).
  61. Sumrall, C. D. & Gregory, A. W. Ontogeny in the fossil record: diversification of body plans and the evolution of “aberrant” symmetry in Paleozoic echinoderms. *Paleobiology* 33, 149–163 (2007).
  62. Mann, K., Wilt, F. H. & Poustka, A. J. Proteomic analysis of sea urchin (*Strongylocentrotus purpuratus*) spicule matrix. *Proteome Sci.* 8, 33 (2010).
  63. Mann, K., Poustka, A. J. & Mann, M. In-depth, high-accuracy proteomics of sea urchin tooth organic matrix. *Proteome Sci.* 6, 33 (2008).
  64. Mann, K., Poustka, A. J. & Mann, M. The sea urchin (*Strongylocentrotus purpuratus*) test and spine proteomes. *Proteome Sci.* 6, 22 (2008).

## Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 31561143016; No. 31621062 and Strategic International Collaborative Research Program (SICORP) of JST. We thank Flore Castellan for help drawing animals and embryos. GMW gratefully acknowledges support in part from NIH 9R01GM125071.

## Author contributions

The study was conceived by N.I., W.Wang, M.Kondo, Y.L., B.L., and C.B. The echinoderm DNA and RNA samples were collected and sequenced by A.C., A.O., A.R., C.B., K.O., M.Kikuchi, M.Kondo, M.U., N.I., R.F., T.B., Y.U., W.Wan, and W.Wang. Genomes were assembled and annotated by G.W., Y.L., M.H., W.Wang, X.L., and Y.R. Bioinformatics analyses were performed by C.B., C.N., F.W., G.W., J.L., S.Satterfield, L.C., N.I., R.F., S.Shigenobu, S.Z., T.Z., Y.D., and Y.L. Hox cluster analysis was performed by H.S., K.I., T.O., T.S., T.T., T.I., and M.Kondo. Proteome analyses were done by B.L., C.N., R.F., and S.Satterfield. WISH analyses were performed by A.O., M.Kondo., and M.Kikuchi. The manuscript was written by A.C., A.O., B.L., C.B., G.W., H.S., N.I., T.O., T.I., W.Wang, and Y.L.

**Competing interests**

The authors declare no competing interests.

**Additional information**

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s42003-020-1091-1>.

**Correspondence** and requests for materials should be addressed to W.W. or N.I.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020