The *Lingula* genome provides insights into brachiopod evolution and the origin of phosphate biomineralization

Yi-Jyun Luo¹, Takeshi Takeuchi¹, Ryo Koyanagi², Lixy Yamada³, Miyuki Kanda¹, Mariia Khalturina¹, Manabu Fujie², Shin-ichi Yamasaki², Kazuyoshi Endo⁴ & Noriyuki Satoh¹

The evolutionary origins of lingulid brachiopods and their calcium phosphate shells have been obscure. Here we decode the 425-Mb genome of *Lingula anatina* to gain insights into brachiopod evolution. Comprehensive phylogenomic analyses place *Lingula* close to molluscs, but distant from annelids. The *Lingula* gene number has increased to ~34,000 by extensive expansion of gene families. Although *Lingula* and vertebrates have superficially similar hard tissue components, our genomic, transcriptomic and proteomic analyses show that *Lingula* lacks genes involved in bone formation, indicating an independent origin of their phosphate biominerals. Several genes involved in *Lingula* shell formation are shared by molluscs. However, *Lingula* has independently undergone domain combinations to produce shell matrix collagens with EGF domains and carries lineage-specific shell matrix proteins. Gene family expansion, domain shuffling and co-option of genes appear to be the genomic background of *Lingula's* unique biomineralization. This *Lingula* genome provides resources for further studies of lophotrochozoan evolution.
Brachiopods are marine invertebrates with calcium phosphate or carbonate shells. Abundant in the fossil record, Darwin first referred to lingulid brachiopods as ‘living fossils,’ because their shell morphology has changed little since the Silurian. Based on molecular phylogeny, brachiopods comprise three subphyla, Linguliformea, Craniiformea and Rhynchocephalia. The Linguliformea, including the extant genus, Lingula, is recognized as the most primitive group, with a fossil record dating back to the early Cambrian and coinciding with the innovation of animal biomineralization. Their shells are composed of calcium phosphate and collagen fibres, characters shared only by evolutionarily distant vertebrates. Morphologically, brachiopods and bivalves superficially resemble each other. However, lingulid brachiopods exhibit several unique features that distinguish them from molluscs. These include hingeless shells that grow along the dorsal–ventral axis, chitinous chaetae, ciliated lophophores and a tail-like pedicle. Since the Permian extinction, bivalves have greatly increased their diversity, but the basic body plan of brachiopods has been constrained, which is still a mystery of metazoan evolution.

It has been proposed that Lingula might have used calcium phosphate, because the phosphorus concentration in seawater was high in the Cambrian. In fact, some Cambrian arthropods, tomotids and various other problematica also used calcium phosphate for their exoskeletons, whereas other extant invertebrates such as corals, molluscs and echinoderms use calcium carbonate. Studies of mollusc mantle transcriptomes and shell proteomes suggest that gene sets responsible for formation of calcium carbonate-based biominerals such as calcite or aragonite have evolved rapidly. Therefore, mineral homology of calcium carbonate-based biominerals such as calcite or aragonite have evolved rapidly. Therefore, mineral homology among molluscs could simply represent parallel evolution. In contrast to mollusc shells and other invertebrate calcified tissues, Lingula shells comprise calcium phosphate, laminated, flexible and rich in organic materials. Despite their palaeontological importance, the evolutionary origin of Lingula shells is still unclear.

More interestingly, although Lingula is a protostome, its embryogenesis exhibits radial cleavage and enterocoelic coelom formation, typical of basal deuterostomes. Despite such unique features, the phylogeny of brachiopods is under debate. Before the 1980s, brachiopods were classified as deuterostomes, based on their mode of development. Then they were grouped within protostomes following an analysis of 18S ribosomal RNAs. This classification was further supported by an analysis of Hox genes in brachiopods and priapulids. However, the phylogenetic position of brachiopods is still controversial, in spite of intensive palaeontological and molecular phylogenetic studies (Supplementary Note 2). For example, whether brachiopods are monophyletic or polyphyletic and whether Brachiopoda is close to Phoronida, Nemertea, Mollusca, Annelida or other lophotrochozoan phyla, remains to be resolved.

Here we present the first brachiopod genome of the lingulid, L. anatina. Our whole-genome phylogenetic analyses support a close relationship between Lingula and molluscs. Unexpectedly, we find that contrary to its reputation as a ‘living fossil,’ the Lingula genome has been actively evolving, with a disorganized Hox cluster and recently expanded gene families. In addition, we show that although Lingula shares shell formation-related genes and mechanisms with molluscs, such as chitin synthase (CHS) and bone morphogenetic protein (BMP) signalling, it uses several domain combinations to produce lineage-specific shell matrix collagens, alanine-rich fibres and novel shell matrix proteins (SMPs). We propose that gene family expansion, domain shuffling and co-option of genes appear to comprise the genomic basis of Lingula’s unique biomineralization. Together with embryonic and adult tissue transcriptomes, as well as a shell proteome, our comparative genomic analyses provide insights into the evolutionary history of this lophotrochozoan and the origin of phosphate biomineralization.

Results

Genome sequencing and assembly. We sequenced the 425-Mb genome of L. anatina (Fig. 1a–i) with ~226-fold coverage using four next-generation sequencers (that is, Roche 454 GS FLX +, Illumina MiSeq and HiSeq 2500, and PacBio RS II). This effort yielded an assembly with a scaffold N50 size of 294 kb, comparable to those of other lophotrochozoan genomes. Here we present the first brachiopod genome of the lingulid, L. anatina. Our whole-genome phylogenetic analyses support a close relationship between Lingula and molluscs. Unexpectedly, we find that contrary to its reputation as a ‘living fossil,’ the Lingula genome has been actively evolving, with a disorganized Hox cluster and recently expanded gene families. In addition, we show that although Lingula shares shell formation-related genes and mechanisms with molluscs, such as chitin synthase (CHS) and bone morphogenetic protein (BMP) signalling, it uses several domain combinations to produce lineage-specific shell matrix collagens, alanine-rich fibres and novel shell matrix proteins (SMPs). We propose that gene family expansion, domain shuffling and co-option of genes appear to comprise the genomic basis of Lingula’s unique biomineralization. Together with embryonic and adult tissue transcriptomes, as well as a shell proteome, our comparative genomic analyses provide insights into the evolutionary history of this lophotrochozoan and the origin of phosphate biomineralization.

Figure 1 | Deuterostomic development of the brachiopod, L. anatina, and its close relationship to molluscs. (a) Adult (shell length ~4–5 cm). (b-i) Embryogenesis: egg (b), embryos at 4-cell (c), 16-cell (d), 32-cell (e) and 128-cell stages (f), blastula (g), late gastrula (h) and 2-pair cirri larva (i). Scale bar, 50 μm. bp, blastopore; cr, cirri; ct, chaeta; gt, gut; ml, mantle lobe; mo, mouth; pd, pedicle; sh, shell; st, stone. (j) Phylogenetic position of Lingula among lophotrochozoans (orange box; molluscs are blue; annelids are green). The tree was constructed using the maximum likelihood method with 150 one-to-one orthologues (46,845 amino-acid positions) with LG + Γ4 model. Circles at all nodes indicate 100% bootstrap support.

© 2015 Macmillan Publishers Limited. All rights reserved.
transcriptomes. The mean size of the Lingula genes is 6.7 kb with an average of 6.6 introns per gene. These numbers are closer to those of the sea snail, Lottia gigantea, than to the leech, Helobdella robusta, or the polychaete, Capitella teleta. A BLAST top-hits search against the NCBI non-redundant (nr) database shows that 28% of Lingula genes are most similar to mollusc genes, but only 12% to annelids, whereas 21% of the genes show no similarity to any known sequence, suggesting that these are unique to the brachiopod lineage (Supplementary Fig. 5). A genome browser, genome and transcriptome assemblies, and related annotation files are available at http://marinegenomics.oi.st.jp.

Phylogenetic position of brachiopods. To resolve the phylogenetic position of brachiopods, we carried out phylogenetic analyses (Supplementary Note 2, Supplementary Figs 6–8 and Supplementary Table 5). Analysis based on 150 one-to-one orthologues with 46,845 amino-acid positions from 15 metazoan genomes supports the assertion that Lingula is closer to Mollusca than to Annelida (Fig. 1j). Comparative analyses of lineage-specific domain losses among Lingula, molluscs and annelids also show that Lingula is closely related to molluscs. There are nearly 20 annelid lineage-specific domain losses, which include chordin, haem-binding protein and Death-associated protein domains (Supplementary Fig. 9 and Supplementary Table 6). In addition, microsyntenic analyses showed that Lingula and Lottia share conservation of a large number of microsyntenic blocks (Supplementary Fig. 10 and Supplementary Tables 7–9), supporting the close phylogenetic relationship between brachiopods and molluscs. Furthermore, intron structures also show similarities between Lingula and molluscs, but not annelids (Supplementary Fig. 11 and Supplementary Table 10). Therefore, it may be concluded that Brachiopoda is closer to Mollusca than Annelida, although the phylogenetic relationships of Brachiopoda to Phoronida and Nemertea remain to be resolved.

The evolving Lingula genome. An abundance of Lingula fossils from the Silurian, with morphology very similar to that of extant species, inspired Darwin with the idea of ‘living fossils.’ Nevertheless, shells of fossilized and living lingulids show considerable diversity in chemical structure. Soft tissue fossils from the Chengjiang fauna reveal morphological changes among lingulid brachiopods. Some findings suggest that lingulid brachiopids have been rapidly evolving. On the other hand, protein-coding genes of the coelacanth, another ‘living fossil,’ are reported to be evolving significantly more slowly than those of other tetrapods. Interestingly, we found that Lingula CHS genes are most similar to mollusc genes, but only 12% to annelids, whereas 21% of the genes show no similarity to any known sequence, suggesting that these are unique to the brachiopod lineage (Supplementary Fig. 5). A genome browser, genome and transcriptome assemblies, and related annotation files are available at http://marinegenomics.oi.st.jp.

A large portion of these young duplicated genes are undergoing negative selection, suggesting a functional constraint on them. We also found that genes related to extracellular matrix are experiencing positive selection (Supplementary Fig. 13), indicating an adaptive need to acquire new functions. These results suggest that the Lingula genome has a unique evolutionary history. Decoupling of molecular and morphological evolution has been also reported in the buthid scorpion, Mesobuthus martensi. We propose that the morphological constraint upon Lingula shells is not due to slow genetic changes. Despite these genomic features, Lingula contains genes for transcription factors (Supplementary Table 11) and signalling molecules (Supplementary Table 12) comparable to those of molluscs.

Expansion of gene families and CHSs. We found lineage-specific expansions of protein domains (Supplementary Table 13) and gene families (Supplementary Table 14). Five of the 20 most expanded families have possible functions in shell formation, including 31 copies of CHS genes and 30 copies of carbohydrate sulfotransferase genes (Supplementary Fig. 14). In addition, microsyntenic analyses showed that Lingula and Lottia share conservation of a large number of microsyntenic blocks (Supplementary Fig. 10 and Supplementary Tables 7–9), supporting the close phylogenetic relationship between brachiopods and molluscs. Furthermore, intron structures also show similarities between Lingula and molluscs, but not annelids (Supplementary Fig. 11 and Supplementary Table 10). Therefore, it may be concluded that Brachiopoda is closer to Mollusca than Annelida, although the phylogenetic relationships of Brachiopoda to Phoronida and Nemertea remain to be resolved.

Comparison of the genome and transcriptome with those of other metazoan clades supports the assertion that Lingula is closer to Mollusca than to Annelida. The proportions of genome and transcriptome alignments that were classified as orthologues with 46,845 amino-acid positions from 15 metazoan genomes supports the assertion that Lingula is closer to Mollusca than to Annelida (Fig. 1j). Comparative analyses of lineage-specific domain losses among Lingula, molluscs and annelids also show that Lingula is closely related to molluscs. There are nearly 20 annelid lineage-specific domain losses, which include chordin, haem-binding protein and Death-associated protein domains (Supplementary Fig. 9 and Supplementary Table 6). In addition, microsyntenic analyses showed that Lingula and Lottia share conservation of a large number of microsyntenic blocks (Supplementary Fig. 10 and Supplementary Tables 7–9), supporting the close phylogenetic relationship between brachiopods and molluscs. Furthermore, intron structures also show similarities between Lingula and molluscs, but not annelids (Supplementary Fig. 11 and Supplementary Table 10). Therefore, it may be concluded that Brachiopoda is closer to Mollusca than Annelida, although the phylogenetic relationships of Brachiopoda to Phoronida and Nemertea remain to be resolved.

The evolving Lingula genome. An abundance of Lingula fossils from the Silurian, with morphology very similar to that of extant species, inspired Darwin with the idea of ‘living fossils.’ Nevertheless, shells of fossilized and living lingulids show considerable diversity in chemical structure. Soft tissue fossils from the Chengjiang fauna reveal morphological changes among lingulid brachiopods.

Comparative genomics of biomineralization-related genes. Animals make hard tissues for protection, support and feeding, mostly in the form of calcified minerals containing carbonate or phosphate. The genome and transcriptome of Lingula has a unique evolutionary history. Decoupling of molecular and morphological evolution has been also reported in the buthid scorpion, Mesobuthus martensi. We propose that the morphological constraint upon Lingula shells is not due to slow genetic changes. Despite these genomic features, Lingula contains genes for transcription factors (Supplementary Table 11) and signalling molecules (Supplementary Table 12) comparable to those of molluscs.

Expansion of gene families and CHSs. We found lineage-specific expansions of protein domains (Supplementary Table 13) and gene families (Supplementary Table 14). Five of the 20 most expanded families have possible functions in shell formation, including 31 copies of CHS genes and 30 copies of carbohydrate sulfotransferase genes (Supplementary Fig. 14). In addition, microsyntenic analyses showed that Lingula and Lottia share conservation of a large number of microsyntenic blocks (Supplementary Fig. 10 and Supplementary Tables 7–9), supporting the close phylogenetic relationship between brachiopods and molluscs. Furthermore, intron structures also show similarities between Lingula and molluscs, but not annelids (Supplementary Fig. 11 and Supplementary Table 10). Therefore, it may be concluded that Brachiopoda is closer to Mollusca than Annelida, although the phylogenetic relationships of Brachiopoda to Phoronida and Nemertea remain to be resolved.

The evolving Lingula genome. An abundance of Lingula fossils from the Silurian, with morphology very similar to that of extant species, inspired Darwin with the idea of ‘living fossils.’ Nevertheless, shells of fossilized and living lingulids show considerable diversity in chemical structure. Soft tissue fossils from the Chengjiang fauna reveal morphological changes among lingulid brachiopods. Some findings suggest that lingulid brachiopids have been rapidly evolving. On the other hand, protein-coding genes of the coelacanth, another ‘living fossil,’ are reported to be evolving significantly more slowly than those of other tetrapods. Interestingly, we found that Lingula genes associated with basic metabolism, such as ribonucleoprtein complex biogenesis and RNA processing, show the slowest evolutionary rate among lophotrochozoans (Fig. 1j). However, we also found a high degree of changes in the genomic structure and gene families (Supplementary Note 3). The Lingula genome contains a disorganized Hox cluster. It is divided into two regions, and Lox2 and Lox4 are missing (Supplementary Fig. 12). Comparison of gene families shared by amphioxus Branchiostoma floridae, Capitella and Lottia show that Lingula has 3,525 unique gene families (Fig. 2a). Further analyses show that the Lingula genome contains 7,263 gains and 8,441 losses of gene families. The turnover rate of gene families in Lingula is the highest among bilaterians (Fig. 2b).

To better understand evolution of Lingula gene families, we further examined the age distribution of duplicated paralogous genes by estimating their non-synonymous substitution rates (Ks). Within the youngest duplicated genes (Ks < 0.1), we found that Lingula genes duplicate at a rate approximately two to four times faster than those of other lophotrochozoans (Fig. 2c). A large portion of these young duplicated genes are undergoing negative selection, suggesting a functional constraint on them. We also found that genes related to extracellular matrix are experiencing positive selection (Supplementary Fig. 13), indicating an adaptive need to acquire new functions. These results suggest that the Lingula genome has a unique evolutionary history. Decoupling of molecular and morphological evolution has been also reported in the buthid scorpion, Mesobuthus martensi. We propose that the morphological constraint upon Lingula shells is not due to slow genetic changes. Despite these genomic features, Lingula contains genes for transcription factors (Supplementary Table 11) and signalling molecules (Supplementary Table 12) comparable to those of molluscs.

Expansion of gene families and CHSs. We found lineage-specific expansions of protein domains (Supplementary Table 13) and gene families (Supplementary Table 14). Five of the 20 most expanded families have possible functions in shell formation, including 31 copies of CHS genes and 30 copies of carbohydrate sulfotransferase genes (Supplementary Fig. 14). In addition, microsyntenic analyses showed that Lingula and Lottia share conservation of a large number of microsyntenic blocks (Supplementary Fig. 10 and Supplementary Tables 7–9), supporting the close phylogenetic relationship between brachiopods and molluscs. Furthermore, intron structures also show similarities between Lingula and molluscs, but not annelids (Supplementary Fig. 11 and Supplementary Table 10). Therefore, it may be concluded that Brachiopoda is closer to Mollusca than Annelida, although the phylogenetic relationships of Brachiopoda to Phoronida and Nemertea remain to be resolved.
expression profiles of genes involved in ribosomal machinery are most similar, while those of genes related to chromosomal and cell cycle regulation are diverse (Supplementary Fig. 18). Genes related to membrane trafficking are expressed in highly similar ways in Lingula and Crassostrea mantles, suggesting that the functional similarity comes mainly from genes involved in secretory machinery. However, it is worth noting that the mantle similarity between Lingula and Crassostrea revealed by our comparative transcriptomics may be the result of sharing common secretory cell types. Whether these two organs share the same evolutionary origin requires more careful examination, although some genes associated with mollusc shell formation, such as calmodulin, calponin and mucin, are also highly expressed in the Lingula mantle (Supplementary Table 19).

To gain further insights into the evolution of biomineralization, we conducted comparative genomics and hierarchical cluster analyses, to examine biomineralization-associated genes among vertebrates31, molluscs19 and Lingula. Given that Lingula and vertebrates share the use of calcium phosphate, we first examined 175 genes associated with bone formation, we found that the number of Lingula homologues to vertebrate bone formation genes is similar to those in other marine invertebrates. There is no specific similarity between Lingula and humans (Fig. 4b and Supplementary Table 20). The innovation of the acidic, secretory, calcium-binding phosphoprotein gene family is essential for vertebrate bone formation51. However, we failed to find orthologues of secretory, calcium-binding phosphoprotein genes in the Lingula genome, although it contains an orthologue of the secreted protein, acidic, cysteine-rich gene family (Supplementary Fig. 19). These analyses show that many of the genes involved in bone formation are derived from genome duplication events in the vertebrate lineage31. Transcriptome analysis of Lingula genes that are associated with bone formation in vertebrates shows that most of these genes are expressed ubiquitously during embryogenesis and in adult tissues, suggesting that they have multiple roles, not limited to biomineralization (Supplementary Tables 21 and 22).

On the other hand, a comparison of 90 genes that are associated with shell formation in molluscs indicates that most of them are shared by bilaterians, whereas mollusc shells contain several lineage-specific proteins (Supplementary Fig. 20). In addition, transcriptome analysis of Lingula adult tissues shows that expression of the shared genes is not limited to the mantle. These results suggest that many mollusc shell formation genes have been co-opted independently in mollusc lineages, while they carry out different functions in other bilaterians (Supplementary Table 23). Notably, genes shared between Lingula and molluscs, such as calcium-dependent protein kinase and CHS, exhibit high expression in larvae and mantle, indicating that they may also be involved in Lingula shell formation (Supplementary Table 24).

Conserved molecular mechanisms in biomineralization. Given that genes associated with biomineralization have diverse functions and have been co-opted in different species, we next tested whether there is a conserved upstream mechanism for this process. We focused on one of the ancient metazoan signalling pathways, BMPs. Previous studies have demonstrated that BMP signalling plays key roles in biomineralization in both molluscs32 and vertebrates33. To explore the possible role of BMP signalling during embryogenesis, we first annotated BMP ligands and receptor-regulated Smad. Lingula has orthologues for one Bmp2/4, one Bmp5-8 and one Smad1/5/9. Our embryonic transcriptome showed that Bmp5-8 and Smad1/5/9 are expressed maternally, whereas Bmp2/4 is expressed after the early blastula stage (Supplementary Fig. 21).

To visualize activation sites of BMP signals, we employed immunostaining of nuclear phosphorylated Smad1/5/9, an activated mediator. In Lingula, embryonic shells are formed on mantle lobes beginning at the one-pair cirri larval stage12,
Interestingly, we discovered that BMP signalling is activated at the anterior margin of the mantle lobe during *Lingula* larval shell formation (Fig. 5 and Supplementary Fig. 21). This suggests that there may be a conserved mechanism for initiating biomineralization in brachiopods and molluscs. Further functional analyses will provide more rigorous testing of this hypothesis.

**SMPs and fibrillar collagens.** Our proteomic analyses of *Lingula* shells identified a total of 65 SMPs (Supplementary Figs 22 and 23). These SMPs are highly expressed in the mantle (Supplementary Fig. 24). Using comparative genomics, we showed that the composition of *Lingula* SMPs share the highest similarities with those of amphioxus and molluscs (Supplementary Fig. 25). Through an examination of amino acid composition, one of the main characteristics of *Lingula* shells compared with other articulate brachiopods or molluscs is that their SMPs contain a large amount of glycine and alanine. We provided here the first molecular evidence that glycine-rich SMPs are collagens (Supplementary Table 25). In addition, we also found that many novel SMPs are alanine-rich and have low molecular weights (that is, amino-acid length ~100–200) (Supplementary Fig. 23 and Supplementary Table 26). Pfam analysis of *Lingula* SMPs shows that the most abundant domains are adi and collagen, whereas the most abundant proteins contain von Willebrand factor type A and epidermal growth factor (EGF) domains (Supplementary Fig. 23 and Supplementary Table 27). The domain composition suggests that the shell matrix is derived from the extracellular matrix. We further examined the expression profile of these SMPs. We found that 26 SMPs are expressed ubiquitously in all adult tissues, indicating that they have functions other than shell formation (Supplementary Fig. 24). On the other hand, 20 SMPs exhibited specific expression in the mantle. These include collagen, chitinase, glutathione peroxidase, hephaestin, hemicentin and peroxidasin (Supplementary Fig. 24 and Supplementary Table 28). Many of these genes function as extracellular enzymes and ion-binding sites in humans, suggesting that they are probably co-opted in *Lingula* for shell formation. Furthermore, it has been reported that secreted acidic proteins play an important role during the calcification process in mollusc shells and coral skeletons. We failed to find secreted acidic proteins among *Lingula* SMPs (Supplementary Tables 25 and 26). Instead, we found that there are novel alanine-rich SMPs with a three-helix bundle structure that may confer elastic properties on the *Lingula* shell (Supplementary Fig. 26).

As the formation of vertebrate bone and *Lingula* shell rely on mineral deposition on fibrillar collagens, we further examined the evolution of fibrillar collagens. Vertebrate fibrillar collagens are subdivided into three major clades carrying COLFI domains (Fig. 6a,b). However, our shell proteomic analyses failed to detect fibrillar collagens with COLFI domains in *Lingula* SMPs. Instead,
further domain combination analyses show that Lingula shell-associated collagens fall into a new group with an EGF domain, which is different from collagens of vertebrate bone (Fig. 6a,b). Intriguingly, some fibrillar collagens probably originated by tandem duplication (Fig. 6c). In addition, we found that Lingula contains the highest number of proteins having both EGF and collagen domains among bilaterians (Supplementary Fig. 27 and Supplementary Tables 29 and 30). These findings suggest that EGF-domain shuffling has occurred more frequently in the Lingula lineage and may result in new types of collagens with novel domain combinations.

Figure 4 | Comparative transcriptomics and genomics reveal different origins of biomineralization-related genes. (a) Spearman’s correlation coefficient (\(\rho\)) and hierarchical clustering analyses of transcriptome data from adult tissues of the brachiopod, Lingula, and Pacific oyster, Crassostrea, in which 6,315 orthologous gene pairs were identified. An adult Lingula is shown with the dorsal shell removed and the anus opening to the right. (b) Genes involved in formation of vertebrate bone, mollusc shell and Lingula shell are compared in biomineralization-capable metazoans. Hierarchical clustering was performed in vertebrate bone formation-associated genes. Numbers of genes analysed are indicated in the parentheses. Shark, C. milii; pearl oyster, Pinctada fucata; sea snail, L. gigantea. BMP, bone morphogenetic protein; FGF, fibroblast growth factor; SCPPs, secreted calcium-binding phosphoproteins; SPARCs, secreted proteins acidic and rich in cysteine.

Figure 5 | BMP signalling may be involved in larval shell formation. (a–h) Confocal images of Lingula larvae from one-pair cirri to three-pair cirri stages. (a–d) Filamentous actin (F-actin) staining shows the cellular structure of larvae. Cytoplasmic membranes and nuclei are labelled with CellMask (grey) and DAPI (blue), respectively. Inset in c shows the cell boundary at higher magnification. (e–h) Activation of BMP signalling is monitored by nuclear signals of phosphorylated Smad1/5/9 (pSmad, red). Inset in e shows nuclear pSmad signals at higher magnification. It is noteworthy that signals are localized at the margin of mantle lobes (arrows). Orientation of embryos is indicated at the bottom-right corner of each panel. cr, cirrus (cirri); mf, muscle fiber; ml, mantle lobe; mo, mouth; sh, embryonic shell; tn, tentacle. Scale bars, 50 \(\mu\)m.

Lingula shell formation and evolution of biomineralization. Mollusc phylogenomic and shell proteomic studies show that mollusc shells may have different origins\(^8,37,38\). Although all molluscs use calcium carbonate, different modes of biomineralization have been adapted among brachiopods. Only the Linguliformea makes shells with calcium phosphate\(^1\). We have shown that Lingula used its own gene set for calcium phosphate biominerals, which is different from those used by vertebrates. Given that mineralized vertebrate bone first appeared (~450 MYA, late Ordovician)\(^31\) much later than lingulid shells (~520 MYA, early Cambrian)\(^22\), it is perhaps not surprising that...
vertebrate bone and Lingula shell have different genetic origins. Although downstream biomineralization-related genes are diverse, we speculate that the metazoan ancestor might use a core set of ancient signalling proteins, such as BMPs and their downstream mediators, to initiate the biomineralization process. We found many calcium-binding and extracellular matrix proteins in Lingula shell and mantle. Those proteins have also been reported to participate in bone and shell formation (Supplementary Table 31). This suggests that metazoan biomineralization probably originated from a calcium-regulated extracellular matrix system. Furthermore, we also discovered that Hox4, tyrosinase, CHS, per lucin, chitinase, peroxidase, mucin and von Willebrand factor type A protein are common shell formation-associated components shared by Lingula and molluscs, suggesting that this fundamental gene set was used by their common ancestor (Supplementary Table 31). There are several Lingula SMPs encoding enzymes such as glutathione peroxidase, hephaestin and hemicentin, which have no reported function in shell or bone formation. However, interestingly, hephaestin and hemicentin are found in the coral skeletal organic matrix34,39. This suggests that these extracellular ion-binding proteins in the biomineral matrix may either have been lost in vertebrate bone and mollusc shell, or that they arose independently in Lingula and corals.

In light of the close phylogenetic relationship between Lingula and molluscs, we hypothesized evolutionary scenarios for the primitive mode of biomineralization in common ancestors of brachiopods and molluscs. By comparing chemical and molecular features, we present three possible primitive modes (that is, Ca-phosphate primitive, Ca-carbonate primitive and chitin scaffold hypotheses; Supplementary Fig. 28). Given different chemistry and genetic components of their shells, we argue that the calcification process might be a derived feature in brachiopods and molluscs. Instead, chitin localized in epithelial cells may be the primitive character, predating biomineralization. A chitinous scaffold may provide the organic framework for interactions between extracellular matrix and mineral ions28. This suggests that metazoan genome has been evolving, instead of remaining static, as one would expect in a genuine ‘living fossil’. Combining transcriptomic and proteomic data, we also show that Lingula has a unique system for calcium phosphate-based biomineralization. Perhaps one of the mysteries in animal evolution is the use of calcium phosphate and fibrillar collagens in the formation of biominerals by the evolutionarily distant lingulid brachiopods and vertebrates1,4. All data presented in this study indicate that Lingula and bony vertebrates have adapted different mechanisms for hard tissue formation. Vertebrates probably evolved calcium phosphate-based biomineralization independently, by duplication
and neofunctionalization of related genes, while extensive expansion of mineralization-related gene families occurred in the *Lingula* genome.

Indeed, many examples of parallel evolution have been reported. For example, studies on collagen evolution among vertebrates and basal chordates show that three different fibrillar collagen clades occurred independently, a co-option in which collagen was used for biomineral formation of chordates. Similarly, studies of biomineralization genes in sea urchins and corals, where co-option, domain shuffling and novel genes are the fundamental mechanisms for metazoan biomineralization.

**Methods**

**Biological materials.** Gravid *L. anatina* adults were collected during July and August in Kasari Bay, Amami Island (28.440583 N 129.667608 E; Supplementary Fig. 1). Mature male gonads were dissected for genomic DNA extraction. Maturation of oocytes was induced by injection of 30 μl of 40 mM dibutyryl-cAMP in PBS into the gonad. Artificial spawning was performed by elevating the temperature to 29°C for 2–6 h, followed by cold shock back to room temperature (~25°C).

**Genome sequencing and assembly.** The *Lingula* genome was sequenced using next-generation sequencing technology with a hybrid approach involving four different platforms: Roche 454 GS FLX +, Illumina (MiSeq and HiSeq 2500) and PacBio RS II. Sequencing quality was checked with FastQC (v0.10.1). Raw Illumina reads were quality filtered and trimmed with Trimmomatic (v0.30)44. Raw mate pair reads were filtered with DeLoxer45 or NextClip (v0.8)46 depending on library preparation. Genome assembly was conducted using Newbler (v2.9) with a hybrid assembly approach using data from 454 and Illumina. First, 17 runs of a 1,750-bp library were sequenced using a Roche GS FLX +. This generated 9.6 Gb data with an average read length of 520 bp (Supplementary Figs 1 and 2, and Supplementary Table 2). Second, taking advantage of the enhancement of the read length in Illumina technology, libraries in size ranging from 500 to 620 bp were prepared and sequenced using an Illumina MiSeq. This generated 32.5 Gb of 250 bp long paired-end data (Supplementary Fig. 2 and Supplementary Table 2). To overcome repetitive regions of the genome, mate pair libraries with 1.5–3 kb lengths were prepared using the Cre-Lox recombination approach. In addition, to produce a long mate pair library, the BluePippin system was applied to prepare 5–17 kb DNA fragments and libraries were constructed using Nextera technology. The long mate pair libraries were sequenced to obtain 45.5 Gb of mate pair data using Illumina MiSeq and HiSeq 2500 platforms (Supplementary Table 2).

Finally, Illumina mate pair reads together with 8.5 Gb of PacBio extra-long reads (7–38 kb) were used for scaffolding. Scaffolding was accomplished by mapping paired-end and mate pair reads (1.5–17 kb) from Illumina using SSPACE (v3.0)49. PacBio long reads (>7 kb) were mapped to the scaffolds generated by Newbler using BLASR (v20141001)50 and upgraded scaffolds were produced with SSPACE-LongRead (v1-1)51 (Supplementary Table 2). Gaps in the scaffolds were filled using GapCloser (v1.12-16) from the SOAPdenovo package52. K-mer analysis was also conducted using Jellyfish (v2.0.0)54 and a custom Perl script. Regions of repetitive sequences were identified with RepeatScout (v1.0.5)55 and then masked with RepeatMasker (v4.0.3). The genome size was estimated by flow cytometry (Supplementary Fig. 1) as well as by K-mer analysis using SOAPec (v2.0.1) and Genomic Character Estimator (GCE; v1.0.0) from the SOAPdenovo package55. K-mer analysis was also conducted using Jellyfish (v2.0.0)54 and a custom Perl script. Completeness of the genome assembly was assessed by searching for the set of 248 core eukaryotic genes using CEGMA (v2.4.010312)55.

**Gene model prediction.** To obtain high-quality gene models, messenger RNA sequencing (RNA-seq) was performed to obtain transcript information (Supplementary Figs 2 and 4). RNA-seq data (369 million read pairs) from embryos and adult tissues were obtained using an Illumina HiSeq 2500 (Supplementary Fig. 4 and Supplementary Table 4). Transcripts assembled de novo with Trinity (v2013.08.14)56 were used as expression evidence for predicting gene models (Supplementary Fig. 2). Gene models were predicted with trained AUGUSTUS (v3.0.2) using hints from spliced alignment of transcripts to the targeted genome assembly produced with BLAT and PASA (v20130907) (Supplementary Fig. 2).
Gene family analyses. To analyze gene family evolution in lophotrochozoans, all-to-all BLASTP analysis was performed followed by Markov clustering, to identify orthologous gene groups with OrthoMCL (v2.0.9)57, according to the standard protocol using a default inflation number of 1.5. Gene family birth and death was estimated by computing the orthologous gene using CAFE (Computational Analysis of gene Family Evolution; v3.1)58. Important transcription factors and signalling components were annotated with Pfam domain search. To identify genes involved in biomineralization that are interesting topics for lineage-specific evolution, the KEGG pathway database was used. Non-synonymous (Ka) and synonymous (Ks) substitution rates of paired-wise paralogues were calculated with KaKsCalculator (v2.0)59.

Phylogenetic analyses. To identify robust phylogenetic markers, two strategies were applied. First, OrthoMCL was used to cluster orthologous gene groups from 22 selected metazoan proteomes (Supplementary Table 1; asterisks) and then orthologues with one-to-one orthologous relationships were selected for further analyses using custom Perl scripts. Second, homology searches using a bidirectional best hits (BBH) approach with BLASTP and custom Bash scripts were used to identify the best orthologous pairs among many-to-many orthologous relationships. Alignments of orthologues were performed with MAFFT (v7.130b)60. Phylogenetic trees were compared with RAxML (v8.0.4)61 and Bayesian trees were constructed with MrBayes (v3.1)62.

Proteomic analyses. SMPs were extracted from adult tissue were rehydrated using PRST (PBS with 0.1% Tween-20) for 10 min and permeabilized in PRSTX (PBS with 0.1% Triton X-100) for 30 min. Afterwards, embryos were blocked in 3% BSA in PBST for at least 1 h followed by incubation in the primary antibody, rabbit anti-phosphorylated Smad1/5/9 (1:200; Cell Signaling, #9511) or BODIPY FL phallacidin (1:50; Invitrogen, B6079) in 1% BSA in PBST at 4°C overnight. It is noteworthy that for phallacidin staining, embryos were from batches without methanol treatment. Alexa Fluor goat anti-rabbit secondary antibody (1:400; Invitrogen, A-11037) was used to visualize signals. Nuclei were counterstained with 4,6-diamidino-2-phenylindole (1:100; Dojinbio, 340–07971) and cytoplasmic membranes were labelled with CellMask Deep Red (1:200; Invitrogen, C10046). Embryos were imaged using a Zeiss LMS 780 inverted confocal system.

Immunostaining and F-actin labelling. Embryos were fixed with 4% paraformaldehyde, dehydrated in cold methanol and stored at −20°C. For immunostaining, embryos were stained with primary antibodies raised against Lophotrochozoa invertebrate Actin (1:100; Sigma, A2066) overnight at 4°C. Secondary antibodies raised against Lophotrochozoa Invertebrate Actin (1:100; Sigma, A2066) were used. The embryo was post-fixed with 1% paraformaldehyde and then paraformaldehyde, dehydrated in cold methanol and stored at −20°C.

References

1. Williams, A., Cusack, M. & Mackay, S. Collagenous chitinophosphatic shell of the brachiopod. *Lingula. Phil. Trans. R. Soc. B* 346, 223–266 (1994).
2. Sperling, E. A., Pisani, D. & Peterson, K. J. Molecular paleobiological insights into the origin of the Brachiopoda. *Evol. Dev.* 13, 290–303 (2011).
3. Williams, A., Carlson, S. J., Brunston, C. H. C., Holmer, L. E. & Popov, L. A supra-ordinal classification of the Brachiopoda. *Phil. Trans. R. Soc. B* 351, 1171–1193 (1996).
4. Knoll, A. H. Biominalization and evolutionary history. *Rev. Mineral. Geochem.* 54, 329–356 (2003).
5. Bitner, M. A. & Cohen, B. L. in *eLS* (John Wiley & Sons, Ltd, 2013).
6. Gould, S. J. & Calloway, C. R. Clams and brachiopods; ships that pass in the night. *Paleobiology* 6, 383–396 (1980).
7. Cook, P. J. & Shergold, J. H. Phosphorus, phosphorites and skeletal evolution at the Precambrian-Cambrian boundary. *Nature* 308, 231–236 (1984).
8. Jackson, D. J. et al. Parallel evolution of nacre building gene sets in molluscs. *Mol. Biol. Evol.* 27, 591–610 (2010).
9. Yatsu, N. On the development of *Lingula anatina*. *J. Coll. Sci. Imp. Univ. Tokyo* 17, 1–112 (1922).
10. Field, K. G. et al. Molecular phylogeny of the animal kingdom. *Science* 239, 748–753 (1988).
11. de Rosa, R. et al. Hox genes in brachiopods and priapulids and protostome evolution. *Nature* 399, 772–776 (1999).
12. Zhang, Z. F. et al. An early Cambrian agglutinated tubular lophophorate with brachiopod characters. *Sci. Rep.* 4, 4862 (2014).
13. Cohen, B. L. Rerooting the rDNA gene tree reveals phorontoids to be ‘brachiopods without shells’; dangers of wide taxon samples in metazoan phylogenetics (*Phoronida; Brachiopoda*). *Zool. J. Linn. Soc.* 167, 82–92 (2013).
14. Dunn, C. W. et al. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* 452, 745–749 (2008).
15. Paps, J., Baguna, J. & Riutort, M. Lophotrochozoa internal phylogeny: new insights from an up-to-date analysis of nuclear ribosomal genes. *Proc. Biol. Sci.* 276, 1245–1254 (2009).
16. Erwin, D. H. et al. The Cambrian conundrum: early divergence and later ecological success in the early history of animals. *Science* 334, 1091–1097 (2011).
17. Simakov, O. et al. Insights into bilaterian evolution from three spiralian genomes. *Nature* 493, 526–531 (2013).
18. Takeuchi, T. et al. Draft genome of the pearl oyster *Pinctada fucata*; a platform for understanding bivalve biology. *DNA Res.* 19, 117–130 (2012).
19. Zhang, G. et al. The oyster genome reveals stress adaptation and differentiation of shell formation. *Nature* 490, 49–54 (2012).
20. Williams, A. & Cusack, M. Evolution of a rhythmic lamination in the organophosphatic shells of brachiopods. *J. Struct. Biol.* 126, 227–240 (1999).
21. Cusack, M., Williams, A. & Buckman, J. O. Chemico-structural evolution of linguloid brachiopod shells. *Palaeontology* 42, 799–840 (1999).
22. Zhang, Z., Shu, D., Han, J. & Liu, J. Morpho-anatomical differences of the Early Cambrian Chengjiang and Recent lingulids and their implications. *Acta Zool. Sin.* 55(4), 277–288 (2009).
23. Ameiniya, C. T. et al. The African coelacanth genome provides insights into tetrapod evolution. *Nature* 496, 311–316 (2013).
24. Putnam, N. H. et al. The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453, 1064–1071 (2008).
25. Cao, L. et al. The genome of *Mesoblastus SUMerosus* reveals a unique adaptation in the evolution of arthropods. *Nat. Commun.* 4, 2602 (2013).
26. Zakrzewski, A. C. et al. Early divergence, broad distribution, and high diversity of animal chitin synthases. *Genome Biol. Evol.* 6, 316–325 (2014).
27. Weiss, I. M., Schonitzer, V., Eichner, N. & Sumpner, M. The chitin synthase involved in marine bivalve mollusk shell formation contains a myosin domain. *FEBS Lett.* 580, 1846–1852 (2006).
28. Schonitzer, V. & Weiss, I. The structure of mollusc larval shells formed in the presence of the chitin synthase inhibitor NIKKOMycin C. *BMC Biol.* 7, 1 (2007).
29. Cusack, M. & Freer, A. Bio-mineralization: elemental and organic influence in carbonate systems. *Chem. Rev.* **108**, 4433–4454 (2008).
30. Marín, V., Luquet, G., Marie, B. & Medakovic, D. Molluscan shell proteins: primary structure, origin, and evolution. *Curr. Top. Dev. Biol.* **80**, 209–276 (2008).
31. Venkatesh, B. et al. Elephant shark genome provides unique insights into gnathostome evolution. *Nature* **505**, 174–179 (2014).
32. Shimizu, K., Sarashina, I., Kagi, H. & Endo, K. Possible functions of Dpp in gastropod shell formation and shell coiling. *Dev. Genes Evol.* **221**, 59–68 (2011).
33. Chen, G., Deng, C. & Li, Y. P. TGF-beta and BMP signaling in osteoblast differentiation and bone formation. *Int. J. Biol. Sci.* **8**, 272–288 (2012).
34. Ramos-Silva, P. et al. The skeletal proteome of the coral *Acropora millepora*: the evolution of calcification by co-option and domain shuffling. *Mol. Biol. Evol.* **30**, 2099–2112 (2013).
35. Nair, A. K., Gautieri, A., Chang, S.-W. & Buehler, M. J. Molecular mechanics of mineralized collagen fibrils in bone. *Nat. Commun.* **4**, 1724 (2013).
36. Kawasaki, K., Buchanan, A. V. & Weiss, K. M. Biomineralization in humans: making the hard choices in life. *Annu. Rev. Genet.* **43**, 119–142 (2009).
37. Smith, S. A. et al. Resolving the evolutionary relationships of molluscs with Biologigraphs. *Nature* **480**, 364–367 (2011).
38. Koott, K. M. et al. Phylegromics reveals deep molluscan relationships. *Nature* **477**, 452–456 (2011).
39. Drake, J. L. et al. Proteomic analysis of skeletal organic matrix from the stony coral *Stylophora pistillata*. *Proc. Natl Acad. Sci. USA* **110**, 3788–3793 (2013).
40. Wada, H., Okuyama, M., Satoh, N. & Zhang, S. Molecular evolution of fibrillar collagen in chordates, with implications for the evolution of vertebrate skeletons and chordate phylogeny. *Dev. Biol.* **8**, 370–377 (2006).
41. Livingston, B. T. et al. A genome-wide analysis of biomineralization-related proteins in the sea urchin *Stronglylocentrotus purpuratus*. *Dev. Biol.* **300**, 335–348 (2006).
42. Nishizawa, A., Sarashina, I., Tsujimoto, Y., Iijima, M. & Endo, K. Artificial fertilization, early development and chromosomal numbers in the brachiopod *Lingula anatina*. *Palaeontologia* **84**, 1–8 (2010).
43. Tagawa, K., Nishino, A., Humphreys, T. & Satoh, N. The spawning and early development of the Hawaiian acorn worm (hemichordate) *Pychoderia flava*. *Zool. Sci.* **15**, 85–91 (1998).
44. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
45. Cheng, G., Deng, C. & Li, Y. P. TGF-beta and BMP signaling in osteoblast differentiation and bone formation. *Int. J. Biol. Sci.* **8**, 272–288 (2012).
46. Liu, A. K., Gautieri, A., Chang, S.-W. & Buehler, M. J. Molecular mechanics of mineralized collagen fibrils in bone. *Nat. Commun* **4**, 1724 (2013).
47. Shinzato, C. *et al.* Resolving the evolutionary relationships of molluscs with Biologigraphs. *Nature* **480**, 364–367 (2011).
48. Koott, K. M. et al. Phylegromics reveals deep molluscan relationships. *Nature* **477**, 452–456 (2011).
49. Drake, J. L. et al. Proteomic analysis of skeletal organic matrix from the stony coral *Stylophora pistillata*. *Proc. Natl Acad. Sci. USA* **110**, 3788–3793 (2013).
50. Wada, H., Okuyama, M., Satoh, N. & Zhang, S. Molecular evolution of fibrillar collagen in chordates, with implications for the evolution of vertebrate skeletons and chordate phylogeny. *Dev. Biol.* **8**, 370–377 (2006).
51. Livingston, B. T. et al. A genome-wide analysis of biomineralization-related proteins in the sea urchin *Stronglylocentrotus purpuratus*. *Dev. Biol.* **300**, 335–348 (2006).
52. Nishizawa, A., Sarashina, I., Tsujimoto, Y., Iijima, M. & Endo, K. Artificial fertilization, early development and chromosomal numbers in the brachiopod *Lingula anatina*. *Palaeontologia* **84**, 1–8 (2010).
53. Tagawa, K., Nishino, A., Humphreys, T. & Satoh, N. The spawning and early development of the Hawaiian acorn worm (hemichordate) *Pychoderia flava*. *Zool. Sci.* **15**, 85–91 (1998).
54. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
55. Van Nieuwerburgh, F. *et al.* Illumina mate-paired DNA sequencing-library preparation using Cre-Lox recombination. *Nucleic Acids Res.* **40**, e24 (2012).
56. Leggett, R. M., Clavijo, B. J., Clissold, L., Clark, M. D. & Caccamo, M. NextClip: an analysis and read preparation tool for Nextera Long Mate Pair libraries. *Bioinformatics* **30**, 566–568 (2014).
57. Shimazono, C. *et al.* Using the *Acropora digitifera* genome to understand coral responses to environmental change. *Nature* **476**, 320–332 (2011).
58. Caruccio, N. Preparation of next-generation sequencing libraries using Nextera technology: simultaneous DNA fragmentation and adaptor tagging by in vitro transcription. *Methods Mol. Biol.* **733**, 241–253 (2011).
59. Bower, M., Henkel, C. V., Jansen, H. I., Butler, D. & Pirovano, W. W. Scaffolding pre-assembled contigs using SSAPSe. *Bioinformatics* **27**, 578–579 (2011).
60. Chaisson, M. J. & Tesler, G. Mapping single molecule sequencing reads using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
61. Boettiger, M. & Pirovano, W. W. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics* **15**, 211 (2014).
62. Luo, R. *et al.* SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**, 18 (2012).
63. Price, A. L., Jones, N. C. & Pevzner, P. A. De novo identification of repeat families in large genomes. *Bioinformatics* **21**, 351–358 (2005).
64. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
65. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
66. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
67. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
68. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
69. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
70. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).