Okinawa Institute of Science and Technology
Graduate University

Thesis submitted for the degree
Doctorate of Philosophy

Insights into lophotrochozoan evolution and the origin of morphological novelties from brachiopod, phoronid, and nemertean genomes

by

Yi-Jyun Luo

Supervisor: Noriyuki Satoh

March, 2017

Declaration of Original and Sole Authorship

I, Yi-Jyun Luo, declare that this thesis entitled "Insights into lophotrochozoan evolution and the origin of morphological novelties from brachiopod, phoronid, and nemertean genomes" and the data presented in it are original and my own work.

I confirm that:

- No part of this work has previously been submitted for a degree at this or any other university.
- References to the work of others have been clearly acknowledged. Quotations from the work of others have been clearly indicated, and attributed to them.
- In cases where others have contributed to part of this work, such contribution has been clearly acknowledged and distinguished from my own work.
- None of this work has been previously published elsewhere, with the exception of the following: (*Corresponding author)

  Luo YJ*, Takeuchi T, Koyanagi R, Yamada L, Kanda M, Khalturina M, Fujie M, Yamasaki S, Endo K, Satoh N* (2015) The *Lingula* genome provides insights into brachiopod evolution and the origin of phosphate biomineralization. *Nature Communications* 6, 8301.

  Luo YJ*, Satoh N, Endo K (2015) Mitochondrial gene order variation in the brachiopod *Lingula anatina* and its implications for mitochondrial evolution in lophotrochozoans. *Marine Genomics* 24, 31-40.

Signature: 

Date: March 23, 2017

Insights into lophotrochozoan evolution and the origin of morphological novelties from brachiopod, phoronid, and nemertean genomes

by
Yi-Jyun Luo

Submitted to the Graduate School in partial fulfillment of the requirements for the degree of Doctorate of Philosophy at the Okinawa Institute of Science and Technology Graduate University

Abstract

Brachiopods, phoronids, and nemerteans are closely related lophotrochozoans, yet they carry distinct feeding apparatuses and lifestyles. They are poorly studied despite their importance in ecology, evolution, and paleontology. As a result, the genetic basis of their evolutionary origins and body plans have been obscure. Since the Cambrian explosion ~540 million years ago, animal forms have greatly diversified. One fundamental question of animal evolution is how these diverse morphologies are formed. While animals share many developmental toolkit genes, they also possess novel genes and expansion of gene families in a lineage-specific manner. How lineage-specific genes and changes of genomic features contribute to morphological novelties is still a challenge in understanding animal evolution. Also, whether common toolkit genes are involved in patterning these novelties at the genomic level is not well understood. Here I present the genomes of the brachiopod *Lingula anatina*, the phoronid *Phoronis australis*, and the nemertean *Notospermus geniculatus*, together with multiple transcriptomes, providing a comparative platform to understand the evolution of animal genomes and the origin of lophotrochozoans.

Using genomic, transcriptomic, and proteomic approaches, I show that although *Lingula* and vertebrates have superficially similar hard tissue components, *Lingula* lacks genes involved in bone formation, suggesting 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 epidermal growth factor 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. Genome-based phylogenetic analyses place Nemertea sister to the group of Brachiopoda and Phoronida. Lophotrochozoans share many gene families with deuterostomes, suggesting that lophotrochozoans retain a core set of bilaterian gene repertoire rather than ecdysozoans or remaining spiralians. Comparative transcriptomics demonstrates that lophophores of brachiopods and phoronids have resemblance not only morphologically but also at the molecular level. Despite lophophores are dissimilar from head structures, lophophores highly express vertebrate head organizer and neuronal marker genes, probably indicating a common origin of bilaterian head patterning. Together, this study reveals a dual nature of lophotrochozoans in which bilaterian-conserved and lineage-specific features shape the evolution of their genomes.

Acknowledgments

I am indebted to my supervisor, Noriyuki Satoh, for providing me excellent resources, absolute freedom, and trust in my scientific ability. In retrospect, those are essential factors that allow me to pursue my graduate training in a whole new field—bioinformatics and genomics of marine invertebrates. Without his insights and his emphasis on a creative environment, I would not have learned so such along this genomic adventure.

I thank my collaborator, Kazuyoshi Endo, for bringing me into the *Lingula* study and for teaching me sampling and artificial spawning. I also thank my collaborators, Takeshi Takeuchi and Lixy Yamada, for their help in proteomic analysis. Thanks also to the collaborators, Tadashi Akiyama, Hirotaka Sakamoto, and Tatsuya Sakamoto, for sampling phoronid and nemertean specimens.

I would like to thank the technical staff, Ryo Koyanagi, Miyuki Kanda, Makiko Tanaka, Mariia Khalturina, Manabu Fujie, and Shinichi Yamasaki, for their assistance in DNA and RNA sequencing. Thanks to Kanako Hisata for preparing genome browsers.

I appreciate the support and stimulating discussion from the Satoh lab members. Mainly thanks to Chuya Shinzato, Yuuri Yasuoka, Kostya Khalturina, Eiichi Shoguchi, Jun Inoue, and Tsai-Ming Lu.

I thank my classmates, roommates, and labmates, Ken Baughman and Keita Ikegami, for their continued friendship as well as sharing our experiences in science and daily life as the pioneer class at OIST.

During my graduate work, I am grateful to have the discussion with Dan Rokhsar, Chris Lowe, Bernie Degnan, Peter Holland, Mark Martindale, and Andy Hejnol. Thanks for their suggestions in the specific studies as well as advice in the scientific career in general.

I wish to thank the administrative staff, Tomomi Teruya and Shoko Yamakawa, for their assistance in daily life and lab-related matters. I thank Steve Aird for English editing and fruitful discussion on scientific writing. Thanks to Jeff Wickens for providing constructive feedback on the first draft.

List of Abbreviations

AP	anterior–posterior
BBH	bidirectional best hits
BLAST	basic local alignment search tool
BMP	bone morphogenetic protein
BSA	bovine serum albumin
CHS	chitin synthase
DAPI	4’,6-diamidino-2-phenylindole
EGF	epidermal growth factor
EST	expressed sequence tag
FACS	fluorescence-activated cell sorting
FGF	fibroblast growth factor
FPKM	fragments per kilobase of transcript per million mapped reads
GAG	glucosaminoglycan gel
GC	guanine-cytosine
GO	gene ontology
GRN	gene regulatory network
LRR	leucine-rich repeats
LSU	large subunit
MHD	myosin-head-domain
ORF	open reading frame
PBS	phosphate-buffered saline
SCPP	secreted calcium-binding phosphoproteins.
SMP	shell matrix protein
SPARC	secreted proteins acidic and rich in cysteine
SSU	small subunit
TGF	transforming growth factor
TIR	toll/interleukin-1 receptor
TLR	toll-like receptor
TMM	trimmed mean of M-values
URF	unassigned open reading frame

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Animal phylogeny

Animals (or metazoans) are a clade of multicellular eukaryotic organisms. Dissimilar to other multicellular eukaryotes, such as plants and fungi, animals lack the plastids and cell walls. While some animals can undergo photosynthesis through sequestering plastids from their algal food (Rumpho et al., 2011), almost all animals are heterotrophic with a few exception. Animals are also motile in most cases, possessing muscle tissues to rapidly respond to the environment. Phylogenetically, animals are closely related to fungi, having emerged as a clade sister to choanoflagellates (King et al., 2008).

Although the positions of early branching animals are under debate (Moroz et al., 2014; Pisani et al., 2015; Ryan et al., 2013; Simion et al., 2017; Whelan et al., 2015), animals can be classified into five monophyletic groups (Dunn et al., 2014; Telford et al., 2015). These groups include (1) Bilateria (humans, flies, and snails); (2) Cnidaria (corals and jellyfish); (3) Placozoa (*Trichoplax*); (4) Porifera (sponges); and (5) Ctenophora (comb jellies). Bilaterians are bilaterally symmetrical animals with three germ layers (ectoderm, mesoderm, and endoderm), which include deuterostomes (ambulacrarians and chordates), protostomes (ecdysozoans and spiralians), and perhaps xenacoelomorphs (Cannon et al., 2016). On the other hand, cnidarians are radially symmetrical animals with two germ layers (ectoderm and endoderm), including hydrozoans (hydra), scyphozoans (jellyfish), cubozoans (box jellies), and anthozoans (corals and sea anemones). Despite beging highly controversial regarding relationships of the clades such as Porifera and Ctenophora, the sister-group relationship of Bilateria and Cnidaria are in a broad consensus (Fig. 1.1a). This hypothesis raises several interesting discussions on the origins of bilateral symmetrical body plan and mesoderm by molecular comparison between bilaterians and cnidarians (Finnerty et al., 2004; Hejnol and Martindale, 2008; Martindale, 2005; Martindale and Hejnol, 2009; Martindale et al., 2004).

Understanding the relationships among animals provides the basic framework for a practical comparison of diverse morphological and genetic features. Only with the knowledge of the history of evolutionary relationships of animals, morphological homology (sameness from shared ancestry) and homoplasy (sameness from independent paths) can be distinguished. The comparative analysis of phenotypic and genomic features based on a well-resolved animal phylogeny provides valuable hypotheses for the origin of morphological

novelties, such as body plan (Lowe et al., 2015), nervous system (Hejnol and Lowe, 2015), and biomineralization (Kocot et al., 2016a).

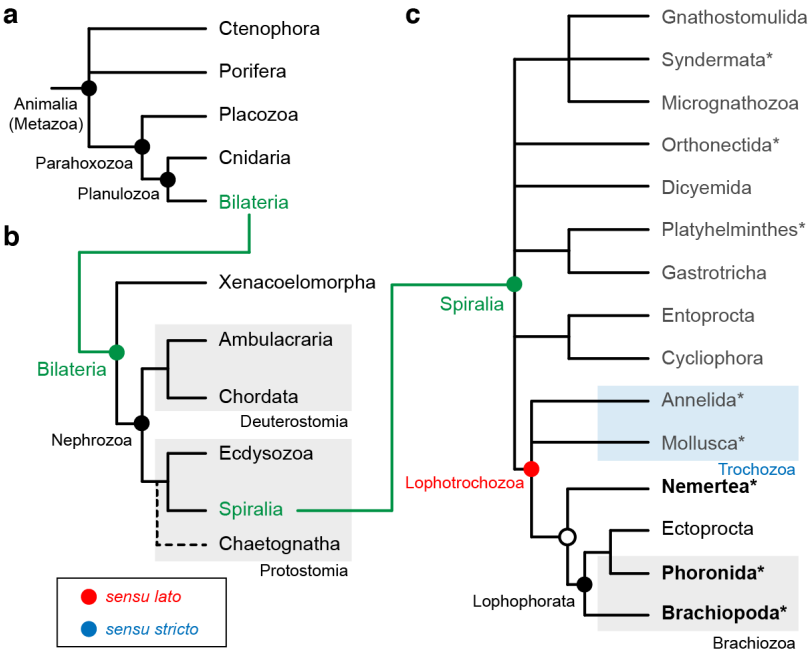


Figure 1.1 | Animal phylogeny with special emphasis on Lophotrochozoa.
(**a**) The current hypothesis of animal phylogeny with five major clades, according to multiple recent genomic and transcriptomic studies. (**b**) Bilaterian phylogeny. Dashed line denotes the uncertain position of Chaetognatha in Protostomia. (**c**) Spiralian phylogeny with a broad definition of Lophotrochozoa. It includes Nemertea and Ectoprocta, where Lophophorata is monophyletic. By contrast, Brachiozoa and Trochozoa (*sensus stricto*) are polyphyletic. Taxa with asterisks represent the availability of genome data. Taxa with genomes newly published in this thesis are in bold. The open circle denotes a clade without formal taxonomic classification. Black circles indicate taxonomic hypotheses.

1.2 Lophotrochozoan relationships

Spiralians (animals that mostly exhibit spiral cleavage) and ecdysozoans (animals that shed their exoskeleton) are sister groups within protostomes (Fig. 1.1b). As a subgroup within spiralians, most lophotrochozoans possess either lophophore or trochophore larvae during the planktonic stage. Lophotrochozoans represent more than one-third of known marine animals and play important ecological roles (Appeltans et al., 2012). Most commonly, lophotrochozoans include annelids (segmented worms), molluscs (snails, octopuses, oysters, and others), nemerteans (ribbon worms), phoronids (horseshoe worms), ectoprocts (bryozoans, or moss animals), and brachiopods (lamp shells). However, many phylogenetic

relationships within the group remain unresolved (Dunn et al., 2008; Laumer et al., 2015; Nesnidal et al., 2013) (Fig. 1.1c).

In fact, the definition of Lophotrochozoa is not in common agreement, although there is no objection to the idea that they are protostomes within Spiralia. Partly due to the unstable phylogenetic positions, different definitions are proposed based on interpretation of various datasets (Halanych, 2016; Kocot, 2016). The original definition of Lophotrochozoa is the animal group including Annelida, Mollusca, Phoronida, Ectoprocta (Bryozoa) and Brachiopoda, where Nemertea is not included (Halanych et al., 1995). Recent phylogenomic studies, however, place Nemertea close to the group of Phoronida and Brachiopoda (Dunn et al., 2008; Hausdorf et al., 2010; Hejnol et al., 2009; Helmkampf et al., 2008). Given relatively early branching of Annelida and Mollusca, the group without trochophore larvae is proposed to be Kryptrochozoa, excluding Ectoprocta (Giribet, 2009).

Because of the morphological disparity of Lophotrochozoa, it has been obscure to define the nomenclature of the clade with their synapomorphies (Nielsen, 2002). As a result, two main definitions are currently used to describe the group Lophotrochozoa depending on *sensu lato* or *sensu stricto* definition of Trochozoa (Annelida and Mollusca). To facilitate the discussion, in this thesis, I defined the nomenclature based on previous studies and my findings as follows (Fig. 1.1c).

Spiralia: Lophotrochozoa, Entoprocta, Cycliohpora, Platyhelminthes, Gastrotricha, Syndemata, Gnathostomulida, Micrognathozoa, Dicyemida and Orthonectida.
Lophotrochozoa (*sensu lato*): Annelida, Mollusca, Nemertea, Phoronida, Ectoprocta and Brachiopoda.
Kryptrochozoa: Nemertea, Phoronida and Brachiopoda.
Lophophorata: Phoronida, Ectoprocta and Brachiopoda.
Brachiozoa: Phoronida and Brachiopoda.

1.3 Lophotrochozoan phylogeny and debates

Historically, phoronids, ectoprocts and brachiopods are grouped into lophophorates based on their morphological similarity in possessing lophophores defined by Hyman (Hyman, 1959). Phoronids and brachiopods are later grouped into deuterostomes because they have features similar to pterobranch hemichordates (Nielsen, 2001). By contrast, brachiopods are placed within protostomes according to the analyses of 18S rDNA (Halanych et al., 1995) and Hox genes (de Rosa et al., 1999). Based on molecular studies, Lophophorata is paraphyletic when

analyzing with ribosomal DNAs (Halanych et al., 1995; Passamaneck and Halanych, 2006). However, EST (expressed sequence tag) and RNA-seq data sets recover the monophyly of Lophophorata (Laumer et al., 2015; Nesnidal et al., 2013). The phylogenetic positions of Nemertea and Lophophorata are largely unresolved.

Within Lophophorata, the phylum Brachiopoda comprises of three major subphyla, Linguliformea, Craniiformea, and Rhynchonelliformea (Sperling et al., 2011), the former including Lingulida and some other orders. Brachiopod phylogenetic position based on molecular phylogeny is still controversial. For example, whether Brachiopoda is monophyletic or paraphyletic is under debate. Analyses of small subunit (SSU) and large subunit (LSU) rRNA sequences from 12 and 21 taxa, respectively, suggest that phoronids are shell-less brachiopods, which are grouped into Inarticulata (i.e. shells have no hinge). Phoronids and inarticulate brachiopods are combined together to form a sister group to Articulata (including brachiopods with calcium carbonate shells) (Cohen, 2013; Cohen and Weydmann, 2005). By contrast, analysis of 7 nuclear housekeeping genes, 3 ribosomal genes, and specific microRNAs suggests that Brachiopoda is a monophyletic group and a sister group to Phoronida (Sperling et al., 2011).

In recent large-scale molecular phylogenetic studies, although Brachiopoda and Phoronida are proposed as sister groups, all studies have used only one brachiopod species, which may yield unresolved results (Dunn et al., 2008; Helmkampf et al., 2008). Therefore, it is still an open question whether Brachiopoda is a monophyletic group. Moreover, another issue that needs to be addressed is the relationship between Brachiopoda and other lophotrochozoan phyla, including Phoronida, Nemertea, Mollusca, and Annelida. Furthermore, whether Brachiopoda and Phoronida are grouped with Ectoprocta by the so-called lophophorate hypothesis, is also unclear (Hausdorf et al., 2010; Nesnidal et al., 2013).

The first comprehensive study addressing these issues, including 168 taxa shows that Brachiopoda (using *Terebratalia*; belonging to Rhrynchonelliformea) and Nemertea are closely related groups (Bourlat et al., 2008). However, in that study, the interpretation of the relationship of brachiopods to other phyla may be problematic, since Mollusca became paraphyletic, which contradicts current understanding (Kocot et al., 2011; Smith et al., 2011). Further studies based on broad sampling proposed that Brachiopoda, Phoronida, and Nemertea are supraphyletic taxa called ‘Kryptrochozoa’ (Dunn et al., 2008; Giribet, 2009; Hausdorf et al., 2010; Helmkampf et al., 2008), but the bootstrap value to support this classification (< 70%) may not be solid enough to exclude other possibilities. Recently, large-scale transcriptome analyses including data from Platyzoa (Struck et al., 2014) and Nemertea

(Andrade et al., 2014) showed that the phylogenetic position of Nemertea is unstable. As a result, the only consistency among these studies is that brachiopods are always grouped with phoronids, which confirms the previously proposed clade Brachiozoa (i.e. Brachipoda + Phoronida) (Cavalier-Smith, 1998). In opposition to the idea of 'Kryptrochozoa,' an analysis based on 11 protein coding genes and 2 ribosomal RNA genes from 96 taxa showed that the sister group of Brachiozoa is Mollusca but not Nemertea (Paps et al., 2009a). In agreement with this, analyses of SSU and LSU from 22 taxa showed similar results, suggesting Nemertea is not close to Brachiozoa (Paps et al., 2009b). In addition, a close relationship between Brachiopoda and Mollusca was supported by a large scale analysis using a 1,487 gene-matrix and a broader sampling with 113 taxa (Erwin et al., 2011; Hejnol et al., 2009). Accordingly, there is still an unresolved phylogenetic issue with brachiopods, phoroinds, and nemerteans.

1.4 Marine invertebrate genomics

Given that our understanding of animal evolution mostly relies on the studies in model systems, such as vertebrates (mice, frogs, and zebrafish), fruit flies, and nematodes, our knowledge may be biased because of uneven sampling in deuterostomes and protostomes. It is particularly the case when one tries to understand the evolution of bilaterians and vertebrates from the perspective of flies and nematodes. It is perhaps no doubt that in most cases, core mechanisms of life at the cellular level, such as transcription, cell cycle, and autophagy, are shared by most eukaryotes. However, understanding the origin of lineage-specific features at the tissue or organ level cannot be achieved without broadly studying other animal groups (Dunn and Ryan, 2015; Hejnol and Pang, 2016).

In this context, sequencing animal genomes that are the outgroup to model systems can provide insights into the mechanisms of animal evolution and the origin of morphological novelties (Canestro et al., 2007). In particular, marine invertebrates that belong to basal deuterostomes such as tunicates, amphioxus, sea urchins, and hemichordates offer a key perspective to understand the origin of deuterostomic features (Lowe et al., 2015). Within chordates, for example, after the release of the draft human genome in 2001, the genome of the tunicate *Ciona* allows us to explore chordate evolution by comparing gene contents between humans and tunicates (Dehal et al., 2002). One the other hand, however, the genome of the pelagic tunicate *Oikopleura* is compact and has lost ancestral genomic organization, showing an example of a disconnection between the genome architecture and morphological features (Denoeud et al., 2010). Thus, the *Oikopleura* genome reveals an example of a high

level of genome plasticity in some chordate lineages. In addition, the amphioxus genome provides evidence to support two rounds of whole-genome duplication in the vertebrate lineage by comparing conserved syntenies between amphioxus scaffolds and human chromosomes (Putnam et al., 2008). Outside of chordates, the sea urchin genome extends our knowledge of gene families that are involved in cell physiology and developmental processes in a deuterostome context (Sodergren et al., 2006). Recently, the hemichordate genomes allow exploring the origin of deuterostome novelties, such as the existence of NK gene cluster for the development of pharyngeal gill slits (Simakov et al., 2015).

Furthermore, exploring genome features in other major animal clades provide insights into the origin of animals. For instance, the sea anemone genome reveals surprising conservation of a core toolkit gene repertoire in Planulozoa (Bilateria + Cnidaria), suggesting that ecdysozoans might lose many eumetazoan (all animals except ctenophores) genes (Putnam et al., 2007) (Fig. 1.1a). Interestingly, although with simple body plan — only two epithelial layers and without body axes, the placozoan genome contains signaling pathway and transcription factor genes that are required for complex developmental patterning and cell-type specification (Srivastava et al., 2008). Together, these genomic studies in basal animals demonstrate the genomic complexity in the parahoxozoan ancestor (Planulozoa + Placozoa) (Fig. 1.1a). In fact, although in a simpler content, many of these genes are also present in the sponge genome, suggesting the emergence of genomic complexity with the origin of metazoan multicellularity (Srivastava et al., 2010). However, recent studies of the ctenophore genomes indicate that ctenophores are possibly the earliest splitting animals (Ryan et al., 2013) and had evolved neural systems independently from other animals (Moroz et al., 2014). Nevertheless, sequencing genomes belonged to phylogenetically important phyla not only broadens our knowledge about how animals evolved, but also provides valuable resources for further studies in comparative genomics and developmental biology.

In contrast to genomic studies of deuterostomes, ecdysozoans, and early branching metazoans, the spiralian genomes are relatively at the lack of research. Within spiralians, recent genomic studies mainly focus on platyhelminthes (e.g., blood flukes and planarians) (Fig. 1.1c). This is perhaps because they are closely related to human health and biomedical research. On the other hand, the earliest sequencing effort for a lophotrochozoan genome is the gastropod *Aplysia dactylomela*, largely for its advantages in studying neuroscience. However, due to its big genome size (~1.8 Gb) and high repetitive sequences (~30%), technical issues have postponed publication of the *Aplysia* genome (Moroz, 2011).

Table 1.1 | Genome assembly statistics of selected spiralians

Common name	Species name	Assembly statistics							CEGMA[b] (%)		Reference
		Ver.[a]	Size (Mb)	Methods	Depth	Assembler	Contig N50	Scaffold N50	C	P	
Brachiopod	*Lingula anatina*	2.0	406	454, MiSeq, HiSeq, PacBio	~226x	Newbler	55 kb	460 kb	83	98	This study
Phoronid	*Phoronis australis*	2.0	498	454, MiSeq, HiSeq	~227x	Platanus	64 kb	655 kb	55	91	This study
Nemertean	*Notospermus geniculatus*	2.0	859	MiSeq, HiSeq	~265x	Platanus	20 kb	239 kb	81	96	This study
Octopus	*Octopus bimaculoides*	2.0	2,372	HiSeq	~154x	Meraculous	6.5 kb	466 kb	33	78	(Albertin et al., 2015)
Limpet	*Lottia gigantea*	1.0	348	Sanger	~9x	JAZZ	96 kb	1.9 Mb	86	98	(Simakov et al., 2013)
Pacific oyster	*Crassostrea gigas*	9.0	558	HiSeq	~155x	SOAPdenovo	33 kb	402 kb	78	95	(Zhang et al., 2012)
Pearl oyster	*Pinctada fucata*	2.0	815	454, GAIIx	~40x	Newbler	22 kb	167 kb	55	94	(Takeuchi et al., 2012)
Polychaete	*Capitella teleta*	1.0	324	Sanger	~8x	JAZZ	22 kb	188 kb	94	97	(Simakov et al., 2013)
Leech	*Helobdella robusta*	1.0	228	Sanger	~8x	JAZZ	52 kb	3.1 Mb	88	96	(Simakov et al., 2013)
Planarian	*Schmidtea mediterranea*	4.0	787	NA	NA	NA	12 kb	80 kb	75	85	(Robb et al., 2015)
Flatworm	*Macrostomum lignano*	ML2	1,040	PacBio	~130x	Celera Assembler	37 kb	NA	90	99	(Wasik et al., 2015)
Blood fluke	*Schistosoma mansoni*	2.0	365	Sanger	~6x	Phusion	77 kb	32 Mb	55	80	(Berriman et al., 2009)
Tapeworm	*Echinococcus multilocularis*	2.0	114	Sanger, 454, HiSeq	~400x	Newbler, Arachne	4.9 Mb	4.9 Mb	87	90	(Tsai et al., 2013)
Rotifer	*Adineta vaga*	2.0	218	454, HiSeq	~465x	MIRA	98 kb	260 kb	98	99	(Flot et al., 2013)

Genomes newly published in this thesis are in bold. [a]Version of the genome assembly at publication. An updated *Lingula* genome is shown here. [b]Completeness of genome assembly is assessed with Core Eukaryotic Genes Mapping Approach (CEGMA) analysis with the percentage of presented complete (C) or partial gene models (P). NA, not available. Lophotrochozoans and other spiralians are separated by a horizontal dashed line.

Until quite recently, the first lophotrochozoan genome was published in 2012, i.e. a pearl oyster genome in the effort to provide an understanding of bivalve biology (Takeuchi et al., 2012). A Pacific oyster genome was released later the same year for understanding the mechanisms of stress adaptation and shell formation in oysters (Zhang et al., 2012). The more comprehensive comparative genomic analysis of lophotrochozoans to other animals was conducted by sequencing the genomes of one mollusc and two annelids (Simakov et al., 2013). This study reveals an overall similarity in gene family size and genomic organization that is shared by mollusc and annelid genomes with deuterostome and basal metazoan genomes (Simakov et al., 2013). Thus, their findings reflect the conserved pan-bilaterian background of genome evolution and raise the question whether molluscs and annelids represent the common feature of other lophotrochozoans. Furthermore, detailed analysis of the leech genome suggests that lineage-specific features, such as disorganized Hox cluster and multiple duplications of Hox genes, also play a role in shaping lophotrochozoan genome evolution. The lineage-specific context is especially the case when looking at the octopus genome, where extensive expansion of protocadherins and zinc-finger transcription factors contributes to their morphological novelties, such as their elaborate nervous system and suckers on their arms (Albertin et al., 2015).

1.5 Deep homology and gene regulatory networks

A major branch of evolutionary developmental biology in zoology searches for the deep homology that behinds the morphological diversity (Shubin et al., 2009), i.e. structures that are considered non-homologous based on traditional morphological criteria may share conserved expression of homologous genes. The idea originated from that all animals evolved from a common ancestor. Although individual phyla may have lineage-specific gene gain and loss, most animals share conserved developmental toolkits and patterning systems for building their body plans (De Robertis, 2008). According to this concept, morphological novelties are not a *de novo* innovation in specific animal lineages, but instead, those novelties are the outcomes through shared common ancestry of using homologous toolkit genes. In most cases, these toolkits refer to signaling pathway and transcription factor genes. The expression and interaction of these genes at the given time and space constitute the gene regulatory networks (GRN) that ultimately regulate specific cell states (Davidson, 2010).

The key concept of deep homology is thus that novel structures arose by the modification of pre-existing GRN derived from the common ancestor deep in the early animal evolution (Erwin and Davidson, 2002). The deep homology often reflects conserved

regulatory sub-circuits located near the top of developmental GRN hierarchy, which are extremely stable and located upstream in the GRN (Erwin and Davidson, 2009). These sub-circuits are so-called GRN kernels, which are required for spatial field specification. In fact, the inflexible regulatory control for certain cell types forms the foundation for the independent evolution of diverse structures. Therefore, the cell-type specification mechanisms by certain common toolkit genes are unexpectedly highly conserved in a broad range of animal phyla and play important roles in controlling animal body plan (Peter and Davidson, 2011). One such example is that *pax6* is required for the development of eyes or light sensing organs across animal phyla (Gehring, 2005).

However, conserved molecular modules at the GRN level might have evolved independently by redeployment to distinct functions in the cell types that have a different evolutionary history (Arendt et al., 2016). Molecular similarities therefore not necessarily reflect phenotypic homology. For example, retinal development sub-circuit composed of *Pax6*, *Six1/2*, *Eya*, and *Dach* are well known for its role in controlling eye development. Unexpectedly, this sub-circuit is redeployed for cell migration in sea urchin embryos (Martik and McClay, 2015). As such, caution would be needed when considering deep homology solely based on the comparative approach on several sub-circuit components.

1.6 Hox genes and morphological novelties

One famous example of deep homology toolkits is the Hox genes (Carroll, 2008; De Robertis, 2008; Duboule, 2007). Hox genes are homeodomain-containing transcription factors, playing an important role in regulating anteroposterior body axis and appendage development (Pearson et al., 2005). They are highly conserved among animals, usually with a fixed gene order on the chromosome and a segmented expression pattern according to its physical location in the genome. This property is so-called ‘colinearity’ (Pearson et al., 2005). However, the evolution of Hox genes also shows some lineage-specific features. For example, the posterior Hox genes in ecdysozoans and lophotrochozoans are found to be distinct from deuterostomes (de Rosa et al., 1999). The posterior Hox genes are highly duplicated in amphioxus (Ferrier et al., 2000) and ambulacrarians (hemichordates + echinoderms) (Freeman et al., 2012), suggesting the term so-called ‘posterior flexibility’ of Hox gene evolution in the deuterostome lineage.

Further genomic studies have shown that structural organizations of Hox clusters are quite diverse in different animals. Instead of being organized, such as in vertebrates, Hox clusters are split in fruit flies (Von Allmen et al., 1996). In addition, Hox clusters are

disorganized in the sea urchin (Sodergren et al., 2006) and atomized in the tunicate *Oikopleura* (Seo et al., 2004). Recent studies have shown that the Hox cluster is surprisingly conserved in bilaterians, suggesting that a single 11-gene Hox cluster is present in the last lophotrochozoan common ancestor (Simakov et al., 2013). This is mainly evident because the beetle *Tribolium* (Richards et al., 2008) and the limpet *Lottia* (Simakov et al., 2013) have a complete Hox cluster. However, fragmented Hox gene clusters have also been reported in many lophotrochozoans, such as the leech *Helobdella,* the polychaete *Capitella* (Simakov et al., 2013), the Pacific oyster *Crassostrea* (Zhang et al., 2012), and the pearl oyster *Pinctada* (Takeuchi et al., 2016). In particular, the *Helobdella* Hox cluster is extensively fragmented and has multiple duplications. The Hox genes, *proboscipedia* and *post1*, are also lost in *Helobdella* (Simakov et al., 2013). Recently, Hox genes are also found to be completely atomized in the octopus genome (Albertin et al., 2015). It is tempting to speculate that lophotrochozoans might experience less selective pressure to keep the intact Hox cluster due to their unique body plan.

Indeed, Hox genes have been co-opted in different squid tissues, such as the brachial crown, funnel tube, and stellate ganglia, reflecting that recruitment of Hox genes contributes to cephalopod morphological novelties (Lee et al., 2003). Similar cases are also found in brachiopods, where new expression pattern of Hox genes are involved in chaeta and shell formation (Schiemann et al., 2017). These findings suggest that the evolution of Hox genes in the lophotrochozoan lineage is more flexible than that of deuterostomes and ecdysozoans. In lophotrochozoans, Hox genes had experienced extensive genomic reorganization and co-opted with novel expression patterns.

1.7 Morphological novelties and similarities in genomic and transcriptomic perspectives

Morphological features are the outcome of gene expression and their interactions with the environments. The tissue or cell identity is in fact grounded in a certain state of GRN, showing expression of particular toolkit genes so-called ‘markers.’ Taking advantage of the development of sequencing technology (Metzker, 2010), it is now possible to understand embryonic development at a new level of resolution. By sequencing genomes and transcriptomes, we can resolve the questions, such as the evolution of GRN (i.e. gain and loss of *cis*-regulatory elements), lineage-specific gene family expansions, and evolution of genome organization in concert with body plan evolution (Tschopp and Tabin, 2017). In particular, gene expression profiling by transcriptome analysis either with microarray or recently RNA-

seq approaches has been successfully applied to study the developmental and evolutionary origins of the cell and tissue types (Tschopp et al., 2014; Wang et al., 2011).

In an evolutionary perspective, for example, temporal developmental transcriptomes together with *in situ* hybridizations suggest that sponges and cnidarians might share similar body plans, reflecting the conservation of ancestral developmental regulatory networks (Leininger et al., 2014). Furthermore, comparing the germ-layer (i.e. ectoderm, mesoderm, and endoderm) temporal transcriptomes of *C. elegans* to frogs, sea anemones, and sponges allows exploring the origin of germ layers. This comparative transcriptomics reveals that the endoderm program might precede that of ectoderm and date back to the origin of multicellularity (Hashimshony et al., 2015).

In addition, during vertebrate development, there is a period at the mid-embryonic stages in which the embryos develop a head, a neural tube, pharyngeal arches, and somites (Duboule, 1994). These stages reflect the constraint for organogenesis, showing highest morphological similarities among vertebrate embryos, which is referred to as the phylotypic stage (i.e. embryos with pharyngeal arch). Interestingly, transcriptome age profiling based on a phylogenetic approach shows that the phylotypic stage expresses the oldest genes, showing strong constraints on developmental regulation (Domazet-Loso and Tautz, 2010). By contrast, newly evolved genes are expressed at the early and late stages. Using microarray analysis, vertebrate pharyngula stages show highest molecular similarities among mice, chickens, frogs, and zebrafish, supporting the hourglass model for the conservation of mid-embryonic stages (Irie and Kuratani, 2011). This embryonic phylotypic period is also observed outside vertebrates, such as fruit flies (Kalinka et al., 2010) and even plants (Quint et al., 2012), suggesting a universal molecular constraint during organogenesis in multicellular organisms.

Although it has been a matter of debate for pairwise comparisons across phyla (Hejnol and Dunn, 2016), a recent large-scale study explored the general characteristic of the phylotypic period in all animals. Surprisingly, the result shows a high diversity of gene expression patterns during mid-embryonic stages in diverse phyla, suggesting the molecular constraints in lineage-specific manner, which disfavors the hourglass model (Levin et al., 2016). Nevertheless, genomic and transcriptomic approaches allow mathematical quantification of molecular similarities, which ultimately enhance our understanding of the evolution of animal body plans and morphological features.

1.8 Biomineralization in shells and bones

From bacteria to vertebrates, biomineralization is employed to make hard tissues, mostly in the form of calcified minerals with carbonate or phosphate, for protection, support, and feeding (Cusack and Freer, 2008; Knoll, 2003; Lowenstam, 1981). Molluscs may be among the most successful animal groups that form hard external tissues. Like most other marine invertebrates, mollusc shells are composed of calcium carbonate ($CaCO_3$). The mineral parts constitute more than 90% of the shell weight, and the mass of organic matrix in the shell is usually less than 5% (Marin et al., 2008; Suzuki and Nagasawa, 2013). Most mollusc shells have three major layers. The outermost layer, the periostracum, is composed of chitin and organic matrix. The middle, or prismatic layer, is a thin sheet composed of crystalline calcite and aragonite, and the inner layer, the nacreous or foliated layer, is the thickest, and is composed of crystalline aragonite (Sun and Bhushan, 2012; Suzuki and Nagasawa, 2013).

In contrast, *Lingula* shells are rich in organic materials which represent about 40% by dry weight (Iwata, 1981), and are made of calcium phosphate (Clarke and Wheeler, 1915) in the form of carbonate-substituted fluorapatite ($Ca_{10}(PO_4)_6F_2$, or francolite). Superficially similar to mollusc shells, brachiopod shells also consist of three major layers. The outermost layer, periostracum (~4 μm), is an organic layer composed of chitin and organic matrix. The primary layer (~40 μm) is composed of rod and botryoid types of apatite and glucosaminoglycan gels (GAGs; with long unbranched polysaccharides). The secondary layer, the laminated layer (variable in thickness), is composed of apatitic laminae (Iwata, 1981; Williams et al., 1994). The laminated structure provides flexibility and fracture resistance, which may benefit burrowing behaviors (Merkel et al., 2009). It is worth mentioning that in *Lingula* there are collagen fibers at the interface of the primary and secondary layers, a feature not shared by molluscs shells but similar to vertebrate bones (Iwata, 1981; Jope, 1977; Williams et al., 1994).

Biomineralization has been extensively studied but the molecular mechanism remains unknown. The process has been termed as ‘biologically induced’ or ‘biologically controlled’ depending on the degree of biological control involved. The minerals are formed by biologically induced processes if their precipitation is the result of interactions between the organism and the environment, in which cell surfaces and compartmentalized fluid cavities catalyze nucleation and growth of the minerals (i.e. mineralization is initiated by an extracellular organic matrix). On the other hand, the biologically controlled process involves direct control of nucleation, growth, morphology, and location of mineral deposition via intracellular regulation (Weiner and Dove, 2003). In humans, for example, cells capable for

making calcified tissues, such as cartilage, bone, and dentin, form so-called matrix vesicles, that bud off from specific regions of the plasma membrane and regulate ion concentration and mineral formation intra-cellularly and intra-vesicularly (Boonrungsiman et al., 2012; Golub, 2009). In sea urchins, larval endoskeletons or spicules are formed intra-cellularly in membrane-delineated compartments generated by multiple skeletogenic cells (Beniash et al., 1999). Skeletogenic cells are able to transform minerals from amorphous calcium carbonate into crystalline calcite (Beniash et al., 1997; Politi et al., 2008).

Two models have been proposed for the mechanism of mollusc shell formation. The matrix-secreted model (biologically induced) suggests that the mantle epithelial cells secrete shell matrix proteins and ions into a compartment (i.e. extrapallial space) where the minerals are formed (Furuhashi et al., 2009; Marin et al., 2008), whereas various tissues may also contribute to this secretion process (Wang et al., 2013). In the cell-mediated model (biologically controlled), cells (e.g., granulocytic hemocytes in case of oysters) form the minerals intra-cellularly, in which crystal nucleation is initiated under cellular regulation (Mount et al., 2004; Zhang et al., 2012). Taken together, it is reasonable to hypothesize that these two models might both be involved in the biomineralization during shell formation.

Even though there is a lot of interest in mollusc shell formation, the evolutionary origin of mollusc shells is unclear. Studies of mollusc mantle transcriptomes and shell proteomes suggest that gene sets responsible for formation of calcium-carbonate-based calcite or aragonite evolved rapidly. Mineral homology among molluscs might be the result of parallel evolution, since their toolkit genes of many species are so diverse (Jackson et al., 2006; Jackson et al., 2010; Sarashina and Endo, 2006). Supporting this view, new shell matrix proteins may have originated from gene duplication events, in which those genes were initially responsible for general functions and were later co-opted for calcification (Aguilera et al., 2017; Sarashina et al., 2006). One interesting proposition is that horizontal gene transfer from bacteria may also have contributed to the rapid neofunctionalization of biomineralization gene sets during early metazoan evolution (Ettensohn, 2014; Jackson et al., 2011), although this idea is still a matter of debate.

In contrast to studies of mollusc shell formation, the origin of the *Lingula* shell is largely unknown. Although some Cambrian arthropods, tommottids, and various other problematica also used calcium phosphate for their skeletons (Bengtson et al., 1992), one intriguing observation is that lingulid brachiopods and craniates (i.e. head vertebrates) are the only two well-characterized groups of extant animals that utilize calcium phosphate minerals (Knoll, 2003). Given that vertebrate bones are made up of hydroxyapatites (i.e.

$Ca_{10}(PO_4)_6(OH)_2$), fibrillar collagens, and GAGs (Kawasaki et al., 2009), which are similar in composition to *Lingula* shell (McConnell, 1963), it is tempting to wonder whether the mechanism of biomineralization between these distant phyla shares a common origin.

However, using solid-state nuclear magnetic resonance spectroscopy and X-ray diffraction, a recent study found that *Lingula* shell has higher mineral crystallinity and shows no GAG-mineral interaction compared to vertebrate bone (Neary et al., 2011). Comparison of ultrastructure by electron diffraction confirmed the higher crystallinity and also determined that carbonate content is lower, in contrast to vertebrate bone (Rohanizadeh and Legeros, 2007). These findings cast doubt on the idea that *Lingula* shell and vertebrate bones involve the same gene sets. Thus, genomic scale comparisons of biomineralization genes among *Lingula*, molluscs, and vertebrates may provide interesting insights into the molecular mechanism and evolutionary origin of the *Lingula* shell.

1.9 The brachiopod *Lingula anatina* and debates on 'living fossil'

Although superficially resembling mussels, lingulid (i.e. tongue-shaped) brachiopods (commonly known as lampshells), including *Lingula anatina,* have several unique features that distinguish them from bivalves (Fig. 1.2a,b). These include flexible, dorsal–ventral shells made of calcium phosphate without hinges, chitinous chaetae on the mantle margins, two arms lined with ciliated tentacles (i.e. lophophores) for filter feeding, and a tail-like structure (i.e. pedicle) to attach to hard substrate (Bitner and Cohen, 2013; Williams et al., 1994). In addition, their early embryonic development is like that of basal deuterostomes (Yatsu, 1902) (i.e. radial cleavage and enterocoely).

With inarticulate shells, *Lingula* has evolved to adapt to an infaunal lifestyle, such as burrowing into the sand in a U-shaped manner, positioning themselves vertically, and living in the intertidal zone (Emig, 1997; Savazzi, 1991). Importantly, their lingulid shell shows some of the very first innovations in animal biomineralization, since the fossil record of lingulid brachiopods dates back more than 520 million years ago (Williams et al., 1996). It seems reasonable that lingulid brachiopods might have taken advantage of calcium phosphate, since the phosphorus concentration in the seawater was ostensibly high during the Precambrian and Cambrian Periods (Cook and Shergold, 1984). Since the Permian extinction, bivalves have rapidly increased their diversity, but the basic body plan of brachiopods has been constrained (Gould and Calloway, 1980).

Darwin first noticed *Lingula* (possibly referring to all then known lingulid brachiopods) while comparing abundant fossils to living species. He concluded that their

shells have changed very little since the early Cambrian, compared to bivalves and referred to them as an example of 'living fossils' (Darwin, 1859). However, this idea is still controversial (Emig, 2003, 2008). Detailed examination of fossilized and living shells of lingulid brachiopods shows that there is a high diversity on their chemical structure (i.e. how the minerals growth and arrange within the shell) (Cusack et al., 1999; Williams and Cusack, 1999). Similar to this line, soft tissue fossils found in the Chengjiang fauna show that there have been morphological changes among lingulid brachiopods, suggesting that they evolved in contrast to the idea of that "the Silurian Lingula differs but little from the living species" by Darwin thought (Zhang et al., 2005). This notion is supported by population genetics of *L. anatina* across the Indo-West Pacific region, which exhibits a high genetic divergence within the same species (Yang et al., 2013).

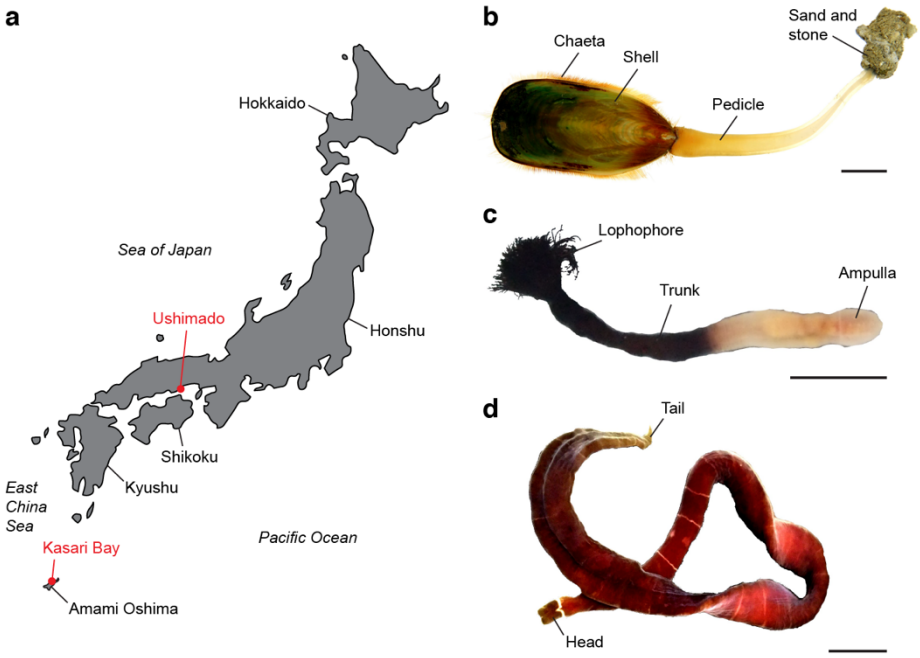


Figure 1.2 | Origins of brachiopod, phoronid, and nemertean samples.
(**a**) Sampling localities of the Seto Inland Sea at Ushimado (Okayama, Japan) and the East China Sea at Kasari Bay (Amami, Japan). (**b**) An adult brachiopod, *Lingula anatina*, with sand and stone attached on the end of the pedicle. (**c**) An adult phoronid, *Phoronis australis*. (**d**) An adult nemertean, *Notospermus geniculatus*. Both anterior ends are to the left. Scale bars, 10 mm. Adult samples of *L. anatina* were collected at Kasari Bay, whereas those of *P. australis* and *N. geniculatus* were collected at Ushimado.

1.10 The phoronid *Phoronis australis*

Phoronids (commonly known as horseshoe worms) are tubular marine invertebrates closely related to ectoprocts and brachiopods (together called lophophorates). The most

distinguishable feature of this group is that they carry the bilaterally symmetrical, ciliated tentacles called lophophores, which are responsible for filter feeding on phytoplanktons. Similar to brachiopods, phoronids have a U-shaped digestive tract, and the anus is close to the mouth (Emig, 2001; Santagata, 2015). There are only two phoronid genera, *Phoronis* and *Phoronopsis*, with about 10 species (Emig, 2001; Santagata, 2015). Phoronids secrete chitinous tubes to which sediment particles adhere (Emig, 2001). There is no evidence of mineralized skeletons; hence they have a poor fossil record. A trace fossil, *Diorygma atrypophilia*, with the affinity to phoronids, has been found dating to the Devonian (Mackinnon and Biernat, 1970). Recently, the discovery of mid-Cambrian hyolith fossils with well-preserved lophophores, suggests that hyoliths are closely related to crown phoronids and brachiopods (Moysiuk et al., 2017). Nevertheless, the evolutionary origin of phoronids is still unclear.

The phoronid, *Phoronis australis*, lives in association with tube-dwelling anemones (e.g., the cerianthid *Cerianthus maua*) (Emig et al., 1972) (Fig. 1.2a,c). However, it is not clear whether their relationship is mutualistic, parasitic, or commensal. Extended, *P. australis* can reach 20 cm and 2–5 mm in diameter. *P. australis* is hermaphroditic, with internal fertilization, brooding its embryos within the tentacle crown until the actinotroch larval stage (Emig, 2001).

1.11 The nemertean *Notospermus geniculatus*

Nemerteans (commonly known as ribbon worms) are flattened unsegmented worms with about 1,400 described species (Moore and Gibson, 2001). They are mostly predators (few scavengers) lurking in crevices between stones and living in benthic predominantly marine inhabitants with a few species in freshwater environments. One of their main features is that they possess an eversible proboscis, which is enclosed a dorsal, fluid-filled cavity called a rhynchocoel (Moore and Gibson, 2001; von Döhren, 2015). The proboscis is armed with calcareous stylets in the species within the order Hoplonemertea (von Döhren, 2015). Armed nemerteans kill their prey by injecting toxins with their stylets; then they suck out the soft body parts, whereas unarmed species usually swallow their prey in whole. The nemertean body is highly extensible. Some species, such as *Lineus longissimus*, can reach 30 m long, making them the longest animal on Earth (Holland, 2011). No fossil nemerteans have been found, due to their lack of hard tissues (Moore and Gibson, 2001).

The nemertean, *Notospermus geniculatus*, belongs to Heteronemertea, which lacks stylets (Fig. 1.2a,d). The length of *N. geniculatus* is about 10–30 cm. They are sexually

dimorphic with external fertilization. No *Notospermus* larvae have been reported, but they possibly carry planktotrophic pilidium larvae (Hiebert and Maslakova, 2015) since that is the feature shared by most pilidiophorans (Hubrechtidae + Heteronemertea) (Andrade et al., 2014; Hiebert and Maslakova, 2015), although adelphophagic intracapsular Schmidt's larvae have also been found in *Lineus ruber* (Martín-Durán et al., 2015).

1.12 Concluding remarks

Based on the foregoing, the aim of my research is to decode genomes of three taxa of Lophotrochozoa, namely, Brachiopoda, Phoronida, and Nemertea. Genome sequencing and comparative analysis might shed light on phylogenetic relationships of these animal groups as well as their evolutionarily specific traits. In this thesis, I explore lophotrochozoan evolution and their morphological novelties, such as brachiopod shells and lophophores, using comparative genomics and transcriptomics. I also report the *Lingula* mitochondrial genome, in which its gene order is unexpectedly highly shuffled.

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

2.1 Introduction

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 (Williams et al., 1994). Based on molecular phylogeny, brachiopods comprise three subphyla, Linguliformea, Craniiformea, and Rhynchonelliformea (Sperling et al., 2011). 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 (Williams et al., 1996). Their shells are composed of calcium phosphate and collagen fibers, characters shared only by evolutionarily distant vertebrates (Knoll, 2003; Williams et al., 1994). Morphologically, brachiopods and bivalves superficially resemble each other. However, lingulid brachiopods exhibit several unique features that distinguish them from molluscs. These include hinge-less shells that grow along the dorsal-ventral axis, chitinous chaetae, ciliated lophophores, and a tail-like pedicle (Bitner and Cohen, 2013; Williams et al., 1994). Since the Permian extinction, bivalves have greatly increased their diversity, but the basic body plan of brachiopods has been constrained (Gould and Calloway, 1980), which is still a mystery of metazoan evolution.

It has been proposed that *Lingula* might have utilized calcium phosphate because the phosphorus concentration in seawater was high in the Cambrian (Cook and Shergold, 1984). In fact, some Cambrian arthropods, tommottids, 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 among molluscs could simply represent parallel evolution (Jackson et al., 2010). In contrast to mollusc shells and other invertebrate calcified tissues, *Lingula* shells are comprised of calcium phosphate, laminated, flexible, and rich in organic materials (Williams et al., 1994). Despite their paleontological 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 (Yatsu, 1902). Despite such unique features, the phylogeny of brachiopods is under debated. Before the 1980s, brachiopods were classified as deuterostomes based upon their mode of development. Then they were grouped within protostomes following an analysis of 18S rRNAs (Field et al., 1988). This classification was further supported by an analysis of Hox genes in brachiopods and priapulids (de Rosa et al., 1999). However, the phylogenetic position of brachiopods is still controversial, in spite of intensive paleontological (Zhang et al., 2014) and molecular phylogenetic studies. For example, whether brachiopods are monophyletic or polyphyletic (Cohen, 2013; Sperling et al., 2011) and whether Brachiopoda is close to Phoronida, Nemertea, Mollusca, Annelida or other lophotrochozoan phyla, remains to be resolved (Dunn et al., 2008; Erwin et al., 2011; Paps et al., 2009).

Here I present the first brachiopod genome of the lingulid, *Lingula anatina.* My whole genome phylogenetic analyses support a close relationship between *Lingula* and molluscs. Unexpectedly, I 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, I show that although *Lingula* shares shell formation-related genes and mechanisms with molluscs, such as chitin synthase and BMP signaling, it utilizes several domain combinations to produce lineage-specific shell matrix collagens, alanine-rich fibers, and novel shell matrix proteins. I 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, my comparative genomic analyses provide insights into the evolutionary history of this lophotrochozoan and the origin of phosphate biomineralization.

2.2 Methods

2.2.1 Biological materials

Gravid *Lingula anatina* adults were collected during July and August in Kasari Bay, Amami Island (28.440583 N 129.667608 E) (Fig. 2.1). 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 phosphate-buffered saline (PBS) into the gonad (Nishizawa, 2010). 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) (Tagawa et al., 1998).

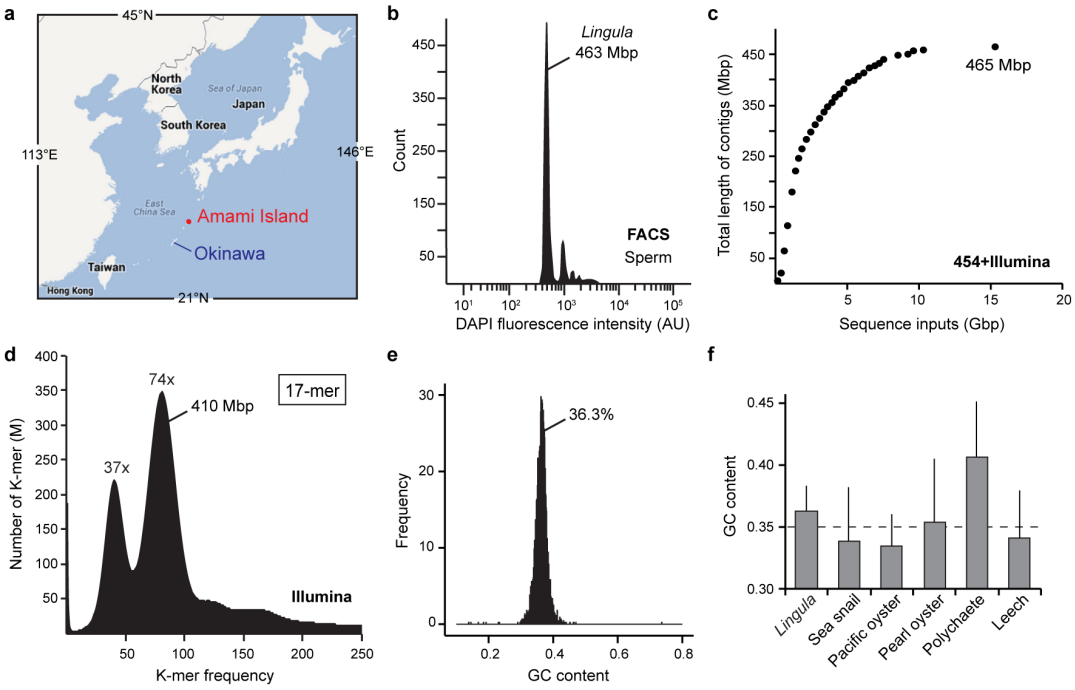
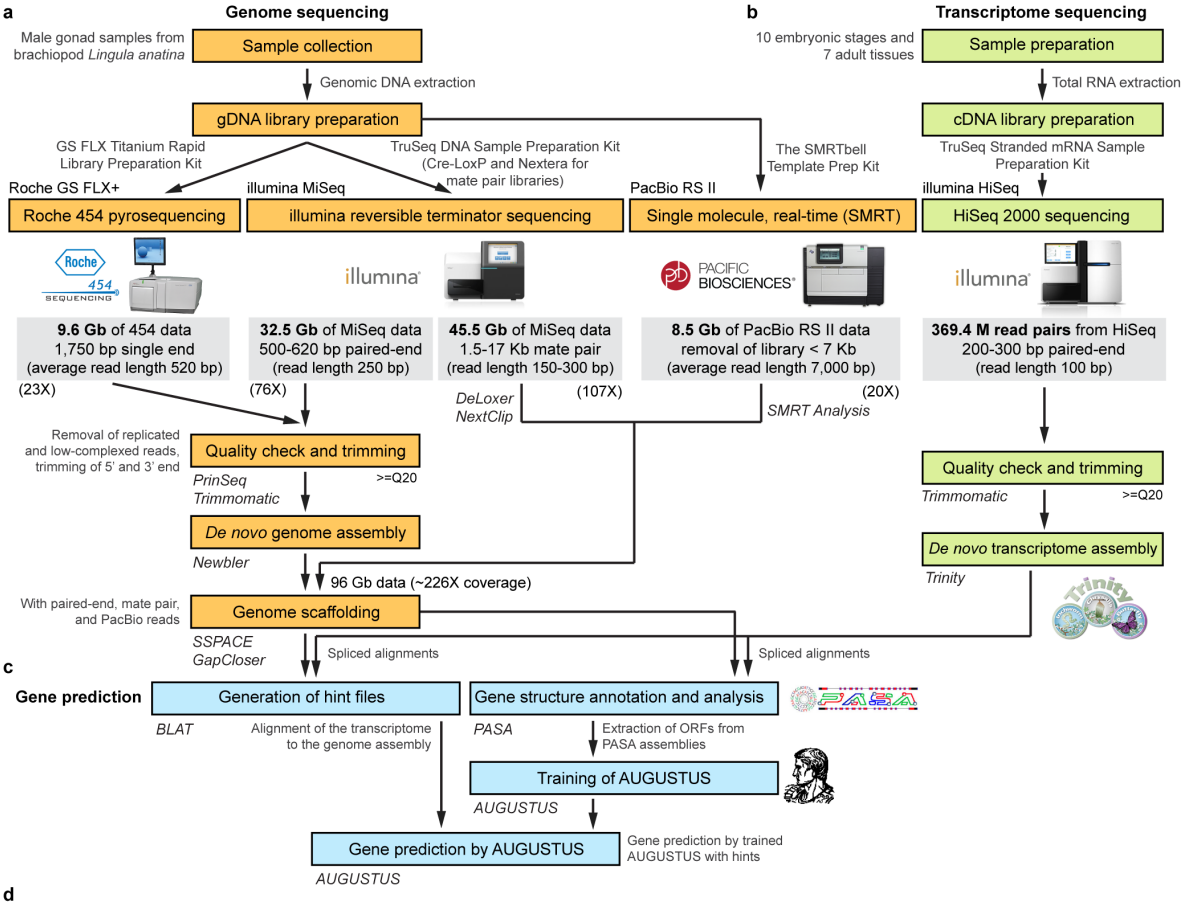


Figure 2.1 | Sampling locality, genome size estimation, and GC content.
(**a**) Sampling locality in Amami Island (i.e., Amami Oshima, Japan) and its relative location to Okinawa are shown with coordinates (adapted from Google Maps). (**b**) Sperm cells collected from gravid male gonads were stained with DAPI and subjected to fluorescence-activated cell sorting (FACS) flow cytometry analysis. Sperm with known genome size from zebrafish (*Danio rerio*) were used as an internal standard to estimate the *Lingula* genome size. (**c**) The analysis of stepwise assembly shows that the saturation point is achieved when input sequences reach 10 Gbp from 454 and Illumina reads. (**d**) *K*-mer analysis (17-mer) using Illumina reads shows two peaks, in which the homozygous peak coverage is twice the heterozygous peak. The estimated heterozygosity rate calculating the ratio of the peaks, is 1.6%. (**e**) Distribution of GC content calculated from 3,830 scaffolds. (**f**) Comparison of GC content in selected lophotrochozoans. Error bars, standard deviation.

2.2.2 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) (Bolger et al., 2014). Raw mate pair reads were filtered with DeLoxer (Van Nieuwerburgh et al., 2012) or NextClip (v0.8) (Leggett et al., 2014) depending on library preparation. Genome assembly was conducted using Newbler (v2.9) with a hybrid assembly approach using data from 454 and Illumina (Shinzato et al., 2011).



d

Genome size (Mbp)	GC (%)	Repeats (%)	Gaps (Mbp)	Gap (%)	Number of contigs	Number of scaffolds	Contig N50 (kbp)	Scaffold N50 (kbp)	Number of genes	Mean gene size (bp)	Mean transcript size (bp)	Mean intron per gene	Mean intron size (bp)
425	36	22	17	4	14,865	3,830	56	294	34,105	6,669	1,425	6.6	787

Figure 2.2 | Schematic flow of sequencing and assembly of the *Lingula* genome.
(**a**) Genomic DNA from a male gonad was extracted for genome sequencing using Roche 454, Illumina, and PacBio platforms. A total of 96-Gb of data was obtained with approximately 226-fold coverage of the 425-Mb *Lingula* genome. (**b**) Ten embryonic stages from egg to larva and seven adult tissues were collected for RNA-seq and reads were assembled *de novo* using Trinity. (**c**) Transcript information from RNA-seq was used to generate hints by spliced alignment with PASA and BLAT. Gene models were predicted with trained AUGUSTUS. (**d**) Summary of the *Lingula* genome assembly and annotation. Programs used here, such as DeLoxer, NextClip, SMRT Analysis, PrinSeq, Trimmomatic, Newbler, SSPACE, GapCloser, Trinity, BLAT, PASA, and AUGUSTUS are marked in italic.

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 (Fig. 2.2). 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. To overcome repetitive regions of the genome, mate pair libraries with 1.5–3 kb lengths were prepared using the Cre-Lox

recombination approach (Van Nieuwerburgh et al., 2012). In addition, in order 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 (Caruccio, 2011). The long mate pair libraries were sequenced to obtain 45.5 Gb of mate pair data using Illumina MiSeq and HiSeq 2500 platforms.

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) (Boetzer et al., 2011). PacBio long reads (>7 kb) were mapped to the scaffolds generated by Newbler using BLASR (v20141001) (Chaisson and Tesler, 2012), and upgraded scaffolds were produced with SSPACE-LongRead (v1-1) (Boetzer and Pirovano, 2014). Gaps in the scaffolds were filled using GapCloser (v1.12-r6) from the SOAPdenovo2 package (r240) (Luo et al., 2012). Redundancy of final scaffolds was removed by calculating BLASTN alignment length and identity using a custom Perl script (Shinzato et al., 2011). Regions of repetitive sequences were identified with RepeatScout (v1.0.5) (Price et al., 2005) and then masked with RepeatMasker (v4.0.3). The genome size was estimated by flow cytometry as well as by K-mer analysis using SOAPec (v2.01) and Genomic Character Estimator (GCE; v1.0.0) from the SOAPdenovo package (Luo et al., 2012). *K*-mer analysis was also conducted using Jellyfish (v2.0.0) (Marçais and Kingsford, 2011) 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) (Parra et al., 2007).

2.2.3 Gene model prediction

To obtain high-quality gene models, mRNA sequencing (RNA-seq) was performed to obtain transcript information (Fig. 2.3). RNA-seq data (369 million read pairs) from embryos and adult tissues were obtained using an Illumina HiSeq 2500. Transcripts assembled *de novo* with Trinity (r2013_08_14) (Haas et al., 2013) were used as expression evidence for predicting gene models. Gene models were predicted with trained AUGUSTUS (v3.0.2) using hints from spliced alignment of transcripts to the masked genome assembly produced with BLAT and PASA (r20130907).

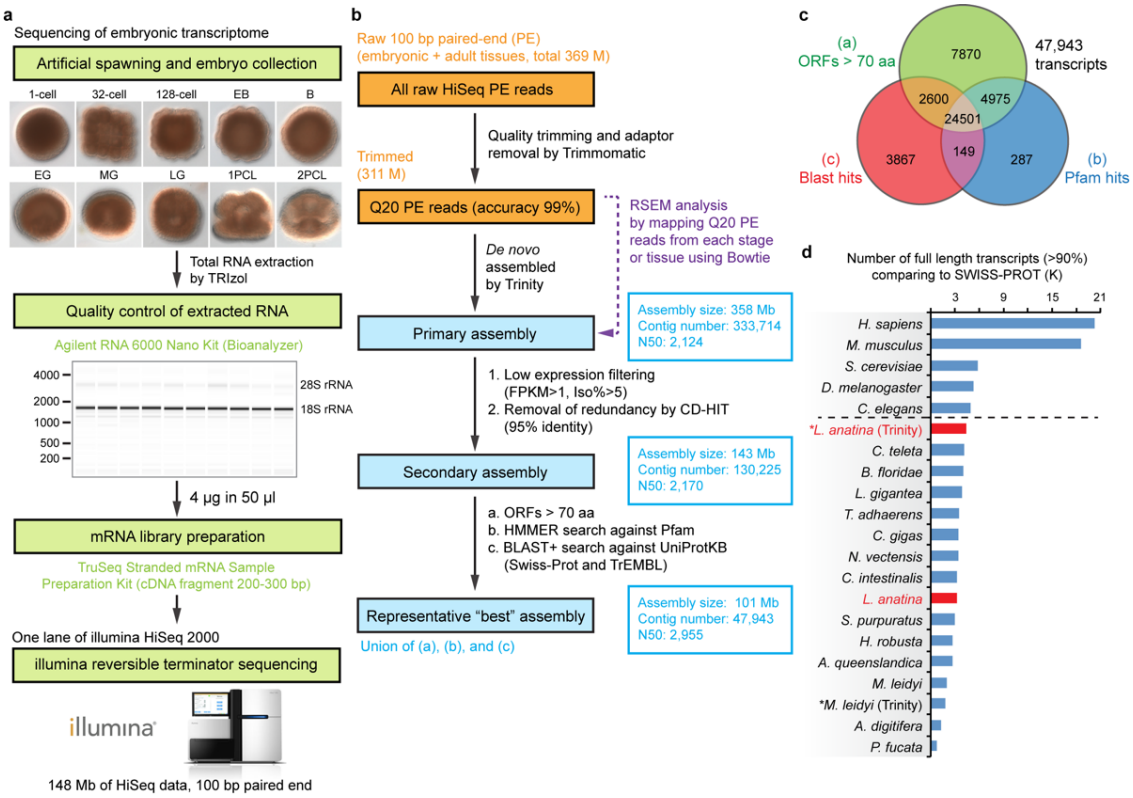


Figure 2.3 | Transcriptome sequencing, assembling, and analyses.
(**a**) Flow chart of transcriptome sequencing with embryonic samples as an example. Extracted RNA is quality checked with a Bioanalyzer to be sure there is no RNA degradation. Note that expression level of 28S rRNA is extremely low in *Lingula*. After mRNA library preparation, samples were subjected to HiSeq sequencing. (**b**) Procedures for assembling the transcriptome. Summary of assembly statistics is given in blue boxes. FPKM, fragments per kilobase of transcript per million mapped reads. ORFs, open reading frames. (**c**) Venn diagram for the final transcriptome assembly containing 47,943 transcripts obtained from three sets of filtering criteria. (**d**) Transcript completeness analysis. Selected gene models predicted from genomes and transcripts assembled with Trinity (marked by asterisks) were mapped to the Swiss-Prot database to estimate the completeness of the given transcripts by checking their sequence alignment rate. Dashed line separates the well-annotated organisms from the others. The *Lingula* gene models and transcriptome are labeled in red.

2.2.4 Gene family analyses

To analyze gene family evolution in lophotrochozoans, all-to-all BLASTP analysis was performed followed by Markov clustering in order to identify orthologous gene groups (OG) with OrthoMCL (v2.0.9) (Fischer et al., 2011), according to the standard protocol using a default inflation number of 1.5. Gene family birth and death was estimated by computing the OG using CAFE (Computational Analysis of gene Family Evolution; v3.1) (De Bie et al., 2006). Important transcription factors and signaling components were annotated with Pfam domain searches using HMMER. To identify genes related to specific pathways, which are

interesting topics for lineage specific evolution, the KEGG pathway database was utilized. Non-synonymous (*Ka*) and synonymous (*Ks*) substitution rates of paired-wise paralogs were calculated with KaKs_Calculator (v2.0) (Wang et al., 2010).

2.2.5 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, and then orthologs 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 orthologs were performed with MAFFT (v7.130b) (Katoh et al., 2002). Unaligned regions were trimmed with TrimAl (v1.2rev59) (Capella-Gutiérrez et al., 2009). The maximum likelihood method with LG+Γ4 and GTR+Γ4 models was used to construct phylogenetic trees with RAxML (v8.0.5) (Stamatakis, 2014). Bayesian trees were constructed with PhyloBayes (v3.3f) (Lartillot et al., 2009) using LG+ Γ4 and GTR+Γ4 models with the first 500 trees as a burn-in. After a run time of ~20 days (with approximately 4,000 generations), convergence of the tree topology was post-analyzed by sampling every 10 trees.

2.2.6 Transcriptome analyses

To make the transcriptome more accessible for downstream analysis, transcript assemblies that contained computation errors, expressed at extremely low levels, and expressed with highly similar isoforms were eliminated. After RNA-seq assembly, raw reads from each embryonic stage and from adult tissue were mapped back to transcript assemblies using Bowtie (v2.1.0) (Langmead and Salzberg, 2012). Transcript abundance was estimated using RSEM (v1.2.5) (Li and Dewey, 2011). Transcripts expressed at less than one fragments per kilobase of transcript per million mapped reads (FPKM) and isoform representing less than 5% of a given transcript were filtered. In addition, redundant isoforms were removed with CD-HIT (v4.6) (Li and Godzik, 2006) using 95% identity as a criterion. Next, three sets of criteria were applied to select transcripts with annotated biological functions. First, open reading frames (ORFs) of transcripts were extracted with the program, getorf, in the EMBOSS package (v6.6.0.0). Transcripts with ORFs longer than 70 amino acids were retained. Next, the transcriptome was searched against the Pfam database (Pfam-A 27.0) with HMMER (v3.1b1) and against UniProtKB database with BLASTP, respectively. The final

representative 'best' assembly is the union of the three sets of transcripts. In order to assess the quality of the transcriptome assembly, full-length transcript analysis was applied using a bundled Perl script 'analyze_blastPlus_topHit_coverage.pl' in the Trinity package (Haas et al., 2013). Venn diagram was plotted with jvenn (Bardou et al., 2014). Gene ontology (GO) enrichment analysis was conducted with DAVID (Huang da et al., 2009) and PANTHER (Mi et al., 2013).

2.2.7 Comparative transcriptomics

To compare with molluscs, RNA-seq raw reads of selected adult tissues from the Pacific oyster *Crassostrea gigas*, which are comparable to those of *Lingula*, were downloaded from OysterDB (http://oysterdb.cn/) and reassembled with Trinity (Haas et al., 2013). Orthologs were identified using a BBH approach. Spearman's ρ is robust when the data set contains extreme values, while Pearson's r is affected by outliers (Mukaka, 2012). To identify transcriptomic similarities between *Lingula* and *Crassostrea* tissues, Spearman's (ρ) and Pearson's (r) correlation coefficients were calculated using custom Bash and Perl scripts. The defined value of the coefficient (ρ) is

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \quad (1)$$

where $d_i = x_i - y_i$ is the difference between the two rank values, and n is the sample size (i.e. the number of BBH orthologs; 6,315 orthologs were identified). In brief, a serial number was given to each orthologous pair. Orthologs were then sorted and ranked by expression level. Afterward, a global comparison was performed. The value of the coefficient (r) is defined by

$$r = \frac{\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y)}{\sigma_x \sigma_y} \quad (2)$$

where between transcriptomes x and y, there are n orthologous pairs, x_i and y_i are the expression levels in FPKM, μ_x and μ_y are the average FPKM values of each transcriptome, and σ_x and σ_y are the corresponding standard deviations.

2.2.8 Comparative genomics

Using recent published resources on bone evolution in elephant shark, *Callorhinchus milii* (Venkatesh et al., 2014), shell formation in molluscs (Zhang et al., 2012), and silk genes in two spiders, *Stegodyphus mimosarum* and *Acanthoscurria geniculate* (Sanggaard et al., 2014), comparative analyses of biomineralization genes associated with bone, shell, and silk formation were conducted. The BBH approach was used to identify orthologous relationships. Genomic scale comparisons of these genes using genomes of humans (*Homo*), sharks (*Callorhinchus*), *Lingula*, and molluscs (pearl oyster, *Pinctada*, Pacific oyster, *Crassostrea*, and sea snail, *Lottia*) were made. The heatmap and clustered matrix were created using R (v3.0.2; http:/www.R-project.org/) with the package Bioconductor (v3.0) and pheatmap (v0.7.7).

2.2.9 Immunostaining and F-actin labeling

Embryos were fixed with 4% paraformaldehyde, dehydrated in cold methanol, and stored at -20°C. For antibody staining, embryos were rehydrated in PBST (PBS with 0.1% Tween-20) for 10 min and permeabilized in PBSTX (PBST with 0.1% Triton X-100) for 30 min. Afterward, embryos were blocked in 3% bovine serum albumin (BSA) in PBST for at least one hour followed by incubation in the primary antibody, rabbit anti-pSmad1/5/9 (1:200; Cell Signaling, 9511S) or BODIPY FL phallacidin (1:50; Invitrogen, B607) in 3% BSA in PBST at 4°C overnight. Note 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 DAPI (1:1,000; Dojindo, 340-07971), and cytoplasmic membranes were labeled with CellMask Deep Red (1:2,000; Invitrogen, C10046). Embryos were imaged using a Zeiss LMS 780 inverted confocal system.

2.2.10 Data availability

This genome project has been registered in NCBI under the BioProject accession PRJNA286275. The genome and transcriptome assemblies have been deposited at DDBJ/EMBL/GenBank under the accessions LFEI00000000 and GDJY00000000, respectively. Sequencing reads of the genome and transcriptome have been deposited in NCBI Sequence Read Archive under the study accession SRP059398. The proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD002652.

2.3 Results

2.3.1 Genome sequencing and assembly

With technical assistance from collaborators, I sequenced the 425-Mb genome of *Lingula anatina* (Fig. 1a–i) with ~226-fold coverage using four next-generation sequencers (i.e. 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 (Simakov et al., 2013; Takeuchi et al., 2012; Zhang et al., 2012). The *Lingula* genome exhibits comparatively high heterozygosity (1.6%) and a low level of repetitive sequences (22.2%).

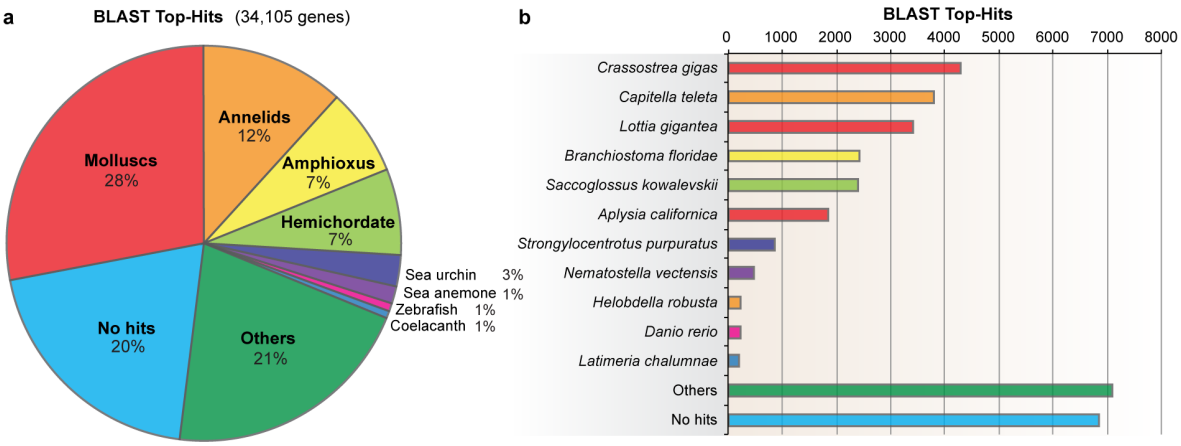


Figure 2.4 | BLAST top-hits analysis against the NCBI nr database.
(**a**) Pie chart of top-hits results among 34,105 gene models in the current *Lingula* genome assembly. *Lingula* has the highest gene similarity to molluscs (28%). A large number of gene models (20%) cannot be assigned to any known sequences. (**b**) More detailed categories for species where the top-hits are distributed. The color code is the same as that of the pie chart. The top-hit species is the Pacific oyster, *Crassostrea gigas* (~4,300 genes). Note that many top-hits are to amphioxus and hemichordate (~5,000 genes). BLAST search was conducted with an *e*-value cutoff of $1e^{-5}$.

Together with a large quantity of transcriptome data from adult tissues and embryonic stages, I estimated that *Lingula* contains 34,105 protein-coding gene models, 91% of which are supported by transcriptomes. The mean size of *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* (Simakov et al., 2013). 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 (Fig. 2.4).

2.3.2 Phylogenetic position of brachiopods

To resolve the phylogenetic position of brachiopods, I carried out phylogenetic analyses. Analysis based on 150 one-to-one orthologs with 46,845 amino-acid positions from 15 metazoan genomes supports the assertion that *Lingula* is closer to Mollusca than to Annelida (Fig. 2.5). 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, heme-binding protein, and Death-associated protein domains (Fig. 2.6).

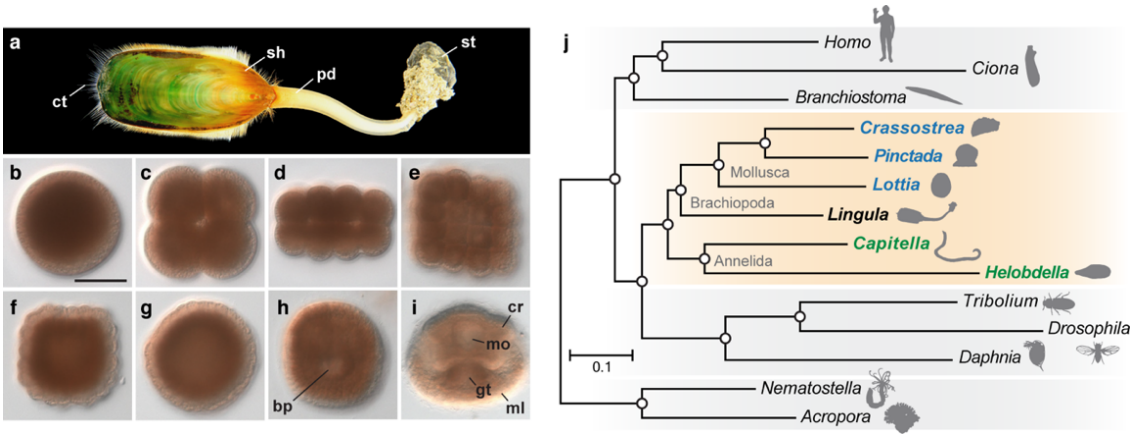


Figure 2.5 | Deuterostomic development of the brachiopod, *Lingula anatina*, and its close relationship to molluscs.

(**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. Abbreviations: ct, chaeta; sh, shell; pd, pedicle; st, stone; bp, blastopore; cr, cirri; mo, mouth; gt, gut; ml, mantle lobe. (**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 orthologs (46,845 amino-acid positions) with LG+Γ4 model. Circles at all nodes indicate 100% bootstrap support.

In addition, microsyntenic analyses showed that *Lingula* and *Lottia* share conservation of a large number of microsyntenic blocks, supporting the close phylogenetic relationship between brachiopods and molluscs. Furthermore, intron structures also show similarities between *Lingula* and molluscs, but not annelids (Fig. 2.7). 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.

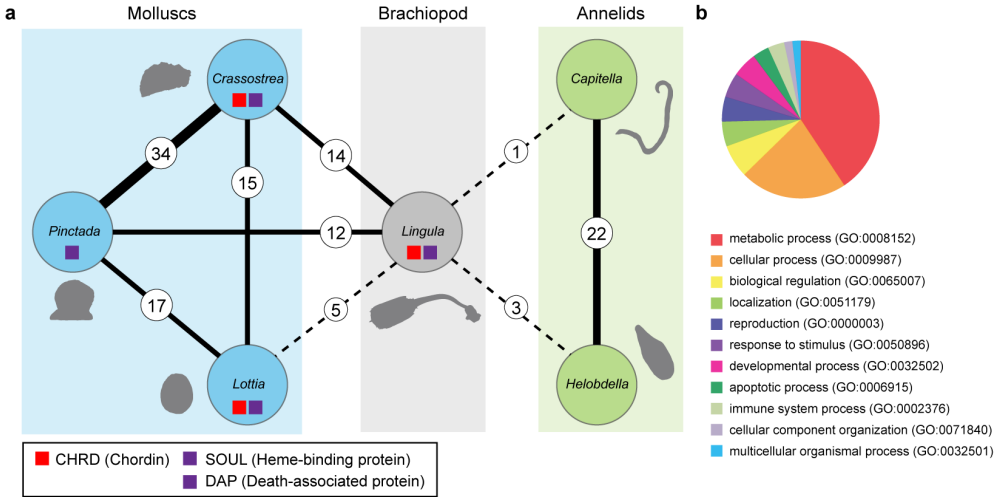


Figure 2.6 | Pairwise comparison of lineage-specific domain loss.
Pairwise comparison of lineage-specific domain loss among *Lingula*, molluscs, and annelids. (**a**) Aanalysis of pairwise lineage-specific domain loss. Numbers of pairwise lineage-specific domain losses are indicated in the circles. Thickened solid lines connecting given pairs are proportional to the value of the loss numbers. Dashed lines indicate low lineage-specific-domain losses between the pairs. CHRD (CHRD domain, PF07452) domain is lost in the pearl oyster (*Pinctada*) and annelids. SOUL (heme-binding protein, PF04832) and DAP (Death-associated protein, PF15228) domains are lost in annelids. (**b**) Functional classification of human genes containing 22 domains lost in annelids, based on GO biological process.

2.3.3 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 (Cusack et al., 1999; Williams and Cusack, 1999). Similarly, soft tissue fossils from the Chengjiang fauna reveal morphological changes among lingulid brachiopods (Zhang et al., 2005). Those findings suggest that lingulid brachiopods 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 (Amemiya et al., 2013). Interestingly, I found that *Lingula* genes associated with basic metabolism, such as ribonucleoprotein complex biogenesis and RNA processing, show the slowest evolutionary rate among lophotrochozoans. However, I also found a high degree of changes in the genomic structure and gene families. The *Lingula* genome contains a disorganized Hox cluster. It is divided into

two regions, and *Lox2* and *Lox4* are missing. Comparison of gene families shared by amphioxus *Branchiostoma floridae* (Putnam et al., 2008), *Capitella,* and *Lottia* show that *Lingula* has 3,525 unique gene families (Fig. 2.8a). 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. 2.8b).

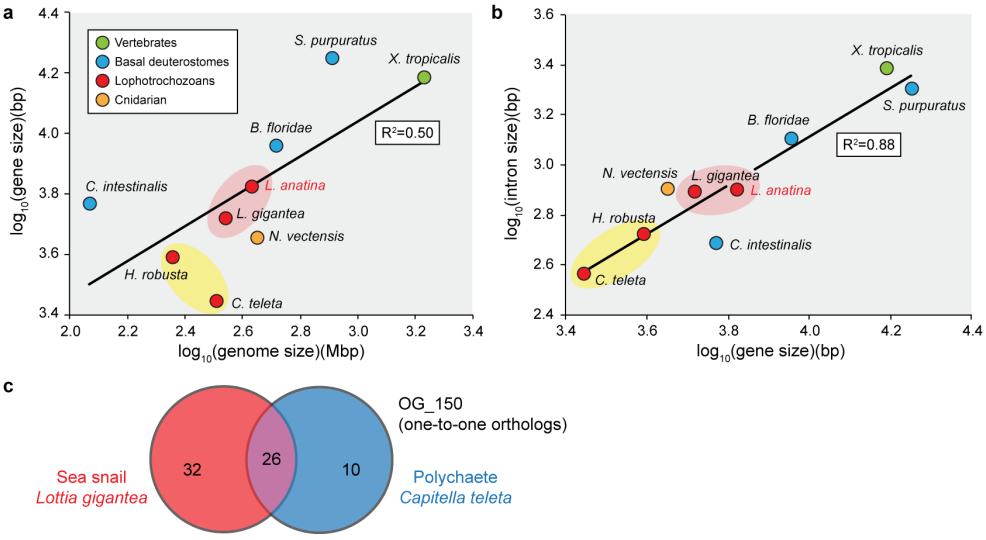


Figure 2.7 | Comparison of intron structure in selected metazoan genomes.
(**a**) Regression analysis of gene size and genome size. R^2, correlation coefficient. (**b**) Regression analysis of intron size and gene size. Close relationships between *Lingula* and sea snails (*Lottia gigantea*) and annelids are circled in red and yellow, respectively. (**c**) Analysis of conserved intron numbers using 150 one-to-one core metazoan gene sets between *Lingula*, *Lottia* and *Capitella*.

To better understand evolution of *Lingula* gene families, I 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), I found that *Lingula* genes duplicate at a rate approximately two to four times faster than those of other lophotrochozoans (Fig. 2.8c). A large portion of these young duplicated genes are undergoing negative selection, suggesting a functional constraint upon them. I also found that genes related to extracellular matrix are experiencing positive selection, 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 martensii* (Cao et al., 2013). I 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 and signaling molecules comparable to those of molluscs.

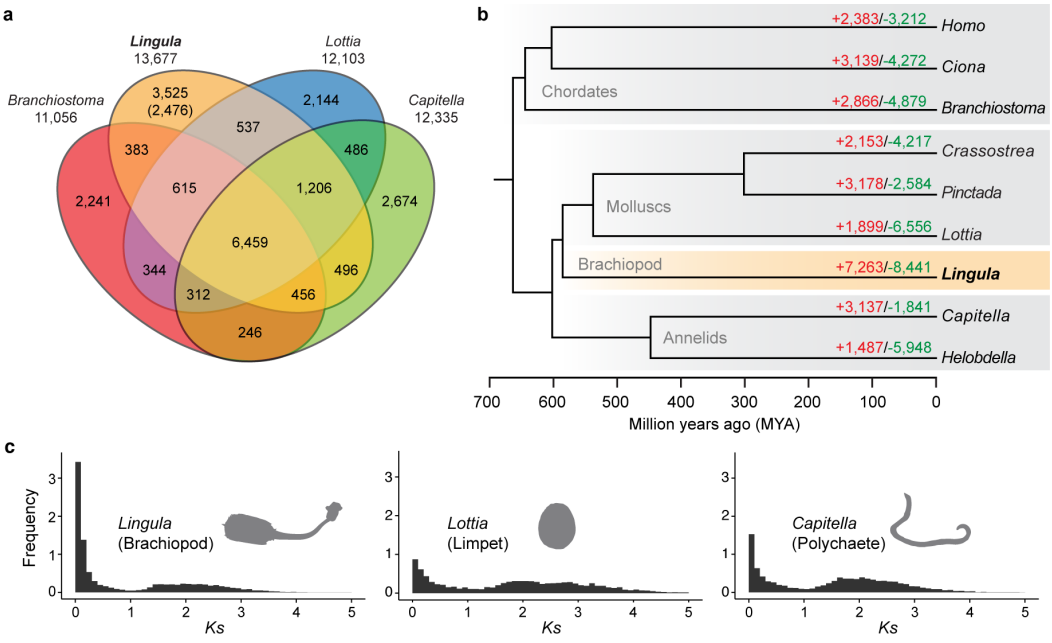


Figure 2.8 | Evolution of the *Lingula* genome is revealed by comparative genomics of lophotrochozoan gene families.

(**a**) Venn diagram of shared and unique gene families in four metazoans. Gene families were identified by clustering of orthologous groups using OrthoMCL. The number in parentheses shows unique gene families compared among 22 selected metazoan genomes. (**b**) Gene family history analyses with CAFE. Divergence times were estimated with PhyloBayes using calibration based on published fossil data. Gene families expanded or gained (red); contracted or lost (green). (**c**) Frequency of pair-wise genetic divergence calculated with synonymous substitution rate (*Ks*) among all possible paralogous pairs in the *Lingula*, *Lottia*, and *Capitella* genomes.

2.3.4 Expansion of gene families and chitin synthases

I found lineage-specific expansions of protein domains and gene families. Five of the 20 most expanded families have possible functions in shell formation, including 31 copies of chitin synthase (CHS) genes and 30 copies of carbohydrate sulfotransferase genes. Chitin, a long-chain polymer of *N*-acetylglucosamine, is a characteristic component of arthropod exoskeletons and mollusc shells. Molecular phylogeny shows that nine *Lingula* CHS genes are included in the lophotrochozoan clade (Fig. 2.9a). In addition, I found that CHS genes of lophotrochozoans contain a myosin-head-domain (MHD). It has been proposed that a MHD might have fused to CHS genes during evolution of lophotrochozoans (Zakrzewski et al., 2014), the only group in which these occur. I found that there is a greater expansion of MHD-

containing CHS genes in molluscs than in *Lingula* or annelids (Fig. 2.9a,b). In molluscs, an MHD-containing CHS gene is expressed specifically in cells that are in close contact with the larval shell (Weiss et al., 2006) and that are probably involved in shell formation (Schonitzer and Weiss, 2007). Its high expression level during larval shell formation and in adult mantle further suggests a role in mollusc shell formation (Zhang et al., 2012).

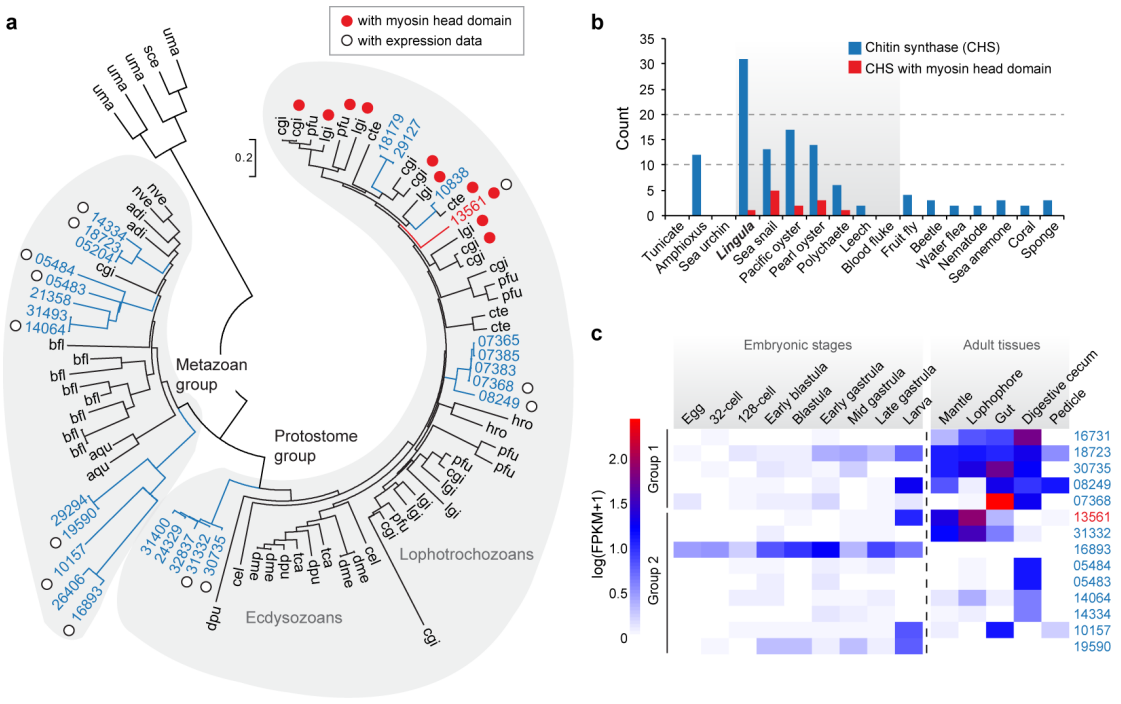


Figure 2.9 | Expansion and expression of *Lingula* chitin synthase genes indicate roles in shell formation and digestion.

(**a**) Phylogenetic analysis of chitin synthase (CHS) genes using the neighbor-joining method with the JTT model (90 genes, 358 amino acids, and 1,000 bootstrap replicates). Three-letter code: sce, baker's yeast (*Saccharomyces cerevisiae*); uma, corn smut fungus (*Ustilago maydis*); bfl, amphioxus (*Branchiostoma floridae*); lgi, sea snail (*Lottia gigantea*); cgi, Pacific oyster (*Crassostrea gigas*); pfu, pearl oyster (*Pinctada fucata*); cte, polychaete (*Capitella teleta*); hro, leech (*Hellobdella robusta*); dme, fly (*Drosophila melanogaster*); tca, beetle (*Tribolium castaneum*); dpu, water flea (*Daphnia pulex*); cel, nematode (*Caenorhabditis elegans*); nve, sea anemone (*Nematostella vectensis*); adi, coral (*Acropora digitifera*); aqu, sponge (*Amphimedon queenslandica*). Numbers are *Lingula* gene IDs. (**b**) CHS genes detected with BLASTP among 17 selected metazoan genomes. Note that CHS genes with myosin head domains are only present among lophotrochozoans (grey area). (**c**) The expression of *Lingula* CHS genes in embryonic stages and adult tissues (separated by a vertical dashed line). FPKM, fragments per kilobase of transcript per million mapped reads.

Transcriptome analysis shows that *Lingula* CHS genes are expressed in all adult tissues and in larvae (Fig. 2.9c). The MHD-containing CHS gene is highly expressed in the larval stage and in mantle, suggesting that it may also play a role in *Lingula* shell formation.

Additionally, CHS genes are highly expressed in the gut and digestive cecum, indicating that a chitinous peritrophic matrix may also be present in the *Lingula* midgut. The expansion of chitin synthase genes in the *Lingula* genome and their different expression profiles suggest that chitins participate in brachiopod biomineralization and digestion.

2.3.5 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 (Cusack and Freer, 2008; Knoll, 2003). Although the shells of *Lingula* and molluscs differ in composition, given that the mantle is the place of shell formation both in brachiopods and molluscs (Marin et al., 2008), I first characterized the molecular nature of the *Lingula* mantle. I found that 2,724 genes are specifically expressed in mantle, including those for signal receptors, adhesion molecules, and metabolic processes. This suggests that the *Lingula* mantle is responsible for extracellular matrix secretion. Next, I performed comparative transcriptome analyses between *Lingula* and the Pacific oyster, *Crassostrea gigas* (Zhang et al., 2012) by calculating Spearman's (ρ) and Pearson's (r) coefficients. My analyses show that the *Lingula* mantle is related to the *Crassostrea* mantle, indicating a functional similarity between these two organs (Fig. 2.10a).

I further found that the expression profiles of genes involved in ribosomal machinery are most similar, while those of genes related to chromosome and cell cycle regulation are diverse. 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 my 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.

To gain further insights into the evolution of biomineralization, I conducted comparative genomics and hierarchical cluster analyses to examine biomineralization-associated genes among vertebrates (Venkatesh et al., 2014), molluscs (Zhang et al., 2012), and *Lingula*.

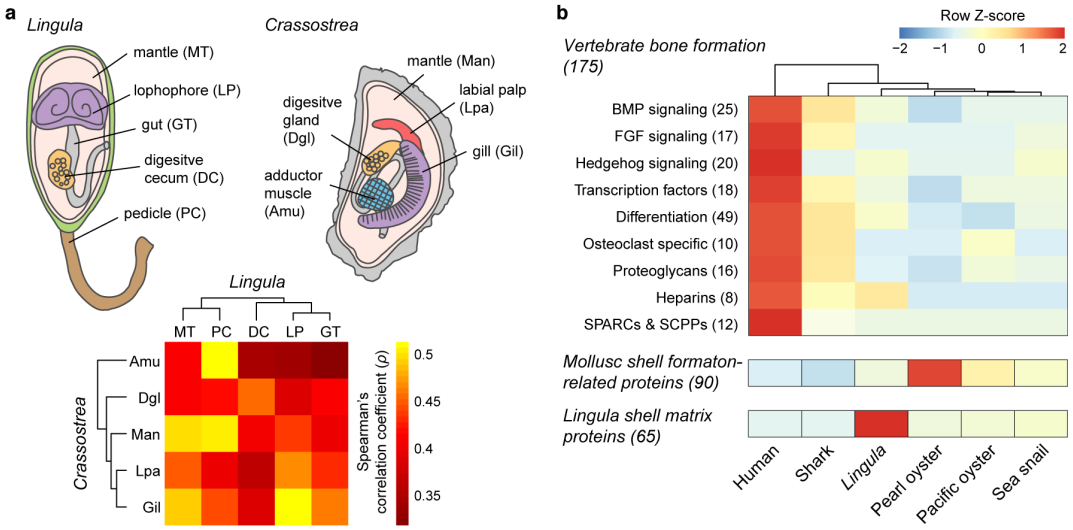


Figure 2.10 | Comparative transcriptomics and genomics reveal different origins of biomineralization-related genes.

(**a**) Spearman's correlation coefficient (ρ) 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 analyzed are indicated in the parentheses. Shark, *Callorhinchus milii*; pearl oyster, *Pinctada fucata*; sea snail, *Lottia gigantea*. BMP, bone morphogenetic protein; FGF, fibroblast growth factor; SPARCs, secreted proteins acidic and rich in cysteine; SCPPs, secreted calcium-binding phosphoproteins.

Given that *Lingula* and vertebrates share the use of calcium phosphate, I first examined 175 genes associated with bone formation. I found that the number of *Lingula* homologs to vertebrate bone formation genes is similar to those in other marine invertebrates. There is no specific similarity between *Lingula* and humans (Fig. 2.10b). The innovation of the acidic, secretory, calcium-binding phosphoprotein (SCPP) gene family is essential for vertebrate bone formation (Venkatesh et al., 2014). However, I failed to find orthologs of SCPP genes in the *Lingula* genome, although it contains an ortholog of the secreted protein, acidic, cysteine-rich (SPARC) gene (Fig. 2.11). These analyses show that many of the genes involved in bone formation are derived from genome duplication events in the vertebrate lineage (Venkatesh et al., 2014). 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.

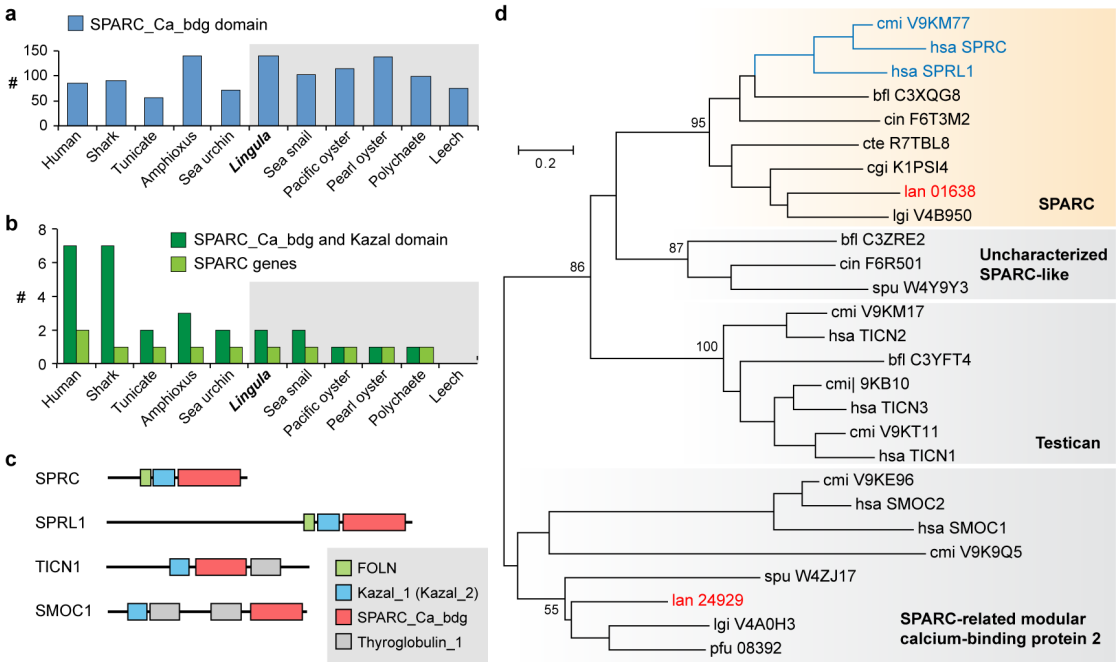


Figure 2.11 | Evolution of SPARC-related genes in *Lingula*.

(**a**) Number of genes with secreted acidic proteins rich in cysteine Ca-binding region domains (SPARC_Ca_bdg, PF10591) in metazoan genomes. Grey box denotes lophotrochozoans. (**b**) Number of proteins with a combination of SPARC_Ca_bdg and Kazal-type serine protease inhibitor domains (Kazal_1, PF00050) (dark green). Number of SPARC genes identified with the BBH approach (light green). (**c**) Domain composition of SPARC-related genes. UniProt ID: SPRC, SPARC; SPRL1, SPARC-like protein 1; TICN1, Testican-1; SMOC1, SPARC-related modular calcium-binding protein 1. Pfam domain: FOLN, Follistatin/Osteonectin-like EGF domain (PF09289); Thyroglobulin_1, Thyroglobulin type-1 repeat (PF00086). (**d**) Phylogeny of SPARC-related genes constructed with 27 genes, Kazal and SPARC_Ca_bdg domains (160 amino acids) using the neighbor-joining method with the JTT model (1,000 bootstrap replicates). Vertebrate lineage with a duplication event of the SPARC gene is labeled in blue. Numbers at the nodes indicate bootstrap support values. Three-letter code: hsa, humans (*Homo sapiens*); cmi, elephant shark (*Callorhinchus milii*); cin, tunicate (*Ciona intestinalis*); bfl, amphioxus (*Branchiostoma floridae*); spu, sea urchin (*Strongylocentrotus purpuratus*); lan, brachiopod (*Lingula anatina*); lgi, sea snail (*Lottia gigantea*); cgi, Pacific oyster (*Crassostrea gigas*); pfu, pearl oyster (*Pinctada fucata*); and cte, polychaete (*Capitella teleta*).

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. 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. Notably, genes

shared between *Lingula* and molluscs, such as calcium-dependent protein kinase and chitin synthase, exhibit high expression in larvae and mantle, indicating that they may also be involved in *Lingula* shell formation.

2.3.6 Conserved molecular mechanisms in biomineralization

Given that genes associated with biomineralization have diverse functions and have been co-opted in different species, I next tested whether there is a conserved upstream mechanism for this process. I focused on one of the ancient metazoan signaling pathways, bone morphogenetic proteins (BMP). Previous studies have demonstrated that BMP signaling plays key roles in biomineralization in both molluscs (Shimizu et al., 2011) and vertebrates (Chen et al., 2012). To explore the possible role of BMP signaling during embryogenesis, I first annotated BMP ligands and receptor-regulated Smad. *Lingula* has orthologs for one *Bmp2/4*, one *Bmp5-8*, and one *Smad/1/5/9*. My embryonic transcriptome showed that *Bmp5-8* and *Smad1/5/9* are expressed maternally, whereas *Bmp2/4* is expressed after the early blastula stage (Fig. 2.12).

To visualize activation sites of BMP signals, I employed immunostaining of nuclear phosphorylated Smad1/5/9 (pSmad), an activated mediator. In *Lingula*, embryonic shells are formed upon mantle lobes beginning at the 1-pair-cirri larval stage (Yatsu, 1902). Interestingly, I discovered that BMP signaling is activated at the anterior margin of the mantle lobe during *Lingula* larval shell formation (Fig. 2.13). 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.

2.3.7 Shell matrix proteins and fibrillar collagens

Proteomic analyses of *Lingula* shells (collaboration with Takeshi Takeuchi) identified a total of 65 shell matrix proteins (SMPs). Using comparative genomics, I showed that the composition of *Lingula* SMPs share the highest similarities with those of amphioxus and molluscs. 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 (Williams et al., 1994). I provided here the first molecular evidence that glycine-rich SMPs are collagens.

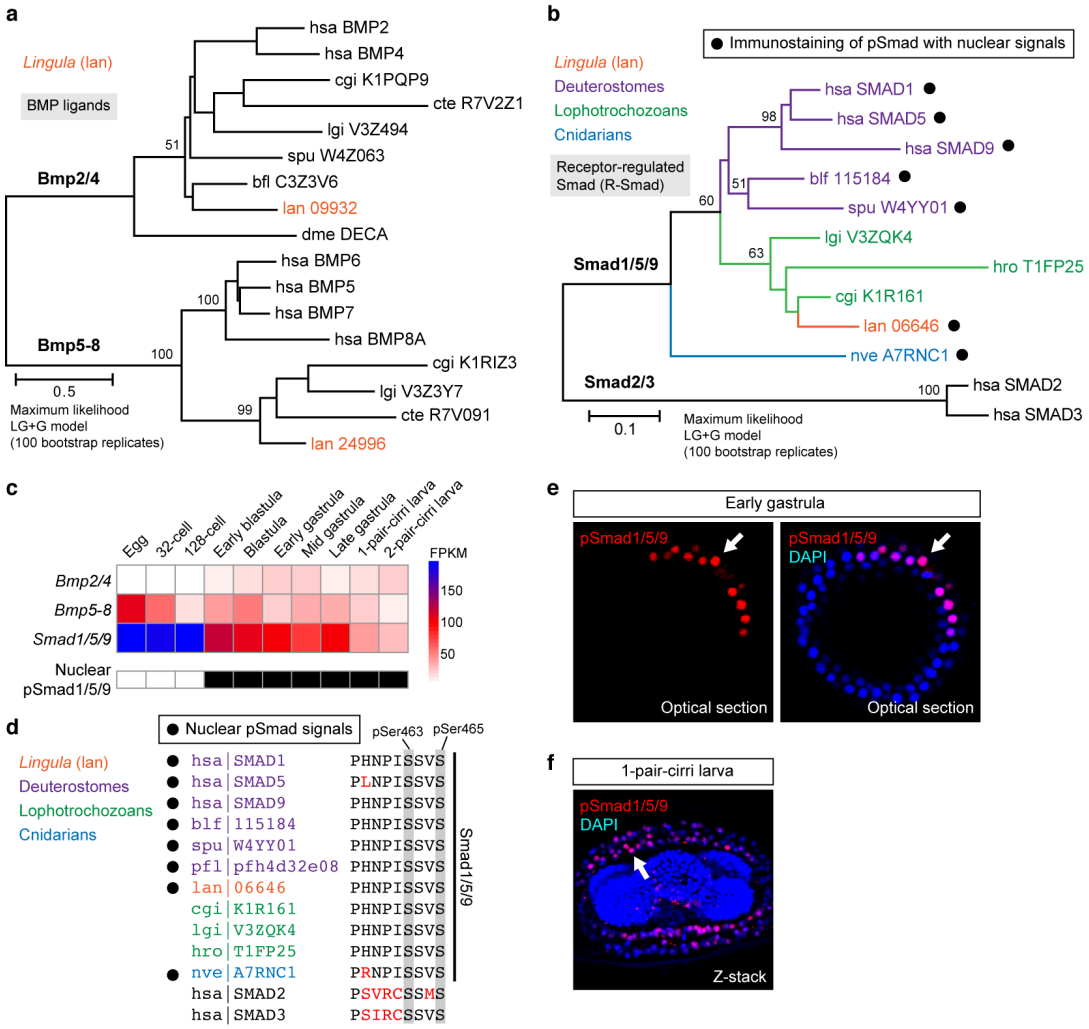


Figure 2.12 | BMP signaling components in *Lingula*.

(**a**) Phylogeny of BMP ligands using 17 genes (364 amino acids). Three-letter code: hsa, humans (*Homo sapiens*); bfl, amphioxus (*Branchiostoma floridae*); spu, sea urchin (*Strongylocentrotus purpuratus*); lgi, sea snail (*Lottia gigantea*); cgi, Pacific oyster (*Crassostrea gigas*); cte, polychaete (*Capitella teleta*); dme, fruit fly (*Drosophila melanogaster*). Proteins are identified by their UniProt IDs. Numbers at the nodes indicate bootstrap support values. (**b**) Phylogeny of receptor-regulated Smad constructed with 12 genes (431 amino acids). The amphioxus sequence is from JGI. hro, leech (*Helobdella robusta*); nve, sea anemone (*Nematostella vectensis*). (**c**) Expression profiles of BMP signaling ligands and mediators. Appearance of nuclear phosphorylated Smad1/5/9 (pSmad) signals is shown in black rectangles. (**d**) Alignment of C-terminus of Smad proteins. Phosphorylated sites of Ser463/465 in human SMAD5 are shaded in grey. Different amino acids compared to SMAD1 are labeled in red. pfl, hemichordate (*Ptychodera flava*; EST ID)(Tagawa et al., 2014). (**e**) Immunostaining of pSmad in early gastrula shows signals with asymmetrical nuclear localization (arrows). Nuclei are labeled with DAPI. (**f**) Nuclear signals of pSmad (arrow) in 1-pair-cirri larva without CellMask staining.

In addition, I also found that many novel SMPs are alanine-rich and have low molecular weights (i.e., amino-acid length ~100-200) (Fig. 2.14a,b). Pfam analysis of *Lingula*

SMPs shows that the most abundant domains are cadherin and collagen, whereas the most abundant proteins contain von Willebrand factor type A (VWA) and epidermal growth factor (EGF) domains (Fig. 2.14c). The domain composition suggests that the shell matrix is derived from extracellular matrix.

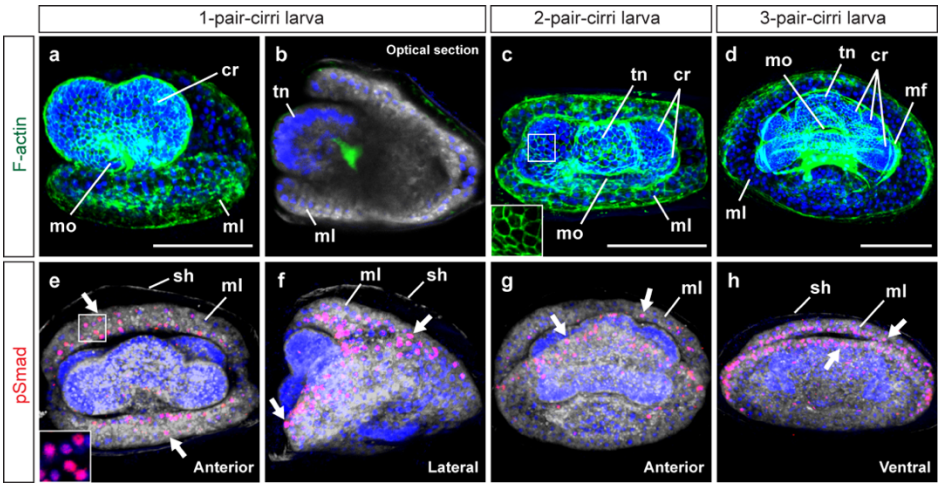


Figure 2.13 | BMP signaling may be involved in larval shell formation.
(**a**–**h**) Confocal images of *Lingula* larvae from 1-pair-cirri to 3-pair-cirri stages. (**a**–**d**) Filamentous actin (F-actin) staining shows the cellular structure of larvae. Cytoplasmic membranes and nuclei are labeled with CellMask (grey) and DAPI (blue), respectively. Inset in (c) shows the cell boundary at higher magnification. (**e**–**h**) Activation of BMP signaling is monitored by nuclear signals of phosphorylated Smad1/5/9 (pSmad, red). Inset in (e) shows nuclear pSmad signals at higher magnification. Note that signals are localized at the margin of mantle lobes (arrows). Orientation of embryos is indicated at the bottom-right corner of each panel. Abbreviations: mo, mouth; cr, cirrus (cirri); ml, mantle lobe; tn, tentacle; mf, muscle fiber; sh, embryonic shell. Scale bars, 50 μm.

I further examined the expression profile of these SMPs. I found that 26 SMPs are expressed ubiquitously in all adult tissues, indicating that they have functions other than shell formation (Fig. 2.15). On the other hand, 20 SMPs exhibited specific expression in the mantle. These include collagen, chitinase, glutathione peroxidase, hephaestin, hemicentin, and peroxidasin. 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 (Marin et al., 2008) and coral skeletons (Ramos-Silva et al., 2013). I failed to find secreted acidic proteins among *Lingula* SMPs. Instead, I found that there are novel alanine-rich SMPs with a three-helix bundle structure that may confer elastic properties upon the *Lingula* shell (Fig. 2.16).

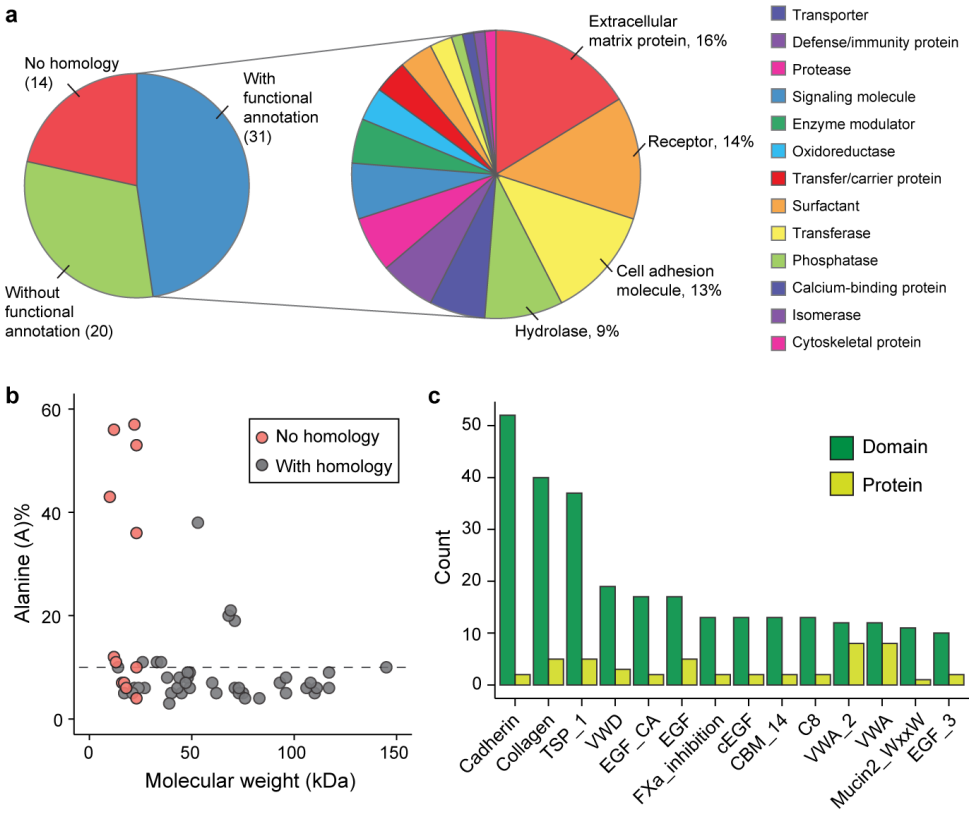


Figure 2.14 | Characterization of *Lingula* SMPs.

(**a**) Distribution of functional classifications of 65 SMPs. Biological processes are shown for the 31 SMPs that have functional annotation data. (**b**) Distribution of alanine composition and molecular weight of *Lingula* SMPs. Seven SMPs with molecular weights greater than 150 kDa are not shown here. The dashed line indicates the 10% in terms of alanine content. (**c**) Top 20% domain distribution of SMPs with significant Pfam hits. Dark green, total number of a detected domains in the SMPs; light green, number of SMPs with that domain shown below. TSP_1, thrombospondins 1; VWD, von Willebrand factor type D domain; EGF_CA, calcium-binding EGF domain; CBM, carbohydrate-binding module; C8, 8 conserved cysteine residues; VWA, Von Willebrand factor type A domain.

Since the formation of vertebrate bone and *Lingula* shell rely on mineral deposition upon fibrillar collagens (Nair et al., 2013; Williams et al., 1994), I further examined the evolution of fibrillar collagens. Vertebrate fibrillar collagens are subdivided into three major clades carrying COLFI domains (Kawasaki et al., 2009). However, my 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. 2.17a,b). Intriguingly, some fibrillar collagens likely originated by tandem duplication (Fig. 2.17c). In addition, I found that *Lingula* contains the highest number of proteins having both

EGF and collagen domains among bilaterians. 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.

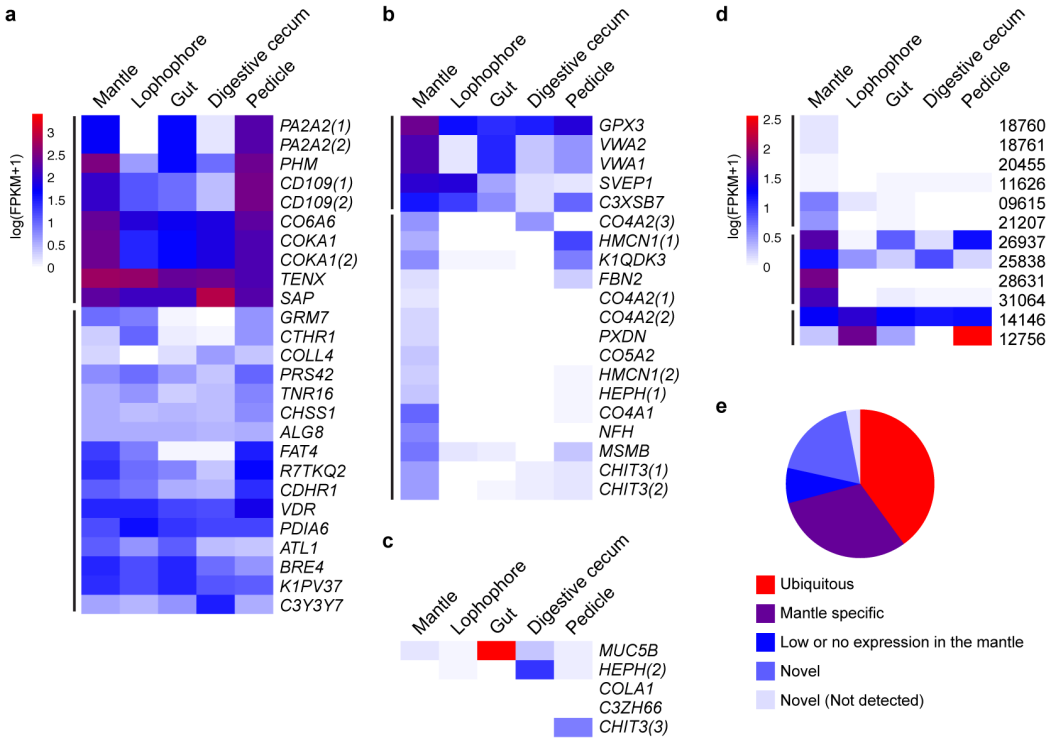


Figure 2.15 | Expression of SMPs in the adult tissues.
(**a**) Expression profile of SMPs with detectable homology, expressed ubiquitously in adult tissues. Vertical lines, clustered groups based on expression pattern. Paralogs are marked by parentheses. (**b**) Expression profile of SMPs with detectable homology, expressed highly or specifically in mantle tissue. (**c**) Expression profiles of SMPs with detectable homology, expressed weakly in mantle tissue. (**d**) Expression profiles of SMPs without detectable homology (novel) shown among *Lingula* gene models. (**e**) Summary of the expression of SMPs. FPKM, fragments per kilobase of transcript per million mapped reads. Gene names are the human entry names in UniProt.

2.3.8 *Lingula* shell formation and evolution of biomineralization

Mollusc phylogenomic and shell proteomic studies show that mollusc shells may have different origins (Jackson et al., 2010; Kocot et al., 2011; Smith et al., 2011). Although all molluscs use calcium carbonate, different modes of biomineralization have been adapted among brachiopods. Only the Linguliformea makes shells with calcium phosphate (Williams et al., 1994). I 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) (Venkatesh et al., 2014) much

later than lingulid shells (~520 MYA, early Cambrian) (Zhang et al., 2005), it is perhaps not surprising that vertebrate bone and *Lingula* shell have different genetic origins. Although downstream biomineralization-related genes are diverse, I speculate that the metazoan ancestor might use a core set of ancient signaling proteins, such as BMPs and their downstream mediators, to initiate the biomineralization process. I 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. This suggests that metazoan biomineralization likely originated from a calcium-regulated, extracellular matrix system.

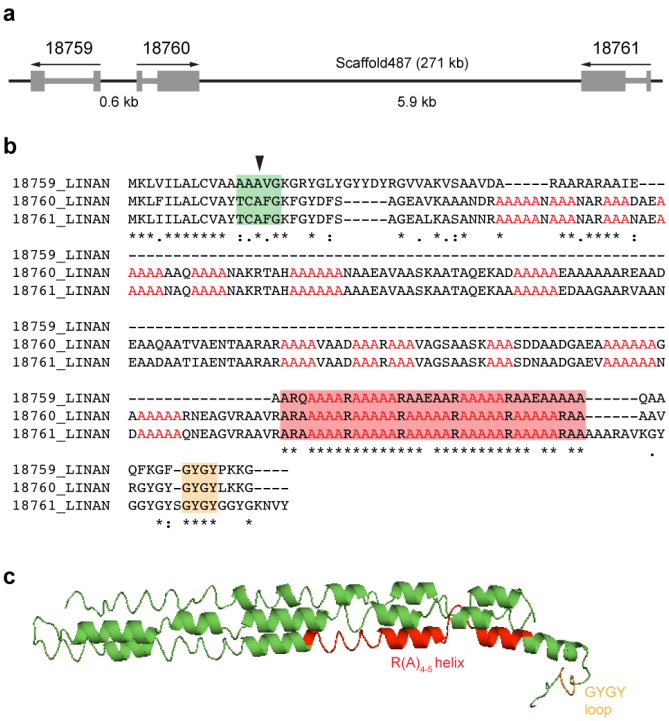


Figure 2.16 | A tandem duplication of novel genes for SMPs.
(**a**) An example of tandem duplicated SMPs. The gene orientation (arrows) and the distance among genes in scale on the scaffold are shown. Grey boxes, exons. (**b**) Multiple alignments were conducted with Clustal Omega. Conserved poly-alanine (>3) is colored in red. Green box, signal peptide predicted by SignalP where the arrowhead indicates the cleavage site. Red box, conserved $R(A)_{4-5}$ domain. Orange box, conserved GYGY motif. Asterisks, fully conserved; colons, strongly similar; periods, weakly similar. (**c**) Predicted three-helix bundle structure of gene model 18761_LINAN by I-TASSER (estimated TM-score, 0.4; RMSD, 12.2 Å) with a TM-score 0.795 to computationally designed three helix bundle (PDB ID: 4TQL). A TM-score >0.5 indicates a model of correct topology not coming from a random similarity. Conserved $R(A)_{4-5}$ helix and GYGY loop are colored in red and orange, respectively.

Furthermore, I also discovered that Hox4, tyrosinase, chitin synthase, perlucin, chitinase, peroxidasin, mucin, and VWA protein are common shell formation-associated

components shared by *Lingula* and molluscs, suggesting that this fundamental gene set was used by their common ancestor. 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 matrix (Drake et al., 2013; Ramos-Silva et al., 2013). 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.

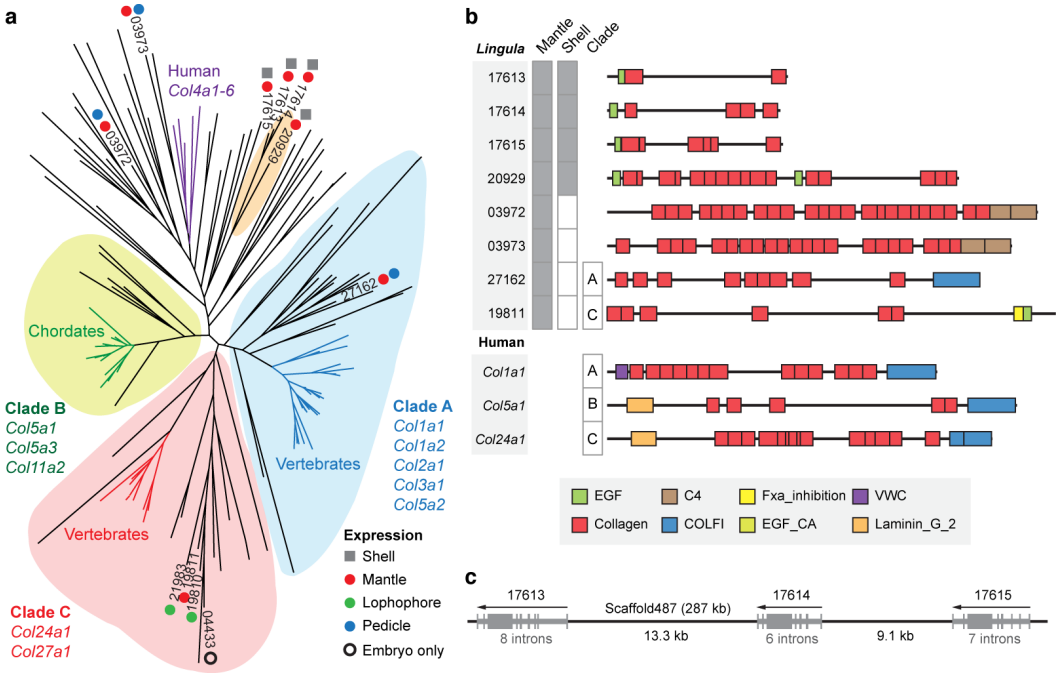


Figure 2.17 | Fibrillar collagens in *Lingula* and vertebrates have different origins.
(**a**) Phylogenetic analysis of the collagen triple helix region using the maximum likelihood method with the LG model (159 genes, 542 amino acids, 100 bootstrap replicates). Expression of gene models is supported by the shell proteome (square) and transcriptome (circle). Numbers indicate *Lingula* gene IDs. (**b**) Domain structure of selected collagens. Expression of proteins is shown in grey boxes. EGF, epidermal growth factor (EGF)-like domain; C4, type IV collagen C4 domain; Fxa_inhibition, coagulation Factor Xa inhibitory site; VWC, von Willebrand factor type C domain; COLFI, fibrillar collagens C-terminal domain; EGF_CA, calcium-binding EGF domain. (**c**) Genomic organization of tandem-duplicated collagen genes expressed in mantle and shell. Arrows indicate the direction of transcription. Grey boxes denote exons.

Given different chemistry and genetic components of their shells, I 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 ions (Schonitzer and Weiss, 2007). In summary, I propose a possible

mechanism for *Lingula* shell formation (Fig. 2.18). First, the interaction of myosin head-containing chitin synthases and actin filaments may translate the cytoskeleton organization into an extracellular chitin scaffold. Chitinase in the shell matrix then possibly remodels the chitin scaffold to facilitate the interaction of chitin and chitin-binding proteins. Calcium-binding proteins likely regulate the calcium concentration in the shell matrix and initiate calcium phosphate deposition, together with other structural proteins, such as EGF domain-containing fibrillar collagens and alanine-rich proteins.

2.4 Discussion

I show that the *Lingula* genome has been evolving, instead of remaining static, as one would expect in a genuine 'living fossil.' Combining transcriptomic and proteomic data, I 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 vertebrates (Knoll, 2003; Williams et al., 1994). All data presented in this study indicate that *Lingula* and bony vertebrates have adapted different mechanisms for hard tissue formation. Vertebrates likely 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 (Wada et al., 2006). Similarly, studies of biomineralization genes in sea urchins and molluscs show that there are extensive differences in their expressed gene sets. Since molluscs, brachiopods, echinoderms, and vertebrates contain different sets of biomineralization genes, biomineral proteins must have arisen independently among metazoans on several occasions (Jackson et al., 2010; Livingston et al., 2006). Taken together, my genomic, transcriptomic, and proteomic analyses of *Lingula* biomineralization show similar patterns to those in molluscs (Jackson et al., 2010) and corals (Ramos-Silva et al., 2013), where co-option, domain shuffling, and novel genes are the fundamental mechanisms for metazoan biomineralization. Finally, although I present data suggesting that Brachiopoda is closer to Mollusca than to Annelida, the phylogenetic position of brachiopods related to other lophotrochozoans remains to be elucidated. The decoded *Lingula* genome provides information essential for such future studies.

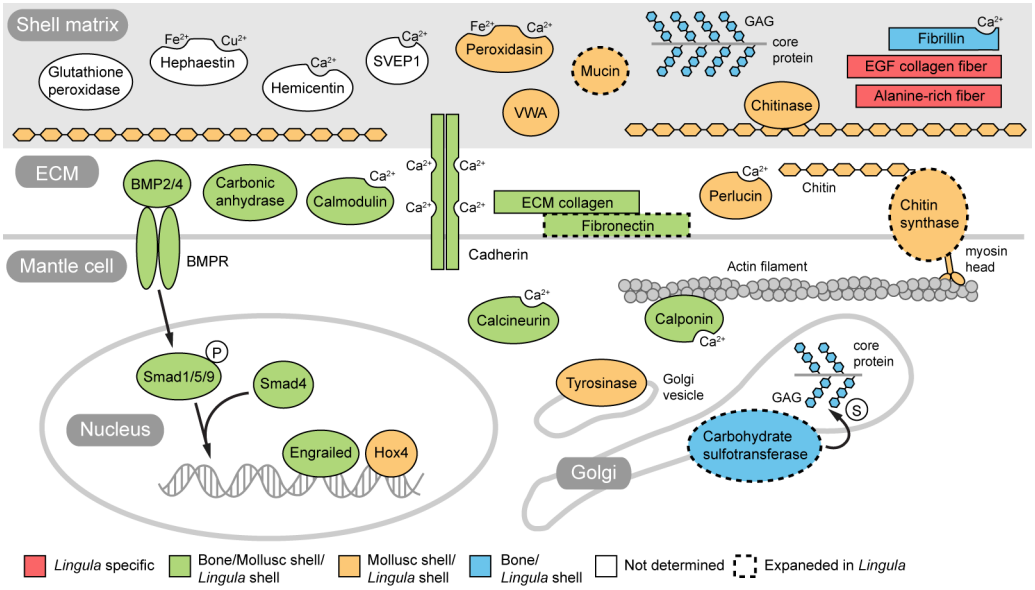


Figure 2.18 | Genes related to biomineralization expressed during *Lingula* shell formation.
A schematic illustration of genes involved in *Lingula* biomineralization identified in the present study. Genes are colored by their known functions in shell or bone formation in molluscs and vertebrates, respectively. Dashed outlines indicate gene families expanded specifically in *Lingula*. Abbreviations: ECM, extracellular matrix; GAG, glucosaminoglycan; SEVP1, Sushi von Willebrand factor type A, EGF and pentraxin domain-containing protein 1; WVA, von Willebrand factor type A domain containing protein; BMPR, bone morphogenetic protein receptor. Proteins with ion-binding domains are labeled with Ca^{2+}, Fe^{2+}, or Cu^{2+}. P and S in white circles indicate phosphate and sulfate groups.

3 Phoronid and nemertean genomes reveal lophotrochozoan evolution and the origin of bilaterian heads

3.1 Introduction

Lophotrochozoans comprise a major clade within protostomes. They represent more than one-third of known marine animals and play important ecological roles (Appeltans et al., 2012). Protostomes consist of two sister groups, spiralians (animals that mostly exhibit spiral cleavage) and ecdysozoans (animals that shed their exoskeleton). As a subgroup within spiralians, most lophotrochozoans possess either lophophore or trochophore larvae during the planktonic stage. Lophotrochozoans include annelids, molluscs, nemerteans (ribbon worms), phoronids (horseshoe worms), ectoprocts (bryozoans, or moss animals) and brachiopods (lamp shells), although many phylogenetic relationships within the group remain unresolved (Dunn et al., 2008; Laumer et al., 2015; Nesnidal et al., 2013) (Fig. 3.1). Molecular phylogenetics suggests that phoronids and nemerteans are closely related (Dunn et al., 2008), yet these two phyla have diverse body plans and exhibit no morphological synapomorphic traits. In particular, they have different lifestyles with distinct larval forms and possessing different types of feeding apparatus. For example, phoronids are sessile filter feeders with ciliated tentacles called lophophores, the feeding apparatus shared by ectoprocts and brachiopods. On the other hand, nemerteans are unsegmented worms. Mostly as predators, they have an eversible proboscis derived from the rhynchocoel (a fluid-filled tubular chamber) for capturing prey and defense. Given the incompatibility of molecular and morphological data, the origins of nemerteans and phoronids have been obscure.

Our genomic understanding of protostomes is largely based on the comparative study of the model systems within ecdysozoans, such as fruit flies and nematodes. Although most developmental toolkit genes are shared between protostomes and deuterostomes, some genes such as *Nodal*, a member of the transforming growth factor-β (TGFβ) superfamily that is required for left–right asymmetry in deuterostomes, are lost in ecdysozoans, but present in lophotrochozoans (Grande and Patel, 2009). Similarly, some gene families, such as innate immunity-related genes, are highly reduced in ecdysozoans, but more complex in lophotrochozoans (Halanych and Kocot, 2014; Zhang et al., 2015). Recent genomic studies have further shown that annelids and molluscs share several genomic features such as gene family size and conserved orthologous gene clusters with invertebrate deuterostomes (e.g., amphioxus and sea urchins) (Simakov et al., 2013). This observation raises the question

whether lophotrochozoans represent some bilaterian ancestral features with invertebrate deuterostomes that are likely lost in ecdysozoans and other lineages during protostome evolution.

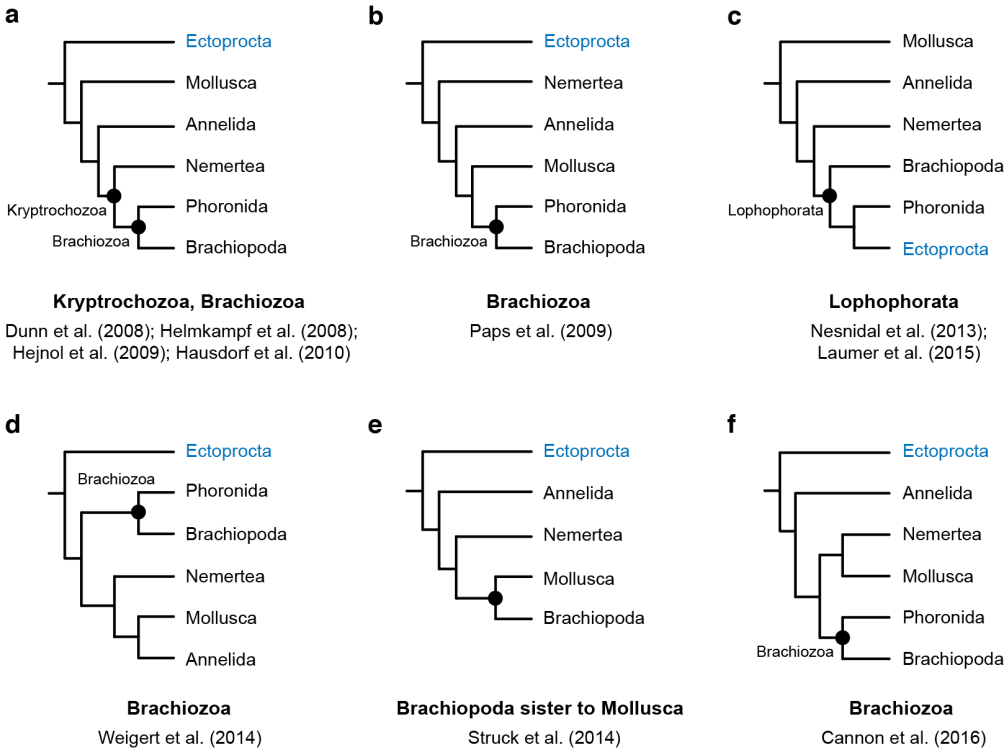


Figure 3.1 | Current hypotheses about relationships of Ectoprocta to other major lophotrochozoans. Hypotheses are listed in chronological order. (**a**) Kryptrochozoa hypothesis (monophyly of Nemertea, Phoronida, and Brachiopoda) with Ectoprocta as an outgroup (Dunn et al., 2008; Hausdorf et al., 2010; Hejnol et al., 2009; Helmkampf et al., 2008). (**b**) Brachiozoa hypothesis (monophyly of Phoronida and Brachiopoda) with Ectoprocta as an outgroup (Paps et al., 2009a). (**c**) Lophophorata hypothesis (monophyly of Phoronida, Ectoprocta, and Brachiopoda) (Laumer et al., 2015; Nesnidal et al., 2013). (**d**) Brachiozoa as a sister group to the rest of lophotrochozoans with Ectoprocta as an outgroup (Weigert et al., 2014). (**e**) Brachiopoda as a sister group to Mollusca with Ectoprocta as an outgroup (Struck et al., 2014). (**f**) Brachiozoa as a sister group to the group of Nemertea and Mollusca with Ectoprocta as an outgroup (Cannon et al., 2016). Major hypotheses are given at the nodes shown as black circles. Ectoprocta is labelled in blue.

Here I present genomes of the phoronid *Phoronis australis* and the nemertean *Notospermus geniculatus*, and explored lophotrochozoan evolution by comparative genomics among metazoans. With both genome and transcriptome data, my phylogenetic analyses provide evidence that nemerteans are sisters to lophophorates (phoronids, ectoprocts and brachiopods). My results clearly show that lophotrochozoans have a different evolutionary

history than remaining spiralians, such as rotifers, flatworms and tapeworms. In particular, lophotrochozoans retain a basic set of bilaterian gene repertoire, which is likely lost in ecdysozoans and other spiralian lineages. Unexpectedly, genes specifically expressed in lophophores of phoronids and brachiopods are strikingly similar to those employed in vertebrate head formation, although novel genes, expanded gene families and redeployment of toolkit genes also contribute to the unique molecular identity of lophophores. Furthermore, I provide examples of lineage-specific genomic features in lophotrochozoans, such as the expansion of innate immunity and toxin-related genes. Taken together, my study reveals a dual nature of lophotrochozoan genomes, showing both conservative and innovative characteristics.

3.2 Methods

3.2.1 Biological materials

Adult phoronids (*Phoronis australis*) were collected in Kuroshima Island near Ushimado town, Okayama, Japan. Adult nemerteans (*Notospermus geniculatus*) were collected in front of the Ushimado Marine Institute, Okayama University, Japan. After starvation, genomic DNA was extracted from intact adults using the phenol/chloroform method.

3.2.2 Genome sequencing and assembly

The *Phoronis* and *Notospermus* genomes were sequenced using next-generation sequencing on Illumina MiSeq, HiSeq 2500 and Roche 454 GS FLX+ platforms (Figs 3.2 and 3.3). Paired-end libraries (286–1,100 bp) were prepared using the NEBNext Ultra DNA Library Prep Kit for Illumina (New England Biolabs). Paired-end reads were sequenced to obtain 71 and 127 Gb of data from *Phoronis* and *Notospermus* samples, respectively, using Illumina MiSeq (read length 250–400 bp). A mate pair library from 3 kb DNA fragments was prepared using the Cre-Lox recombination approach. Other mate pair libraries generated from 1.5 to 20 kb DNA fragments were size selected with the automated electrophoresis platforms SageELF or BluePippin (Sage Science) and prepared using the Nextera Mate Pair Sample Prep Kit. Mate pair libraries were sequenced to obtain 38 and 100 Gb of data from *Phoronis* and *Notospermus* samples, respectively, using Illumina HiSeq 2500 and MiSeq platforms.

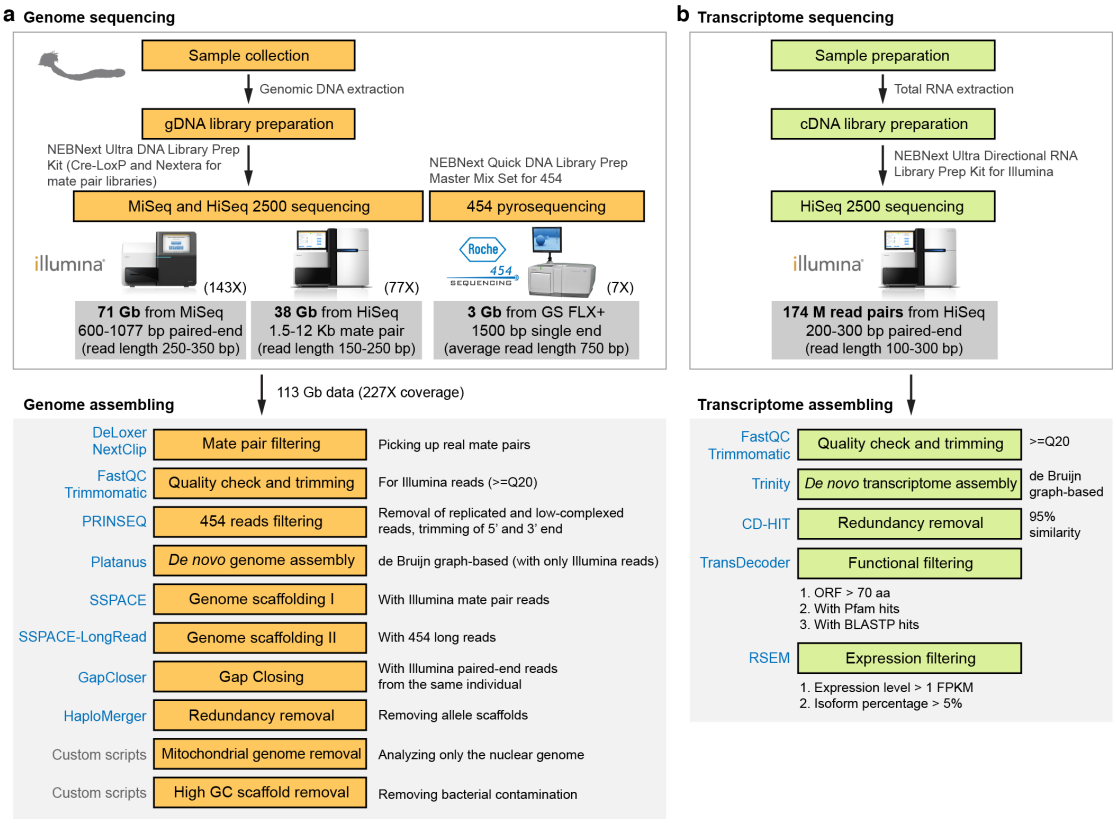


Figure 3.2 | Sequencing and assembly of the *Phoronis* genome and transcriptome.
(**a**) Genomic DNA from a whole adult was extracted for genome sequencing using Illumina MiSeq and HiSeq as well as Roche 454 platforms. 113 Gb of data were obtained with approximately 227-fold coverage of the 498-Mb *Phoronis* genome. The genome was assembled with Platanus and SSPACE. After further scaffolding with 454 long reads, allele scaffolds were removed with HaploMerger. (**b**) One embryonic and five adult tissues were collected for RNA-seq and 174 million read pairs were assembled *de novo* using Trinity. Redundant transcripts were removed with CD-HIT and then functionally filtered with TransDecoder. FPKM, fragments per kilobase of transcript per million mapped reads. ORFs, open reading frames. Programs used here are listed in blue on the left. Comments are on the right.

After quality control checks with FastQC (v0.10.1), Illumina reads were quality filtered (Q score ≥ 20) and trimmed with Trimmomatic (v0.33). Roche 454 reads were filtered with PRINSEQ (v0.20.3) to remove duplicated and low-complexity sequences. Mate pair reads prepared from Cre-LoxP and Nextera were filtered with DeLoxer (http://genomes.sdsc.edu/downloads/deloxer/) and NextClip (v0.8) (Leggett et al., 2014), respectively. To overcome high heterozygosity, genomes were assembled using a de Bruijn graph-based assembler, Platanus (v1.2.4) (Kajitani et al., 2014). Scaffolding was conducted by mapping Illumina paired-end and mate pair reads to contigs using SSPACE (v3.0) (Boetzer et al., 2011). For the *Phoronis* genome, a set of long 454 reads (750 bp) with 3 Gb of data was used for scaffolding with SSPACE-LongRead (v1-1) (Boetzer and Pirovano, 2014).

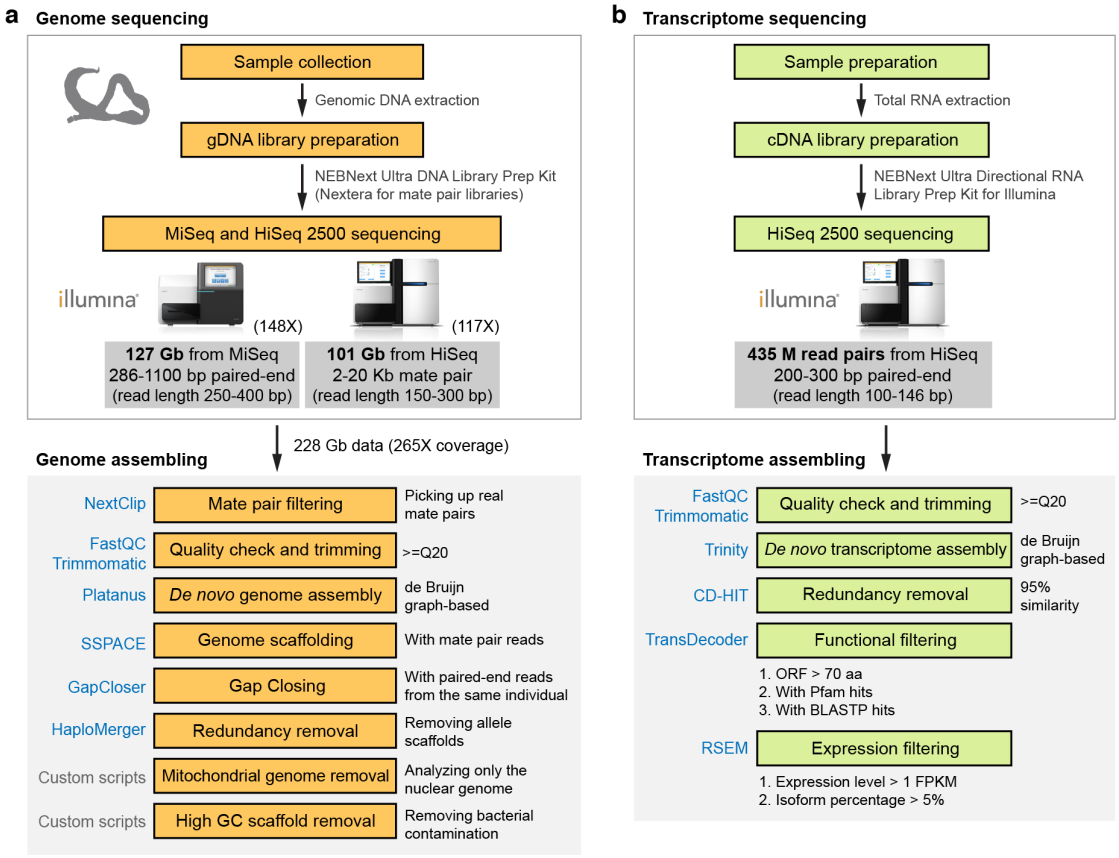


Figure 3.3 | Sequencing and assembly of the *Notospermus* genome and transcriptome.
(**a**) Genomic DNA from a whole adult was extracted for genome sequencing using Illumina MiSeq and HiSeq platforms. 228 Gb of data were obtained with approximately 265-fold coverage of the 859-Mb *Notospermus* genome. The genome was assembled with Platanus and SSPACE. Allele scaffolds were removed with HaploMerger. (**b**) One embryonic and fifteen adult tissues were collected for RNA-seq and 435 million read pairs were assembled *de novo* using Trinity. Redundant transcripts were removed with CD-HIT and then functionally filtered with TransDecoder. FPKM, fragments per kilobase of transcript per million mapped reads. ORFs, open reading frames. Programs used here are listed in blue on the left. Comments are on the right.

Gaps in scaffolds were filled with GapCloser (v1.12-r6) (Luo et al., 2012). Redundant allele scaffolds were removed with HaploMerger (2_20151106) (Huang et al., 2012). The quality of genome assemblies was assessed with N(X) graphs using QUAST (v3.1). Mitochondrial genomes and high GC scaffolds possibly derived from bacterial contamination were removed using custom Perl scripts. Genome sizes and heterozygosity rates were estimated by *k*-mer analysis using SOAPec (v2.01) and GCE (v1.0.0) (Luo et al., 2012) as well as JELLYFISH (v2.0.0) (Marçais and Kingsford, 2011) and a custom Perl script. Genome assembly completeness was assessed with CEGMA (v2.5) (Parra et al., 2007).

3.2.3 Transcriptome sequencing and assembly

RNA sequencing (RNA-seq) of adult tissues and embryonic stages was performed using the Illumina HiSeq 2500 platform. In total, 174 and 435 million RNA-seq read pairs from 6 *Phoronis* and 15 *Notospermus* samples, respectively, were generated (read length 100–300 bp). After quality checking and trimming of raw sequencing reads, transcripts were *de novo* assembled with Trinity (v2.1.0) (Haas et al., 2013). Transcript isoforms with high similarity (≥ 95%) were removed with CD-HIT-EST (Li and Godzik, 2006). Transcript abundance was estimated with Bowtie (v2.1.0) (Langmead and Salzberg, 2012) and RSEM (v1.2.26) (Li and Dewey, 2011) by mapping reads back to the transcript assembly. The trimmed mean of M-values (TMM)-normalized expression values in fragments per kilobase of transcript per million mapped reads (FPKM) were used to estimate relative expression levels across samples.

To reduce the data complexity, functional filtering with TransDecoder (v2.0.1) (Haas et al., 2013) was applied with the following three criteria: (1) open reading frames (ORFs) larger than 70 amino acids; (2) sequences with HMMER (v3.1b2) hits against Pfam database (Pfam-A 29.0; 16,295 families); and (3) sequences with BLASTP (v2.2.29+) hits against Swiss-Prot database (20160122; 550,299 sequences). An expression filtering was applied with two criteria: (1) expression levels ≥ 1 FPKM in at least one sample; and (2) transcript isoforms with abundances > 5%.

3.2.4 Repeat analysis

Regions of repetitive sequences in the genomes were identified with RepeatScout (v1.0.5) (Price et al., 2005) using default settings (i.e. sequence length larger than 50 bp and occurring over 10 times). Repetitive sequences were masked with RepeatMasker (http://www.repeatmasker.org/; v4.0.6). Transposable elements were annotated with TBLASTX and BLASTN searches against Repbase for RepeatMasker (v20150807). Repeat landscape (Kimura genetic distance) was calculated with the Perl script RepeatLandscape.pl bundled within RepeatMasker (v4.0.5+).

3.2.5 Gene prediction and annotation

Non-exon (i.e. repeat) hints were generated with RepeatScout and RepeatMasker. Intron hints from spliced alignment of RNA-seq reads were generated using TopHat (v2.0.9) (Kim et al., 2013) and Bowtie (v2.1.0) (Langmead and Salzberg, 2012) with the two-step method: (1)

genome assembly mapping and (2) exon-exon junction mapping. Exon hints were generated from a spliced alignment of transcriptome assembly using BLAT (v.35). Gene structure was annotated by extraction of ORFs with PASA (v2.0.2). Gene models were predicted with trained AUGUSTUS (v3.2.1) (Stanke et al., 2008) with repeat, intron and exon hints on the soft-masked genome assemblies. KEGG orthology was assigned using the KEGG Automatic Annotation Server. Gene models were annotated with protein identity and domain composition by BLASTP and HMMER searches against Swiss-Prot and Pfam databases, respectively (Fig. 3.4).

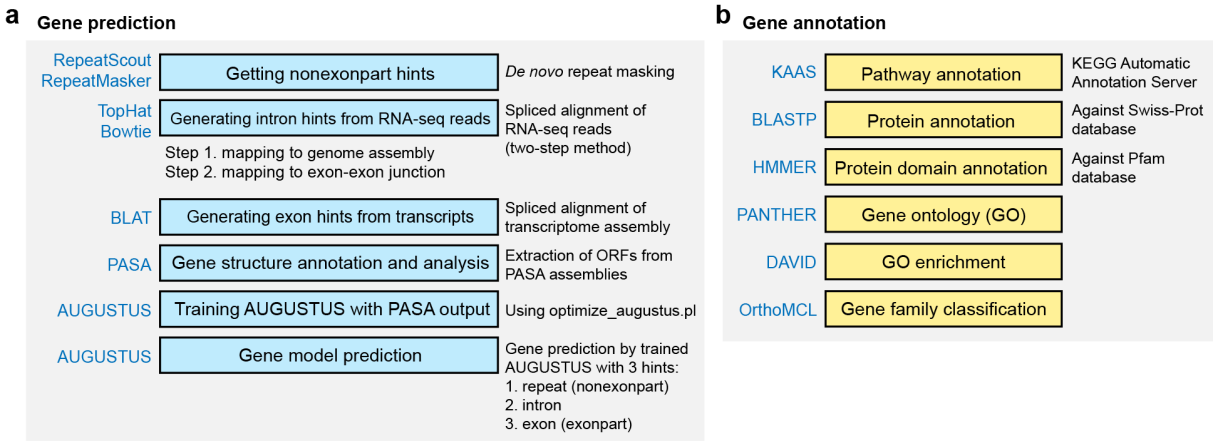


Figure 3.4 | Gene prediction and annotation pipeline.

(**a**) Non-exon-part hints (i.e. repetitive sequences) were generated during repeat masking. Transcript information from RNA-seq was used to produce hints by spliced alignment (Bowtie for intron hints; BLAT for exon hints). Gene models were predicted with trained AUGUSTUS and hints from repeat, intron, and exon information. (**b**) Gene annotation was performed based on protein sequence homology. Programs used here are listed in blue on the left. Comments are on the right.

3.2.6 Gene family analysis

After all-to-all BLASTP searches against 31 selected metazoan proteomes, orthologous groups (OG) were identified with OrthoMCL (v2.0.9) (Fischer et al., 2011) using a default inflation number (I = 1.5). Venn diagrams were plotted with jvenn (Bardou et al., 2014). Gene ontology (GO) annotation was performed with PANTHER (v10.0) (Mi et al., 2013) using the PANTHER HMM scoring tool (pantherScore.pl). GO enrichment analysis was conducted with DAVID (v6.8) (Huang da et al., 2009). Gene family gain-and-loss was estimated using CAFE (v3.1) (De Bie et al., 2006). Principal component analysis was performed using the R package, prcomp.

3.2.7 Phylogenetic analysis

Genome-based orthologs with one-to-one relationships were selected with custom Perl scripts from OrthoMCL orthologous groups. Orthologs identified from transcriptome data with many-to-many relationships were selected with HaMStR (v13.2.3) (Ebersberger et al., 2009). Sequence alignments were performed with MAFFT (v7.271) (Katoh et al., 2002). Unaligned regions were trimmed with TrimAl (v1.2rev59) (Capella-Gutiérrez et al., 2009). Species trees were constructed with RAxML (v8.2.4) (Stamatakis, 2014) using the maximum likelihood method with the LG, LG4M and LG4X models. Bayesian analyses were performed with PhyloBayes (v3.3f) (Lartillot et al., 2009) using CAT+GTR model with the first 1,000 trees as a burn-in.

3.2.8 Microsynteny analysis

At least three orthologs on the same scaffold shared between two species were considered as microsyntenic blocks as previously described (Luo et al., 2015). In brief, after assigning orthologs with a universal OG identifier using OrthoMCL, genomic locations of orthologs among different species were compared. All-to-all pairwise comparison was conducted with genome GFF (general feature format) files and OrthoMCL outputs using custom Perl scripts.

3.2.9 Transcriptome analysis

To identify transcriptomic similarities between tissues, orthologs among species were identified using the bidirectional best hits (i.e. reciprocal BLAST) approach. To capture the full set of differentially expressed orthologues in the *Notospermus*, *Phoronis* and *Lingula* genomes, two approaches based on different strategies of orthology assignment and statistical criteria were conducted. The first method identifies differentially expressed genes ($p < 0.001$, fold change > 4) regardless of initial orthology assignment. Tissue-specific genes based on orthology to human proteomes were then assigned, using OrthoMCL and the orthology assignment pipeline (Fig. 3.5). This approach was applied because methods like BBH often misses orthologues if gene families in the genome are highly expanded (Dalquen and Dessimoz, 2013). Also, this method allows the identification of lineage-specific (i.e. genes with no annotation) differentially expressed genes.

The second approach utilizes the 8,650 orthologues used for calculating transcriptomic similarity identified through BBH (i.e. the approach with orthology assignment first and then differential expression analysis). In the second method, we focused on shared orthologues (i.e. non-lineage-specific genes) expressed in *Notospermus* in both a head-specific and

lophophore-specific manner. Tissue-specific genes (fold change > 2 to all other tissues) were further analyzed to discover common elements using Venn diagrams. The core gene set shared by the head of *Notospermus* and lophophores of *Phoronis* and *Lingula* were then identified. The final results presented here are the combination of both approaches.

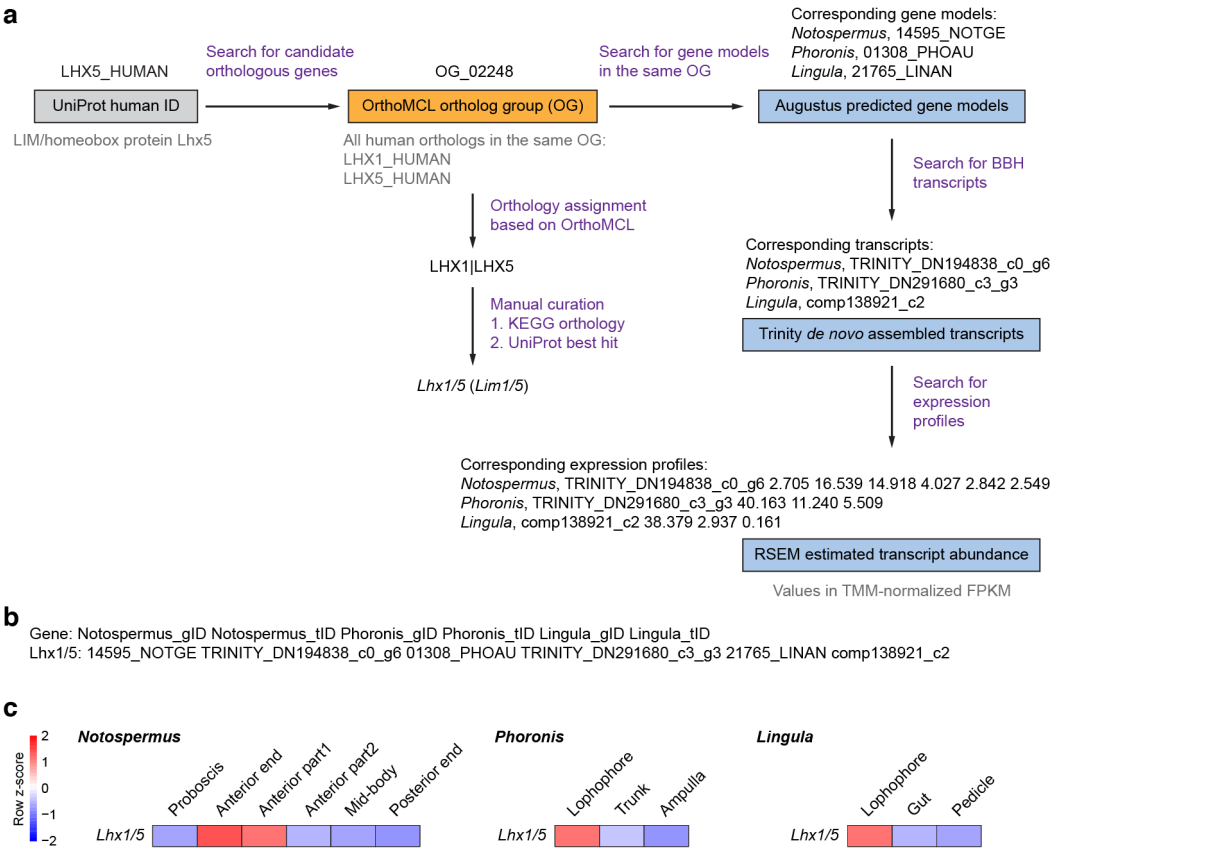


Figure 3.5 | Schematic workflow of orthology assignment in selecting head marker genes.
(**a**) Pipeline of orthology assignment. (**b**) Example of the output format. (**c**) Final results in heatmap format. Searching for Lhx1/5 orthologue in the *Notospermus*, *Phoronis* and *Lingula* genomes is shown as an example here. BBH, bidirectional best hits. Orthology is assigned based on the results of OrthoMCL, KEGG and UniProt best hits.

Spearman's and Pearson's correlation coefficients were calculated as previously described (Luo et al., 2015). Differential expression analysis was conducted with a Trinity bundled Perl script (run_DE_analysis.pl). Heat maps and clustered matrices were created using R (v3.2.4) with the package Bioconductor (v3.0) and pheatmap (v1.0.8).

3.3 Results

3.3.1 Genome characterization

With the technical assistance from the DNA sequencing section, I sequenced two lophotrochozoan genomes using random shotgun approaches with at least 220-fold coverage using Illumina MiSeq and HiSeq, as well as Roche 454 platforms. The haploid assembly sizes of the phoronid *Phoronis australis* and the nemertean *Notospermus geniculatus* genomes are 498 and 859 Mb, respectively, with the N50 lengths of the assembled scaffolds of 655 and 239 kb, respectively (Table 3.1).

Table 3.1 | Summary of nemertean, phoronid, and brachiopod genomic features

Species	*Notospermus geniculatus*	*Phoronis australis*	*Lingula anatina*†
Phylum	Nemertea	Phoronida	Brachiopoda
Common name	Ribbon worms	Horseshoe worms	Lamp shells
Genome size (Mb)	859	498	406
Sequencing coverage	265-fold	227-fold	226-fold
Number of scaffolds	11,108	3,984	2,677
Scaffold N50 (kb)	239	655	460
GC content (%)	42.9	39.3	36.4
Repeats (%)	37.5	39.4	23.3
Number of genes	43,294	20,473	29,907
Gene density (per Mb)	50.4	41.1	73.7
Mean gene size (bp)	8,223	14,590	7,725
Mean transcript size (bp)	1,448	1,587	1,551
Mean intron per gene	5.2	7.4	7.3
Mean intron size (bp)	1,308	1,744	840

†An updated *Lingula* genome with improved scaffolding and gene model prediction was included for comparison with phoronids. The *Notospermus* and *Phoronis* genomes are newly published in this study.

The genome size and assembly quality are comparable to those of other lophotrochozoans, such as the polychaete *Capitella teleta* (324 Mb) (Simakov et al., 2013), the Pacific oyster *Crassostrea gigas* (558 Mb) (Zhang et al., 2012) and the brachiopod *Lingula anatina* (406 Mb) (Luo et al., 2015). I estimated that the *Phoronis* and *Notospermus* genomes contain 20,473 and 43,294 protein-coding genes, respectively, with the support of deep RNA sequencing (RNA-seq) data obtained from 21 libraries, including embryonic stages and adult tissues. Both *Phoronis* and *Notospermus* genomes exhibit high heterozygosity (1.2 and 2.4%, respectively). The abundance of repetitive sequences contributes to the increased size of their genomes (39.4 and 37.5%, respectively). In particular, although the intron-exon structure (8 exons and 7 introns, on average) is similar

between *Phoronis* and *Lingula*, the insertions of transposable elements (TEs) into the introns result in doubling of the *Phoronis* gene size (14,590 bp) compared to that of *Lingula* (7,725 bp) (Fig. 3.6).

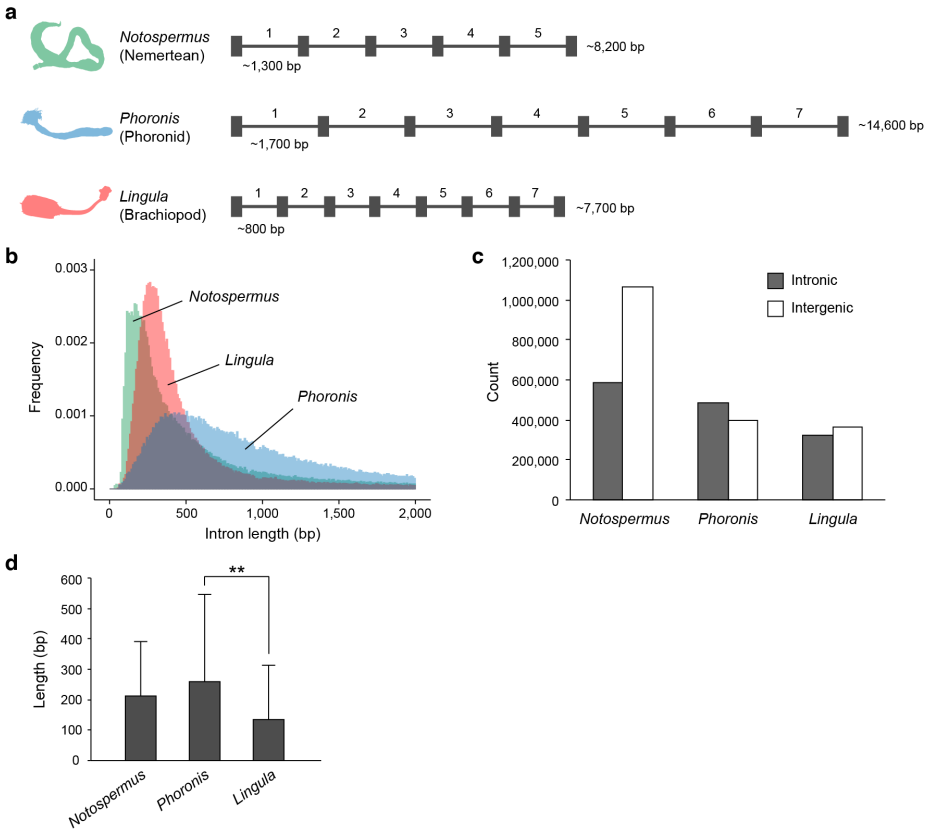


Figure 3.6 | Expansion of transposable elements in the *Phoronis* introns have doubled its gene size compared to those of *Lingula*.

(**a**) Gene structures in the three genomes. Black boxes denote exons. Introns are numbered on the lines. *Phoronis* and *Lingula* have the same gene structure (8 exons and 7 introns in average). Approximate mean gene sizes are shown on the right. Intron sizes are indicated at the bottom. (**b**) Distribution of intron lengths. (**c**) Distribution of transposable elements in introns and intergenic regions. (**d**) Lengths of transposable elements located in introns. *Phoronis* has significantly (*P* value < 0.01**) larger intronic transposable elements than *Lingula*.

The evolutionary histories of TEs in the *Notospermus* and *Phoronis* genomes are very different. *Notospermus* has experienced two TE expansion events, an ancient expansion of *Pao*-type long terminal repeats (LTRs) and relatively recent one of *Penelope* retrotransposons, a type of long interspersed elements (LINEs) (Fig. 3.7). Despite the close relationship of phoronids and brachiopods, *Phoronis* has undergone three waves of expansion,

including two early runs contributed by the DNA transposons, *Sola* and *P* elements, and a more recent wave by *Gypsy*-type LTRs.

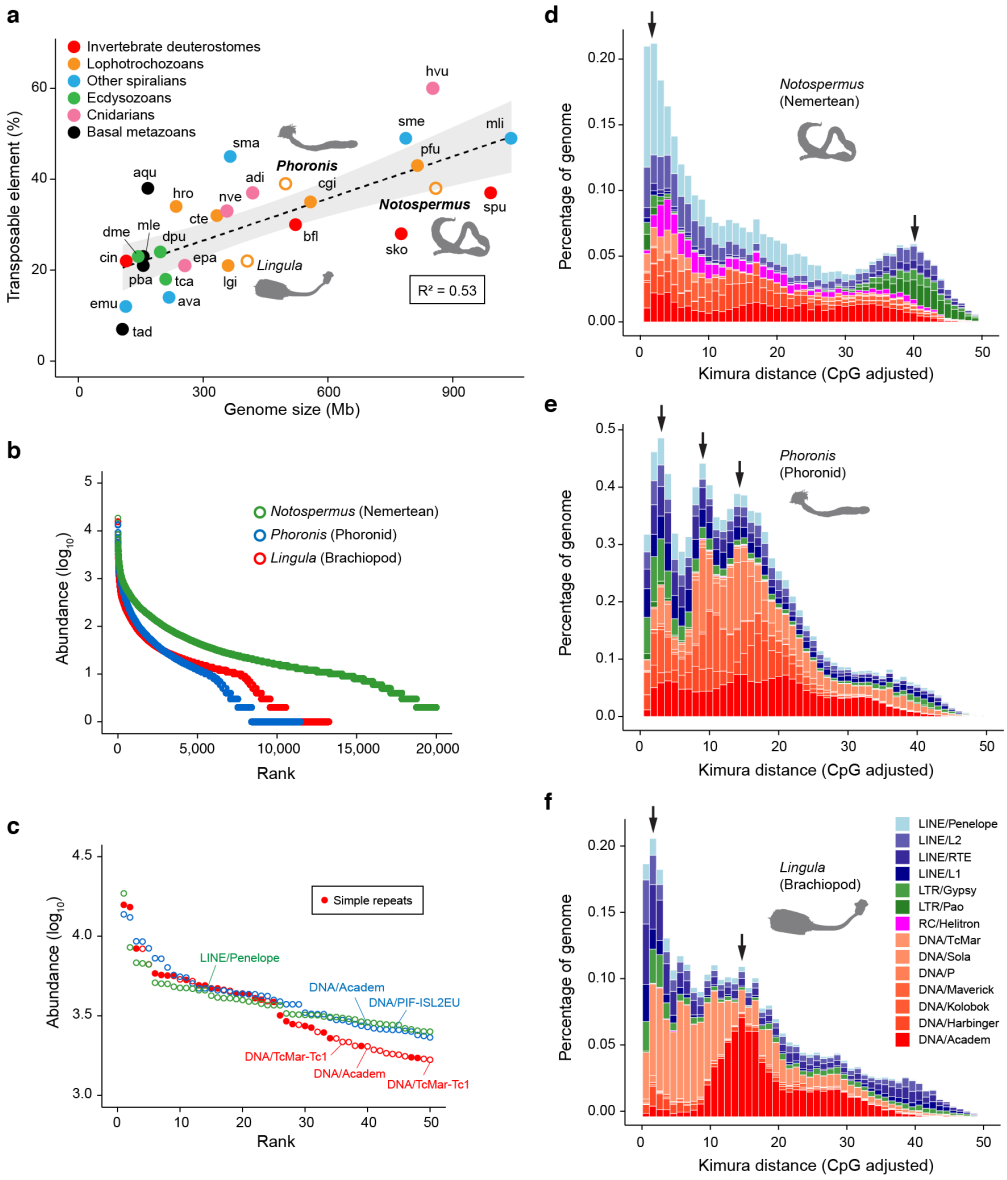


Figure 3.7 | Lineage-specific expansion of lophotrochozoan transposable elements.
(**a**) Regression analysis (indicated by a dashed line; shaded area denotes 95% confidence region; R^2, correlation coefficient) of transposable element (TE) content and genome size for a representative set of metazoan genomes. A simple rule of thumb in this panel shows that larger genomes have more TEs in general. (**b**) Rank abundance curve of TE species (i.e. unique repeats identified by RepeatScout) showing relative TE species abundance fitted with a log-normal distribution (based on the lowest value of Akaike information criterion). (**c**) The 50 most abundant TE species. Known TEs with annotated class/family are labelled. Many abundant TEs are unknown. (**d**–**f**, TE landscapes of known TE families in *Notospermus* (**d**), *Phoronis* (**e**) and *Lingula* (**f**). Arrows indicate TE expansion events.

On the other hand, *Lingula* has had two expansion events contributed by the DNA transposons, *Academ* and *TcMar*. The expanded TEs are differentially expressed in certain tissues such as *Penelope* (LINE) and *Gypsy* (LTR) in the *Notospermus* posterior end and *Phoronis* ampulla, as well as *Academ* in the *Lingula* lophophore, suggesting that these TEs play roles in regulating homeostasis of body structures (Feschotte, 2008). Further analyses of TE species showed that many lophotrochozoans, invertebrate deuterostomes and cnidarians share a similar pattern of high copy numbers of certain TE families, such as *Proto2*, *Sola* and *Academ*. This finding suggests a conserved pool of TEs likely emerged at the stem of metazoans, despite the fact that many lophotrochozoan TEs (70–90%) are novel and difficult to study under current annotation limits.

3.3.2 Phylogeny of lophotrochozoans

Given that nemerteans possess few morphological features compared to other lophotrochozoans, the phylogenetic position of Nemertea within Lophotrochozoa is highly controversial (Cannon et al., 2016; Dunn et al., 2008; Hausdorf et al., 2010; Helmkampf et al., 2008; Paps et al., 2009a; Weigert et al., 2014). Some phylogenomic studies place Nemertea sister to Phoronida and Brachiopoda (Dunn et al., 2008; Laumer et al., 2015; Nesnidal et al., 2013) (Fig. 3.8a).

However, others propose different hypotheses based on various marker sets and mathematical models, placing Nemertea in a variety of phylogenetic positions (Cannon et al., 2016; Kocot et al., 2016b; Paps et al., 2009a; Weigert et al., 2014) (Fig. 3.8b–d). To resolve this issue, I applied genome-based phylogenetic analysis. Using 173 one-to-one orthologous genes from available lophotrochozoan genomes (Albertin et al., 2015; Luo et al., 2015; Simakov et al., 2013; Takeuchi et al., 2016; Zhang et al., 2012), I showed that Nemertea is close to Phoronida and Brachiopoda (Fig. 3.8e). Phylogenetic trees based on gene content inferred from both parsimonious and maximum likelihood methods also support this relationship.

Besides the position of Nemertea, several issues about the lophotrochozoan phylogeny remain a matter of debate. For example, it has been unclear whether Phoronida is nested within Brachiopoda (Cohen, 2013; Hausdorf et al., 2010; Sperling et al., 2011) and whether Ectoprocta belongs to the historical superphylum, Lophophorata (animals with lophophores) (Cannon et al., 2016; Dunn et al., 2008; Hausdorf et al., 2010; Helmkampf et al., 2008; Laumer et al., 2015; Nesnidal et al., 2013; Paps et al., 2009a; Weigert et al., 2014).

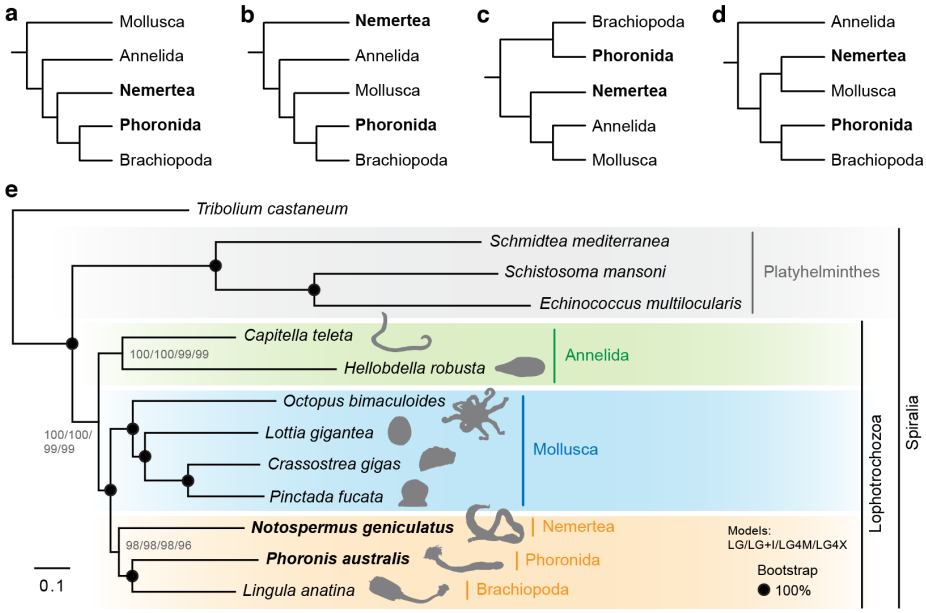


Figure 3.8 | Genome-based phylogenetics supports a close relationship between Nemertea and Phoronida.

(**a**–**d**) Proposed relationships of the five major clades in lophotrochozoans: (**a**) Kryptrochozoa hypothesis (monophyly of Nemertea, Phoronida and Brachiopoda); (**b**) Nemertea as sister group to the rest of lophotrochozoans; (**c**) Phoronida and Brachiopoda as sister group to the rest of lophotrochozoans; and (**d**) Nemertea as sister to Mollusca. (**e**) Phylogeny of lophotrochozoans inferred from 173 one-to-one orthologous genes (62,928 amino acid positions with 92% overall matrix completeness). Maximum-likelihood trees were obtained using LG+Γ, LG+I+Γ, LG4M+Γ and LG4X+Γ models with 1,000 bootstrap replicates. Black circles on nodes indicate 100% bootstrap support from all four models.

To test these hypotheses, I retrieved deep RNA-seq reads from 14 taxa, including nemerteans (Egger et al., 2015; Halanych and Kocot, 2014; Whelan et al., 2014), phoronids (Halanych and Kocot, 2014), ectoprocts (Laumer et al., 2015; Wong et al., 2014) and brachiopods (Cannon et al., 2016; Halanych and Kocot, 2014; Luo et al., 2015). After *de novo* assembling the transcriptomes, I retained those with high quality and then performed phylogenetic analyses with both genomic and transcriptomic data. My analysis supports the monophyly of Brachiopoda, in which Linguliformea and Craniiformea are sisters to Rhynchonelliformea (Fig. 3.9). Furthermore, Phoronida is a distinct group not nested within Brachiopoda, but sister to Ectoprocta. My results thus support the traditional classification of Lophophorata (Phoronida, Ectoprocta and Brachiopoda) rather than Brachiozoa (only Brachiopoda and Phoronida). The discrepancy between this and previous studies is possibly due to the rapid evolutionary rates of ectoproct genes, highlighting the importance of careful selection of genes with strong phylogenetic signals (Salichos and Rokas, 2013).

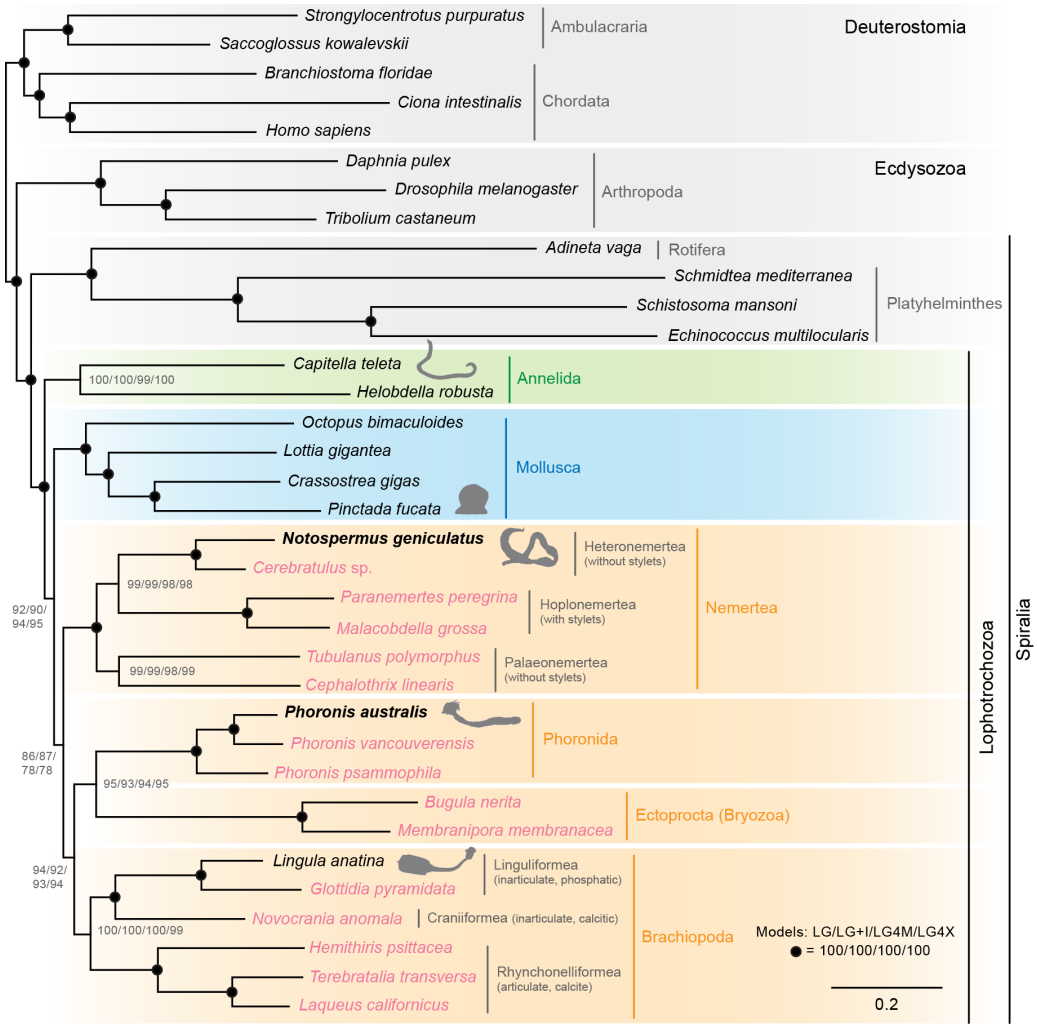


Figure 3.9 | Phylogenetic relationships of nemerteans and phoronids within lophotrochozoans using genomic and transcriptomic data.

Lophotrochozoan phylogeny inferred from 141 genes and 35 taxa (83,970 amino acid positions with 85% overall matrix completeness). Maximum-likelihood trees were obtained using LG+Γ, LG+I+Γ, LG4M+Γ, and LG4X+Γ models with 100 bootstrap replicates and with the constraint topology of all lophotrochozoans in the same clade. Black circles on the nodes indicate 100% bootstrap support from all four models. Species highlighted in bold are new genomes published with this study. Taxa in black indicate genomic data, whereas those in pink are transcriptomic data.

3.3.3 Bilaterian gene repertoire and gene family evolution

To gain insight into bilaterian gene family evolution, I compared lophotrochozoan proteomes with those of other metazoans. The *Notospermus* genome experienced high turnover rate and a recent expansion of gene families compared to that of *Phoronis* (Fig. 3.10). Comparing gene families among four lophotrochozoans including *Lingula* (Luo et al., 2015) and *Octopus* (Albertin et al., 2015), I identified 7,007 lophotrochozoan core gene families, with 1,127 gene families shared only by nemerteans, phoronids and brachiopods, reflecting their relatively

close phylogenetic relationships (Fig. 3.11a). Principle component analysis of gene family size and protein domain shows that lophotrochozoans are consistently clustered with invertebrate deuterostomes, such as amphioxus, acorn worms and sea urchins (Fig. 3.11b).

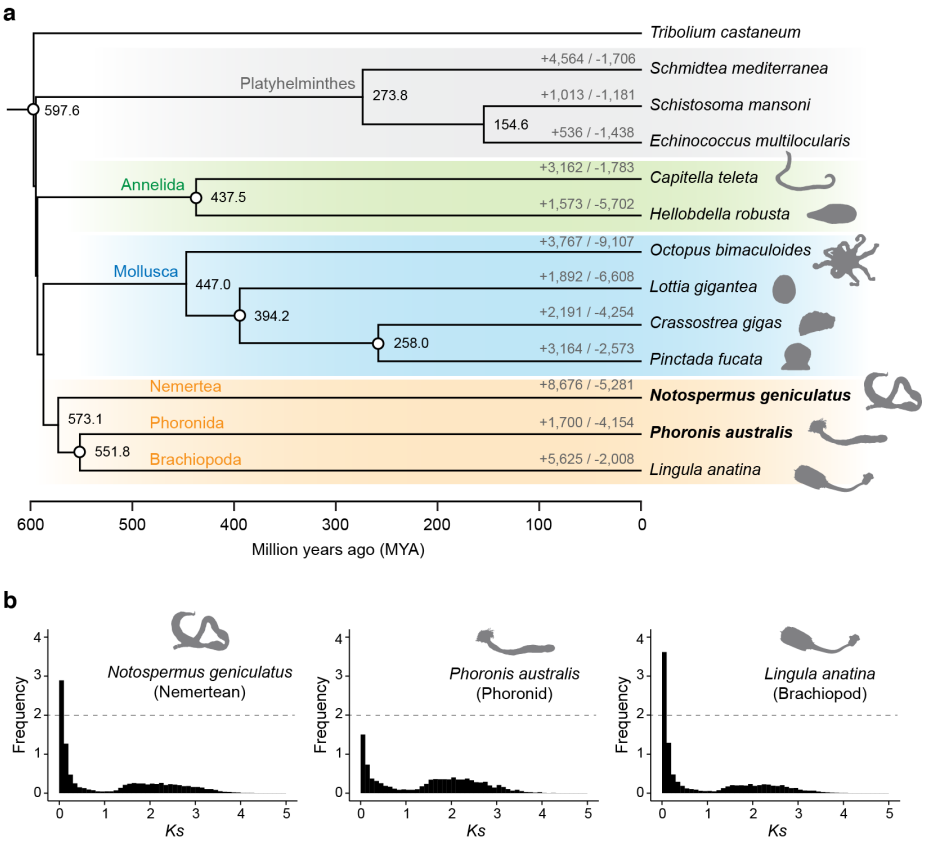


Figure 3.10 | Gene family divergence time and gene gain-and-loss.
(**a**) Divergence time estimated using PhyloBayes using the log-normal relaxed clock model together with available calibrations based on published fossil data (Erwin et al., 2011; Simakov et al., 2015) (open circles at the nodes). The age of the root was specified as 600 MYA for the divergence of ecdysozoans and spiralians. The time tree was obtained after 20,669 cycles with the first 2,500 cycles discarded as burn-ins. Gene family history of gain (plus) and loss (minus) was analyzed with CAFE. (**b**) Frequency of pair-wise genetic divergence calculated with synonymous substitution rate (*Ks*) among all possible paralogous pairs in the *Notospermus*, *Phoronis* and *Lingula* genomes. Note that peaks at smaller *Ks* value indicate relatively recent expansion of gene families.

I further determined that lophotrochozoans and deuterostomes share 4,662 gene families that are not found in ecdysozoans or other spiralians, such as rotifers and planarians. In particular, except those belonged to eumetazoan genes (Putnam et al., 2007), 2,870 gene families are bilaterian-specific, which cannot be found in cnidarians and sponges (Fig. 3.11c). Many of these gene families carry EGF-like, zinc finger and fibronectin domains, which are

related to regulation of cell cycle, biological adhesion, and immune response (Fig. 3.12). My data thus suggest that an ancestral bilaterian gene repertoire retained in lophotrochozoans and deuterostomes is related to control of homeostasis and multicellularity (Srivastava et al., 2010).

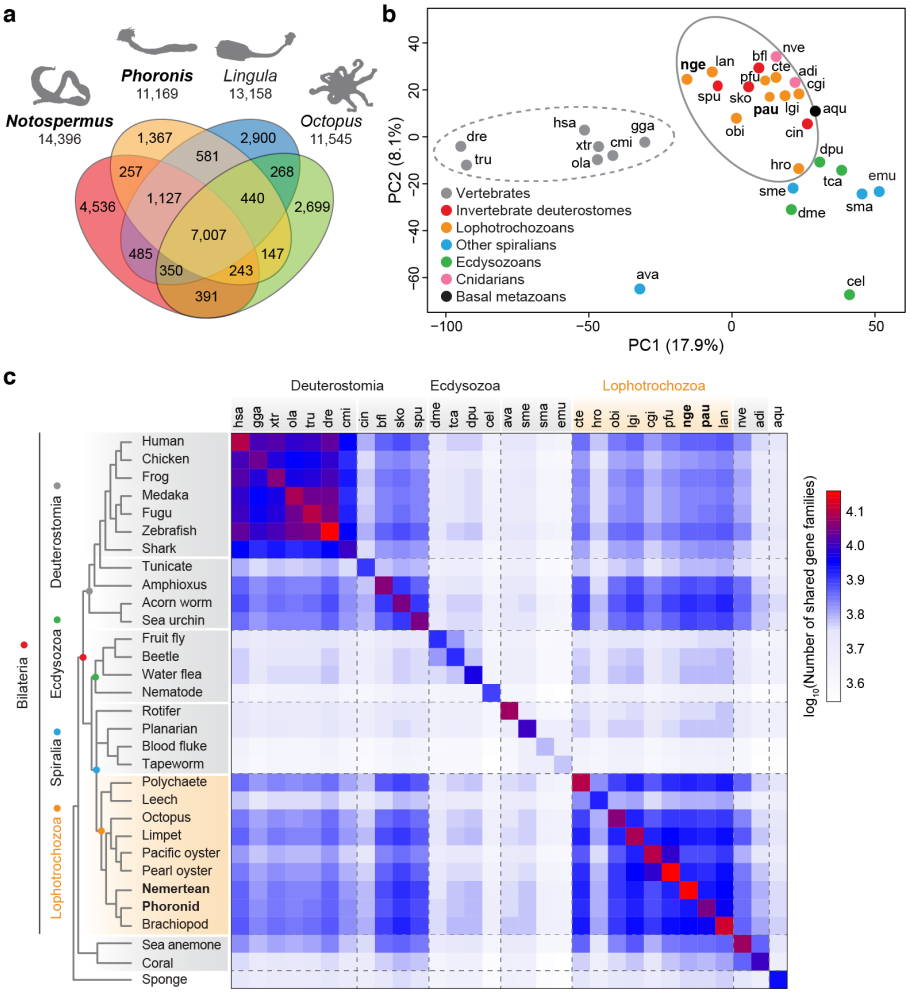


Figure 3.11 | Lophotrochozoans share an ancestral bilaterian gene repertoire with deuterostomes.

(**a**) Venn diagram of shared and unique gene families in four selected lophotrochozoans. Gene families were identified by clustering of orthologous groups using OrthoMCL. (**b**) Principal component (PC) analysis of PANTHER gene family sizes. Note the clustering of invertebrate deuterostomes (bfl, sko, and spu) with lophotrochozoans (solid-lined circle). Dashed-lined circle denotes the clustering of vertebrates. (**c**) Matrix of shared gene families among selected metazoans. A cladogram is denoted on the left based on their phylogenetic positions inferred from this study. Dashed lines separate the major clades. Note that tunicates (cin) and leeches (hro) share fewer genes with other bilaterians, likely because of their relatively high evolutionary rates and gene loss in each lineage.

Species code in Figure 3.11: adi, *Acropora digitifera*; aqu, *Amphimedon queenslandica*; ava, *Adineta vaga*; bfl, *Branchiostoma floridae*; cel, *Caenorhabditis elegans*; cgi, *Crassostrea gigas*; cin, *Ciona intestinalis*; cmi, *Callorhynchus milii*; cte, *Capitella teleta*; dme, *Drosophila melanogaster*; dpu, *Daphnia pulex*; dre, *Danio rerio*; emu, *Echinococcus multilocularis*; gga, *Gallus gallus*; hro, *Helobdella robusta*; hsa, *Homo sapiens*; lan, *Lingula anatina*; lgi, *Lottia gigantea*; nge, *Notospermus geniculatus*; nve, *Nematostella vectensis*; obi, *Octopus bimaculoides*; ola, *Oryzias latipes*; pau, *Phoronis australis*; pfu, *Pinctada fucata*; sko, *Saccoglossus kowalevskii*; sma, *Schistosoma mansoni*; sme, *Schmidtea mediterranea*; spu, *Strongylocentrotus purpuratus*; tca, *Tribolium castaneum*; tru, *Takifugu rubripes*; xtr, *Xenopus tropicalis*.

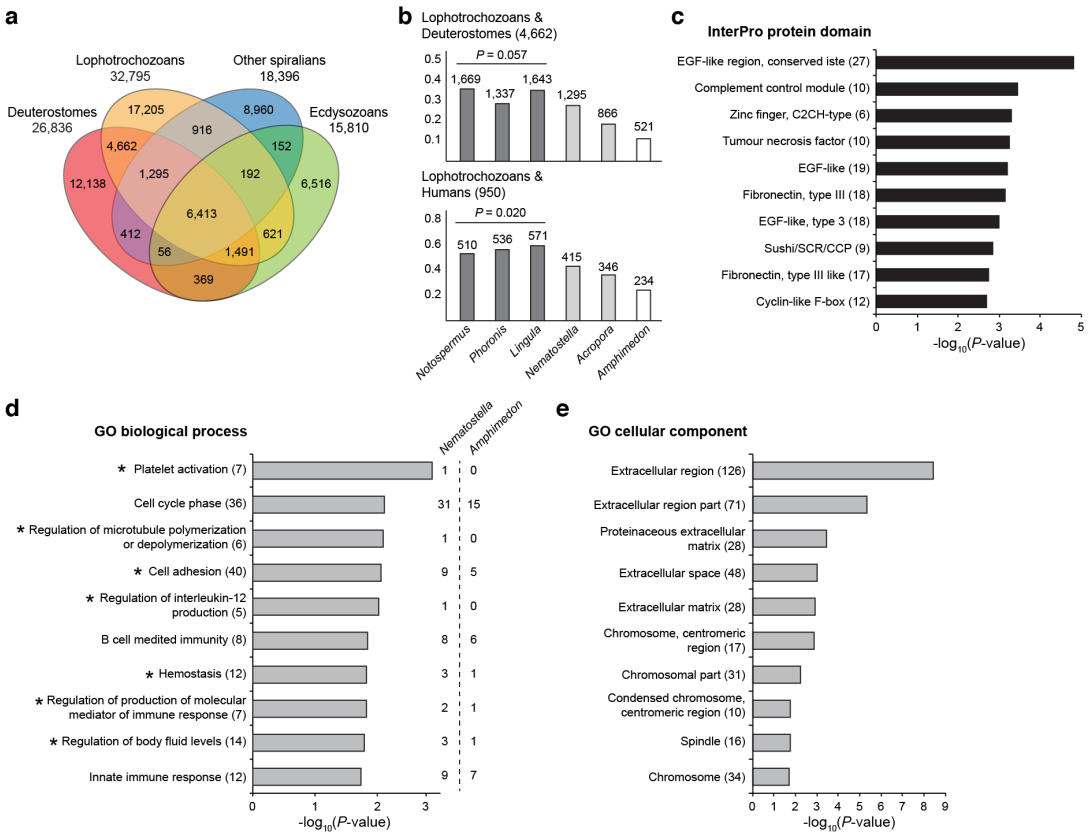


Figure 3.12 | Gene families shared by lophotrochozoans and deuterostomes.
(**a**) Venn diagram of shared gene families in four categorized bilaterian groups. (**b**) Distribution of gene families (in parentheses) shared by lophotrochozoans (dark grey) with deuterostomes (upper panel) or with humans (lower panel) in selected metazoans (cnidarians in light grey, sponges in white). (**c**–**e**) Gene ontology (GO) enrichment terms of InterPro protein domain (**c**), biological process (**d**), and cellular component (**e**). Numbers of genes with annotated human orthologues are indicated in parentheses. Asterisks in (d) indicate GO functional categories largely containing bilaterian-specific genes (< 30% genes can be found in *Nematostella* and *Amphimedon*, where numbers of genes are shown on the right). Note that the top GO terms are mostly related to cell cycle and biological adhesion located at chromosomes or extracellular regions, indicating the possible roles of these genes in controlling multicellularity and cellular homeostasis in bilaterians.

To investigate the evolution of toolkit gene content, I annotated transcription factor and signaling pathway-related genes. The *Phoronis* genome has a smaller number of genes with homeobox and helix-loop-helix binding domains, compared to those of other lophotrochozoans. Specifically, *Phoronis* has 98 homeobox genes, considerably fewer than *Capitella* (183), *Lottia* (141), *Notospermus* (157) and *Lingula* (125).

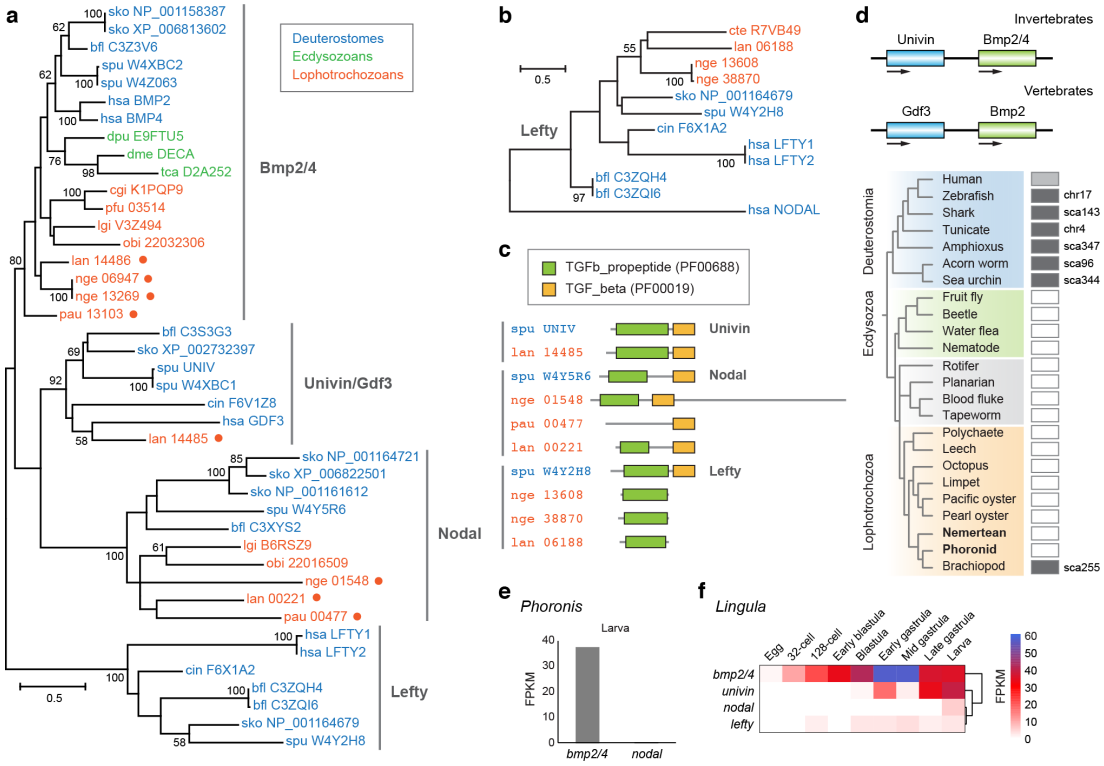


Figure 3.13 | Conservation of bilaterian TGFβ superfamily ligand genes among lophotrochozoans. (**a**) Phylogeny of selected TGFβ ligands based on the TGF_beta domain and surrounding regions (44 genes, 204 amino acids) using the maximum likelihood method with LG+Γ model (100 bootstrap replicates). Solid circles indicate the gene models presented in this study. Numbers at the nodes denote the percentage of bootstrap support. Species codes are defined in Figure 3.11. (**b**) Phylogeny of Lefty based on TGFb_propeptide domain (12 genes, 112 amino acids) using the maximum likelihood method with the LG+Γ model (100 bootstrap replicates). (**c**) Domain composition of selected TGFβ ligands analyzed with the Pfam database. (**d**) Genomic organization of linked Univin and Bmp2/4 (arrows denote the direction of transcripts) among bilaterians. Univin is tightly linked (8.5 kb) to Bmp2/4 in the *Lingula* genome. Light grey box, two genes separated; dark grey boxes, tightly linked; white boxes, no Univin gene found; chr, chromosome; sca, scaffold. **e**, Expressions of *bmp2/4* and *nodal* in the *Phoronis* larva. FPKM, fragments per kilobase of transcript per million mapped reads. (**f**) Expressions of selected TGFβ ligands during *Lingula* early development.

Nevertheless, both the *Notospermus* and *Phoronis* genomes contain considerable numbers of zinc finger (C2H2-type) as well as TGFβ and Wnt genes, compared to those of

other lophotrochozoans. Notably, the CHRD domain for the gene *Chordin*, which is essential for body patterning by modulating BMP signaling, and which is shared by all other lophotrochozoans, is not found in annelids, rotifers, or planarians.

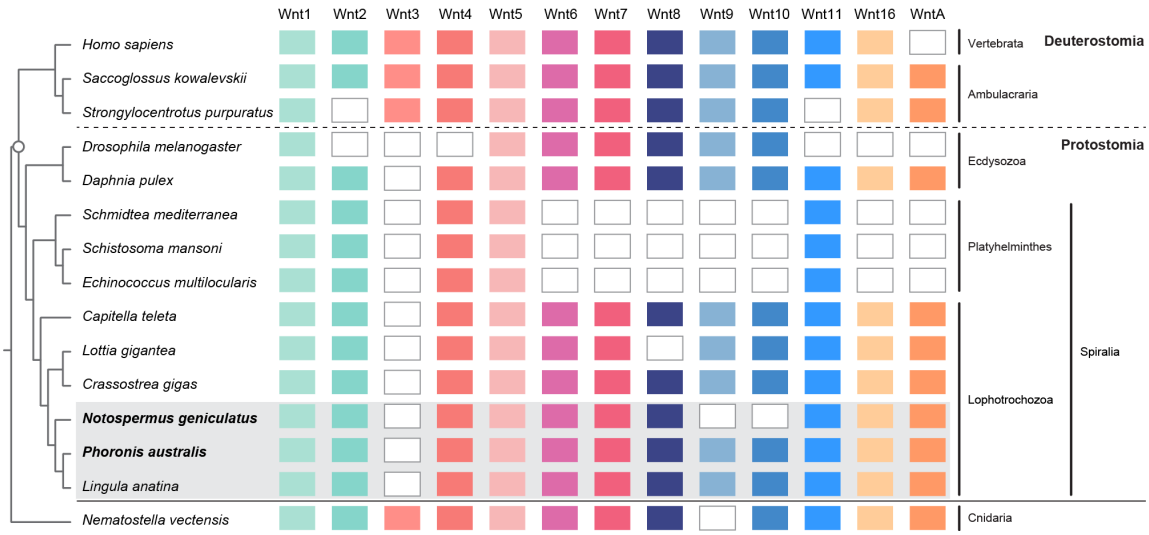


Figure 3.14 | Distribution of Wnt genes in selected metazoans.
Comparison of Wnt genes in bilaterians with Cnidaria as an outgroup. The cladogram is based on phylogenetic positions inferred from this study. The circle at one node denotes the clade of bilaterians. Species highlighted in bold have new genomes published in this study. Coloured boxes indicate the distribution of Wnt subfamilies. Empty boxes represent their absence. Genomes analyzed in this study are highlighted in grey. The dashed line separates Deuterostomia and Protostomia. Protostomia has lost Wnt3. Platyhelminthes has lost Wnt3, Wnt6-10, Wnt16, and WntA. Chordata (including Vertebrata) has lost WntA.

TGFβ and Wnt signalings play important roles in axial patterning, cell specification and control of cell behaviour during embryonic development (Massague, 2012; Niehrs, 2012). Some TGFβ genes modulating Nodal signals, such as *Lefty* and *Univin*, are considered as a deuterostome novelty (Simakov et al., 2015). The *Notospermus* and *Phoronis* genomes have 15 and 10 TGFβ genes, respectively. Interestingly, in addition to *Nodal*, which can be found in the *Notospermus*, *Phoronis* and *Lingula* genomes, I discovered the syntenic linkage of *Univin* and *Bmp2/4* in the *Lingula* genome, despite its absence in other protostomes. Thus, this finding suggests that the linkage of *Univin* and *Bmp2/4* is likely a bilaterian ancestral feature that is lost in some vertebrates and protostomes (Fig. 3.13).

Transcriptome analysis shows that *nodal* is either not expressed or is expressed at very low levels during early development in *Phoronis* and *Lingula*. The *Notospermus* and *Phoronis* genomes have 17 and 12 Wnt genes, respectively. In *Notospermus* and *Phoronis*, I identified all Wnt genes (*Wnt1*, *Wnt2*, *Wnt4–11*, *Wnt16* and *WntA*) except *Wnt3*, which has

been likely lost in all protostomes. I failed to find *Wnt9* and *Wnt10* in *Notospermus* (Fig. 3.14). Unlike lophotrochozoans, extensive loss of Wnt genes may be a common feature in Platyhelminthes (Riddiford and Olson, 2011) and Pancrustacea (Kao et al., 2016).

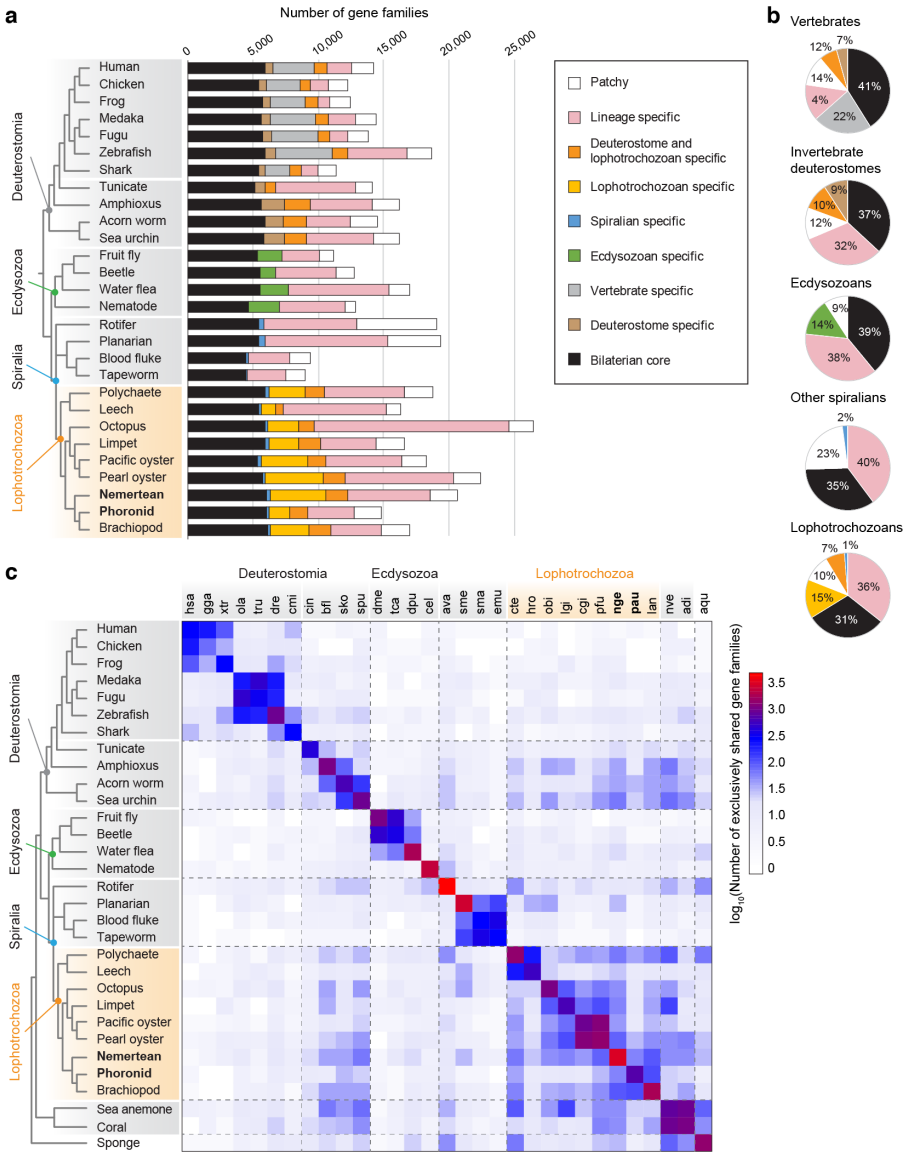


Figure 3.15 | Compositions of bilaterian gene families.

(**a**) Comparison of bilaterian genes showing the bilaterian core, clade-specific, lineage-specific (within species), and patchy gene families (i.e. genes retained in certain lineages but unevenly lost in others). (**b**) Pie charts showing percentages of gene families in an average bilaterian genome. (**c**) Matrix of patchy gene families exclusively shared by two given species pairs among 31 selected metazoan genomes. The cladogram is based on their phylogenetic positions inferred from this study. Dashed lines separate the major clades. Numbers of exclusively shared gene families are correlated with phylogenetic distance. Remarkably, despite the long phylogenetic distance, lophotrochozoans share many patchy gene families with invertebrate deuterostomes (i.e. amphioxus, acorn worms and sea urchins). Species codes are defined in Figure 3.11.

Remarkably, I also observed many gene families that are lineage-specific (10–30%) and patchy (~10%; i.e. genes retained in certain lineages, but unevenly lost in others) among bilaterians (Albalat and Canestro, 2016) (Fig. 3.15). Together with lineage-specific gene family expansion, these features reflect the dynamics of genome evolution (Fig. 3.16).

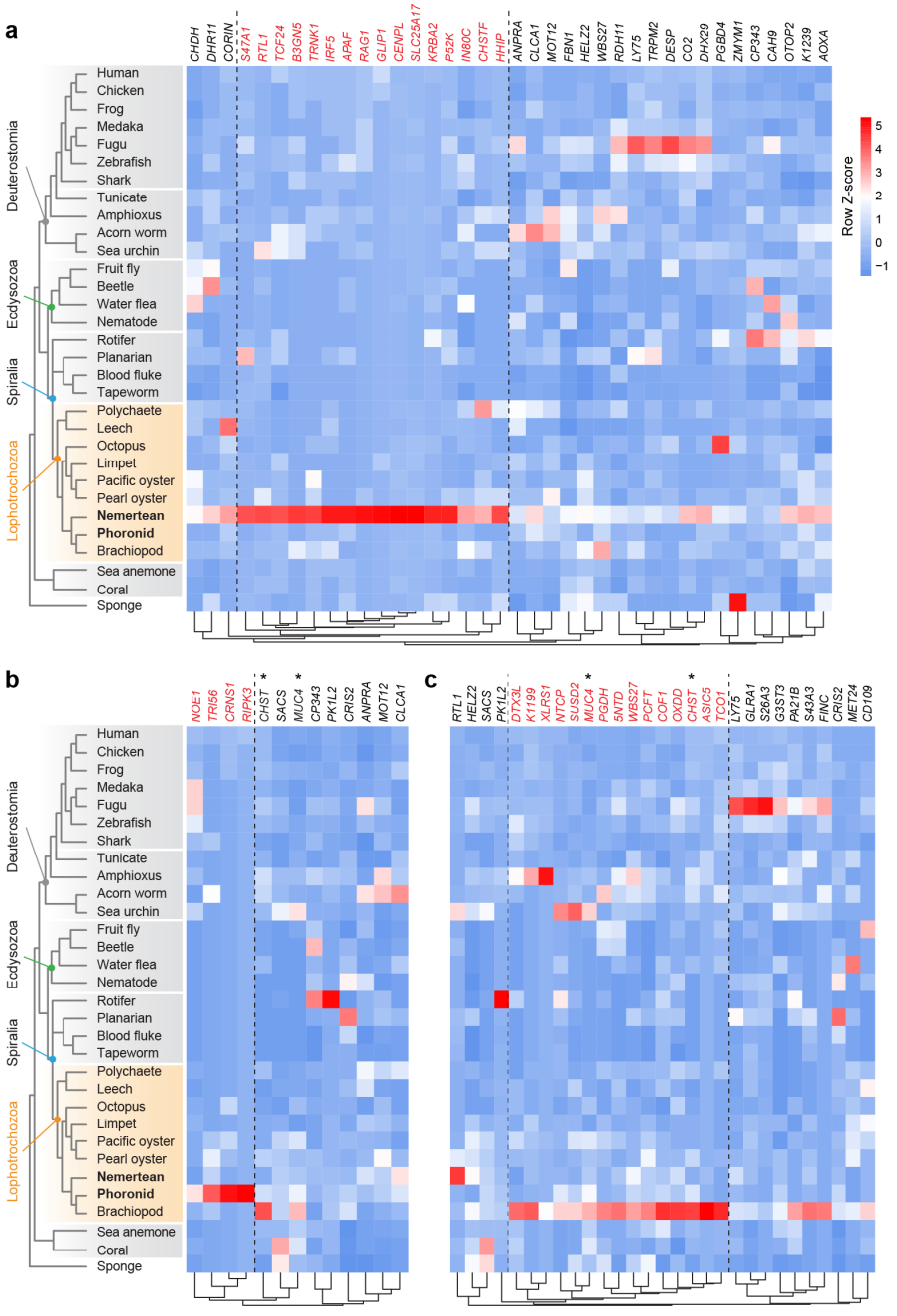


Figure 3.16 | Expansion of lineage-specific gene families.

(**a**–**c**) The most expanded gene families with detectable homology and functional annotation compared to 31 selected metazoan genomes in *Notospermus* (**a**), *Phoronis* (**b**) and *Lingula* (**c**). Gene names are based upon best hits in the human proteome from UniProt. Significantly expanded gene families are shown in red ($P < 0.001$). Asterisks indicate the gene families that are expanded in both phoronids and brachiopods.

For instance, the most expanded gene family in *Notospermus* belongs to retrotransposon-like protein (*RTL1*). The role of this gene is not clear, but it has been shown to be neofunctionalized for developmental processes (Sekita et al., 2008). Other expanded gene families in *Notospermus* are mostly related to toxin metabolism (*SLC25A17* and *S47A1*) and immune response (*APAF*, *IRF5* and *IN80C*). The most expanded gene families in *Phoronis* are also related to immunity and programmed cell death (*TRI56* and *RIPK3*). In fact, further analysis shows that both *Notospermus* and *Phoronis* genomes have more genes with apoptosis-related domains, indicating more complex regulation of cell death programs.

Notably, gene families related to mucus production such as mucin-4 (*MUC4*) and carbohydrate sulfotransferase (*CHST*) are expanded independently in *Phoronis* and *Lingula*, and are highly expressed in the lophophores. This finding indicates possible independent adaptation within each lophophorate lineage, where *Phoronis* may adapt to live with tube-dwelling anemones by protecting themselves with mucus layers. Altogether, my results suggest that both conservations (e.g., conserved gene repertoire) and innovations (e.g., lineage-specific gene gain-and-loss and gene family expansion) are the fundamental processes shaping the evolution of bilaterian gene families.

3.3.4 Hox genes and conserved bilaterian microsyntenies

Hox genes play essential roles during metazoan development, especially for body patterning and appendage formation (Pearson et al., 2005). *Notospermus* contains 16 Hox genes and two ParaHox genes, although *Xlox* may have been absent. The *Notospermus* Hox cluster is disorganized, with Hox genes dispersed in 10 different scaffolds (Fig. 3.17a). On the other hand, *Phoronis* has eight Hox genes in one single Hox cluster and three ParaHox genes. I failed to find *Src* and *Antp* in *Phoronis*. In brachiopods, *Scr* and *Antp* are expressed in the shell-forming epithelium (Schiemann et al., 2017). Possible gene loss of *Scr* and *Antp* in the phoronid lineage implies that the common lophophorate ancestors have mineralized shells, which might be secondarily lost in crown phoronids (Moysiuk et al., 2017).

With improved scaffolding, I discovered *Lox4* in *Lingula* which is linked between *Post2* and *Antp*. Both *Notospermus* and *Phoronis* have only one posterior Hox (*Post2*). My phylogenetic analysis shows that *Post2* is shared by all spiralians and has a different evolutionary origin from ecdysozoan *AbdB*, whereas *Post1* might be specific to lophotrochozoans. *Notospermus* Hox genes are expressed along the adult anterior–posterior (AP) axis with *Hox1* and *Hox2* expressed anteriorly, *Lox2* and *Lox4* mid-posteriorly and *Post2* posteriorly but with no strict spatial collinearity. On the other hand, Hox gene

expression in *Phoronis* and *Lingula* does not exhibit apparent spatial polarity (Fig. 3.18). Remarkably, Hox genes are not expressed in the proboscis and head of *Notospermus* or lophophores of *Phoronis* and *Lingula*. This anterior Hox-free region is also found in juveniles of hemichordates (Gonzalez et al., 2016), nemerteans (Hiebert and Maslakova, 2015) and annelids (Fröbius et al., 2008), suggesting that absence of Hox gene expression at the anterior end is a common adult body plan for all bilaterians.

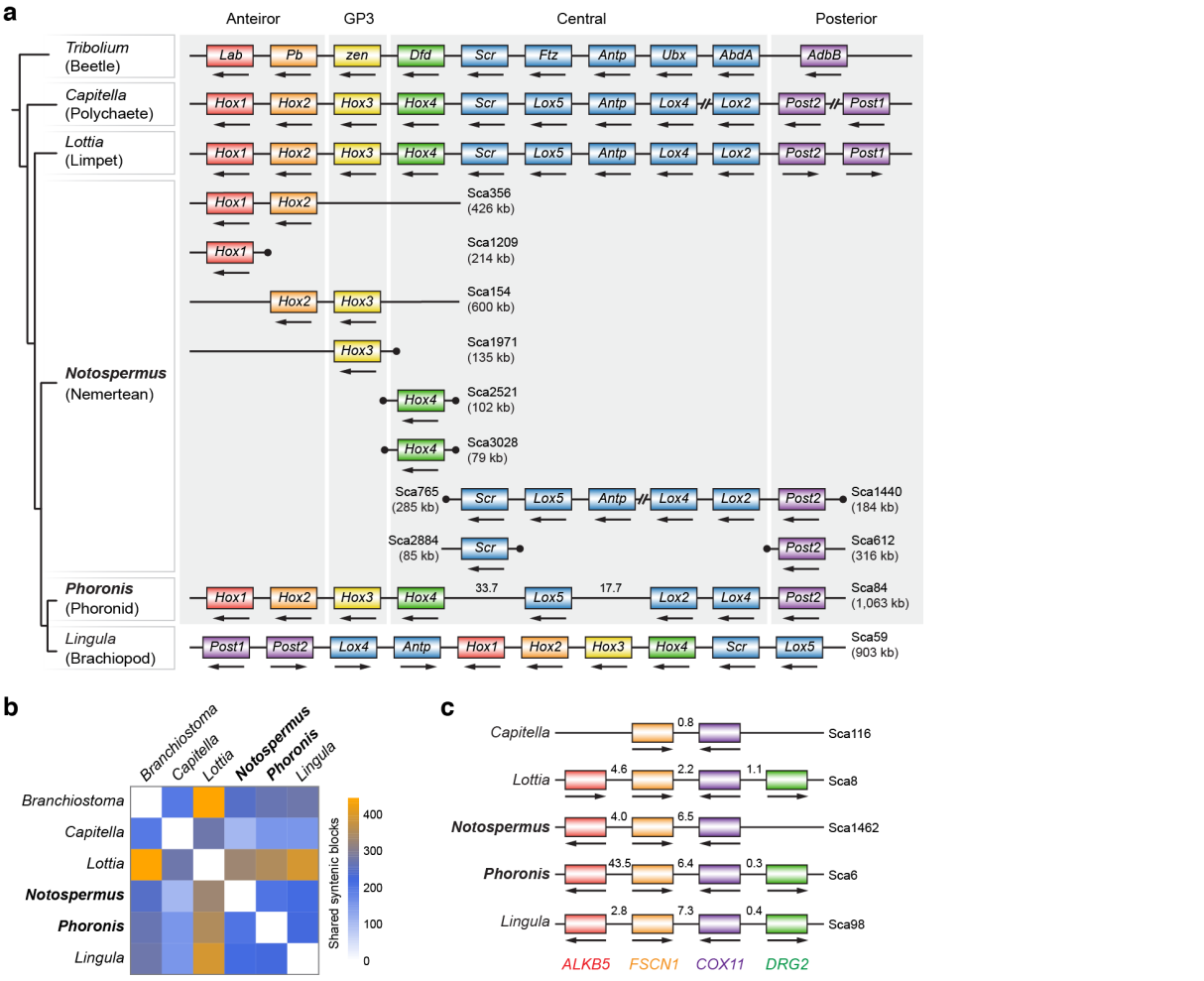


Figure 3.17 | Disorganized Hox gene clusters in *Notospermus* and conserved microsyntenies among lophotrochozoans.

(**a**) The Hox clusters in selected protostomes. Arrows indicate the direction of the transcripts. Double slashes denote non-continuous linkage between two genes. Black dots signify the end of the scaffolds. (**b**) Matrix of microsyntenic blocks (clustered orthologous genes ≥ 3) among bilaterians. (**c**) An example of neighbouring tightly-linked (< 20 kb) orthologous genes shared by lophotrochozoans. Numbers on the scaffolds indicate genomic distance (kb). ALKB5, alpha-ketoglutarate-dependent dioxygenase alkB homolog 5; FSCN1, fascin; COX11, mitochondrial cytochrome c oxidase assembly protein; DRG2, developmentally-regulated GTP-binding protein 2. Sca, scaffold.

I identified ~300–400 ancestral microsyntenic blocks (i.e. clusters of three or more conserved orthologues with close physical linkage) among lophotrochozoans and amphioxus (Fig. 3.17b). However, most of gene clusters associated with embryonic development, such as Wnt (*Wnt9*, *Wnt1*, *Wnt6* and *Wnt10*), ParaHox (*Gsx*, *Xlox* and *Cdx*) and NK (*Msxlx*, *Nkx2.2* and *Nkx2.1*; *Msx*, *Nkx4*, *Nkx3*, *Lbx* and *Tlx*) clusters, are disorganized in *Notospermus* and *Phoronis*, although they are retained intact in *Lingula* (Fig. 3.19). In contrast to the Hox cluster, where transcriptional direction among Hox genes is often the same, neighbouring tightly-linked genes (distance < 20 kb) in the microsyntenic blocks are mostly in opposite directions (Figs. 3.17c and 3.20). Interestingly, I found that tightly-linked genes show significantly lower evolutionary rates, suggesting that they are under strong negative selection. Also, tightly-linked genes within the microsyntenic blocks tend to be expressed constantly across different species and tissue types (Fig. 3.21).

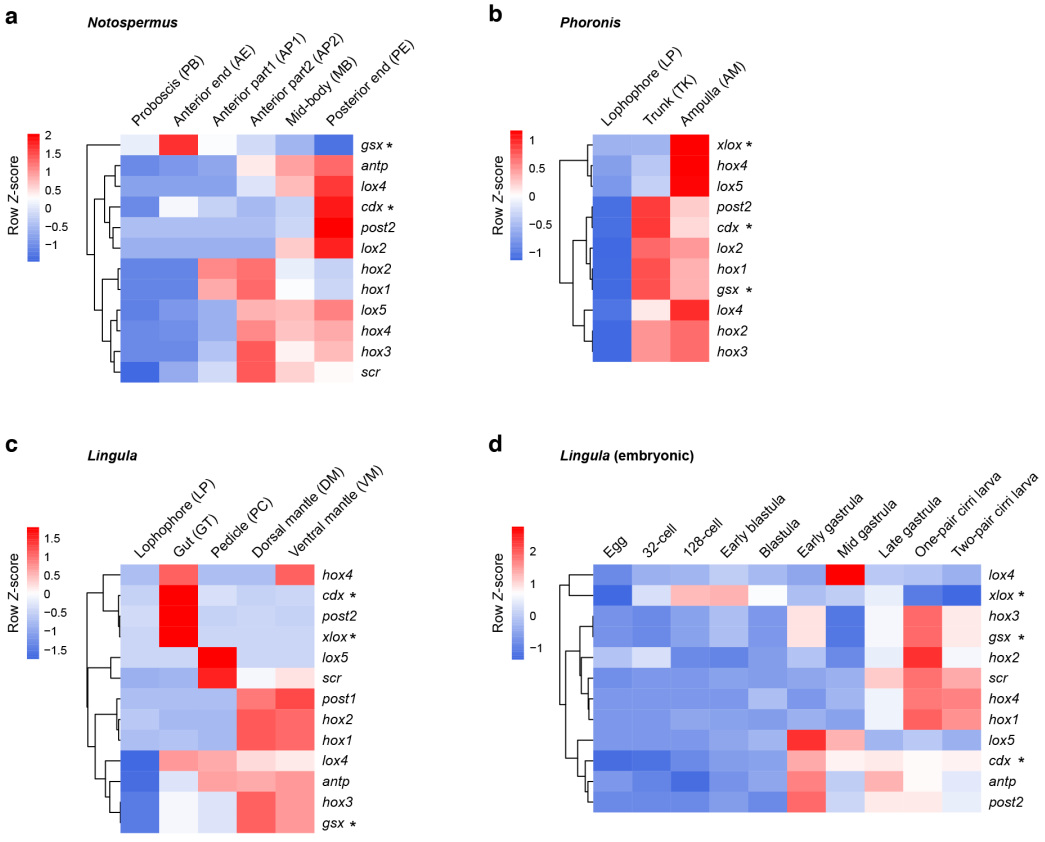


Figure 3.18 | Expression profiles of Hox and ParaHox genes.

(**a**–**d**) Hierarchical clustering heatmaps of Hox and ParaHox gene expressions in tissues of *Notospermus* (**a**), *Phoronis* (**b**) and *Lingula* (**c**) as well as the embryonic stages of *Lingula* (**d**). ParaHox genes are marked with asterisks. There is no expression of Hox genes in the proboscis, the anterior end, or anterior part1 in *Notospermus*, or in lophophores of *Phoronis* and *Lingula*.

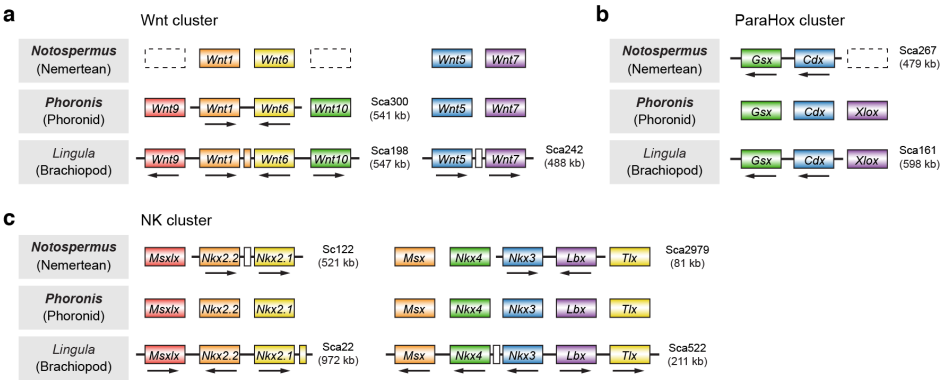


Figure 3.19 | Examples of gene clusters related to development.
(**a**–**c**) Clusters of Wnt (**a**), ParaHox (**b**) and NK (**c**) genes that play important roles during embryonic development in many animal lineages. Arrows indicate the direction of the transcripts. Same coloured boxes next to the gene (e.g., *Wnt1* and *Nkx2.1*) represent fragmented gene model with homology to the neighbouring gene, which possibly results from incorrect annotation. White boxes indicate non-Hox gene models. Dashed boxes denote possible gene loss.

To further explore the dynamics of microsyntenies, I compared tightly-linked orthologues between *Phoronis* and *Lingula*. First, I showed that numbers of shared microsyntenies are not related to scaffold size ($r = 0.27$), suggesting that microsynteny analysis is independent of assembly quality (Fig. 3.21a). Next, I tested evolutionary rates of tightly-linked orthologues by calculating their substitution mutations (*Ka*) and silent mutations (*Ks*). Interestingly, I found that tightly-linked orthologues within syntenic blocks show significantly lower evolutionary rates, suggesting that they are under strong negative selection (Fig. 3.21b–c).

Furthermore, tightly-linked orthologues within microsyntenic blocks tend to be constantly expressed across different species and tissue types ($r > 0.99$) (Fig. 3.21e,f). By contrast, orthologues outside of syntenic blocks are expressed variously. In summary, neighbouring, tightly-linked orthologues have conserved fixed intergenic distances, arranged in opposite directions and are constantly expressed in different tissues. Further studies will be needed to understand regulatory mechanisms controlling tightly-linked orthologues.

3.3.5 Molecular signature of the lophophore and bilaterian head patterning

The lophophore is a lineage-specific feeding apparatus in Lophophorata (Helmkampf et al., 2008). A recent immunohistochemical and ultrastructural study has shown that the lophophore is enriched with neural cells (Temereva and Tsitrin, 2015), yet the molecular signature of the lophophore remains unclear. To explore the origin of the lophophore, I

applied molecular profiling by comparing different tissues among *Notospermus*, *Phoronis* and *Lingula* using RNA-seq (Fig. 3.22a).

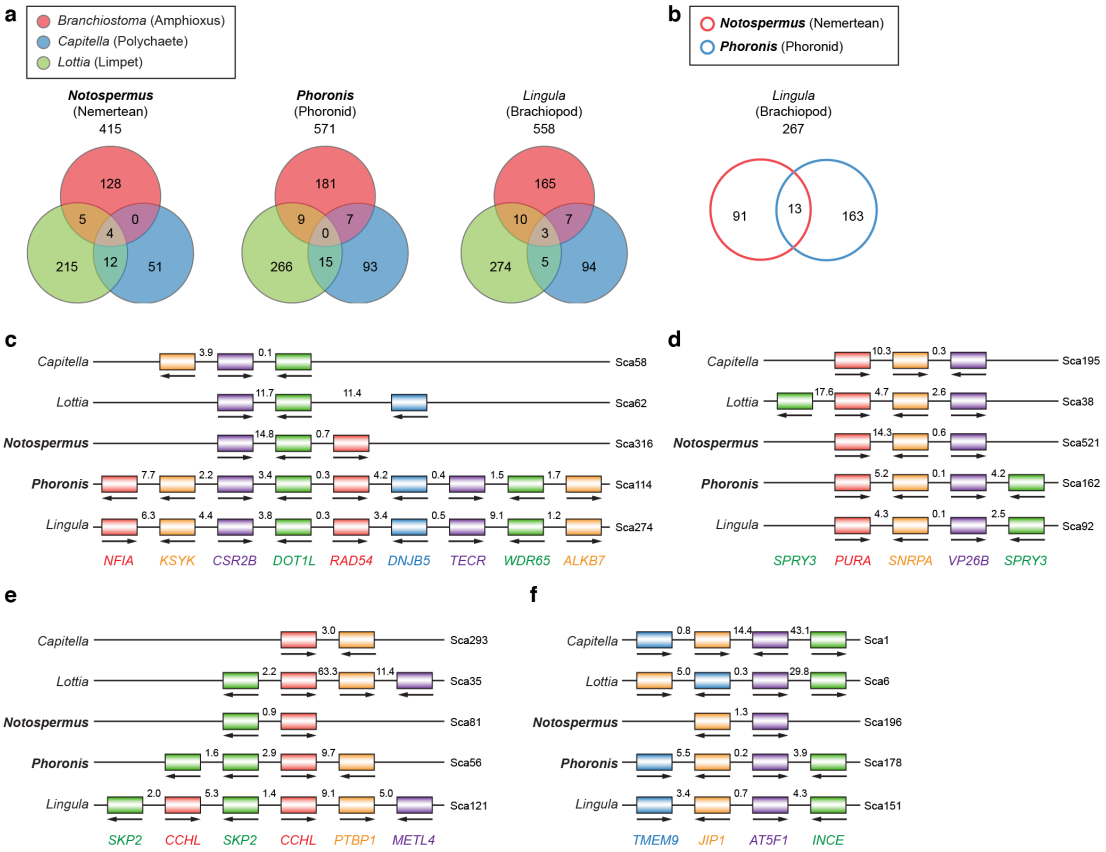


Figure 3.20 | Conserved bilaterian gene linkages.
(**a** and **b**) Venn diagrams of the numbers of microsyntenic blocks (genes ≥ 3) shared by *Notospermus*, *Phoronis* and *Lingula* with *Branchiostoma*, *Capitella* and *Lottia*, respectively (**a**) as well as those of *Notospermus* and *Phoronis* shared with *Lingula* (**b**). (**c** and **d**) Examples of long (nine genes) (**c**) and short (four genes) (**d**) tightly-linked syntenic blocks (<20 kb) shared by selected lophotrochozoans. **e**, An example of tandem duplicated genes within tightly-linked syntenic blocks (*SKP2* in *Phoronis*; *SKP2* and *CCHL* in *Lingula*). **f**, An example of gene inversion within the tightly-linked syntenic block (*TMEM9* and *JIP1* in *Lottia*). Arrows indicate the direction of transcripts. Gene names are given according to human UniProt entries. Numbers on the scaffolds represent genomic distances in kilobases. Tightly-linked genes in the syntenic blocks are always oriented in opposite directions. Sca, scaffold.

I first conducted comparative transcriptomics by calculating the Spearman's correlation coefficient (ρ) based on the expression levels of 8,650 orthologs shared by all three genomes. The *Notospermus* proboscis is molecularly distinct from other types of *Notospermus* tissues (Fig. 3.23) and dissimilar to the *Phoronis* lophophore ($\rho = 0.31$) (Fig. 3.22b).

The *Phoronis* lophophore is considerably more similar to the *Notospermus* head (anterior end and anterior part1; $\rho = 0.46$) (Fig. 3.22a,b). Further comparison of *Phoronis* and *Lingula* lophophores confirms the shared origin of their feeding apparatus ($\rho = 0.61$) (Fig. 3.22c). Next, to investigate the molecular nature of lophophores, I performed expression profiling based on differentially expressed genes. I identified 2,572 and 1,591 genes that are specifically expressed in the lophophores of *Phoronis* and *Lingula*, respectively. Approximately 40% of these genes have no available annotation, reflecting the contribution of a large number of lineage-specific genes to the tissue-specific functions (Fig. 3.24).

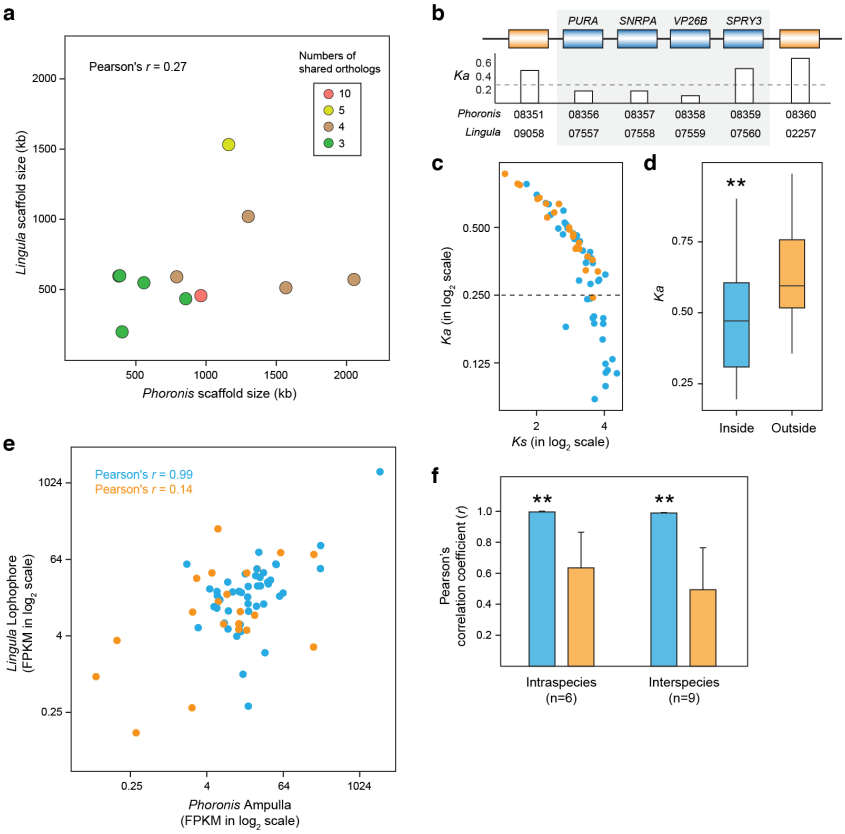


Figure 3.21 | Dynamics of tightly-linked orthologous genes between *Phoronis* and *Lingula*. (**a**) Scatter plot of scaffold size between where *Phoronis* and *Lingula* shared tightly-linked orthologues. (**b**–**d**) Tightly-linked orthologues inside the syntenic block are under negative selective pressure. (**b**) An example of non-synonymous substitution rates (*Ka*) of *Phoronis* and *Lingula* orthologues inside (blue) or outside (orange) the syntenic block (grey box). Human gene names are shown on the top. Gene model IDs are shown at the bottom of the bar graph. (**c**) Scatter plot of synonymous substitution rates (*Ks*) and *Ka* for orthologous gene pairs. Dashed lines mark where *Ka* = 0.25. (**d**) Box plot of *Ka*. Each box represents the interquartile range (IQR). Lines denote medians and whiskers indicate the most disparate data points from the median within 1.5×IQR (*P* value < 0.005**). (**e** and **f**) Transcriptome similarity of orthologues between *Phonoris* and *Lingula* tissues. (**e**) Scatter plot of expression levels between the *Phoronis* ampulla and *Lingula* lophophore. (**f**) Comparison of Pearson's correlation among tissues (*P* value < 0.005**).

Many annotated genes in the lophophores are related to neural development, like those expressed in the *Notospermus* head (Fig 3.24). Unexpectedly, I found that vertebrate head markers such as *otx*, *lhx1/5*, *foxG*, *pax6* and *six3/6* are specifically expressed in both the *Notospermus* head and lophophores (Figs. 3.22d and 3.25).

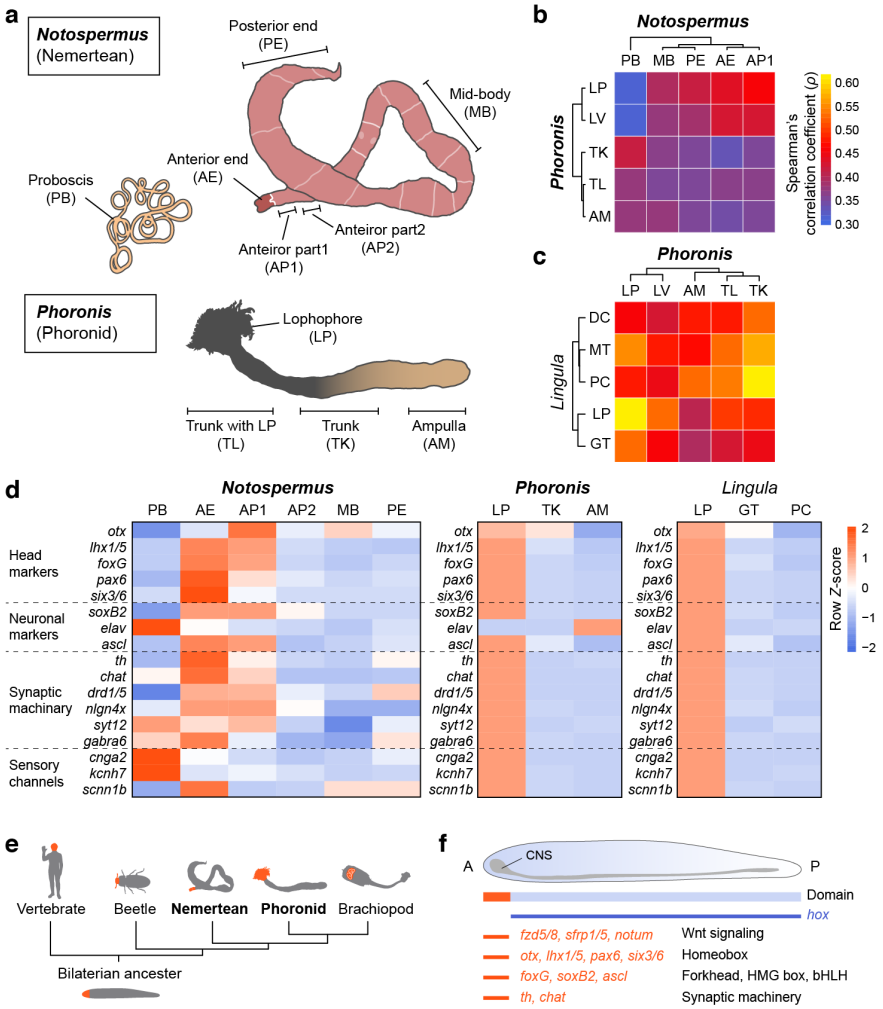


Figure 3.22 | Comparative transcriptomics reveals the molecular similarity between lophophores and bilaterian heads.

(**a**) Cartoon illustration of an adult *Notospermus* and *Phoronis* with the anterior end facing to the left. (**b** and **c**) Analyses of Spearman's correlation coefficient (*ρ*) and hierarchical clustering with the expression levels of 8,650 orthologous genes from larvae and adult tissues of *Notospermus* versus *Phoronis* (**b**) and *Phoronis* versus *Lingula* (**c**). DC, digestive cecum; GT, gut; LV, larva; MT, mantle; PC, pedicle. (**d**) Expression profiles of head patterning-related and neuronal genes in the head of *Notospermus* and lophophores of *Phoronis* and *Lingula*. (**e** and **f**) Schematic representation of the anterior-posterior patterning in bilaterians. (**e**) A simplified phylogeny of bilaterians and regions of anterior positioning heads (highlighted in orange). (**f**) Domain map for the conserved signaling components, transcription factors and genes associated with synaptic machinery along the anterior-posterior axis in the last common bilaterian ancestor. CNS, central nervous system; A, anterior; P, posterior; HMG, high mobility group; bHLH, basic helix-loop-helix.

Neuronal markers such as *soxB2* and achaete-scute (*ascl*), as well as genes associated with synaptic machinery, such as tyrosine monooxygenase (*th*) and choline acetyltransferase (*chat*), are also highly and specifically expressed in lophophores (Fig. 3.22d).

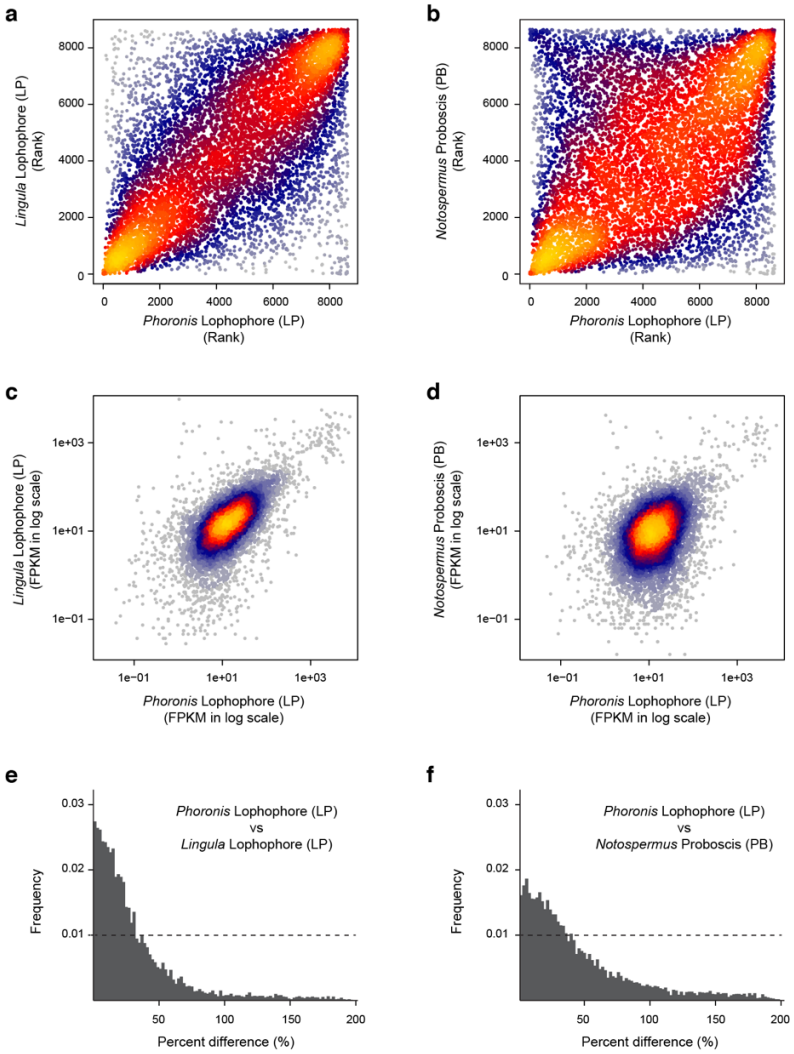


Figure 3.23 | Comparative transcriptomics of lophophores and nemertean proboscises. (**a**–**f**) Comparative transcriptomics with the expression levels of 8,650 orthologues shared by *Notospermus*, *Phoronis* and *Lingula*. (**a**–**d**) Analyses of Spearman's (ρ) and Pearson's (r) correlation coefficients between *Phoronis* and *Lingula* lophophores (ρ, 0.61; r, 0.61) (**a** and **c**) and between the *Phoronis* lophophore and the *Notospermus* proboscis (ρ, 0.31; r, 0.34) (**b** and **d**). FPKM, fragments per kilobase of transcript per million mapped reads. Data points are colour-coded by density (high, yellow; medium, red; low, blue). (**e** and **f**) Distributions of percent difference in transcript expression levels between lophophores (**e**) and between lophophore and proboscis (**f**). Dashed lines mark the diagonals (c and d) and the frequency at 1% (e and f).

In addition, I also found specific expression of genes for sensory ion channels, such as cyclic nucleotide-gated olfactory channel (*cnga2*) and amiloride-sensitive sodium channel

subunit beta (*scnn1b*) in lophophores, suggesting their roles in taste perception and environmental response (Fig. 3.22d). These results indicate that the lophophore shares the molecular nature of the head and anterior centralized nervous system (or the brain).

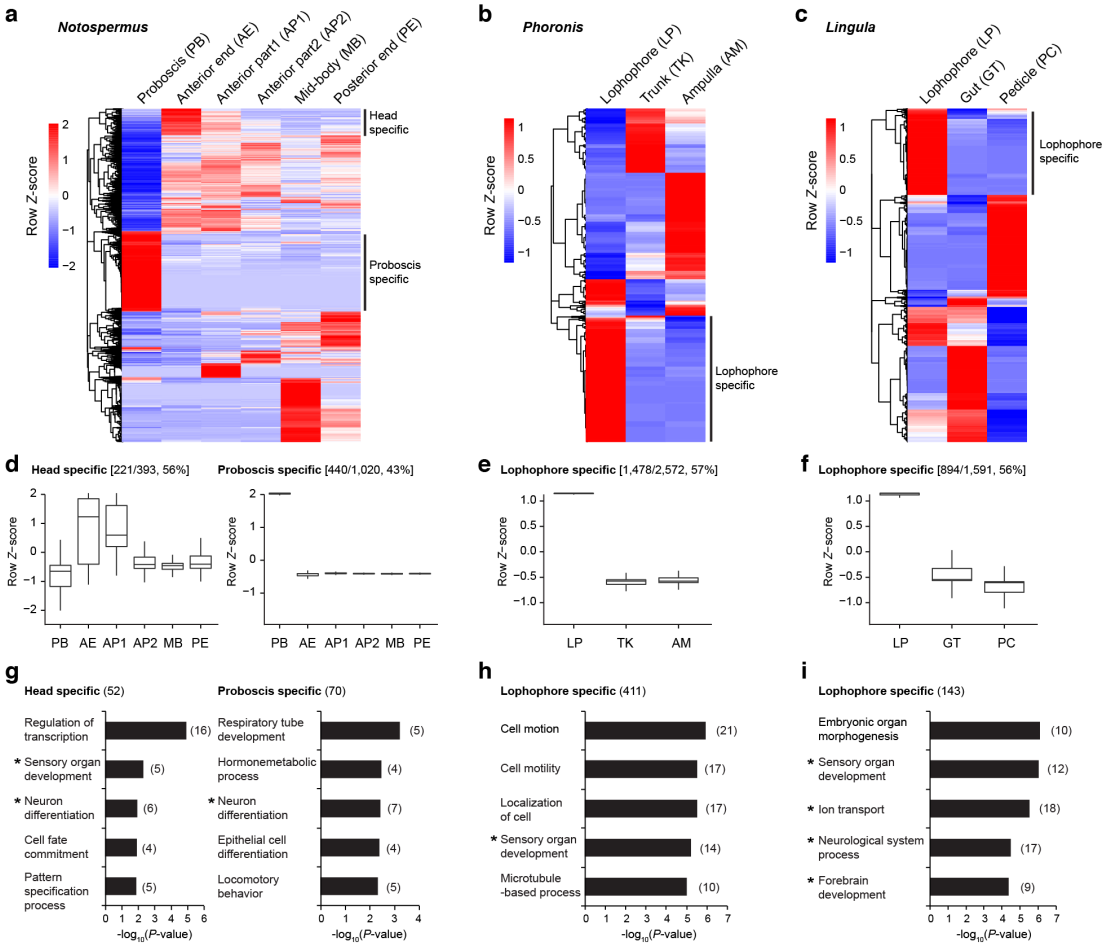


Figure 3.24 | Differential expression profiling of tissue-specific genes.
(**a**–**c**) Hierarchical clustering heatmaps of 6,880 (**a**), 7,319 (**b**) and 6,881 (**c**) differentially expressed genes ($P < 0.001$, fold change > 4). (**d**–**f**) Box plots of differentially expressed genes. Each box represents the interquartile range (IQR). Lines denote medians and whiskers indicate the most disparate data points from the median within 1.5×IQR. Numbers of tissue-specific genes are indicated in square brackets (numbers with PANTHER annotation/total numbers, the percentage of genes with annotation). (**g**–**i**) Enriched gene ontology (GO) terms of tissue-specific differentially expressed genes. Asterisks indicate GO terms that are associated with neural systems. Numbers of genes with annotated human orthologues are indicated in parentheses. Numbers of enriched genes are indicated in parentheses next to the bars.

Interestingly, many of these 'head/lophophore' genes overlap with those that are conservatively expressed during the organogenesis stage in vertebrates, called the phylotypic period (Irie and Kuratani, 2011), including *foxG1*, *pax6*, *klf2*, *emx2* and *islet1*. Most of these

genes are associated with neuron differentiation, sensory organ development and forebrain development. Thus, the vertebrate phylotypic period likely reflects the importance of the head patterning step during the evolution of bilaterian development.

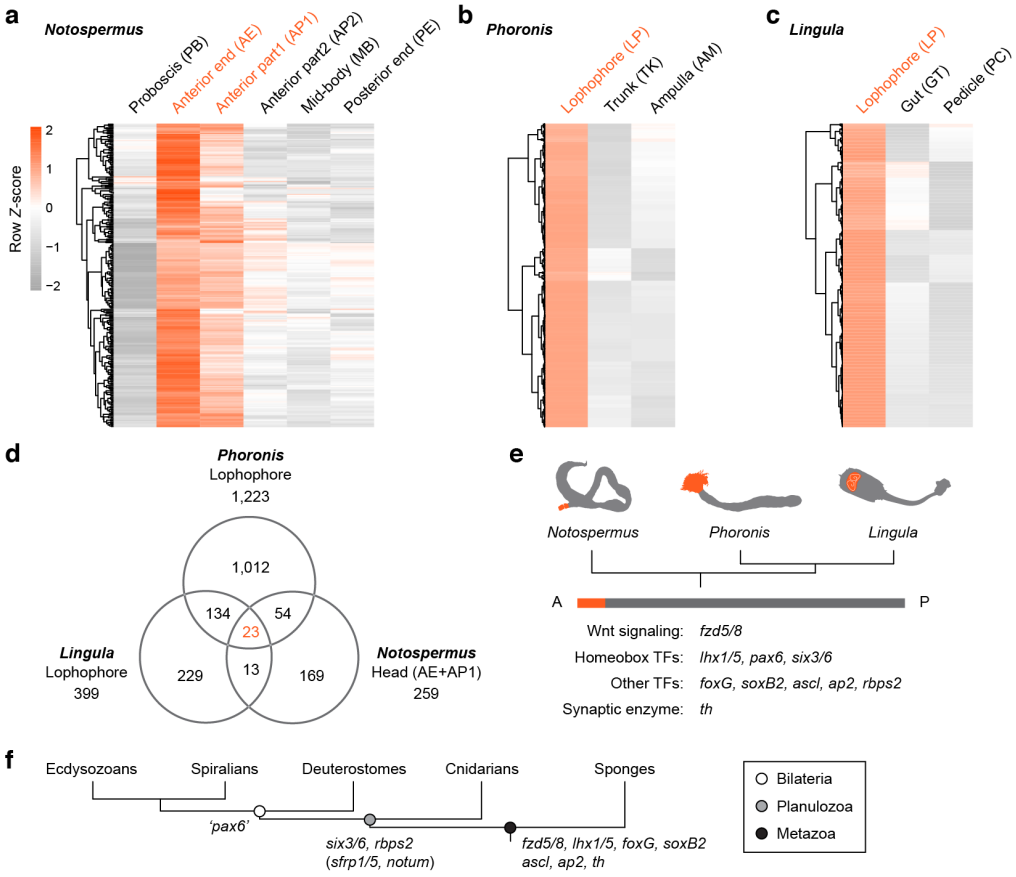


Figure 3.25 | Transcriptomic similarity of tissue-specific orthologues in anterior regions. (**a**–**c**) Hierarchical clustering heatmaps of 8,650 orthologues identified with bidirectional best hits that are specifically expressed in the head of *Notospermus* (259 genes) (**a**) and lophophores of *Phoronis* (1,223) (**b**) and *Lingula* (399) (**c**). (**d**) Venn diagram of tissue-specific genes (fold change > 2 relative to all other tissues). Numbers of genes shared by the head of *Notospermus* and lophophores of *Phoronis* and *Lingula* are shown in orange. (**e**) Schematic representation of the anterior region (shown in orange) with transcriptomic similarity shared by *Notospermus*, *Phoronis* and *Lingula*. A, anterior; P, posterior; TF, transcription factor. Genes specifically expressed in the anterior region (or head) with available annotation are listed. (**f**) Origins of head patterning genes during metazoan evolution. Genes specific to clades are listed at the nodes. Quotation marks highlight the *bona fide* bilaterian-type *pax6*. Anterior genes in parentheses indicate that they are relatively broadly expressed compared to other anterior-specific genes. The Xenacoelomorpha is excluded from the discussion due to the lack of genome resources for comparison.

In bilaterians, the AP axis is patterned by the gradient of canonical Wnt signaling through β-catenin (Petersen and Reddien, 2009). Along the axis, the bilaterian head develops

at the anterior end, characterized by centralization of the nervous system, where the Wnt signaling is down-regulated (Glinka et al., 1997).

Intriguingly, Wnt signaling genes are differentially expressed along the AP axis with the Wnt receptor *fzd5/8*, as well as Wnt antagonists, *sfrp1/5* and *notum*, which are expressed in the head of *Notospermus* and lophophores of *Phoronis* and *Lingula* (Fig. 3.26).

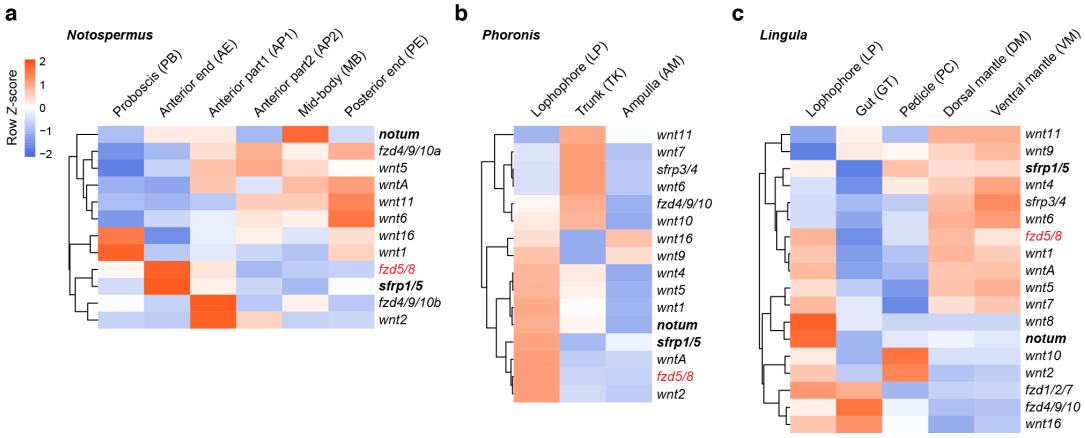


Figure 3.26 | Expression of Wnt signalling components along the anterior-posterior axis. (**a**–**c**) Hierarchical clustering heatmaps of Wnt signalling gene expressions in the tissues of *Notospermus* (**a**), *Phoronis* (**b**) and *Lingula* (**c**). Wnt antagonists are marked in bold. Shared genes with specific expression in the head of *Notospermus* and lophophores of *Phoronis* and *Lingula* are shown in red. Tissue polarity along the anterior–posterior axis in adult *Phoronid* and *Lingula* is not clearly defined because of their U-shaped guts, although the lophophore can be considered the anterior end.

Thus, it is tempting to speculate the existence of a conserved AP patterning mechanism in which inactivation of Wnt signaling at the anterior end is essential for bilaterian head formation. Superimposed on the conserved patterning system, I found 10 homeobox genes (*uncx*, *pou4*, *six4/5*, *barx*, *prox*, *arx*, *vsx*, *alx*, *msx* and *nkx1*) that are specifically expressed in both *Phoronis* and *Lingula* lophophores but not in the *Notospermus* head, suggesting a redeployment of toolkit genes in patterning lineage-specific structures (Fig. 3.27).

Taken together, the lophophore is a structure at the anterior end without Hox gene expression. It expresses Wnt antagonists, head and neuronal markers and genes that are associated with synaptic machinery and sensory functions. These features thus resemble the head patterning systems and identities seen in other deuterostomes, ecdysozoans and lophotrochozoans (Holland et al., 2013; Lowe et al., 2015; Steinmetz et al., 2010). Therefore, despite the lack of morphological similarity, lophophores bear a molecular resemblance to the

heads of other bilaterians. My findings thus suggest a possible common origin of bilaterian head patterning in the bilaterian ancestor of protostomes and deuterostomes, although distinct corresponding structures are apparently formed in different lineages (Pani et al., 2012; Santagata et al., 2012).

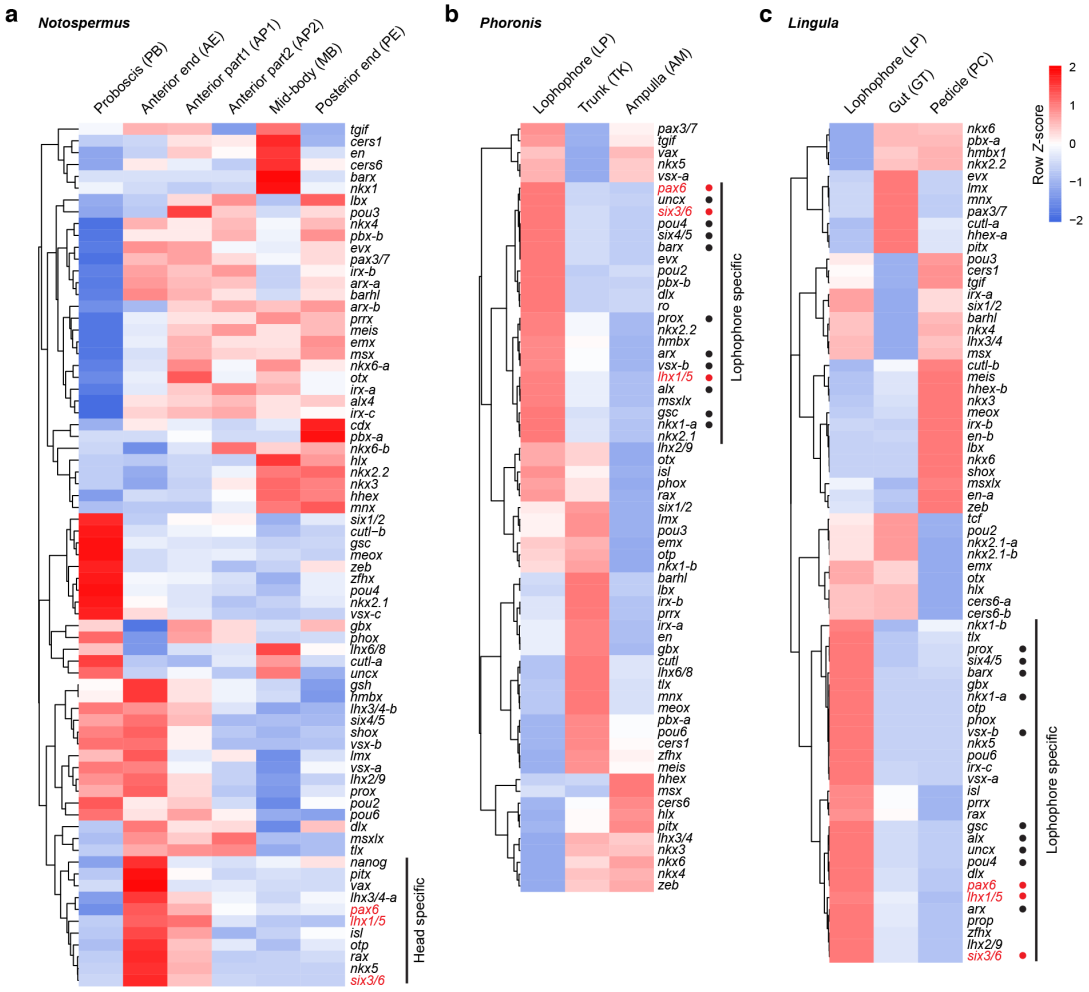


Figure 3.27 | Tissue-specific expression of non-Hox homeobox genes.
(**a**–**c**) Hierarchical clustering heatmaps of non-Hox (both Hox and ParaHox included) homeobox gene expression in tissues of *Notospermus* (**a**), *Phoronis* (**b**) and *Lingula* (**c**). Shared genes with specific expression in the head of *Notospermus* and lophophores of *Phoronis* and *Lingula* are shown in red. Solid circles indicate lophophore-specific genes shared between *Phoronis* and *Lingula*. Gene names here are manually curated with HomeoDB2 (Zhong and Holland, 2011). Gene names with '-a' (or '-b' and '-c') represent lineage-specific duplications.

3.3.6 Lineage-specific features and adaptations

One notable observation about the *Notospermus*, *Phoronis* and *Lingula* genomes is that they all have extremely low levels of 28S rRNA expression compared to those of other animals

(Fig. 3.28). In eukaryotes, 28S rDNA is usually physically linked with 18S rDNA. However, I found several cases of disconnected 18S and 28S rDNA in lophotrochozoans, including polychaetes, pearl oysters and brachiopods. The reason for the low 28S rRNA expression is not clear and whether it is due to transcriptional regulation or post-transcriptional processing will require further research. Nevertheless, recent studies in the naked mole-rat have shown that processing of 28S rRNA increases translational fidelity (Fang et al., 2014). It is tempting to speculate that the low level of 28S rRNA may contribute to lophotrochozoan survival in the intertidal environment by stabilizing their proteomes.

Invertebrates defend themselves against infection by viruses, bacteria, fungi, or other parasites through innate immune responses that involve pattern recognition and signaling (Hibino et al., 2006) (Fig. 3.29a). I found that toll-like receptor (TLR) genes are absent in rotifers, planarians and blood flukes, but are expanded in most lophotrochozoans with numbers of genes comparable with those of deuterostomes (Fig. 3.29b). The *Notospermus* and *Phoronis* genomes contain 8 and 25 TLR genes, respectively. Most TLR genes show lineage-specific expansion through tandem duplications (Fig. 3.29c,d). Although TLR genes are mostly intronless, I observed several that carry introns (Fig. 3.29d).

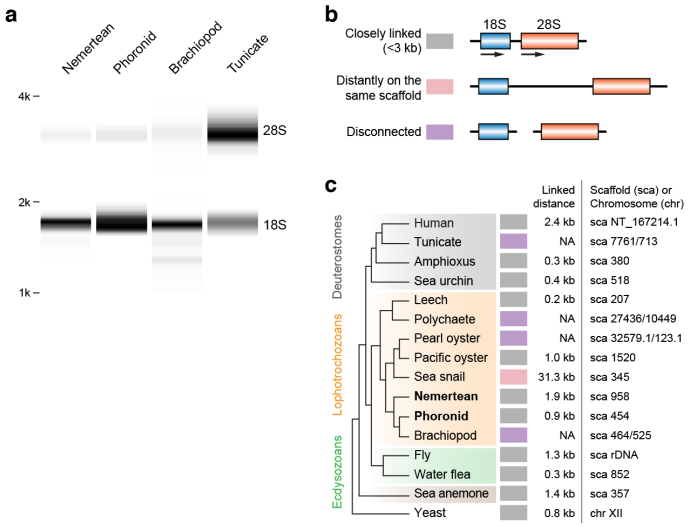


Figure 3.28 | Reduced expression levels of 28S rRNAs and separation of 18S and 28S rDNAs. (**a**) Compositions of total RNAs from selected invertebrate samples analyzed with RNA 6000 Nano chips (Agilent Bioanalyzer 2100). In *Notospermus*, *Phoronis* and *Lingula*, the amount of 28S rRNAs is relatively much less than 18S rRNA. (**b**) Genomic organization of 18S and 28S rDNA. (**c**) Genomic organization of 18S and 28S rDNA among selected eukaryotes. NA, not available.

In humans, TLR genes with low numbers of leucine-rich repeats (LRRs) (< 10) such as TLR1, TLR2 and TLR6 recognize glycolipids or lipopeptides, whereas those with high numbers of LRRs (10–18) usually target to nucleic acids (Medzhitov, 2001).

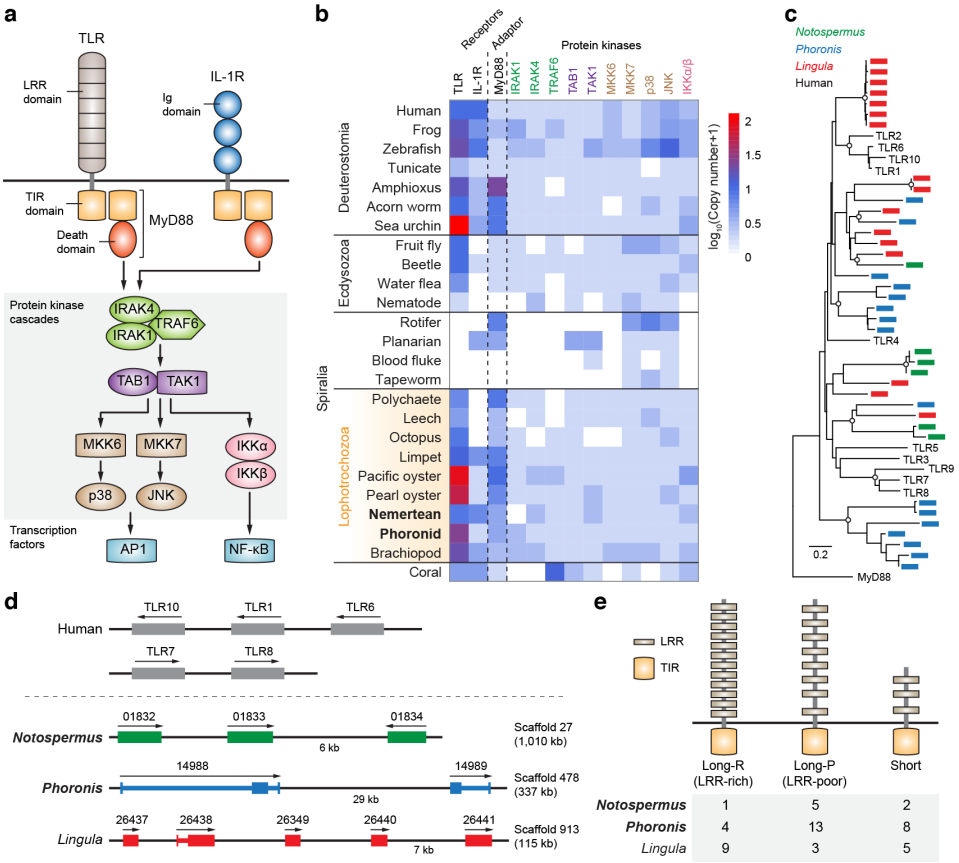


Figure 3.29 | Lineage-specific expansion of toll-like receptor genes in lophotrochozoans.
(**a**) Schematic representation of components in the MyD88-dependent pathway of toll-like receptor (TLR) and interleukin-1 receptor (IL-1R) signaling. LRR, leucine-rich repeat; TIR, toll/interleukin-1 receptor; MyD88, myeloid differentiation primary response protein 88; IRAK, interleukin-1 receptor-associated kinase; TRAF, TNF receptor-associated factor. TAB1, TAK1-binding protein 1; TAK1, TGF-beta-activated kinase 1; MKK, mitogen-activated protein kinase kinase; JNK, c-Jun N-terminal kinases; IKK, inhibitor of κB kinase; AP1, activator protein 1; NF-κB, nuclear factor-κB. (**b**) Distribution of TLR/IL-1R signaling components among selected metazoans. (**c**) Phylogenetic analysis of TLR genes among *Notospermus* (green), *Phoronis* (blue) and *Lingula* (red) with TIR domains using the neighbor-joining method with the JTT model (49 genes, 133 amino acids, 1,000 bootstrap replicates). Open circles on nodes denote bootstrap support > 60%. (**d**) Genomic organization of TLR genes. Arrows show the direction of transcription. Rectangles indicate exons. (**e**) Gene structure of TLR genes.

Expanded *Notospermus* and *Phoronis* TLR genes are mostly long and have low numbers of LRRs (Fig. 3.29e). Some TLR genes are specifically expressed in the *Phoronis*

and *Lingula* lophophores, whereas many of them have low expression across tissues, indicating that they may be triggered by infection (Zhang et al., 2015) (Fig. 3.30).

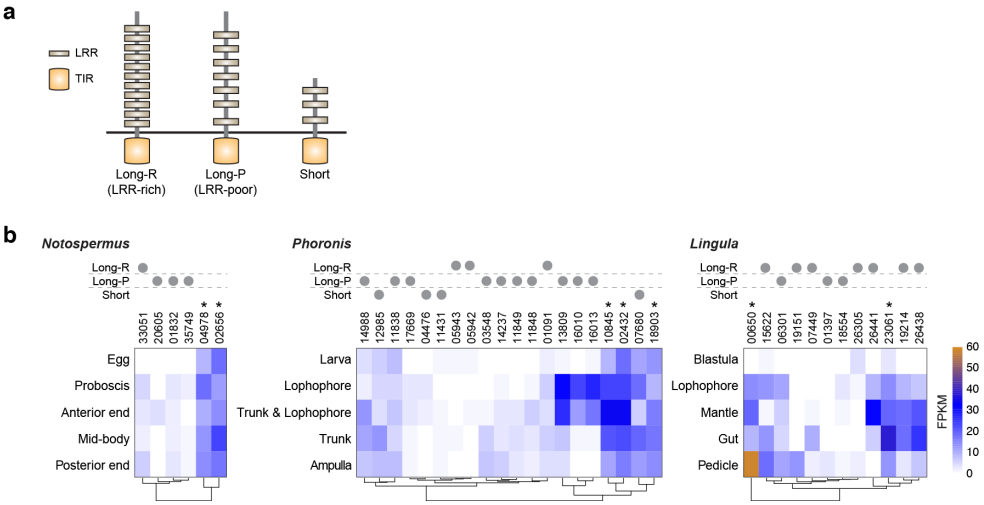


Figure 3.30 | Expression profiles of TIR-domain-containing genes.

(**a**) Gene structure of TLR genes (long, >700 amino acids; short, ≤700 amino acids; rich, ≥10 LRRs; poor, <10 LRRs). LRR, leucine-rich repeat; TIR, toll/interleukin-1 receptor. (**b**) Expression profiles of TIR-domain-containing genes (TLR and MyD88). Asterisks on gene IDs indicate MyD88. Gene structure of each gene is indicated above the heatmaps. FPKM, fragments per kilobase of transcript per million mapped reads.

3.3.7 Nemertean toxins and lophophorate biomineralization

Nemerteans produce peptide toxins to capture preys and for defense (Whelan et al., 2014). To investigate the possible origin of nemertean toxin genes, I annotated putative toxin genes with the UniProt database. I downloaded 6,592 proteins with molecular function annotated as the keyword, toxin (KW-0800). These UniProt entries belong to 11 major categories, including: (1) cardiotoxin; (2) enterotoxin; (3) neurotoxin; (4) ion channel-impairing toxin; (5) myotoxin; (6) dermonecrotic toxin; (7) hemostasis-impairing toxin; (8) G-protein coupled receptor-impairing toxin; (9) complement system-impairing toxin; (10) cell adhesion-impairing toxin; and (11) viral exotoxin.

I first assigned the orthology using reciprocal BLASTP searches (i.e. BBH) against UniProt toxin sequences followed by manual curation with OrthoMCL output (i.e. orthologous groups). Using this method, I annotated 63 putative toxin genes in *Notospermus*. To obtain putative nemertean-specific toxin genes, I filtered out possible non-toxin genes with those shared with brachiopods, phoronids and octopuses, in which those genes might have other functions.

In total, I found 15 putative toxin genes that are shared by all selected lophotrochozoans that exhibit no reported toxicity (Fig. 3.31a). These include a plancitoxin-1-like gene (i.e. DNase II with hepatotoxicity), which is first purified from the crown-of-thorns starfish, *Acanthaster planci* (Shiomi et al., 2004), which has also been found in nemerteans (Whelan et al., 2014) and jellyfish (Li et al., 2014). The fact that a plancitoxin-1-like gene is widely distributed in metazoans likely suggests that this gene may have other functions. Similarly, many putative toxin genes, such as metalloproteases and phospholipases, are shared by non-toxic lophotrochozoans, suggesting possible co-option of metabolic genes for toxic functions in certain lineages. However, I cannot exclude the possibility that these 15 genes also have toxic functions that are not yet reported in lophotrochozoans.

Here I focused on 32 putative toxin genes that are specifically present in the *Notospermus* genome, but not in other lophotrochozoans (Fig. 3.31a). Among those genes, I identified 26 genes that are differentially expressed (Fig. 3.31b). In particular, I showed that C-type lectins (snaclec 27, *SL27*; snaclec bothroinsularin subunit alpha, *SLA*) and L-amino-acid oxidase (*OXLA*), as well as a serine protease inhibitor (U-actitoxin-Avd3i, *VKT6*) associated with inhibition of platelet aggregation and hemolysis, respectively, are highly and specifically expressed in the proboscis (Fig. 3.31b). Among these toxin genes, we also found several genes that have high sequence similarities to the stonefish toxin, stonustoxin (Ueda et al., 2006). Stonustoxin is a pore-forming protein of the Membrane Attack Complex-Perforin/Cholesterol-Dependent Cytolysin (MACPF/CDC) superfamily, widely distributed among eukaryotes (Ellisdon et al., 2015). This fact suggests that they may play a broader role than envenomation. Stonustoxin-like genes are expanded in *Notospermus* (Fig. 3.31c). However, I cannot detect their expression from my transcriptome data, possibly indicating that they are only expressed in certain circumstances.

For known nemertean-specific toxin genes, I could not find neurotoxin B-II or neurotoxin B-IV in the *Notospermus* genome, indicating they might be lineage-specific in *Cerebratulus lacteus* (Blumenthal et al., 1981). Instead, I found the cytolytic protein, cytotoxin A-III, which is expanded in *Notospermus* (Fig. 3.31d). Cytotoxin A-III is a polypeptide cytotoxin first isolated from *C. lacteus* mucus (Kem and Blumenthal, 1978) that has also been found in another heteronemertean, *Parborlasia corrugatus* (Butala et al., 2015). Using published transcriptome data, I also examined whether cytotoxin A-III is present in the hoplonemerteans, *Paranemertes peregerina* and *Malacobdella grossa* (Whelan et al., 2014), as well as the palaeonemerteans *Tubulanus polymorphus* (Halanych and Kocot, 2014) and *Cephalothrix linearis* (Egger et al., 2015). I failed to find cytotoxin A-III in these nemerteans.

Given that *Notospermus* is also within the group of Heteronemertea, cytotoxin A-III might be specific to that lineage. I showed that some cytotoxin A-III genes are tandemly duplicated (Fig. 3.31e). Expanded cytotoxins A-III exhibit diverse expression. They are ubiquitously expressed among tissues or specifically expressed in eggs or the proboscis (Fig. 3.31f).

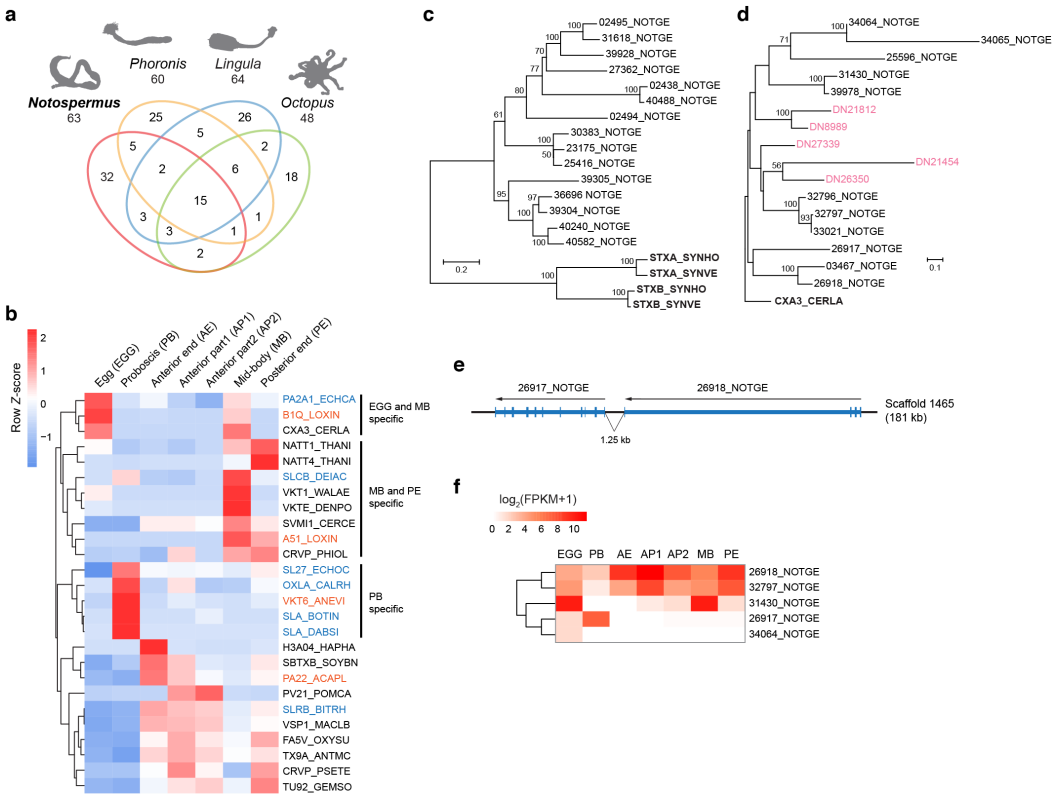


Figure 3.31 | Identification of putative toxin genes and expansion of toxin genes in the *Notospermus* genome.

(**a**) Venn diagram of shared and unique putative toxin genes in *Notospermus* and selected lophotrochozoans. (**b**) Differential expression profiles of putative toxin genes that are not shared with selected lophotrochozoans (26/32). Best hits UniProt entries are listed on the right side. Genes associated with inhibition of platelet aggregation (blue) and hemolysis (red) are highlighted in different colours. (**c** and **d**) Phylogeny of stonustoxin-like genes (19 genes, 593 amino acids) (**c**) and cytotoxin A-III genes (17 genes, 118 amino acids) (**d**) using the neighbour-joining method with the JTT model (1,000 bootstrap replicates). Protein sequences from transcriptome data of *Cerebratulus* sp. are shown in pink. (**e**) Genomic organization of cytotoxin A-III genes. Arrows indicate the direction of transcription. Rectangles represent exons. (**f**) Expression of cytotoxin A-III genes. FPKM, fragments per kilobase of transcript per million mapped reads.

Among lophophorates, phoronids are the only group without mineralized tissues. Chitin synthase genes which are required for biomineralization are reduced in *Phoronis* (6) compared to *Lingula* (31) (Luo et al., 2015). Some chitin synthase genes present in molluscs

and brachiopods with close orthology cannot be found in *Phoronis*, likely indicating loss of these genes in the phoronid lineage.

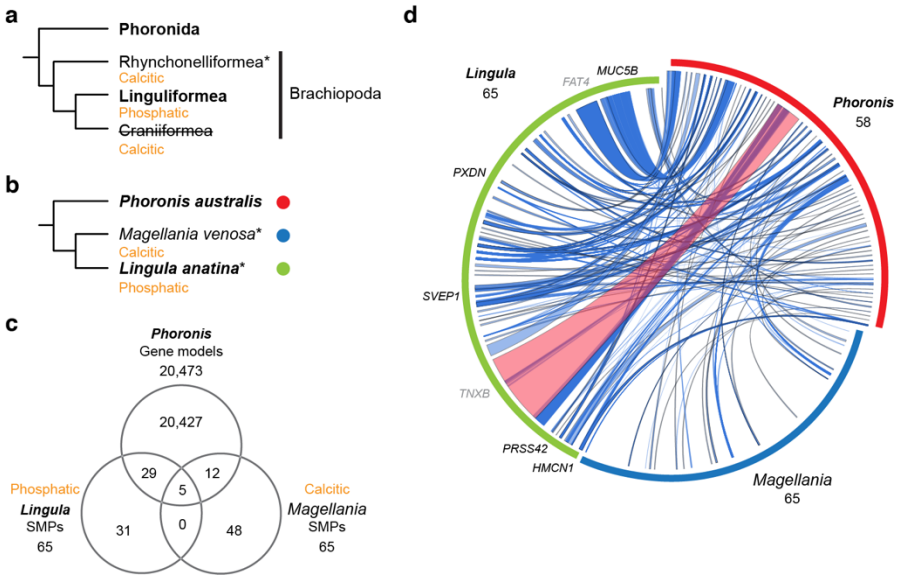


Figure 3.32 | Comparison of brachiopod shell matrix proteins with the *Phoronis* genome. (**a**) Simplified phylogenetic positions of Phoronida and Brachiopoda. Chemical compositions of mineralized shells are labelled in orange. Taxa with at least one available genome are highlighted in bold. Asterisks indicate the availability of shell proteomes and mantle transcriptomes. Craniiformea has no sequence data for comparison. (**b**) Species used for genomic and transcriptomic comparison in this study. (**c**) Venn diagram of shared shell matrix proteins (SMPs) found in mantle transcriptomes among phoronids and brachiopods. (**d**) Circos plot showing sequence comparison based on BLAST (*e*-value = $1e^{-10}$) visualized with Circoletto (Darzentas, 2010) (color codes for ribbons: blue, sequence identity <= 50%; red, > 99.9999%). Genes shared by *Phoronis*, *Lingula* and *Magellania* are labelled in black next to the *Lingula* sequences. Genes with high sequence similarity shared between only *Phoronis* and *Lingula* are shown in grey. Most brachiopod SMP genes have no detectable sequence homology in *Phoronis*.

Next, to explore the origin of mineralized tissues in lophophorates, I compared biomineralization-related genes among phoronids and brachiopods. To achieve a comprehensive comparison, besides *Lingula*, I downloaded the mantle transcriptome of the brachiopod *Magellania venosa* (http://dx.doi.org/10.5282/ubm/data.67) (Jackson et al., 2015). I retrieved 65 unique *Magellania* shell matrix proteins (SMPs) (contig 20189 has two protein IDs: F20121130 and F30121131), and then compared those genes with 65 *Lingula* SMPs (Luo et al., 2015) and the *Phoronis* genome (Fig. 3.32a–c). Using reciprocal BLAST searches, I found only five SMP genes that are shared by *Phoronis*, *Lingula* and *Magellania* (Fig. 3.32c). These genes include *peroxidasin* (*PXDN*), *mucin-5B* (*MUC5B*), *serine protease 42* (*PRS42*), *SVEP1* and *hemicentin-1* (*HMCN1*) (Fig. 3.32d).

Some other genes that are shared between *Phoronis* and *Lingula* (29), as well as *Phoronis* and *Magellania* (12), are mostly extracellular matrix structural proteins or enzymes. Most of these genes can also be found in other metazoans with other functions besides biomineralization. On the other hand, I cannot find any *Lingula* or *Magellania* lineage-specific SMPs in the *Phoronis* genome, suggesting the possible independent evolution of biomineralization in different brachiopod lineages, similar to that of molluscs (Aguilera et al., 2017; Jackson et al., 2010; Kocot et al., 2016a).

Therefore, my results show that lineage-specific *CHS* gene expansion, novel SMP genes and redeployment of extracellular matrix genes are involved in the evolution of biomineralization in brachiopods, and possibly also in ectoprocts. Thus, my findings suggest that lineage-specific gene expansions, gain of novel genes and redeployment of extracellular matrix genes are involved in the evolution of lophophorate biomineralization.

3.4 Discussion

I have decoded two lophotrochozoan genomes representing two unexplored phyla, the Nemertea and Phoronida. I have also provided an updated brachiopod genome. Despite being phylogenetically closely related, nemerteans, phoronids and brachiopods diverged early, perhaps before the Cambrian explosion (Erwin et al., 2011). During more than 540 million years of evolution, they have evolved many lineage-specific features, and yet retained unexpected elements in terms of the bilaterian gene repertoire and head patterning system.

One remarkable finding is that the same developmental head organizer genes are expressed in the adult anterior structure, which may highlight their roles in maintaining tissue identity and homeostasis in all bilaterians. I argue that the molecular basis of morphological features is the combination of the conserved gene repertoire and patterning system together with lineage-specific gene family expansions and novel genes (Albertin et al., 2015; Luo et al., 2015; Zhang et al., 2012). However, co-option and redeployment of the developmental toolkit and structural genes in different lineages also contribute to specialization and functions of body structures (Jandzik et al., 2015). Although my phylogenetic analysis based on transcriptome data suggests the monophyly of lophophorates, an ectoproct genome will be needed for a comprehensive understanding of lophophorate origins. Given Xenacoelomorpha as the earliest branching bilaterians (Cannon et al., 2016), the origins of the bilaterian gene repertoire and heads will be further clarified with the available genomes from Acoela, Nemertodermatida and *Xenoturbella*. The draft *Notospermus* and *Phoronis* genomes

presented here together with my comparative genomics and transcriptomics provide insights into conservations and dynamics during lophotrochozoan evolution.

4 Mitochondrial gene order variation in the brachiopod *Lingula anatina* and its implications for mitochondrial evolution in lophotrochozoans

4.1 Introduction

Mitochondria are essential eukaryotic organelles that generate energy and participate in metabolism, thermoregulation, apoptosis, and aging. They carry their own genomes, which are typically 14 to 18 kb in most animals. The genome encodes mitochondrial translational machinery (rRNAs and tRNAs) as well as genes for oxidative phosphorylation involving ATP production and heat generation (Boore, 1999). Because of the elevated mutation rate of mitochondrial genes and their relatively stable gene arrangement, mitochondrial sequences are often used as markers for deciphering phylogenetic relationships among metazoans, especially among vertebrates (Bernt et al., 2013a; Gissi et al., 2008; Perseke et al., 2013). However, these features are variable in marine invertebrates. In tunicates, mitochondrial gene arrangement is highly variable, even at the interspecific levels (Gissi et al., 2010). For example, *nad1* and tRNA genes are rearranged between *Ciona intestinalis* and *C. savignyi* (Gissi et al., 2004). More strikingly, gene order is complete shuffled within the genus *Phallusia*, suggesting that tunicates have high mitochondrial genome plasticity (Iannelli et al., 2007a). On the other hand, similar cases have been reported in molluscs, where rearrangement of tRNA occurs within the oyster genus, *Crassostrea* (Milbury and Gaffney, 2005). High variance among congeneric species has been also shown in gastropods of the genus *Dendropoma*, in which *nad6* and tRNA genes are transposed (Rawlings et al., 2001). Although mitochondria are transmitted maternally in most metazoans, doubly uniparental inheritance (DUI) is found in some bivalves (Breton et al., 2007). In this context, homologous recombination has occurred and mitochondrial genomes are thus heteroplasmic, leading to high sequence variation (Breton et al., 2014). While more than 3,000 metazoan mitochondrial genomes have been sequenced (>2,000 for vertebrates) (D'Onorio de Meo et al., 2012), only four are available from brachiopods (Endo et al., 2005; Helfenbein et al., 2001; Noguchi et al., 2000; Stechmann and Schlegel, 1999). Mitochondrial genomes of congeneric brachiopods have been poorly explored.

Brachiopods are lophotrochozoans, protostomes that manifest deuterostomic features during development, such as radial cleavage and enterocoelic coelom formation (Yatsu, 1902). Although brachiopods superficially resemble bivalves, they have dorsoventral shells, lophophores for feeding, and a pedicle for attachment (Bitner and Cohen, 2013). Brachiopods

are well known from their rich fossil record. The similar shell morphology of extinct and extant lingulid brachiopods inspired Darwin with the idea of 'living fossils.' Among brachiopods, the subphylum Linguliformea has been recognized as one of the most primitive groups, with fossils dating back to the early Cambrian (Williams et al., 1996). Populations of the lingulid, *Lingula anatina*, are widely distributed in the western Pacific (Williams et al., 2000). However, in contrast to the 'living fossil' concept, high heterogeneity of Cox1 amino acid sequences has been observed, indicating rapid mitochondrial evolution (Endo et al., 2001).

Previously, the mitochondrial genome of *L. anatina* from Yanagawa (Kyushu, Japan) was shown to have large repetitive sequences and a high substitution rate compared with those of other brachiopods (Endo et al., 2005). These features resemble those of some bivalves that exhibit DUI (Passamonti et al., 2011). Using an Illumina MiSeq, I sequenced the *L. anatina* mitochondrial genome from a different locality (Amami Island, Japan), as part of the nuclear genome project (Luo et al., 2015). Unexpectedly, the mitochondrial gene arrangement is completely shuffled between these two localities. I performed large-scale analyses of gene order and molecular phylogeny using 101 metazoan mitochondrial genomes. In addition, pairwise comparisons of non-synonymous substitution rates at specific, generic, and phylum levels were conducted. My results suggest that the evolutionary history of mitochondrial genomes in *Lingula*, mussels, and oysters is unique. Analyses of whole mitochondrial genomes might be useful to address taxonomic issues in *L. anatina*.

4.2 Methods

4.2.1 Mitochondrial genome sequencing, assembly, and characterization

Gravid adults were collected in Kasari Bay, Amami Island, Japan (hereafter Amami *Lingula*). DNA samples were obtained from male gonads to avoid contamination from ingested algae and surface-attached microbes. Paired-end and Cre-LoxP mate pair libraries were made from a single male specimen, while a Nextera mate pair library was prepared from a different sample. Libraries were then sequenced using an Illumina MiSeq. Since both nuclear and mitochondrial DNAs were included during library preparation, after genome assembly with Newbler (v2.9), the preliminary mitochondrial scaffold was identified with TBLASTN searches (v2.2.29+) (Camacho et al., 2009) against mitochondrial proteins, Cox1 (UniProt entry name: COX1_CAEEL) and Nad5 (NU5M_BRAFL). To improve the quality of the mitochondrial assembly, a more accurate assembly was performed using the de Bruijn graph

approach. Illumina MiSeq reads were mapped to the preliminary mitochondrial scaffold with Bowtie2 (v2.1.0) (Langmead and Salzberg, 2012). Mapped reads were retrieved to reassemble the mitochondrial genome. Assembly, contiging, scaffolding, and gap-closing were conducted using the Platanus pipeline (Kajitani et al., 2014) with a setting of 47-mer. (v1.2.1; platanus assemble -k 47 -s 1 -n 1500 -c 1500 -a 5.0 -u 0.2 -d 0.3; platanus scaffold -u 0.2).

Sequencing coverage was estimated by k-mer analysis using Jellyfish (v2.0.0) (Marcais and Kingsford, 2011). Per base coverage was calculated from the mapping result using SAMtools (v1.1; samtools mpileup -BQ0 -d10000000) (Li et al., 2009). Hypervariable regions (HVs) and insertions and deletions (indels) were identified with a sorted binary sequence alignment/map (BAM) file using the Integrative Genomics Viewer (IGV; v2.3.36) (Robinson et al., 2011). Reads mapped to HVs were retrieved with a tool in JVARKIT, SAM4WebLogo (https://github.com/lindenb/jvarkit). Nucleotide composition percentages were generated using WebLogo (v3.4) (Crooks et al., 2004).

4.2.2 Sequence annotation

Annotation of the Amami *Lingula* mitochondrial genome was performed using Dual Organellar GenoMe Annotator (DOGMA) (Wyman et al., 2004) and MITOS (v671) (Bernt et al., 2013b) with the invertebrate mitochondrial setting (start, ttg/att/ata/atg; stop, taa/tag). To improve annotation accuracy, manual examination of all open reading frames (ORFs) obtained from getorf (EMBOSS package v6.6.0.0; getorf -table 5 -minsize 100 -find 1) was also conducted (Rice et al., 2000). Mitochondrial rRNAs were identified using BLASTN searches against Yanagawa *Lingula* mitochondrial rRNAs. Mitochondrial tRNA genes were identified using tRNAscan-SE (v1.21) (Lowe and Eddy, 1997) and ARWEN (v1.2.3.c; arwen -mtx -gcmet) (Laslett and Canback, 2008). Transfer RNA secondary structures were predicted with a thermodynamics-based program RNAstructure (v5.7) (Reuter and Mathews, 2010), and tRNA sequences and structures were manually curated. Unassigned open reading frames (URF) with no detectable primary sequence homology to known proteins were subjected to prediction of possible structures and functions with I-TASSER (Yang et al., 2015).

4.2.3 Transcriptome analyses

cDNA libraries were prepared from embryonic stages and adult tissues using a TruSeq stranded mRNA sample prep kit. In total, 369 million pairs of 100-bp paired-end reads were sequenced using an Illumina HiSeq2500. After quality trimming, sequence reads were *de novo* assembled using a de Bruijn graph-based assembler Trinity (r2013_08_14) (Haas et al.,

2013). Mitochondrial protein-coding genes, together with rRNA and tRNA genes were searched against the transcriptome assembly using BLASTN to identify polycistronic transcripts. To estimate transcript abundance, sequence reads were mapped back to assembled transcripts with Bowtie 2 (Langmead and Salzberg, 2012). Expression levels of transcripts were estimated using RSEM (v1.2.5) (Li and Dewey, 2011) by calculating fragments per kilobase of transcript per million mapped reads (FPKM).

4.2.4 Gene rearrangement analyses

One hundred metazoan mitochondrial genomes in GenBank format were downloaded from the NCBI database. Gene names retrieved from GenBank files were then modified for nomenclatural consistency (i.e., ATP6, COB, COX1, COX2, COX3, ND1, ND2, ND3, ND4, ND4L, ND5, and ND6). Gene orders were also obtained from GenBank files with *cox1* reoriented as the first gene. Gene rearrangement and breakpoint distance analyses were performed using Common interval Rearrangement Explorer (CREx) (Bernt et al., 2007). Heatmap matrixes were created using R (v3.0.2; http:/www.R-project.org/) with the package Bioconductor (v3.0) (Gentleman et al., 2004) and pheatmap (v0.7.7).

4.2.5 *Ka* and *Ks* analyses

Protein and nucleotide sequences with species names in the header were retrieved from GenBank files. The Perl script, ParaAT (v1.0) (Zhang et al., 2012), which incorporates NAL2PAL (v13) (Suyama et al., 2006) and KaKs_Calculator (v2.0) (Wang et al., 2010) in its pipeline, was used for analyses of non-synonymous (*Ka*) and synonymous (*Ks*) substitution rates. After resolving inconsistencies of gene names, *Ka* and *Ks* were computed using ParaAT with a list of homolog pairs as well as corresponding protein and nucleotide sequences.

4.2.6 Molecular evolution and phylogenetic analyses

Mitochondrial orthologs in different species were pooled into separate files for further analyses. Sequence alignments for each protein were conducted separately with MAFFT (v7.130b) (Katoh and Standley, 2013) and unaligned regions were trimmed with Gblocks (v0.91b) (Castresana, 2000). Aligned conserved sites of 12 protein-coding genes (except *atp8*) were concatenated for phylogenetic analysis. Fasta format was converted into phylip format using a Perl script catfasta2phyml.pl (https://github.com/nylander/catfasta2phyml). Maximum likelihood analyses were performed using PhyML (v20120412; -q -d aa -c 4 -a e -b 100) (Guindon et al., 2010) with mtREV (based on amino acid substitution matrix in

vertebrate mitochondria) (Adachi and Hasegawa, 1996), mtART (arthropod mitochondria) (Abascal et al., 2007), and rtREV (retroviruses) (Dimmic et al., 2002) models with gamma distribution. Nucleotide sequences of 18S rDNA and *cox1* were downloaded from NCBI and aligned with MAFFT. For nucleotide analyses, the best-fit model was determined by the lowest values of Bayesian Information Criterion and Akaike Information Criterion estimated with MEGA6 (v6.06) (Tamura et al., 2013). The GTR+G+I model was selected and maximum likelihood analysis was conducted using MEGA6.

4.3 Results

4.3.1 Mitochondrial genome sequencing and assembly

Next-generation sequencing (NGS) was used to obtain complete mitochondrial genomes, using Roche 454 and Illumina platforms (Arquez et al., 2014; Williams et al., 2014). These PCR-free methods generate complete mitochondrial sequences without prior knowledge of mitochondrial genes, providing a powerful strategy for broader sampling. Recently, an Illumina short-read (100 bp)-based pipeline was developed for sequencing novel mitochondrial genomes (Hahn et al., 2013).

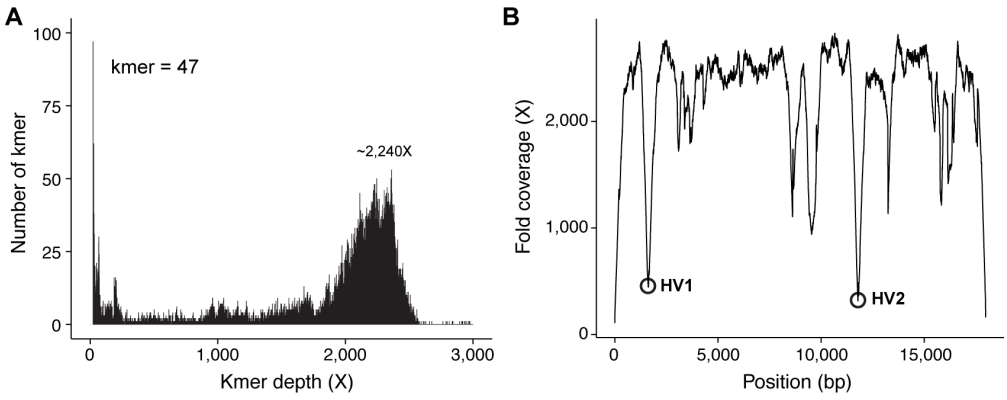


Figure 4.1 | High sequencing coverage of the Amami *Lingula* mitochondrial genome.
(**A**) *K*-mer analysis using a size of 47-mer. (**B**) Mapping of Illumina paired-end reads to the mitochondrial genome (with *cox1* as the first gene). Circles denote the hypervariable regions (HV) for which coverage was the lowest (~400-fold). Mean coverage was >2,000-fold.

However, here I applied a different approach. As part of the nuclear genome project (Luo et al., 2015), I obtained 64.0 Gb of long read data (mainly 250 bp) from the Illumina MiSeq platform. Taking advantage of the longer read-length of the MiSeq, reads containing both nuclear and mitochondrial sequences were first assembled using Newbler. The

preliminary mitochondrial scaffold was identified by searching for mitochondrial genes. I then mapped all raw reads back to the preliminary mitochondrial scaffold, which gave rise to 74.4 Mb data (i.e. 0.12% of mitochondrial reads). Sequencing coverage of these data is ~2,240-fold as estimated by k-mer analysis (Fig. 4.1A). Finally, I performed *de novo* assembly with these data using Platanus to obtain the final mitochondrial assembly.

To investigate heterogeneity of mitochondrial copies from the same individual, I mapped back mitochondrial reads from paired-end libraries. I found that average coverage was above 2,000-fold, while there were two regions with lower coverage of ~400-fold (Fig. 4.1B). Careful examination showed that these were hypervariable regions (HV) ~70 bp in length. There are 35-bp sequences denoted as "N" (gaps or lack of a consensus sequence) in the final mitochondrial assembly within the HV1 (1,617–1,651 bp) where polymorphisms are too numerous to yield consensus sequences. In addition, the regions flanking HV1 showed high insertion and deletion (indel) rates. On the other hand, in HV2, indels occurred together within the region of high polymorphism. My data suggest that there are different mitochondria within the same individual. This implies that *Lingula* mitochondria may be inherited biparentally and undergo recombination or that there is a high mutation rate during mitochondrial replication. Further studies will be needed to resolve the origin of this sequence variation.

4.3.2 Mitochondrial genome organization and transcriptome

The final assembly of the Amami *Lingula* mitochondrial DNA (mtDNA) was a circular genome, 17,970 bp with a 35-bp highly polymorphic region between *cox1* and *nad4l* (Fig. 4.2A). HV1 and HV2 are situated in intergenic regions between *cox1-nad4l* (1,600–1,670 bp) and *trnE-trnT* (11,760–11,830 bp), respectively (Fig. 4.2A). The genome contains 13 protein-coding genes, 2 rRNAs, and 23 tRNAs. All genes are transcribed from the same strand of mtDNA. Amami *Lingula* has two copies of *trnF*, *trnL*, *trnW*, and *trnY*, whereas I failed to find *atp8* and *trnK* (Fig. 4.2A).

Interestingly, I found an unassigned open reading frame (URF) located between *trnW-uca* and *trnW-cca* that shares a high similarity with Yanagawa *Lingula* URF2 (Endo et al., 2001). Further analysis based on possible 3D structure similarities suggests that this protein may be related to oxidation-reduction processes. GC content (40.3%) is comparable to that of other terebratulide brachiopods, but higher than that of phoronids (Helfenbein and Boore, 2004; Helfenbein et al., 2001; Noguchi et al., 2000; Stechmann and Schlegel, 1999).

Unexpectedly, the mtDNA length, GC content, and gene order of Amami *Lingula* are strikingly different from those reported for Yanagawa *Lingula* (Endo et al., 2005).

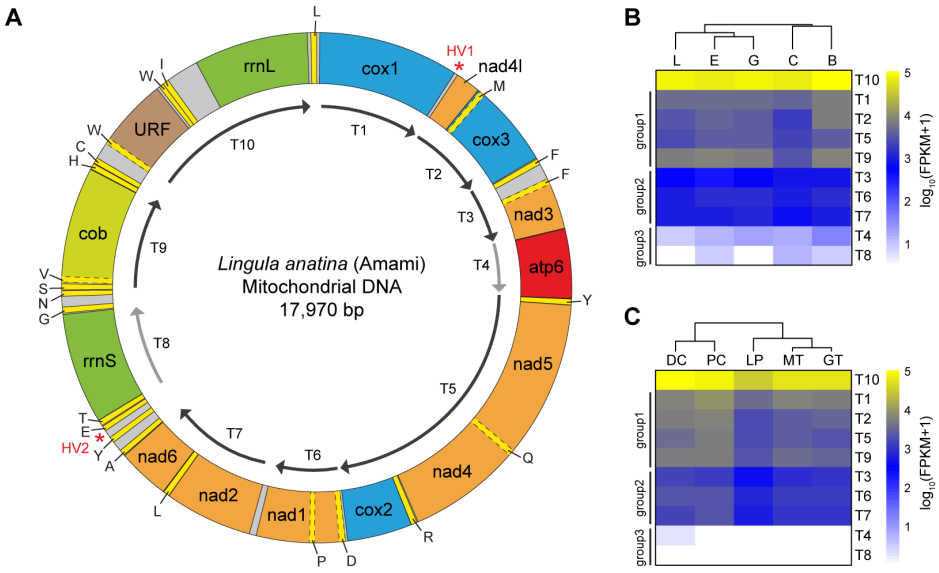


Figure 4.2 | The Amami *Lingula* mitochondrial genome is polycistronic and is differentially expressed in embryos and adult tissues.

(**A**) Organization of the genome. All genes are encoded in the same strand. Intergenic regions are in grey. Letters denote amino acids transferred by the tRNA encoded. Boxes in dashed outlines indicate that tRNA genes have an overlap with other transcripts. Asterisks mark hypervariable regions (HV). Transcripts (curved arrows denoted by T+number) are transcribed in the clockwise direction. Weakly expressed transcripts are shown as grey curved arrows. URF, unassinged open reading frame. Mitochondrial transcript expression varies with embryonic stage (**B**) and with adult tissue (**C**), although relative expression levels of transcripts follow the same general pattern. FPKM, fragments per kilobase of transcript per million mapped reads. Abbreviations: L, larva; E, egg; G, gastrula; C, cleavage; B, blastula; DC, digestive cecum; PC, pedicle; LP, lophophore; MT, mantle; GT, gut.

In humans, mitochondrial transcripts are polyadenylated and polycistronic; polyadenylation may play a role in mitochondrial RNA degradation and regulation (Slomovic et al., 2005). To examine whether brachiopod mitochondrial transcripts have these characteristics, I prepared polyadenylated RNA libraries and performed RNA sequencing (RNA-seq). Transcriptomes assembled *de novo* from RNA-seq showed that mitochondrial transcripts in Amami *Lingula* are polycistronic (Fig. 4.2A). In addition, I further showed that those transcripts are differentially expressed in embryos and adult tissues. I found that a single transcript, including 16S rRNA (*rrnL*), *trnW*, and URF was the most highly expressed, while expression levels of transcripts with *atp6* and 12S rRNA (*rrnS*), respectively, were extremely poorly expressed. Higher expression of mitochondrial transcripts was observed during the

blastula stage, and in digestive cecum and pedicle, suggesting different metabolic requirements compared to other embryonic stages and tissues (Fig. 4.2B and C).

Figure 4.3 | Gene order is stable and relatively invariant in vertebrates, but highly variable in invertebrates.

The Amami and Yanagawa *Lingula* populations have significantly different gene orders, and the Amami population appears to have lost two genes. Interestingly, the mollusc, *Oncomelania*, the nemertean, *Lineus*, and the brachiopod, *Terebratulina* all share the same gene order (red). Transposition (light grey) and inversion (dark grey) in invertebrates are compared with those of vertebrates. Taxa with high gene order variation among species are marked with asterisks. Gene order of *Ciona intestinalis* is shown for tunicates. A, Amami; Y, Yanagawa.

4.3.3 Mitochondrial gene order in lophotrochozoans

Given that the mitochondrial gene order of Amami *Lingula* is completely different from that of Yanagawa *Lingula*, I examined whether this is a general feature in other marine

invertebrates. A total of 100 complete mitochondrial genomes, mostly of lophotrochozoans were downloaded for analyses. Since tRNA genes are highly mobile, I first analyzed gene order of protein-coding genes. Abalones, chitons, nautilus, and octopuses share exactly the same protein-coding gene order. I found that among these 101 species, gene order is highly conserved in the vertebrate lineage (Fig. 4.3). However, tunicates, oysters, mussels, and brachiopods showed highly variant gene orders compared with other taxa. In oysters and mussels, *atp8* was lost, and I could not detect *atp8* in Amami *Lingula* (Fig. 4.3). The only conservation between Amami and Yanagawa *Lingula* is the *nad5-nad4* gene block.

In fact, this block was conserved in all metazoans except tunicates and annelids (Fig. 4.3). In Amami *Lingula*, *nad5* and *nad4* are transcribed together with *cox2* as a single transcript (Fig. 4.2A), suggesting that there may be a transcriptional basis for this conservation. Interestingly, freshwater snails (*Oncomelania*) (Zhao et al., 2010), ribbon worms (*Lineus* and *Zygeupolia*) (Chen et al., 2012; Podsiadlowski et al., 2009), and terebratulide brachiopods (*Terebratulina*) (Stechmann and Schlegel, 1999) have exactly the same gene order, suggesting that they acquired this feature from a common ancestor. Furthermore, the gene order in phoronids differs only at one point from that of octopuses, in which *atp6* is transposed between *cox3* and *nad3,* compared to the standard vertebrate gene order (Fig. 4.3). Collectively, this observation may reflect the close phylogenetic relationship of these groups.

To better understand evolution of gene order in bilaterians, I first analyzed the breakpoint distance of protein-coding genes (Blanchette et al., 1999). Distances of genome rearrangement are estimated by numbers of breakpoint, where lower breakpoint distances indicate greater similarity between species. Among 101 species, I found that there were 36 different gene orders. The vertebrate gene order had smallest total breakpoint distance (Fig. 4.4), suggesting that vertebrate gene order may retain the ancestral bilaterian condition. Gene orders from other taxa can be explained by converting them to the vertebrate order with the smallest possible number of gene or gene-block transpositions and inversions.

When comparing the breakpoint distance with a phylogenetic framework, I found that there were three major conserved matrix blocks, including: (1) deuterostomes (vertebrates and ambulacrarians); (2) fruit flies and cephalopods; and (3) annelids, nemerteans, and phoronids (Fig. 4.4). These matrix blocks also showed very high similarity. In addition, gene orders of the nemerteans, *Lineus viridis* and *Emplectonema gracile*, were very close to those of deuterostomes, while the phoronid, *Phoronis psammophila*, was highly similar to the cephalopod, *Nautilus macromphalus*. On the other hand, gene orders in tunicates, molluscs

(except cephalopods), and brachiopods were very diverse, sharing little with other groups (Fig. 4.4).

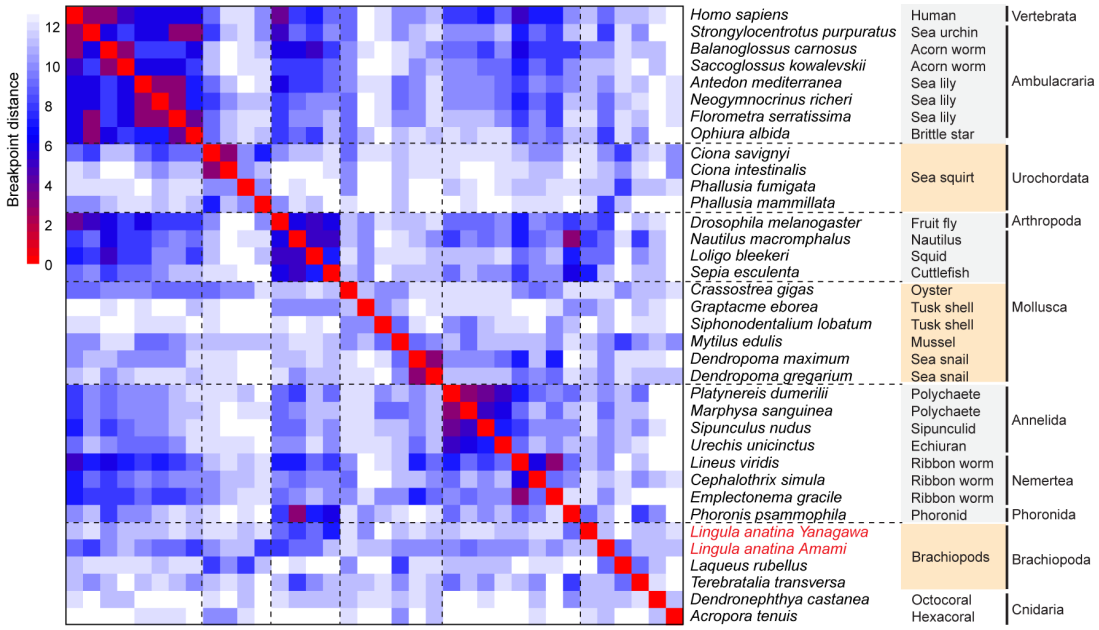


Figure 4.4 | A breakpoint distance matrix shows that some metazoan groups, highlighted in grey (right), have highly conserved gene orders among group members.
Groups highlighted in orange have more disparate gene orders. Groups are separated by dashed lines. Two individuals of *Lingula anatina* from different sampling localities are in red.

When comparing the total breakpoint distance by sorting invertebrate values compared to those of humans, I found that tunicates, brachiopods, sea snails, and tusk shells were the most distant from vertebrates. The small breakpoint distance of the ribbon worm, *Lineus viridis*, to deuterostomes and fruit flies, suggests that they share the ancestral bilaterian mitochondrial gene order. Furthermore, broad-scale analysis based on the whole gene set, including not only protein-coding genes, but also rRNA and tRNA genes, was also conducted. There were 57 different gene orders among 101 species. In deuterostomes, a group of chordates and hemichordates and a group of echinoderms shared the most similar in-group gene order.

Moreover, I showed that gene orders in annelids, nemerteans, and phoronids highly resembled each other. These three groups also shared similarities with deuterostomes. Nemerteans and phoronids were closest to a group containing fruit flies and cephalopods. Although tunicates and bivalves showed diverse gene orders compared to other taxa, they

were highly similar at the congeneric level. In contrast, gene orders of two *Lingula anatina* collected at different localities were complete shuffled.

Gene order is diverse in brachiopods and some marine invertebrates, such as tunicates and bivalves. In fact, it is not rare to have very different gene orders in the same genus, as has been reported in tunicates *Phallusia* (Iannelli et al., 2007a) and sea snails *Dendropoma* (Rawlings et al., 2001). However, to the best of my knowledge, there are no examples of different gene orders within a species. My findings reveal a *Lingula anatina* species complex and suggest that evolutionary rates of nuclear and mitochondrial genomes are uncoupled. Yanagawa *Lingula* carries additional repeats and two copies of *atp8*, which are absent in Amami *Lingula*. Breakpoint distance analysis also shows that gene order of Yanagawa *Lingula* is more diverse. Therefore, the *Lingula* mitochondrial genome has experienced dramatic evolutionary changes, exhibiting very high gene order shuffling rates at the species level, and rapidly evolving protein-coding genes, as in bivalves (Wu et al., 2010).

The reason why the *Lingula* mitochondrial genome shows a high gene rearrangement rate is unclear. One major mechanism to explain the gene rearrangement in vertebrates is the tandem duplication-random loss model. In this model, slipped strand mispairing during replication generates an additional copy, and the duplicated gene is subsequently lost by deletion (Boore and Brown, 1998). However, this cannot explain the highly shuffled gene orders I observed between Amami and Yanagawa *Lingula*. My findings suggest that other mechanisms may be involved in the *Lingula* mitochondrial genome rearrangement. I argue that an unknown transposition mechanism involving recombination with double-stranded break repair might be required.

Indeed, intramolecular recombination is common in most plants and has been reported in some animals, although the molecular mechanism is still unknown (Rokas et al.). There are about 1,500 nuclear genes that are imported into mitochondria, including proteins involved in mitochondrial DNA replication, such as DNA polymerase gamma, single-stranded DNA binding protein, and DNA helicase (Chan and Copeland, 2009). Mutations in these genes have been reported to increase the mutation rate and to produce deletions in the mitochondrial genome (Zeviani et al., 2003). Further studies on these nuclear genes may shed light on this question.

4.3.4 Molecular phylogeny of *Lingula anatina*

Although the species boundary of *Lingula anatina* requires further study, it is believed to be widely distributed in the western Pacific (Williams et al., 2000). Because mitochondrial gene

order in Amami and Yanagawa *Lingula* is completely different, I attempted to test whether their genetic distance is also high. One important marker to determine the species boundary is 18S rDNA. It has been used to resolve phylogenetic relationships at the phylum and species levels (Halanych et al., 1995; Wada, 1998). Since 18S rDNA has been reported to be polyadenylated in humans (Slomovic et al., 2006), I checked 18S rDNA in the *Lingula* transcriptome assembly, and found its complete sequence.

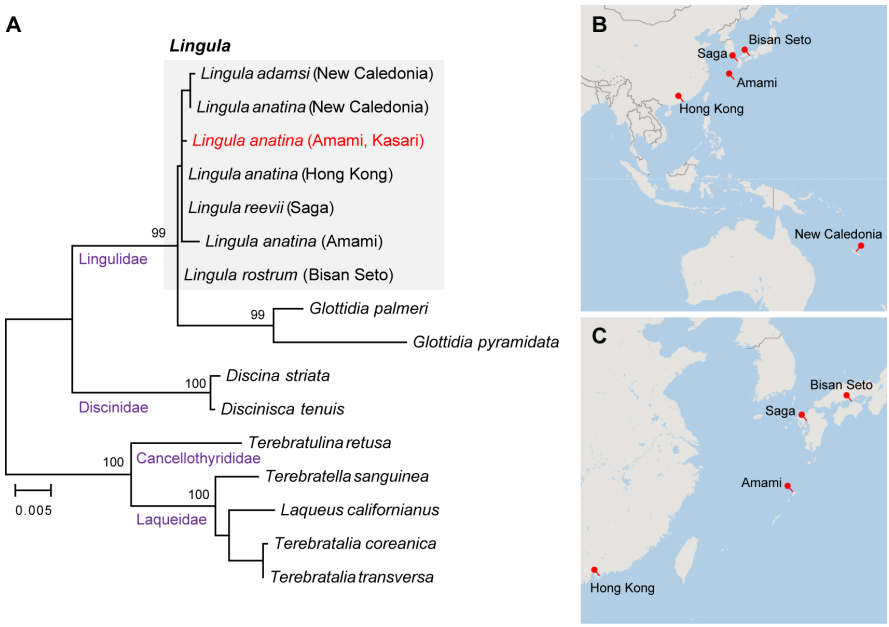


Figure 4.5 | Genetic distances, shown by phylogenetic analysis of 18S rDNA sequences, are shorter within the genus *Lingula* than among other brachiopods.
(**A**) Maximum likelihood tree based on an alignment of 16 brachiopod 18S rDNA sequences using the GTR+G+I model (1,757 position, 1,000 bootstrap replications). Sampling localities are shown in parentheses. The individual used for genome sequencing is in red. Numbers at the nodes indicate bootstrap values. (**B**) Map with sampling localities (red pins) in the western Pacific region. (**C**) Close-up map of northeast Asia region in (B).

To determine genetic distances among *Lingula* species, I first analyzed brachiopod 18S rDNAs. From all available 18S rDNA sequences, I showed that congeneric differences in *Lingula* are very low compared to the genus *Glottidia*, in the family Linguidae (Fig. 4.5). This result is consistent with a previous report (Cohen et al., 1998).

Among marine invertebrates, tunicates are another example of completely different gene orders within the same genus, *Phallusia*. To compare the genetic differences between brachiopods and tunicates, I analyzed 18S rDNAs in tunicates, mainly in the Ascidiidae, Cionidae, and Styelidae. Analysis of branch length of 18S rDNA trees showed that the

longest distance within two *Lingula* species was 0.0039, while tunicates *P. mammilata* and *P. fumigata* had the distance of 0.0087. This result suggested that the genetic distance of two individuals of *Lingula anatina* from Amami and Yanagawa might be shorter than that of tunicates *Phallusia*. On the other hand, based on *cox1* nucleotide sequences, northern Pacific populations of *Lingula anatina* are genetically heterogeneous (Endo et al., 2001). Indeed, my analysis of *cox1* using new data from Amami showed that the population from Amami might be closer to that from Hong Kong, which can be separated from that of Yanagawa.

4.3.5 *Ka* and *Ks* analyses of mitochondrial genomes

To test the idea of evolutionary rate differences in mitochondrial genomes, I next examined sequence divergence by calculating non-synonymous (*Ka*) and synonymous (*Ks*) substitution rates. Intra-familial comparisons among terrestrial vertebrates show that snakes and lizards have higher *Ka* values than birds, in which evolutionary rates of mitochondrial genomes have proven correlated with speciation rates (Eo and DeWoody, 2010). In addition, in bony fishes, those living in colder regions have lower *Ka* values with fewer mutations and stronger selection constraints than those of tropical species (Sun et al., 2011).

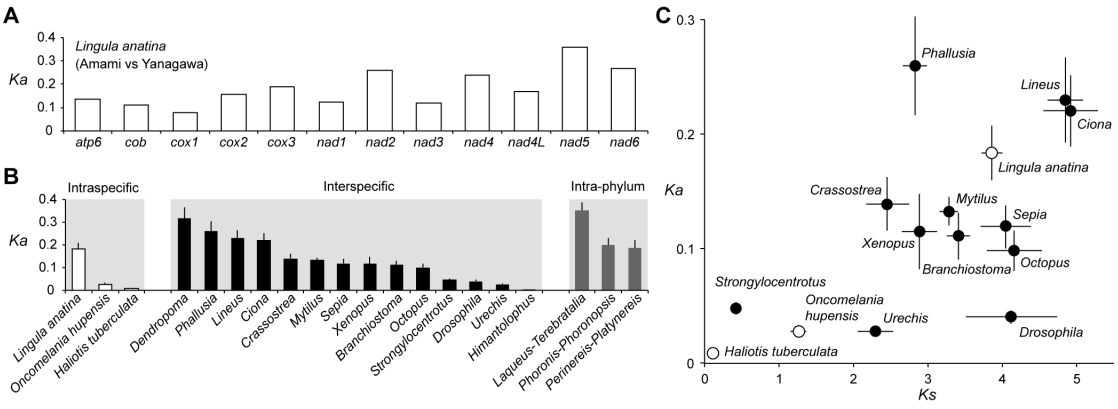


Figure 4.6 | Sequence divergence and evolutionary rates among mitochondrial protein-coding genes between two *Lingula* populations exceed those between many bilaterian species.
(**A**) The non-synonymous substitution rate (*Ka*) of 12 mitochondrial protein-coding genes (except *atp8*) between Amami and Yanagawa *Lingula anatina*. (**B**) Comparisons of *Ka* in 20 selected bilaterian pairs. (**C**) A scatter plot of synonymous substitution rate (*Ks*) versus *Ka* in 15 selected bilaterian pairs. Solid circles represent interspecific pairs. Open circles denote intraspecific pairs. Error bars, standard error of the mean.

When comparing 12 protein-coding genes between Amami and Yanagawa *Lingula*, I found that there was a *Ka* difference across all coding genes (mean *Ka*, 0.183; standard deviation, 0.082), where *cox1* had the lowest rate of 0.078 and *nad5* had the fastest (0.357)

(Fig. 4.6A). I further conducted pairwise comparisons of 12 protein-coding genes at three taxonomic levels: intraspecific, interspecific, and intra-phylum. I found that species pairs from different genera had *Ka* values larger than 0.185 (intra-phylum), while interspecific comparisons had *Ka* values ranging from 0.005 to 0.316 with a mean of 0.133 (Fig. 4.6B). On the other hand, comparisons at the subspecies level (snails, *Oncomelania hupensis* and *Haliotis tuberculata*) showed a mean *Ka* of 0.018. In contrast, the two *Lingula* had a *Ka* of 0.183, which was comparable to interspecific variation (Fig. 4.6B, intraspecific).

Similar results were obtained when comparing both *Ka* and *Ks*, where the pair of *Lingula anatina* was within interspecific pairs (Fig. 4.6C). In *Phallusia*, *Ciona*, *Lineus*, *Lingula*, *Crossostrea*, *Mytilus*, *Xenopus*, and *Branchiostoma*, variations of *Ka* were higher than *Ks*, suggesting they have experienced weaker selection constraints, whereas *Drosophila* and *Urechis* had higher selection pressure with lower *Ka* and higher variation of *Ks* (Fig. 4.6C). Mitochondrial genomes have been utilized to identify cryptic species by applying intra- and inter-specific *Ka* comparisons (Caputi et al., 2007; Iannelli et al., 2007b). My finding suggests that there might be different cryptic species within current classification of *Lingula anatina*. Further studies would be important to resolve this issue. In addition, it has been proposed that the rates of gene rearrangement and molecular evolution are positively correlated in arthropods (Xu et al., 2006). The fact that *Lingula anatina* has both high gene rearrangement and substitution rate fits well to this notion.

4.3.6 Phylogenetic analyses of mitochondrial genomes

Mitochondrial genes are often used for phylogenetic analyses, especially in vertebrates (Perseke et al., 2013), but mitochondrial genes in *Lingula anatina* have evolved rapidly (Endo et al., 2005; Saito et al., 2000). Accordingly, I examined whether it is appropriate to use mitochondrial genes to assess phylogenetic relationships in lophotrochozoans. I performed extensive analyses of protein-coding genes among metazoans at the mitochondrial genome level. Since *atp8* is not found in oysters and mussels, only 12 protein-coding genes were used for the analyses. In an unrooted tree generated with mtREV, bivalves and *Lingula* showed extraordinarily long branch-lengths, suggesting that their mitochondrial protein-coding genes have high evolutionary rates (Fig. 4.7A).

More strangely, positions of bivalves and *Lingula* were out of the groups of molluscs and terebratulide brachiopods, respectively. *Lingula* grouped within bivalves, most likely because of long-branch attraction (Fig. 4.7A). When comparing branch lengths, the distance between Amami and Yanagawa *Lingula* (0.35) was twice as long as the largest interspecific

differences in *Mytilus* (0.13, *M. edulis* to *M. californianus*) (Breton et al., 2006; Ort and Pogson, 2007) and *Crassostrea* (0.18, *C. virginica* to *C. iredalei*) (Milbury and Gaffney, 2005; Wu et al., 2010) (Fig. 4.7B).

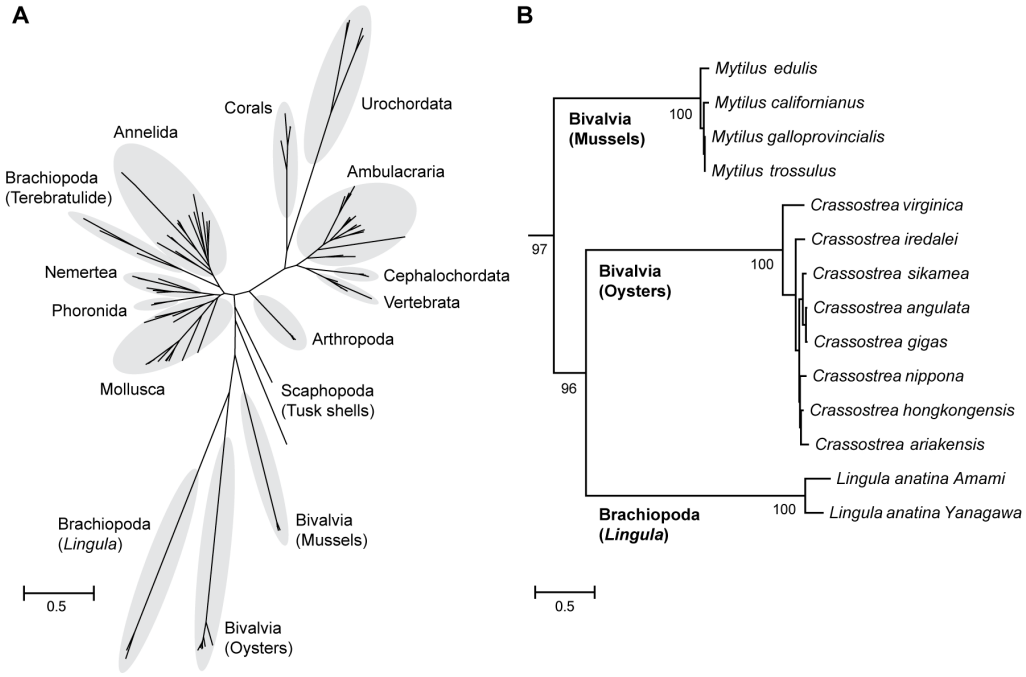


Figure 4.7 | Disparate evolutionary rates among invertebrate taxa result in divergent and unreliable phylogenetic trees.

(**A**) An unrooted tree obtained using maximum likelihood analysis with 12 protein-coding genes (except *atp8*; 101 metazoans; 2,318 amino-acid positions, bootstrap 100 replications; mtREV model). (**B**) A closer look at relationships among mussels, oysters, and two specimens of *Lingula anatina*. Numbers at the nodes indicate bootstrap values.

In addition, phylogenetic trees using the mtREV, mtART, and rtREV models showed that topologies for annelids, molluscs, terebratulide brachiopods, phoronids, and nemerteans were unstable. Each of the three models generated a different phylogenetic relationship among these five groups. Excluding bivalves, only the mtREV model supported monophyly of the Mollusca, including the Polyplacophora (i.e., chitons). Odd enough, the mtART model placed terebratulide brachiopods as an outgroup. Only the rtREV model supported the sister group of phoronids and terebratulide brachiopods, but with low bootstrap support. These results suggest that marine invertebrate mitochondrial genomes have a unique evolutionary history that cannot be resolved by current amino acid substitution models.

Furthermore, environmental factors, such as temperature, affect selection pressure, contributing to mitochondrial sequence variation (Ruiz-Pesini et al., 2004). Interestingly,

mussels, oysters, and *Lingula* are all filter feeders, living in the intertidal zone. Therefore, it is reasonable to speculate that temperature, salinity, and irradiation stresses may facilitate mitochondrial mutation rates. Indeed, it has been shown that reactive oxygen species and other stress factors increase these rates (Wallace and Chalkia, 2013). It would be interesting to test this hypothesis with species living in different habitats.

Finally, it has been proposed that mitochondrial genomes are inappropriate to resolve high-level phylogenetic debates, because they lack clock-like neutral mutation rates (Galtier et al., 2009). My data confirm that the *Lingula* mitochondrial genome cannot be used to resolve problems of phylogenetic relationships among annelids, brachiopods, phoronids, nemerteans, and molluscs.

4.4 Discussion

In this study, using the NGS approach, I presented a mitochondrial genome of the brachiopod, *Lingula anatina* from a new locality, Amami Island, Japan. Through analyses of transcriptomes, gene arrangements, non-synonymous substitution rates, and molecular phylogeny, I demonstrated that the mitochondrial genome of currently defined *L. anatina* is highly variable. My study thus implies that there may be cryptic species of *L. anatina* in the northern Pacific. Further mitochondrial genome studies in brachiopods, especially in *Lingula*, will provide better understanding of mitochondrial genome evolution in brachiopods and lophotrochozoans.

5 Conclusions

5.1 General features of lophotrochozoan genomes

Given the general characteristics among nine available lophotrochozoan genomes, I attempt to provide the features in a general lophotrochozoan genome. Haploid nuclear genome sizes of early branching animals, such as sponges (Srivastava et al., 2010), ctenophores (Moroz et al., 2014; Ryan et al., 2013), and placozoans (Srivastava et al., 2008), are relatively small in ~100–200 Mb. However, genome sizes of cnidarians, such as sea anemones (Putnam et al., 2007) and corals (Shinzato et al., 2011), are ~400–500 Mb. In general, animal genome size is positively correlated with the numbers of protein-coding genes, intron contents, and repetitive contents (i.e. transposable elements) (Elliott and Gregory, 2015). Thus, gene family expansions, gain of introns, and expansions of transposable elements contribute to the increased genome size during early animal evolution, leading to the lineage of Planulozoa (Bilateria + Cnidaria). In most cases, lophotrochozoan genomes share ~7,000 gene families, reflecting a conserved core set of bilaterian ancestor genes.

Interestingly, most lophotrochozoan genomes are in the range of 300–500 Mb, suggesting that some lophotrochozoans represent an ancient genome architecture from the bilaterian and cnidarian common ancestor. It is particularly evident that lophotrochozoan genomes share ~300–400 microsyntenic linkage groups to deuterostomes and some basal metazoans. The average exon number per gene of a lophotrochozoan genome is 6–8, which is comparably larger than that of an ancestral metazoan genome (i.e. 3–4) (Simakov and Kawashima, 2016). On the other hand, the mean intron length of lophotrochozoan genomes varies from ~300–1,700 bp, suggesting lineage-specific expansions of transposable elements in the intron regions. The repetitive content comprises ~20–40% of lophotrochozoan genomes. Overall, lophotrochozoan genomes retain a core set of planulozoan gene repertoire and genomic organization, whereas also exhibit diverse patterns of repetitive content. The similarity between lophotrochozoan and cnidarian genomes may reflect the early origin of the developmental toolkits for body patterning (Erwin, 2009).

5.2 Phylogeny and evolution of lophotrochozoans

My analysis, combining genomic and transcriptomic data, has shown that (1) Nemertea is sister to Lophophorata (Phoronida, Ectoprocta, and Brachiopoda) and (2) Phoronida is sister to Ectoprocta, supporting the monophyly of Lophophorata. In Nemertea, my results suggest

that Anopla (unarmed, without stylets), including Palaeonemertea and Heteronemertea, is not a monophyletic group. Thus, the possession of stylets might be a derived feature that originated during nemertean evolution. Within Lophophorata, the phylogenetic relationships of Phoronida and Brachiopoda are much debated (Sperling et al., 2011). Based on analysis of rDNA genes, phoronids are considered as a subtaxon within brachiopods, suggesting that phoronids are 'brachiopods without shells' (Cohen, 2013; Cohen and Weydmann, 2005). My analysis supports the monophyly of Brachiopoda, in which Linguliformea and Craniiformea (both inarticulate) are sisters to Rhynchonelliformea (articulate). My results thus argue that Linguliformea and Craniiformea are not sister groups to Phoronida. Together with the fact that Phoronida is sister to Ectoprocta, my data do not support the shell-less brachiopod hypothesis (or paraphyletic brachiopods) for the origins of phoronids.

In summary, my phylogenetic analyses with high-quality transcriptome data using both maximum likelihood and Bayesian methods support the close relationship of Ectoprocta and Phoronida. Further analysis of orthologue subsets with strong phylogenetic signals using the maximum likelihood method also supports the same conclusion. Thus, I may conclude that Ectoprocta is sister to Phoronida. Together, Ectoprocta, Phoronida, and Brachiopoda constitute the traditional animal group, Lophophorata. While my data strongly support the grouping of Lophophorata, an ectoproct genome will be needed to address its evolutionary origin.

5.3 Biomineralization mechanisms in *Lingula*

I have demonstrated that *Lingula* used its own gene sets to originate their calcium phosphate chemistry that is different from the set used by vertebrates. In addition, I have shown that there are lineage-specific SMPs in *Lingula* and molluscs, respectively. I propose that the metazoan ancestor used a core of ancient signaling proteins to initiate the biomineralization process. I speculate that this involves canonical BMP signaling, in which BMP ligands bind to its receptor, from which a signal is transduced by the regulatory and co-mediator, pSmad1/5/9 and Smad4, respectively. They then act as transcription factors, interacting with other proteins to activate the expression of downstream biomineralization genes. The other conserved transcription factor is *engrailed*, which is involved in both bone and shell formation. Furthermore, many calcium-binding proteins (e.g., calcineurin, calponin, and calmodulin) and extracellular matrix proteins (e.g., cadherin, collagen, and fibronectin) have been reported to participate in bone and shell formation. This implies that metazoan biomineralization likely originated from a calcium-regulated extracellular matrix system.

Taken together, genomic, transcriptomic, and proteomic analyses of *Lingula* biomineralization show similar patterns to those in molluscs (Aguilera et al., 2017; Jackson et al., 2010) and corals (Ramos-Silva et al., 2013), where co-option, domain shuffling, and novel genes are the fundamental mechanisms for metazoan biomineralization.

5.4 Evolution of morphological novelties: brachiopod shells

In fossil records, the first vertebrate mineralized bones (i.e. endoskeletons) appeared in the late Ordovician (~450 million years ago) (Venkatesh et al., 2014) much later than lingulid shells (~520 million years ago, early Cambrian) (Zhang et al., 2005). Together with the distant phylogenetic relationship of vertebrates and *Lingula*, it is perhaps not surprising that vertebrate bones and *Lingula* shells shared different genetic origins.

Recent discoveries from Cambrian fossils have changed our ideas about evolution of early molluscs and animal biomineralization. For example, a non-mineralized cephalopod fossil, *Nectocaris*, found in Burgess Shale (~508 million years ago, middle Cambrian) suggests that a mineralized shell is a derived character of cephalopods (Smith and Caron, 2010). On the other hand, phylogenomic studies of mollusc phylogeny show that shells may have multiple origins (Kocot et al., 2011; Smith et al., 2011), which is in agreement with the proteomic studies of mollusc shells (Jackson et al., 2006; Jackson et al., 2010; Sarashina and Endo, 2006).

Extant molluscs can be divided into two major groups, Conchifera (shell-bearing; Gastropoda, Bivalvia, Scaphopoda, Cephalopoda, and Monoplacophora) and Aculifera (worm-like; Neomeniomorpha, Chaetodermomorpha, and Polyplacophora) (Kocot et al., 2011). Although conchiferans make shells and aculiferans have only sclerites, both of them use calcium carbonate. Despite my transcriptome analysis shows that mantles of the Pacific oyster and *Lingula* share some molecular similarities, particularly in genes that are associated with secretion and extracellular matrix, it is not clear whether these similarities come from convergent evolution. Therefore, it is at the moment difficult to test whether mollusc and brachiopod shells share the same origin. Functional analysis of the gene regulatory networks that are required to specify shell fields and produce mineralized shells would be needed to tackle this problem.

While brachiopods have adopted different modes of biomineralization, only the Linguliformea makes shells with calcium phosphate (Williams et al., 1994). Comparative genomics among other brachiopods which possess calcium carbonate shells would be necessary to understand the origin of phosphate biomineralization in the lingulid lineage.

Nevertheless, my data suggest that lineage-specific expansions of chitin synthases, co-option of Hox genes and extracellular matrix proteins, together with novel genes contribute to the unique phosphate shells in *Lingula*.

5.5 Evolution of morphological novelties: lophophores

Lophophores have been considered a key morphological novelty in lophophorates (Moysiuk et al., 2017; Temereva et al., 2015). However, their origins are largely unknown. Through comparative transcriptomics, my findings indicate that lophophores exhibit molecular identity similar to that of a bilaterian head. These include sharing specific expressions of several head toolkit genes, such as *six3/6* and *lhx1/5*. In addition, similar to the heads of other animals, the expression of Hox gene is also absent in lophophores. Thus, a common anterior patterning system required for bilaterian head formation might regulate the development of lophophores, although a functional study during embryogenesis would be required to test this hypothesis. On top of conserved patterning system, there are also lineage-specific features that shape the unique evolution of lophophores. Lophophores had experienced a redeployment of several homeobox genes and expansions of mucins and carbohydrate sulfotransferases. Lophophores also express a large number of novel genes. Together, my results suggest that lophophorates originated from the anterior patterning system that is shared deeply with the bilaterian ancestor, although the bilaterian heads independently evolved in different animals depending on several lineage-specific features.

5.6 Concluding remarks

I have demonstrated that the whole genome sequencing of non-model systems with important phylogenetic positions can provide insights into animal evolution in many different perspectives, such as phylogenetic relationships, evolution of gene content and synteny, and the genomic origin of morphological novelties. Together with embryonic and tissue transcriptomes, decoded brachiopod, phoronid, and nemertean genomes provide resources for future studies. These studies include to understand the origin of lophotrochozoan morphological features and development by comparative genomics and functional experiments (e.g. CRISPR-based genome editing and epigenomic profiling).

6 Reference List

Aguilera, F., McDougall, C., and Degnan, B.M. (2017). Co-option and de novo gene evolution underlie molluscan shell diversity. Mol. Biol. Evol. *34*, 779-792.

Albalat, R., and Canestro, C. (2016). Evolution by gene loss. Nat. Rev. Genet. *17*, 379-391.

Albertin, C.B., Simakov, O., Mitros, T., Wang, Z.Y., Pungor, J.R., Edsinger-Gonzales, E., Brenner, S., Ragsdale, C.W., and Rokhsar, D.S. (2015). The octopus genome and the evolution of cephalopod neural and morphological novelties. Nature *524*, 220-224.

Andrade, S.C., Montenegro, H., Strand, M., Schwartz, M.L., Kajihara, H., Norenburg, J.L., Turbeville, J.M., Sundberg, P., and Giribet, G. (2014). A transcriptomic approach to ribbon worm systematics (nemertea): resolving the pilidiophora problem. Mol. Biol. Evol. *31*, 3206-3215.

Appeltans, W., Ahyong, S.T., Anderson, G., Angel, M.V., Artois, T., Bailly, N., Bamber, R., Barber, A., Bartsch, I., Berta, A.*, et al.* (2012). The magnitude of global marine species diversity. Curr. Biol. *22*, 2189-2202.

Arendt, D., Musser, J.M., Baker, C.V., Bergman, A., Cepko, C., Erwin, D.H., Pavlicev, M., Schlosser, G., Widder, S., Laubichler, M.D.*, et al.* (2016). The origin and evolution of cell types. Nat. Rev. Genet. *17*, 744-757.

Bardou, P., Mariette, J., Escudie, F., Djemiel, C., and Klopp, C. (2014). jvenn: an interactive Venn diagram viewer. BMC Bioinformatics *15*, 293.

Bengtson, S., Farmer, J.D., Fedonkin, M.A., Lipps, J.H., and Runnegar, B.N. (1992). The Proterozoic-Early Cambrian evolution of metaphytes and metazoans. In The Proterozoic Biosphere: A Multidisciplinary Study, Schopf J. W., and K. C., eds. (Cambridge).

Beniash, E., Addadi, L., and Weiner, S. (1999). Cellular control over spicule formation in sea urchin embryos: A structural approach. J. Struct. Biol. *125*, 50-62.

Beniash, E., Aizenberg, J., Addadi, L., and Weiner, S. (1997). Amorphous calcium carbonate transforms into calcite during sea urchin larval spicule growth. Proc. R. Soc. B *264*, 461-465.

Berriman, M., Haas, B.J., LoVerde, P.T., Wilson, R.A., Dillon, G.P., Cerqueira, G.C., Mashiyama, S.T., Al-Lazikani, B., Andrade, L.F., Ashton, P.D.*, et al.* (2009). The genome of the blood fluke *Schistosoma mansoni*. Nature *460*, 352-358.

Bitner, M.A., and Cohen, B.L. (2013). Brachiopoda. In eLS (John Wiley & Sons, Ltd).

Blumenthal, K.M., Keim, P.S., Heinrikson, R.L., and Kem, W.R. (1981). Structure and action of heteronemertine polypeptide toxins. Amino acid sequence of *Cerebratulus lacteus* toxin B-II and revised structure of toxin B-IV. J. Biol. Chem. *256*, 9063-9067.

Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D., and Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. Bioinformatics *27*, 578-579.

Boetzer, M., and Pirovano, W. (2014). SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. BMC Bioinformatics *15*, 211.

Boonrungsiman, S., Gentleman, E., Carzaniga, R., Evans, N.D., McComb, D.W., Porter, A.E., and Stevens, M.M. (2012). The role of intracellular calcium phosphate in osteoblast-mediated bone apatite formation. Proc. Natl. Acad. Sci. USA *109*, 14170-14175.

Bourlat, S.J., Nielsen, C., Economou, A.D., and Telford, M.J. (2008). Testing the new animal phylogeny: a phylum level molecular analysis of the animal kingdom. Mol. Phylogenet. Evol. *49*, 23-31.

Butala, M., Sega, D., Tomc, B., Podlesek, Z., Kem, W.R., Kupper, F.C., and Turk, T. (2015). Recombinant expression and predicted structure of parborlysin, a cytolytic protein from the Antarctic heteronemertine *Parborlasia corrugatus*. Toxicon *108*, 32-37.

Canestro, C., Yokoi, H., and Postlethwait, J.H. (2007). Evolutionary developmental biology and genomics. Nat. Rev. Genet. *8*, 932-942.

Cannon, J.T., Vellutini, B.C., Smith, J., 3rd, Ronquist, F., Jondelius, U., and Hejnol, A. (2016). Xenacoelomorpha is the sister group to Nephrozoa. Nature *530*, 89-93.

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

Carroll, S.B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell *134*, 25-36.

Cavalier-Smith, T. (1998). A revised six-kingdom system of life. Biol. Rev. Camb. Philos. Soc. *73*, 203-266.

Clarke, F.W., and Wheeler, W.C. (1915). The composition of brachiopod shells. Proc. Natl. Acad. Sci. USA *1*, 262-266.

Cohen, B.L. (2013). Rerooting the rDNA gene tree reveals phoronids to be 'brachiopods without shells'; dangers of wide taxon samples in metazoan phylogenetics (Phoronida; Brachiopoda). Zool. J. Linn. Soc. *167*, 82-92.

Cohen, B.L., and Weydmann, A. (2005). Molecular evidence that phoronids are a subtaxon of brachiopods (Brachiopoda: Phoronata) and that genetic divergence of metazoan phyla began long before the early Cambrian. Org. Divers. Evol. *5*, 253-273.

Cook, P.J., and Shergold, J.H. (1984). Phosphorus, phosphorites and skeletal evolution at the Precambrian-Cambrian boundary. Nature *308*, 231-236.

Cusack, M., and Freer, A. (2008). Biomineralization: elemental and organic influence in carbonate systems. Chem. Rev. *108*, 4433-4454.

Cusack, M., Williams, A., and Buckman, J.O. (1999). Chemico-structural evolution of linguloid brachiopod shells. Palaeontology *42*, 799-840.

Dalquen, D.A., and Dessimoz, C. (2013). Bidirectional best hits miss many orthologs in duplication-rich clades such as plants and animals. Genome Biol. Evol. *5*, 1800-1806.

Darwin, C. (1859). On the Origin of Species by Means of Natural Selection (London: Murray London).

Darzentas, N. (2010). Circoletto: visualizing sequence similarity with Circos. Bioinformatics *26*, 2620-2621.

Davidson, E.H. (2010). Emerging properties of animal gene regulatory networks. Nature *468*, 911-920.

De Bie, T., Cristianini, N., Demuth, J.P., and Hahn, M.W. (2006). CAFE: a computational tool for the study of gene family evolution. Bioinformatics *22*, 1269-1271.

De Robertis, E.M. (2008). Evo-devo: variations on ancestral themes. Cell *132*, 185-195.

de Rosa, R., Grenier, J.K., Andreeva, T., Cook, C.E., Adoutte, A., Akam, M., Carroll, S.B., and Balavoine, G. (1999). Hox genes in brachiopods and priapulids and protostome evolution. Nature *399*, 772-776.

Dehal, P., Satou, Y., Campbell, R.K., Chapman, J., Degnan, B., De Tomaso, A., Davidson, B., Di Gregorio, A., Gelpke, M., Goodstein, D.M*., et al.* (2002). The draft genome of *Ciona intestinalis*: Insights into chordate and vertebrate origins. Science *298*, 2157-2167.

Denoeud, F., Henriet, S., Mungpakdee, S., Aury, J.M., Da Silva, C., Brinkmann, H., Mikhaleva, J., Olsen, L.C., Jubin, C., Canestro, C*., et al.* (2010). Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate. Science *330*, 1381-1385.

Domazet-Loso, T., and Tautz, D. (2010). A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. Nature *468*, 815-818.

Duboule, D. (1994). Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. Dev. Suppl., 135-142.

Duboule, D. (2007). The rise and fall of Hox gene clusters. Development *134*, 2549-2560.

Dunn, C.W., Giribet, G., Edgecombe, G.D., and Hejnol, A. (2014). Animal phylogeny and its evolutionary implications. Annu. Rev. Ecol. Evol. Syst. *45*, 371-395.

Dunn, C.W., Hejnol, A., Matus, D.Q., Pang, K., Browne, W.E., Smith, S.A., Seaver, E., Rouse, G.W., Obst, M., Edgecombe, G.D*., et al.* (2008). Broad phylogenomic sampling improves

resolution of the animal tree of life. Nature *452*, 745-749.
Dunn, C.W., and Ryan, J.F. (2015). The evolution of animal genomes. Curr. Opin. Genet. Dev. *35*, 25-32.
Ebersberger, I., Strauss, S., and von Haeseler, A. (2009). HaMStR: profile hidden markov model based search for orthologs in ESTs. BMC Evol. Biol. *9*, 157.
Egger, B., Lapraz, F., Tomiczek, B., Muller, S., Dessimoz, C., Girstmair, J., Skunca, N., Rawlinson, K.A., Cameron, C.B., Beli, E.*, et al.* (2015). A transcriptomic-phylogenomic analysis of the evolutionary relationships of flatworms. Curr. Biol. *25*, 1347-1353.
Elliott, T.A., and Gregory, T.R. (2015). What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. Philos. Trans. R. Soc. Lond. B Biol. Sci. *370*, 20140331.
Ellisdon, A.M., Reboul, C.F., Panjikar, S., Huynh, K., Oellig, C.A., Winter, K.L., Dunstone, M.A., Hodgson, W.C., Seymour, J., Dearden, P.K.*, et al.* (2015). Stonefish toxin defines an ancient branch of the perforin-like superfamily. Proc. Natl. Acad. Sci. USA *112*, 15360-15365.
Emig, C.C. (1997). Ecology of the inarticulated brachiopods. In Treatise on Invertebrate Paleontology. Part H. Brachiopoda., R.L. Kaesler, ed. (Boulder, Colorado, and Lawrence, Kansas: Geological Society of America and University of Kansas), pp. 473-495.
Emig, C.C. (2001). Phoronida. In eLS (John Wiley & Sons, Ltd).
Emig, C.C. (2003). Proof that *Lingula* (Brachiopoda) is not a living fossil, and amended diagnoses of the Family Lingulidae. Carnets de Géologie / Notebooks on Geology letter, 1-8.
Emig, C.C. (2008). On the history of the names *Lingula, anatina*, and on the confusion of the forms assigned them among the Brachiopoda. Carnets de Géologie / Notebooks on Geology letter, 1-13.
Emig, C.C., Herberts, C., and Thomassin, B.A. (1972). Sur l'association de *Phoronis australis* (Phoronida) avec *Cerianthus maua* (Ceriantharia) dans les zones récifales de Madagascar. Mar. Biol. *15*, 304-315.
Erwin, D.H. (2009). Early origin of the bilaterian developmental toolkit. Philos. Trans. R. Soc. Lond. B Biol. Sci. *364*, 2253-2261.
Erwin, D.H., and Davidson, E.H. (2002). The last common bilaterian ancestor. Development *129*, 3021-3032.
Erwin, D.H., and Davidson, E.H. (2009). The evolution of hierarchical gene regulatory networks. Nat. Rev. Genet. *10*, 141-148.
Erwin, D.H., Laflamme, M., Tweedt, S.M., Sperling, E.A., Pisani, D., and Peterson, K.J. (2011). The Cambrian conundrum: early divergence and later ecological success in the early history of animals. Science *334*, 1091-1097.
Ettensohn, C.A. (2014). Horizontal transfer of the *msp130* gene supported the evolution of metazoan biomineralization. Evol. Dev. *16*, 139-148.
Fang, X., Seim, I., Huang, Z., Gerashchenko, M.V., Xiong, Z., Turanov, A.A., Zhu, Y., Lobanov, A.V., Fan, D., Yim, S.H.*, et al.* (2014). Adaptations to a subterranean environment and longevity revealed by the analysis of mole rat genomes. Cell Rep. *8*, 1354-1364.
Ferrier, D.E., Minguillon, C., Holland, P.W., and Garcia-Fernandez, J. (2000). The amphioxus Hox cluster: deuterostome posterior flexibility and Hox14. Evol. Dev. *2*, 284-293.
Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet. 9*, 397-405.
Finnerty, J.R., Pang, K., Burton, P., Paulson, D., and Martindale, M.Q. (2004). Origins of bilateral symmetry: *Hox* and *dpp* expression in a sea anemone. Science *304*, 1335-1337.
Fischer, S., Brunk, B.P., Chen, F., Gao, X., Harb, O.S., Iodice, J.B., Shanmugam, D., Roos, D.S., and Stoeckert, C.J., Jr. (2011). Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. Curr. Protoc. Bioinformatics *Chapter 6*, Unit 6 12 11-19.

Flot, J.-F., Hespeels, B., Li, X., Noel, B., Arkhipova, I., Danchin, E.G.J., Hejnol, A., Henrissat, B., Koszul, R., Aury, J.-M*., et al.* (2013). Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta vaga*. Nature *500*, 453-457.

Freeman, R., Ikuta, T., Wu, M., Koyanagi, R., Kawashima, T., Tagawa, K., Humphreys, T., Fang, G.C., Fujiyama, A., Saiga, H*., et al.* (2012). Identical genomic organization of two hemichordate hox clusters. Curr. Biol. *22*, 2053-2058.

Fröbius, A.C., Matus, D.Q., and Seaver, E.C. (2008). Genomic organization and expression demonstrate spatial and temporal Hox gene colinearity in the lophotrochozoan *Capitella* sp. I. PLOS ONE *3*, e4004.

Furuhashi, T., Schwarzinger, C., Miksik, I., Smrz, M., and Beran, A. (2009). Molluscan shell evolution with review of shell calcification hypothesis. Comp. Biochem. Physiol. B Biochem. Mol. Biol. *154*, 351-371.

Gehring, W.J. (2005). New perspectives on eye development and the evolution of eyes and photoreceptors. J. Hered. *96*, 171-184.

Giribet, G., Dunn, C. W., Edgecombe, G. D., Hejnol, A., Martindale, M. Q., Rouse, G. W. (2009). Assembling the spiralian tree of life. In Animal Evolution: Genomes, Fossils, and Trees, M.J. Telford, ed., pp. 52-64.

Glinka, A., Wu, W., Onichtchouk, D., Blumenstock, C., and Niehrs, C. (1997). Head induction by simultaneous repression of Bmp and Wnt signalling in *Xenopus*. Nature *389*, 517-519.

Golub, E.E. (2009). Role of matrix vesicles in biomineralization. Biochim. Biophys. Acta. *1790*, 1592-1598.

Gonzalez, P., Uhlinger, K.R., and Lowe, C.J. (2016). The adult body plan of indirect developing hemichordates develops by adding a Hox-patterned trunk to an anterior larval territory. Curr. Biol. *27*, 87-95.

Gould, S.J., and Calloway, C.B. (1980). Clams and brachiopods; ships that pass in the night. Paleobiology *6*, 383-396.

Grande, C., and Patel, N.H. (2009). Nodal signalling is involved in left-right asymmetry in snails. Nature *457*, 1007-1011.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M*., et al.* (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. *8*, 1494-1512.

Halanych, K.M. (2016). How our view of animal phylogeny was reshaped by molecular approaches: lessons learned. Org. Divers. Evol. *16*, 319-328.

Halanych, K.M., Bacheller, J.D., Aguinaldo, A.M., Liva, S.M., Hillis, D.M., and Lake, J.A. (1995). Evidence from 18S ribosomal DNA that the lophophorates are protostome animals. Science *267*, 1641-1643.

Halanych, K.M., and Kocot, K.M. (2014). Repurposed transcriptomic data facilitate discovery of innate immunity toll-like receptor (TLR) genes across Lophotrochozoa. Biol. Bull. *227*, 201-209.

Hashimshony, T., Feder, M., Levin, M., Hall, B.K., and Yanai, I. (2015). Spatiotemporal transcriptomics reveals the evolutionary history of the endoderm germ layer. Nature *519*, 219-222.

Hausdorf, B., Helmkampf, M., Nesnidal, M.P., and Bruchhaus, I. (2010). Phylogenetic relationships within the lophophorate lineages (Ectoprocta, Brachiopoda and Phoronida). Mol. Phylogenet. Evol. *55*, 1121-1127.

Hejnol, A., and Dunn, C.W. (2016). Animal evolution: are phyla real? Curr. Biol. *26*, R424-426.

Hejnol, A., and Lowe, C.J. (2015). Embracing the comparative approach: how robust phylogenies and broader developmental sampling impacts the understanding of nervous system evolution. Philos. Trans. R. Soc. Lond. B Biol. Sci. *370*.

Hejnol, A., and Martindale, M.Q. (2008). Acoel development indicates the independent evolution of the bilaterian mouth and anus. Nature *456*, 382-386.

Hejnol, A., Obst, M., Stamatakis, A., Ott, M., Rouse, G.W., Edgecombe, G.D., Martinez, P., Baguna, J., Bailly, X., Jondelius, U.*, et al.* (2009). Assessing the root of bilaterian animals with scalable phylogenomic methods. Proc. Biol. Sci. *276*, 4261-4270.

Hejnol, A., and Pang, K. (2016). Xenacoelomorpha's significance for understanding bilaterian evolution. Curr. Opin. Genet. Dev. *39*, 48-54.

Helmkampf, M., Bruchhaus, I., and Hausdorf, B. (2008). Phylogenomic analyses of lophophorates (brachiopods, phoronids and bryozoans) confirm the Lophotrochozoa concept. Proc. Biol. Sci. *275*, 1927-1933.

Hibino, T., Loza-Coll, M., Messier, C., Majeske, A.J., Cohen, A.H., Terwilliger, D.P., Buckley, K.M., Brockton, V., Nair, S.V., Berney, K.*, et al.* (2006). The immune gene repertoire encoded in the purple sea urchin genome. Dev. Biol. *300*, 349-365.

Hiebert, L.S., and Maslakova, S.A. (2015). Hox genes pattern the anterior-posterior axis of the juvenile but not the larva in a maximally indirect developing invertebrate, *Micrura alaskensis* (Nemertea). BMC Biol. *13*, 23.

Holland, L.Z., Carvalho, J.E., Escriva, H., Laudet, V., Schubert, M., Shimeld, S.M., and Yu, J.-K. (2013). Evolution of bilaterian central nervous systems: a single origin? EvoDevo *4*, 27.

Holland, P. (2011). The Animal Kingdom: A Very Short Introduction (New York: Oxford University Press).

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44-57.

Huang, S., Chen, Z., Huang, G., Yu, T., Yang, P., Li, J., Fu, Y., Yuan, S., Chen, S., and Xu, A. (2012). HaploMerger: reconstructing allelic relationships for polymorphic diploid genome assemblies. Genome Res. *22*, 1581-1588.

Hyman, L.H. (1959). The Invertebrates, Vol 5 (New York: McGraw-Hill).

Irie, N., and Kuratani, S. (2011). Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. Nat. Commun. *2*, 248.

Iwata, K. (1981). Ultrastructure and mineralization of the shell of *Lingula unguis* Linne, (inarticualte brachiopod). J. Faculty Sci. Hokkaido Univ. 4, Geol. Mineral. *20*, 35-65.

Jackson, D.J., Macis, L., Reitner, J., and Worheide, G. (2011). A horizontal gene transfer supported the evolution of an early metazoan biomineralization strategy. BMC Evol. Biol. *11*, 238.

Jackson, D.J., Mann, K., Haussermann, V., Schilhabel, M.B., Luter, C., Griesshaber, E., Schmahl, W., and Worheide, G. (2015). The *Magellania venosa* biomineralizing proteome: A window into brachiopod shell evolution. Genome Biol. Evol. *7*, 1349-1362.

Jackson, D.J., McDougall, C., Green, K., Simpson, F., Worheide, G., and Degnan, B.M. (2006). A rapidly evolving secretome builds and patterns a sea shell. BMC Biol. *4*, 40.

Jackson, D.J., McDougall, C., Woodcroft, B., Moase, P., Rose, R.A., Kube, M., Reinhardt, R., Rokhsar, D.S., Montagnani, C., Joubert, C.*, et al.* (2010). Parallel evolution of nacre building gene sets in molluscs. Mol. Biol. Evol. *27*, 591-608.

Jandzik, D., Garnett, A.T., Square, T.A., Cattell, M.V., Yu, J.-K., and Medeiros, D.M. (2015). Evolution of the new vertebrate head by co-option of an ancient chordate skeletal tissue. Nature *518*, 534-537.

Jope, M. (1977). Brachiopod shell proteins: their functions and taxonomic significance. Amer. Zool. *17*, 133-140.

Kajitani, R., Toshimoto, K., Noguchi, H., Toyoda, A., Ogura, Y., Okuno, M., Yabana, M., Harada, M., Nagayasu, E., Maruyama, H.*, et al.* (2014). Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome Res. *24*, 1384-1395.

Kalinka, A.T., Varga, K.M., Gerrard, D.T., Preibisch, S., Corcoran, D.L., Jarrells, J., Ohler, U., Bergman, C.M., and Tomancak, P. (2010). Gene expression divergence recapitulates the developmental hourglass model. Nature *468*, 811-814.

Kao, D., Lai, A.G., Stamataki, E., Rosic, S., Konstantinides, N., Jarvis, E., Di Donfrancesco, A.,

Pouchkina-Stancheva, N., Semon, M., Grillo, M.*, et al.* (2016). The genome of the crustacean *Parhyale hawaiensis*, a model for animal development, regeneration, immunity and lignocellulose digestion. eLife *5*.

Katoh, K., Misawa, K., Kuma, K.i., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. *30*, 3059-3066.

Kawasaki, K., Buchanan, A.V., and Weiss, K.M. (2009). Biomineralization in humans: making the hard choices in life. Annu. Rev. Genet. *43*, 119-142.

Kem, W.R., and Blumenthal, K.M. (1978). Purification and characterization of the cytotoxic *Cerebratulus* A toxins. J. Biol. Chem. *253*, 5752-5757.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

King, N., Westbrook, M.J., Young, S.L., Kuo, A., Abedin, M., Chapman, J., Fairclough, S., Hellsten, U., Isogai, Y., Letunic, I.*, et al.* (2008). The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. Nature *451*, 783-788.

Knoll, A.H. (2003). Biomineralization and evolutionary history. Rev. Mineral. Geochem. *54*, 329-356.

Kocot, K.M. (2016). On 20 years of Lophotrochozoa. Org. Divers. Evol. *16*, 329-343.

Kocot, K.M., Aguilera, F., McDougall, C., Jackson, D.J., and Degnan, B.M. (2016a). Sea shell diversity and rapidly evolving secretomes: insights into the evolution of biomineralization. Front. Zool. *13*, 23.

Kocot, K.M., Cannon, J.T., Todt, C., Citarella, M.R., Kohn, A.B., Meyer, A., Santos, S.R., Schander, C., Moroz, L.L., Lieb, B.*, et al.* (2011). Phylogenomics reveals deep molluscan relationships. Nature *477*, 452-456.

Kocot, K.M., Struck, T.H., Merkel, J., Waits, D.S., Todt, C., Brannock, P.M., Weese, D.A., Cannon, J.T., Moroz, L.L., Lieb, B.*, et al.* (2016b). Phylogenomics of Lophotrochozoa with consideration of systematic error. Syst. Biol. *66*, 256-282.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357-359.

Lartillot, N., Lepage, T., and Blanquart, S. (2009). PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics *25*, 2286-2288.

Laumer, C.E., Bekkouche, N., Kerbl, A., Goetz, F., Neves, R.C., Sorensen, M.V., Kristensen, R.M., Hejnol, A., Dunn, C.W., Giribet, G.*, et al.* (2015). Spiralian phylogeny informs the evolution of microscopic lineages. Curr. Biol. *25*, 2000-2006.

Lee, P.N., Callaerts, P., De Couet, H.G., and Martindale, M.Q. (2003). Cephalopod Hox genes and the origin of morphological novelties. Nature *424*, 1061-1065.

Leggett, R.M., Clavijo, B.J., Clissold, L., Clark, M.D., and Caccamo, M. (2014). NextClip: an analysis and read preparation tool for Nextera Long Mate Pair libraries. Bioinformatics *30*, 566-568.

Leininger, S., Adamski, M., Bergum, B., Guder, C., Liu, J., Laplante, M., Brate, J., Hoffmann, F., Fortunato, S., Jordal, S.*, et al.* (2014). Developmental gene expression provides clues to relationships between sponge and eumetazoan body plans. Nat. Commun. *5*, 3905.

Levin, M., Anavy, L., Cole, A.G., Winter, E., Mostov, N., Khair, S., Senderovich, N., Kovalev, E., Silver, D.H., Feder, M.*, et al.* (2016). The mid-developmental transition and the evolution of animal body plans. Nature *531*, 637-641.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323.

Li, R., Yu, H., Xue, W., Yue, Y., Liu, S., Xing, R., and Li, P. (2014). Jellyfish venomics and venom gland transcriptomics analysis of *Stomolophus meleagris* to reveal the toxins associated with sting. J. Proteomics *106*, 17-29.

Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of

protein or nucleotide sequences. Bioinformatics *22*, 1658-1659.
Lowe, C.J., Clarke, D.N., Medeiros, D.M., Rokhsar, D.S., and Gerhart, J. (2015). The deuterostome context of chordate origins. Nature *520*, 456-465.
Lowenstam, H. (1981). Minerals formed by organisms. Science *211*, 1126-1131.
Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y.*, et al.* (2012). SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. Gigascience *1*, 18.
Luo, Y.J., Takeuchi, T., Koyanagi, R., Yamada, L., Kanda, M., Khalturina, M., Fujie, M., Yamasaki, S., Endo, K., and Satoh, N. (2015). The *Lingula* genome provides insights into brachiopod evolution and the origin of phosphate biomineralization. Nat. Commun. *6*, 8301.
Mackinnon, D.I., and Biernat, G. (1970). The probable affinities of the trace fossil *Diorygma atrypophilia*. Lethaia *3*, 163-172.
Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. Bioinformatics *27*, 764-770.
Marin, F., Luquet, G., Marie, B., and Medakovic, D. (2008). Molluscan shell proteins: primary structure, origin, and evolution. Curr. Top. Dev. Biol. *80*, 209-276.
Martik, M.L., and McClay, D.R. (2015). Deployment of a retinal determination gene network drives directed cell migration in the sea urchin embryo. eLife *4*.
Martín-Durán, J.M., Vellutini, B.C., and Hejnol, A. (2015). Evolution and development of the adelphophagic, intracapsular Schmidt's larva of the nemertean *Lineus ruber*. EvoDevo *6*, 28.
Martindale, M.Q. (2005). The evolution of metazoan axial properties. Nat. Rev. Genet. *6*, 917-927.
Martindale, M.Q., and Hejnol, A. (2009). A developmental perspective: changes in the position of the blastopore during bilaterian evolution. Dev. Cell. *17*, 162-174.
Martindale, M.Q., Pang, K., and Finnerty, J.R. (2004). Investigating the origins of triploblasty: 'mesodermal' gene expression in a diploblastic animal, the sea anemone *Nematostella vectensis* (phylum, Cnidaria; class, Anthozoa). Development *131*, 2463-2474.
Massague, J. (2012). TGFβ signalling in context. Nat. Rev. Mol. Cell Biol. *13*, 616-630.
McConnell, D. (1963). Inorganic constituents in the shell of the linving brachiopod *Lingula*. Geol. Soc. Am. Bull. *74*, 363-364.
Medzhitov, R. (2001). Toll-like receptors and innate immunity. Nat. Rev. Immunol. *1*, 135-145.
Merkel, C., Deuschle, J., Griesshaber, E., Enders, S., Steinhauser, E., Hochleitner, R., Brand, U., and Schmahl, W.W. (2009). Mechanical properties of modern calcite- (*Mergerlia truncata*) and phosphate-shelled brachiopods (*Discradisca stella* and *Lingula anatina*) determined by nanoindentation. J. Struct. Biol. *168*, 396-408.
Metzker, M.L. (2010). Sequencing technologies - the next generation. Nat. Rev. Genet. *11*, 31-46.
Mi, H., Muruganujan, A., and Thomas, P.D. (2013). PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Res. *41*, D377-386.
Moore, J., and Gibson, R. (2001). Nemertea (Ribbon Worms). In eLS (John Wiley & Sons, Ltd).
Moroz, L.L. (2011). *Aplysia*. Curr. Biol. *21*, R60-R61.
Moroz, L.L., Kocot, K.M., Citarella, M.R., Dosung, S., Norekian, T.P., Povolotskaya, I.S., Grigorenko, A.P., Dailey, C., Berezikov, E., Buckley, K.M.*, et al.* (2014). The ctenophore genome and the evolutionary origins of neural systems. Nature *510*, 109-114.
Mount, A.S., Wheeler, A.P., Paradkar, R.P., and Snider, D. (2004). Hemocyte-mediated shell mineralization in the eastern oyster. Science *304*, 297-300.
Moysiuk, J., Smith, M.R., and Caron, J.-B. (2017). Hyoliths are Palaeozoic lophophorates. Nature *541*, 394-397.
Mukaka, M.M. (2012). Statistics corner: A guide to appropriate use of correlation coefficient in medical research. Malawi Med. J. *24*, 69-71.

Neary, M.T., Reid, D.G., Mason, M.J., Friscic, T., Duer, M.J., and Cusack, M. (2011). Contrasts between organic participation in apatite biomineralization in brachiopod shell and vertebrate bone identified by nuclear magnetic resonance spectroscopy. J. R. Soc. Interface *8*, 282-288.

Nesnidal, M.P., Helmkampf, M., Meyer, A., Witek, A., Bruchhaus, I., Ebersberger, I., Hankeln, T., Lieb, B., Struck, T.H., and Hausdorf, B. (2013). New phylogenomic data support the monophyly of Lophophorata and an Ectoproct-Phoronid clade and indicate that Polyzoa and Kryptrochozoa are caused by systematic bias. BMC Evol. Biol. *13*, 253.

Niehrs, C. (2012). The complex world of WNT receptor signalling. Nat. Rev. Mol. Cell Biol. *13*, 767-779.

Nielsen, C. (2001). Animal evolution: Interrelationships of the living phyla, 2nd edn (Oxford: Oxford Univ. Press).

Nielsen, C. (2002). The phylogenetic position of entoprocta, ectoprocta, phoronida, and brachiopoda. Integr. Comp. Biol. *42*, 685-691.

Pani, A.M., Mullarkey, E.E., Aronowicz, J., Assimacopoulos, S., Grove, E.A., and Lowe, C.J. (2012). Ancient deuterostome origins of vertebrate brain signalling centres. Nature *483*, 289-294.

Paps, J., Baguna, J., and Riutort, M. (2009a). Bilaterian phylogeny: a broad sampling of 13 nuclear genes provides a new Lophotrochozoa phylogeny and supports a paraphyletic basal acoelomorpha. Mol. Biol. Evol. *26*, 2397-2406.

Paps, J., Baguna, J., and Riutort, M. (2009b). Lophotrochozoa internal phylogeny: new insights from an up-to-date analysis of nuclear ribosomal genes. Proc. Biol. Sci. *276*, 1245-1254.

Parra, G., Bradnam, K., and Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics *23*, 1061-1067.

Passamaneck, Y., and Halanych, K.M. (2006). Lophotrochozoan phylogeny assessed with LSU and SSU data: evidence of lophophorate polyphyly. Mol. Phylogenet. Evol. *40*, 20-28.

Pearson, J.C., Lemons, D., and McGinnis, W. (2005). Modulating Hox gene functions during animal body patterning. Nat. Rev. Genet. *6*, 893-904.

Peter, I.S., and Davidson, E.H. (2011). Evolution of gene regulatory networks that control embryonic sevelopment of the body plan. Cell *144*, 970-985.

Petersen, C.P., and Reddien, P.W. (2009). Wnt signaling and the polarity of the primary body axis. Cell *139*, 1056-1068.

Pisani, D., Pett, W., Dohrmann, M., Feuda, R., Rota-Stabelli, O., Philippe, H., Lartillot, N., and Worheide, G. (2015). Genomic data do not support comb jellies as the sister group to all other animals. Proc. Natl. Acad. Sci. USA *112*, 15402-15407.

Politi, Y., Metzler, R.A., Abrecht, M., Gilbert, B., Wilt, F.H., Sagi, I., Addadi, L., Weiner, S., and Gilbert, P.U.P.A. (2008). Transformation mechanism of amorphous calcium carbonate into calcite in the sea urchin larval spicule. Proc. Natl. Acad. Sci. USA *105*, 17362-17366.

Price, A.L., Jones, N.C., and Pevzner, P.A. (2005). *De novo* identification of repeat families in large genomes. Bioinformatics *21*, i351-i358.

Putnam, N.H., Butts, T., Ferrier, D.E.K., Furlong, R.F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M., Shoguchi, E., Terry, A., Yu, J.-K.*, et al.* (2008). The amphioxus genome and the evolution of the chordate karyotype. Nature *453*, 1064-1071.

Putnam, N.H., Srivastava, M., Hellsten, U., Dirks, B., Chapman, J., Salamov, A., Terry, A., Shapiro, H., Lindquist, E., Kapitonov, V.V.*, et al.* (2007). Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science *317*, 86-94.

Quint, M., Drost, H.G., Gabel, A., Ullrich, K.K., Bonn, M., and Grosse, I. (2012). A transcriptomic hourglass in plant embryogenesis. Nature *490*, 98-101.

Ramos-Silva, P., Kaandorp, J., Huisman, L., Marie, B., Zanella-Cleon, I., Guichard, N., Miller, D.J., and Marin, F. (2013). The skeletal proteome of the coral *Acropora millepora*: the evolution of calcification by co-option and domain shuffling. Mol. Biol. Evol. *30*, 2099-2112.

Richards, S., Gibbs, R.A., Weinstock, G.M., Brown, S.J., Denell, R., Beeman, R.W., Gibbs, R., Beeman, R.W., Brown, S.J., Bucher, G.*, et al.* (2008). The genome of the model beetle and pest *Tribolium castaneum*. Nature *452*, 949-955.

Riddiford, N., and Olson, P.D. (2011). Wnt gene loss in flatworms. Dev. Genes Evol. *221*, 187-197.

Robb, S.M., Gotting, K., Ross, E., and Sanchez Alvarado, A. (2015). SmedGD 2.0: The Schmidtea mediterranea genome database. Genesis *53*, 535-546.

Rohanizadeh, R., and Legeros, R.Z. (2007). Mineral phase in linguloid brachiopod shell: *Lingula adamsi*. Lethaia *40*, 61-68.

Rumpho, M.E., Pelletreau, K.N., Moustafa, A., and Bhattacharya, D. (2011). The making of a photosynthetic animal. J. Exp. Biol. *214*, 303-311.

Ryan, J.F., Pang, K., Schnitzler, C.E., Nguyen, A.D., Moreland, R.T., Simmons, D.K., Koch, B.J., Francis, W.R., Havlak, P., Smith, S.A.*, et al.* (2013). The genome of the ctenophore *Mnemiopsis leidyi* and its implications for cell type evolution. Science *342*, 1242592.

Salichos, L., and Rokas, A. (2013). Inferring ancient divergences requires genes with strong phylogenetic signals. Nature *497*, 327-331.

Santagata, S. (2015). Phoronida. In Evolutionary Developmental Biology of Invertebrates 2: Lophotrochozoa (Spiralia), A. Wanninger, ed. (Vienna: Springer Vienna), pp. 231-245.

Santagata, S., Resh, C., Hejnol, A., Martindale, M.Q., and Passamaneck, Y.J. (2012). Development of the larval anterior neurogenic domains of *Terebratalia transversa* (Brachiopoda) provides insights into the diversification of larval apical organs and the spiralian nervous system. EvoDevo *3*, 3.

Sarashina, I., and Endo, K. (2006). Skeletal matrix proteins of invertebrate animals: Comparative analysis of their amino acid sequences. Paleontol. Res. *10*, 311-336.

Sarashina, I., Yamaguchi, H., Haga, T., Iijima, M., Chiba, S., and Endo, K. (2006). Molecular evolution and functionally important structures of molluscan Dermatopontin: implications for the origins of molluscan shell matrix proteins. J. Mol. Evol. *62*, 307-318.

Savazzi, E. (1991). Burrowing in the inarticulate brachiopod *Lingula anatina*. Palaeogeogr. Palaeocl. *85*, 101-106.

Schiemann, S.M., Martin-Duran, J.M.M., Borve, A., Vellutini, B.C., Passamaneck, Y.J., and Hejnol, A. (2017). Clustered brachiopod Hox genes are not expressed collinearly and are associated with lophotrochozoan novelties. Proc. Natl. Acad. Sci. USA *114*, E1913-E1922.

Sekita, Y., Wagatsuma, H., Nakamura, K., Ono, R., Kagami, M., Wakisaka, N., Hino, T., Suzuki-Migishima, R., Kohda, T., Ogura, A.*, et al.* (2008). Role of retrotransposon-derived imprinted gene, *Rtl1*, in the feto-maternal interface of mouse placenta. Nat. Genet. *40*, 243-248.

Seo, H.C., Edvardsen, R.B., Maeland, A.D., Bjordal, M., Jensen, M.F., Hansen, A., Flaat, M., Weissenbach, J., Lehrach, H., Wincker, P.*, et al.* (2004). Hox cluster disintegration with persistent anteroposterior order of expression in *Oikopleura dioica*. Nature *431*, 67-71.

Shinzato, C., Shoguchi, E., Kawashima, T., Hamada, M., Hisata, K., Tanaka, M., Fujie, M., Fujiwara, M., Koyanagi, R., Ikuta, T.*, et al.* (2011). Using the *Acropora digitifera* genome to understand coral responses to environmental change. Nature *476*, 320-323.

Shiomi, K., Midorikawa, S., Ishida, M., Nagashima, Y., and Nagai, H. (2004). Plancitoxins, lethal factors from the crown-of-thorns starfish *Acanthaster planci*, are deoxyribonucleases II. Toxicon *44*, 499-506.

Shubin, N., Tabin, C., and Carroll, S. (2009). Deep homology and the origins of evolutionary novelty. Nature *457*, 818-823.

Simakov, O., and Kawashima, T. (2016). Independent evolution of genomic characters during major metazoan transitions. Dev. Biol. *S0012-1606*, 30480-30488.

Simakov, O., Kawashima, T., Marlétaz, F., Jenkins, J., Koyanagi, R., Mitros, T., Hisata, K., Bredeson, J., Shoguchi, E., Gyoja, F.*, et al.* (2015). Hemichordate genomes and

deuterostome origins. Nature *527*, 459-465.
Simakov, O., Marletaz, F., Cho, S.-J., Edsinger-Gonzales, E., Havlak, P., Hellsten, U., Kuo, D.-H., Larsson, T., Lv, J., Arendt, D*., et al.* (2013). Insights into bilaterian evolution from three spiralian genomes. Nature *493*, 526-531.
Simion, P., Philippe, H., Baurain, D., Jager, M., Richter, D.J., Di Franco, A., Roure, B., Satoh, N., Quéinnec, É., Ereskovsky, A*., et al.* (2017). A large and consistent phylogenomic dataset supports sponges as the sister group to all other animals. Curr. Biol. *27*, 1-10.
Smith, M.R., and Caron, J.B. (2010). Primitive soft-bodied cephalopods from the Cambrian. Nature *465*, 469-472.
Smith, S.A., Wilson, N.G., Goetz, F.E., Feehery, C., Andrade, S.C.S., Rouse, G.W., Giribet, G., and Dunn, C.W. (2011). Resolving the evolutionary relationships of molluscs with phylogenomic tools. Nature *480*, 364-367.
Sodergren, E., Weinstock, G.M., Davidson, E.H., Cameron, R.A., Gibbs, R.A., Angerer, R.C., Angerer, L.M., Arnone, M.I., Burgess, D.R., Burke, R.D*., et al.* (2006). The genome of the sea urchin *Strongylocentrotus purpuratus*. Science *314*, 941-952.
Sperling, E.A., Pisani, D., and Peterson, K.J. (2011). Molecular paleobiological insights into the origin of the Brachiopoda. Evol. Dev. *13*, 290-303.
Srivastava, M., Begovic, E., Chapman, J., Putnam, N.H., Hellsten, U., Kawashima, T., Kuo, A., Mitros, T., Salamov, A., Carpenter, M.L*., et al.* (2008). The *Trichoplax* genome and the nature of placozoans. Nature *454*, 955-960.
Srivastava, M., Simakov, O., Chapman, J., Fahey, B., Gauthier, M.E., Mitros, T., Richards, G.S., Conaco, C., Dacre, M., Hellsten, U*., et al.* (2010). The *Amphimedon queenslandica* genome and the evolution of animal complexity. Nature *466*, 720-726.
Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics.
Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. Bioinformatics *24*, 637-644.
Steinmetz, P.R., Urbach, R., Posnien, N., Eriksson, J., Kostyuchenko, R.P., Brena, C., Guy, K., Akam, M., Bucher, G., and Arendt, D. (2010). *Six3* demarcates the anterior-most developing brain region in bilaterian animals. EvoDevo *1*, 14.
Struck, T.H., Wey-Fabrizius, A.R., Golombek, A., Hering, L., Weigert, A., Bleidorn, C., Klebow, S., Iakovenko, N., Hausdorf, B., Petersen, M*., et al.* (2014). Platyzoan paraphyly based on phylogenomic data supports a noncoelomate ancestry of spiralia. Mol. Biol. Evol. *31*, 1833-1849.
Sun, J., and Bhushan, B. (2012). Hierarchical structure and mechanical properties of nacre: a review. RSC Advances *2*, 7617-7632.
Suzuki, M., and Nagasawa, H. (2013). Mollusk shell structures and their formation mechanism. Can. J. Zool. *91*, 349-366.
Tagawa, K., Arimoto, A., Sasaki, A., Izumi, M., Fujita, S., Humphreys, T., Fujiyama, A., Kagoshima, H., Shin, I.T., Kohara, Y*., et al.* (2014). A cDNA resource for gene expression studies of a hemichordate, *Ptychodera flava*. Zoolog. Sci. *31*, 414-420.
Takeuchi, T., Kawashima, T., Koyanagi, R., Gyoja, F., Tanaka, M., Ikuta, T., Shoguchi, E., Fujiwara, M., Shinzato, C., Hisata, K*., et al.* (2012). Draft genome of the pearl oyster *Pinctada fucata*: a platform for understanding bivalve biology. DNA Res. *19*, 117-130.
Takeuchi, T., Koyanagi, R., Gyoja, F., Kanda, M., Hisata, K., Fujie, M., Goto, H., Yamasaki, S., Nagai, K., Morino, Y*., et al.* (2016). Bivalve-specific gene expansion in the pearl oyster genome: implications of adaptation to a sessile lifestyle. Zool. Lett. *2*, 3.
Telford, M.J., Budd, G.E., and Philippe, H. (2015). Phylogenomic insights into animal evolution. Curr. Biol. *25*, R876-887.
Temereva, E.N., Gebruk, A.A., and Malakhov, V.V. (2015). Demonstration of the preoral coelom in the brachiopod *Lingula anatina* with consideration of its phylogenetic significance. Zool. Anz. *256*, 22-27.

Temereva, E.N., and Tsitrin, E.B. (2015). Modern data on the innervation of the lophophore in *Lingula anatina* (Brachiopoda) support the monophyly of the lophophorates. PLOS ONE *10*, e0123040.

Tsai, I.J., Zarowiecki, M., Holroyd, N., Garciarrubio, A., Sanchez-Flores, A., Brooks, K.L., Tracey, A., Bobes, R.J., Fragoso, G., Sciutto, E.*, et al.* (2013). The genomes of four tapeworm species reveal adaptations to parasitism. Nature *496*, 57-63.

Tschopp, P., Sherratt, E., Sanger, T.J., Groner, A.C., Aspiras, A.C., Hu, J.K., Pourquie, O., Gros, J., and Tabin, C.J. (2014). A relative shift in cloacal location repositions external genitalia in amniote evolution. Nature *516*, 391-394.

Tschopp, P., and Tabin, C.J. (2017). Deep homology in the age of next-generation sequencing. Philos. Trans. R. Soc. Lond. B Biol. Sci. *372*.

Ueda, A., Suzuki, M., Honma, T., Nagai, H., Nagashima, Y., and Shiomi, K. (2006). Purification, properties and cDNA cloning of neoverrucotoxin (neoVTX), a hemolytic lethal factor from the stonefish *Synanceia verrucosa* venom. Biochim. Biophys. Acta *1760*, 1713-1722.

Venkatesh, B., Lee, A.P., Ravi, V., Maurya, A.K., Lian, M.M., Swann, J.B., Ohta, Y., Flajnik, M.F., Sutoh, Y., Kasahara, M.*, et al.* (2014). Elephant shark genome provides unique insights into gnathostome evolution. Nature *505*, 174-179.

Von Allmen, G., Hogga, I., Spierer, A., Karch, F., Bender, W., Gyurkovics, H., and Lewis, E. (1996). Splits in fruitfly Hox gene complexes. Nature *380*, 116.

von Döhren, J. (2015). Nemertea. In Evolutionary Developmental Biology of Invertebrates 2: Lophotrochozoa (Spiralia), A. Wanninger, ed. (Vienna: Springer Vienna), pp. 155-192.

Wang, X., Li, L., Zhu, Y., Du, Y., Song, X., Chen, Y., Huang, R., Que, H., Fang, X., and Zhang, G. (2013). Oyster shell proteins originate from multiple organs and their probable transport pathway to the shell formation front. PLoS One *8*, e66522.

Wang, Z., Young, R.L., Xue, H., and Wagner, G.P. (2011). Transcriptomic analysis of avian digits reveals conserved and derived digit identities in birds. Nature *477*, 583-586.

Wasik, K., Gurtowski, J., Zhou, X., Ramos, O.M., Delas, M.J., Battistoni, G., El Demerdash, O., Falciatori, I., Vizoso, D.B., Smith, A.D.*, et al.* (2015). Genome and transcriptome of the regeneration-competent flatworm, *Macrostomum lignano*. Proc. Natl. Acad. Sci. USA *112*, 12462-12467.

Weigert, A., Helm, C., Meyer, M., Nickel, B., Arendt, D., Hausdorf, B., Santos, S.R., Halanych, K.M., Purschke, G., Bleidorn, C.*, et al.* (2014). Illuminating the base of the annelid tree using transcriptomics. Mol. Biol. Evol. *31*, 1391-1401.

Weiner, S., and Dove, P. (2003). An overview of biomineralization processes and the problem of the vital effect. Rev. Mineral. Geochem. *54*, 1-29.

Whelan, N.V., Kocot, K.M., Moroz, L.L., and Halanych, K.M. (2015). Error, signal, and the placement of Ctenophora sister to all other animals. Proc. Natl. Acad. Sci. USA *112*, 5773-5778.

Whelan, N.V., Kocot, K.M., Santos, S.R., and Halanych, K.M. (2014). Nemertean toxin genes revealed through transcriptome sequencing. Genome Biol. Evol. *6*, 3314-3325.

Williams, A., Carlson, S.J., Brunton, C.H.C., Holmer, L.E., and Popov, L. (1996). A supra-ordinal classification of the Brachiopoda. Phil. Trans. R. Soc. B *351*, 1171-1193.

Williams, A., and Cusack, M. (1999). Evolution of a rhythmic lamination in the organophosphatic shells of brachiopods. J. Struct. Biol. *126*, 227-240.

Williams, A., Cusack, M., and Mackay, S. (1994). Collagenous chitinophosphatic shell of the brachiopod *Lingula*. Phil. Trans. R. Soc. B *346*, 223-266.

Wong, Y.H., Ryu, T., Seridi, L., Ghosheh, Y., Bougouffa, S., Qian, P.Y., and Ravasi, T. (2014). Transcriptome analysis elucidates key developmental components of bryozoan lophophore development. Sci. Rep. *4*, 6534.

Yang, S., Lai, X., Sheng, G., and Wang, S. (2013). Deep genetic divergence within a “living fossil” brachiopod *Lingula anatina*. J. Paleontol. *87*, 902-908.

Yatsu, N. (1902). On the development of *Lingula anatina*. J. Coll. Sci. Imp. Univ. Tokyo *17*, 1-112.

Zhang, G., Fang, X., Guo, X., Li, L., Luo, R., Xu, F., Yang, P., Zhang, L., Wang, X., Qi, H.*, et al.* (2012). The oyster genome reveals stress adaptation and complexity of shell formation. Nature *490*, 49-54.

Zhang, L., Li, L., Guo, X., Litman, G.W., Dishaw, L.J., and Zhang, G. (2015). Massive expansion and functional divergence of innate immune genes in a protostome. Sci. Rep. *5*, 8693.

Zhang, Z., Shu, D., Han, J., and Liu, J. (2005). Morpho-anatomical differences of the Early Cambrian Chengjiang and Recent lingulids and their implications. Acta Zool. *86*, 277-288.

Zhong, Y.F., and Holland, P.W. (2011). HomeoDB2: functional expansion of a comparative homeobox gene database for evolutionary developmental biology. Evol. Dev. *13*, 567-568.