Asymmetry in Spiralian Development

Erica K.O. Namigai

St. Hilda's College

Submitted for the degree of Doctor of Philosophy in Zoology

Trinity Term 2015

Asymmetry in Spiralian Development

Erica K.O. Namigai

St. Hilda's College

Submitted for the degree of Doctor of Philosophy in Zoology

Trinity Term 2015

Abstract

In animal development, how an animal builds its body has been a fundamental and central question. Spanning different fields, we have a good understanding of how animals build their body axes (anterior-posterior, dorsal-ventral) and this understanding is now a routine concept when learning about animal development. There are a number of mysteries left when considering early development, and the establishment of left-right (LR) asymmetry remains one of them. One key problem is how initial symmetry is broken early in animal development and how this translates into molecular asymmetry. Our lack of understanding is based on the difficulty of adapting available techniques to non-model species. This thesis uses an alternative approach to the current LR asymmetry models, and reveals that by adapting techniques to non-model organisms and by using newly available technologies, there is still much to be learned and discovered in embryology.

For my family

Table of Contents

List of Figures

Chapter 1: Introduction

Chapter 2: Molecular resources for studies of annelid development

Chapter 3: Nodal signalling in annelid development

List of Tables

Chapter 1

General Introduction

Bilateral symmetry in the Bilateria

Most living animals have a bilaterally symmetric body plan. Bilateral symmetry is achieved when an animal has two body axes, anterior-posterior (AP) and dorsal-ventral (DV), that result in mirror images on either side of the medio-lateral line. Animals with this characteristic have been classically put into a large grouping called the Bilateria (Figure 1.1), which consists of three major Superphyla: Ecdysozoa (insects and allies), Lophotrochozoa (molluscs, annelids, and allies), and Deuterostomia (humans, sea squirts, and allies) (Dunn et al., 2008; Pick et al., 2010). These bilaterian animals contrast with the Diploblastica (formerly the Radiata), a paraphyletic grouping, which consists of animals with radial symmetry such as the cnidarians and the ctenophores. Of course, there are caveats to this; for example, evidence has emerged that one cnidarian species could exhibit molecular bilaterality (Matus et al., 2006a, 2006b). It is widely accepted that bilaterality evolved once in the animal kingdom, and that the last common ancestor of the Bilateria was bilaterally symmetric. However, what is often overlooked is that in most animals, bilateral symmetry is not truly bilaterally symmetric. Such differences can be subtle, such as the coiling of internal organs in the human and the fly, or readily apparent, such as the directional coiling of snail shells (Palmer, 2004). These morphological asymmetries are termed left-right (LR) asymmetry.

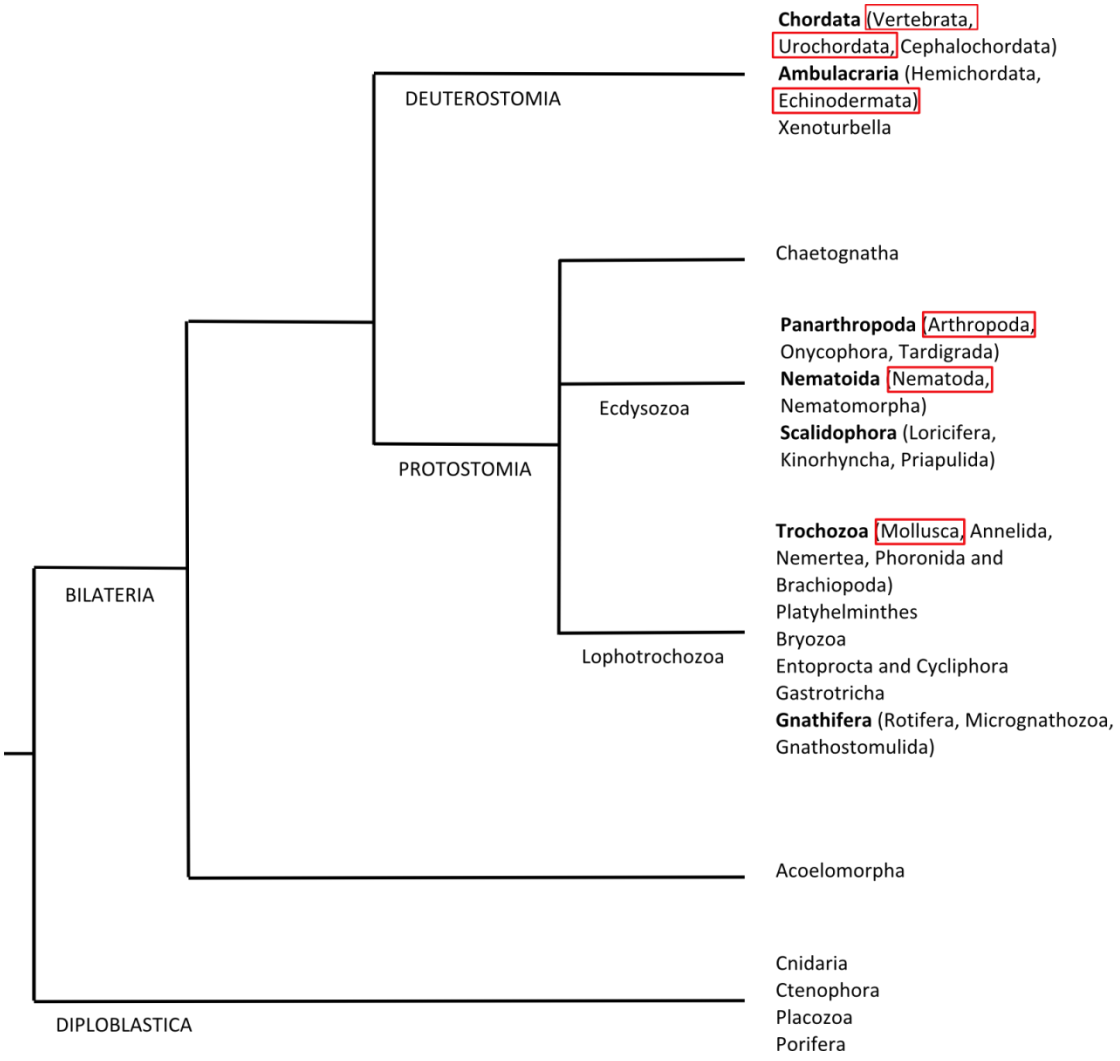


Figure 1.1 Phylogenetic relationships within the animal kingdom. Red boxes indicate animals that have been used in mechanistic LR asymmetry studies so far. In bold are names of large groupings. Tree adapted from (Dunn et al., 2014) and Jordi Paps (personal communication).

Caveats for the LR axis and LR asymmetry

It can be said that there is a third, LR, axis that is established during development. However, caution should be noted in this case, as the LR axis is conceptually different in comparison to the AP and DV axes. Rather than an axis that spans the left and right sides (as the AP and DV axes span their respective given names), the LR axis is not a continuous axis. There is one axis

on either side of the midline (Palmer, 2004). This is significant when considering developmental mechanisms of how LR asymmetry is established, as there are two relatively autonomous axes being considered that would in effect require two patterning mechanisms. However, for the purpose of this thesis, we will refer to the LR axis in its singular form, while keeping in mind that there is this underlying duality.

LR asymmetry itself can be the result of both random and non-random events, and has been subdivided into three categories: fluctuating asymmetry, anti-symmetry, and directional asymmetry. LR asymmetry can be the result of developmental noise, which would imply that small differences of growth or patterning during development result in asymmetry that is not heritable (termed fluctuating asymmetry). Asymmetry among individuals in a population can also be completely stochastic (termed anti-symmetry). What we will be considering in this thesis is when asymmetries are non-random (termed directional symmetry). This is defined as asymmetries that are directionally fixed in a population (Palmer, 2004), and as such are heritable. A lingering question is: how is directional asymmetry introduced and communicated during development, and how does this happen despite initial symmetry of the AP and DV axes?

Three steps to LR asymmetry

Before moving further, it is important to understand how LR asymmetry establishment is viewed. The aforementioned morphological asymmetries are really the endpoint of a developmental process to establish LR asymmetry. This

process consists of three steps: symmetry breaking, asymmetric gene expression, and morphological asymmetry (Figure 1.2). In short, symmetry is generally first broken at the embryonic stage, and this asymmetry is read into a molecular mechanism (as seen by asymmetric gene expression), which then cascades and produces adult morphological asymmetry. In order to understand how morphological asymmetries are generated, the mechanisms underlying symmetry breaking and asymmetric gene expression must be studied. As the realm of gene expression has been the best studied so far, I will begin with asymmetric gene expression and later move to the underlying mechanisms of symmetry breaking.

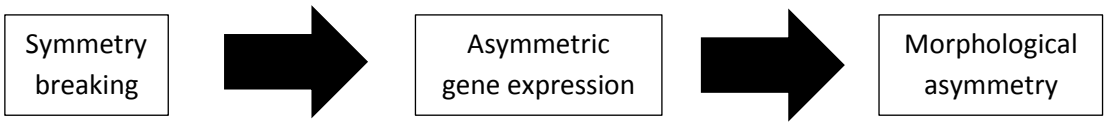


Figure 1.2 Three-step process to establish LR asymmetry.

Asymmetric gene expression: Nodal

In 1990, Brown and Wolpert identified the key problem in considering LR asymmetry (Brown and Wolpert, 1990): the search for the chiral molecule. In their model, in order to generate differences between the left and right sides, an inherently chiral molecule would need to be oriented along the AP and DV axes, generating, by its chirality, LR asymmetry. This model catalyzed a renewed interest in LR asymmetry, and in 1993 the gene *Nodal* was discovered as a molecular marker for asymmetry in vertebrates (Zhou et al., 1993). *Nodal* has since been found to have left-sided expression in all vertebrates studied so far.

Nodal is a transforming growth factor beta (TGF-beta) ligand that participates in mesendoderm induction, and is later expressed on the left side of the node (a midline structure found in most vertebrates). This left-sided expression induces left-sided expression of the homeobox transcription factor *pitx2*, and results in left-sided adult morphological asymmetry. With an exploitable molecular mechanism, interest in LR asymmetry was refuelled, and Nodal signalling and sided *Nodal* expression were revealed to be conserved in both vertebrate and invertebrate deuterostomes as a key participant in establishing LR asymmetry (Beck et al., 2002; Boorman and Shimeld, 2002; Duboc et al., 2005; Levin et al., 1995; Yu et al., 2002), where mouse and chicken have one copy of *Nodal*, *Xenopus* has five paralogs (*Xnr1, 2, 4, 5, 6*), zebrafish has three paralogs (*cyclops, squint,* and *southpaw*), and at least one copy of *Nodal* is found in invertebrate deuterostomes such as sea urchins and sea squirts.

It quickly became apparent that there were differences in the role of Nodal in establishing LR asymmetry. One difference lay in the sidedness of *Nodal* expression, where it is right-sided in echinoderms and left-sided in chordates. This can likely be attributed to the DV inversion event (for further discussion, see Gerhart, 2000). Another key difference is that the Ecdysozoa are lacking *Nodal* altogether (Blum et al., 2014 and references therein), and with Nodal signalling markedly missing in the Ecdysozoa, it was possible that sided Nodal signalling was confined to the deuterostomes, and that this mechanism was not conserved across the bilaterian Superphyla.

Mechanisms of symmetry breaking

As an increasing amount of evidence accumulated supporting the conserved role of Nodal in LR asymmetry establishment, the search expanded to the mechanism by which this asymmetric expression was achieved. This produced two models primarily studied in vertebrates: the cilial flow model and the ion channel model, and their ramifications and problems will be discussed below. Two other models, chromatid segregation and planar cell polarity (PCP), will not be considered here as the evidence supporting these models is not as strong (if interested, see Vandenberg and Levin, 2012, 2013).

Cilial flow model

Once Nodal and other molecular markers of asymmetry were discovered, there was a push to understand how this molecular asymmetry was established. One such model emerged that centered around cilia (Figure 1.3). This model proposes that molecular components are asymmetrically expressed because of the asymmetric beating of cilia, which transports certain molecules to one side of the embryo by generating a LR directional flow at the node (reviewed in Capdevila et al., 2000; Nonaka et al., 1998; Tabin, 2006).

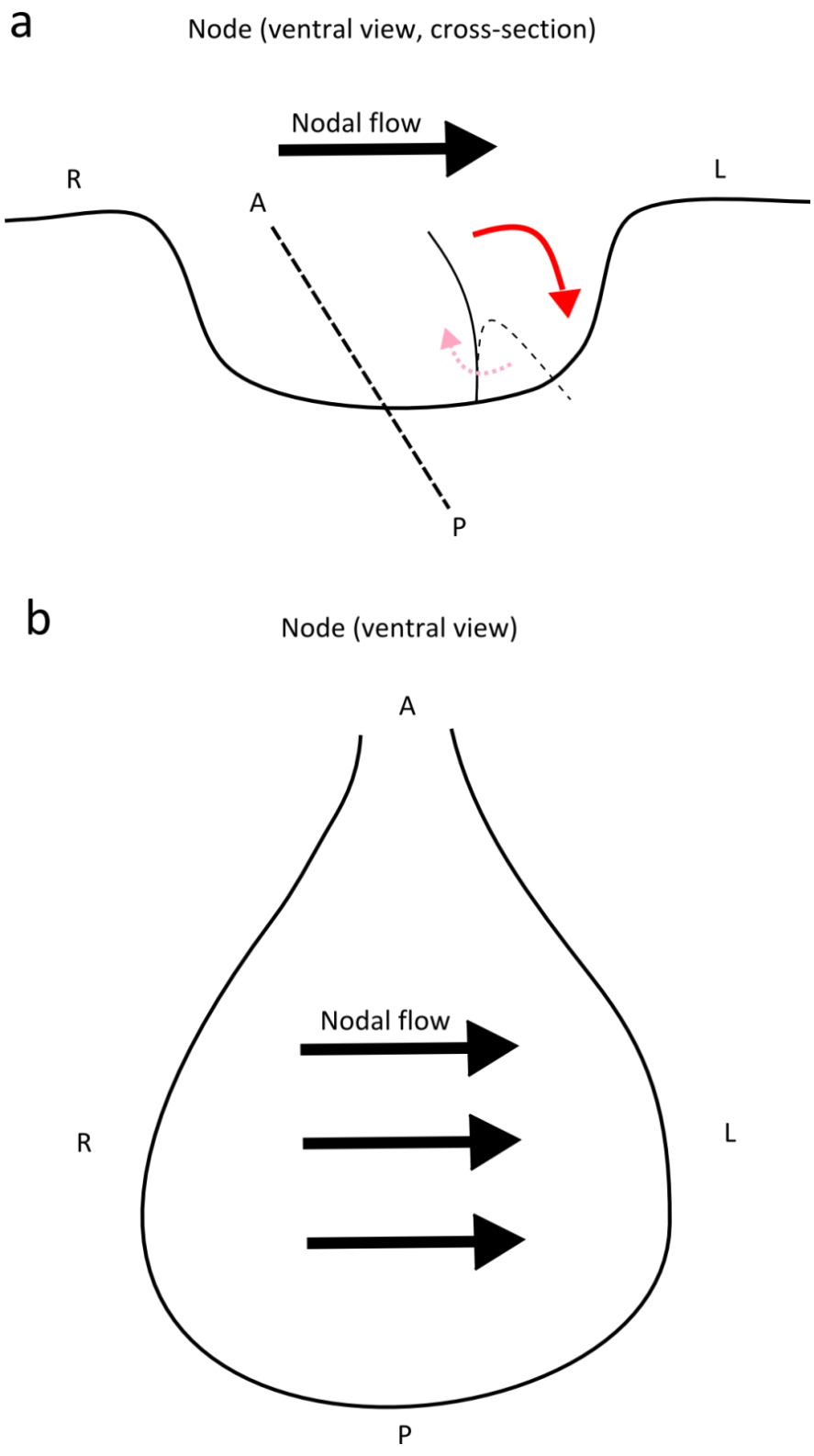


Figure 1.3 Nodal flow hypothesis. (A) Leftward flow is generated across the node by cilia, which rotate clockwise and generate a stronger flow towards the left side (red arrow) rather than the right side (pink, dotted arrow) due to the posterior tilt of the cilia, (B) the cilia-filled node has a general leftward flow due to the above mechanism (adapted from Nakamura and Hamada, 2012).

However, it has emerged that despite the conservation of *Nodal* and *pitx* expression in deuterostomes, the conservation of the cilial flow model is contentious and requires further investigation (discussed in Hirokawa et al., 2006; Nakamura and Hamada, 2012; Namigai et al., 2014; see Appendix 1). For example, this model does not necessarily apply to other deuterostomes. Rather, the cilia-driven nodal flow has been suggested to be a mammal-specific, or even a mouse and rabbit –specific, innovation (Raya and Izpisúa Belmonte, 2006). In frog, fish, and chicken, there is asymmetric gene expression prior to the formation of nodal cilia, implying that cilia may be an amplification step of an already established asymmetric signal (Capdevila et al., 2000; Hamada et al., 2002; Levin, 2005; Raya and Izpisúa Belmonte, 2006; Tabin and Vogan, 2003). Chicken does not have nodal cilia, establishing LR asymmetry through cell migration and changing the shape of the node itself (Cui et al., 2009; Gros et al., 2009), and pig also does not have nodal cilia (Gros et al., 2009). In addition, recent findings have revealed that there is no evidence of a motile function for cilia in the sea squirt *Ciona intestinalis*. Rather than a motile function, the cilia likely play a sensory role in the sea squirts, although this has recently been subject to debate (Nishide et al., 2012; Thompson et al., 2012). Such findings point to a re-evaluation of the cilial flow hypothesis to further consider two cilial flow models with two types of primary cilia, motile and sensory (Tabin and Vogan, 2003), in driving asymmetry, and to also consider mechanisms that do not involve cilia at the first instance, such as cell movement.

Ion channel model

The ion channel model comes from a different angle (Figure 1.4). Levin and colleagues considered the electrical potential of membranes, and proposed that differential charge across the early cleavage stage membrane generates asymmetry (symmetry breaking). This could be generated by H^+/K^+ATPase ion pumps which differentially transport charged molecules through gap junctions (Levin and Nascone, 1997). H^+/K^+ ATPase alpha has been found to be asymmetrically expressed as early as the four-cell stage in *Xenopus laevis* (Levin et al., 2002), and serotonin has been proposed as the charged molecule (Fukumoto et al., 2005).

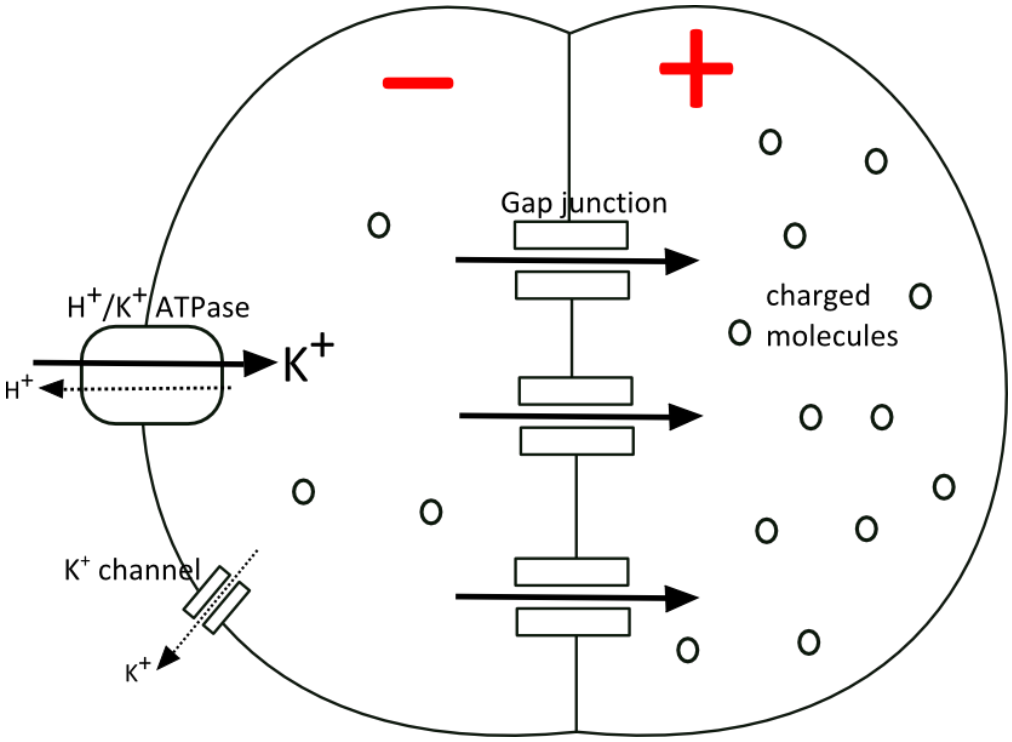


Figure 1.4 Ion channel hypothesis. Note: this takes place at the 4-cell stage in *X. laevis*, but is drawn as a 2-cell embryo for clarity (adapted from Levin et al., 2002).

There are problems with this model. The proposed charged molecule, serotonin, has recently been contested (Beyer et al., 2012), with doubts expressed as to whether serotonin is really asymmetrically distributed in the early *X. laevis* embryo. Early asymmetric expression of ion channels has not been seen in other organisms, although a role for these ion channels in asymmetry is suggested in sea urchins and sea squirts (Duboc et al., 2005; Shimeld and Levin, 2006). It remains possible that ion channels play a central role in the establishment of the LR axis, however it is clear that this model needs further testing before conjectures are made as to its conservation as a LR asymmetry establishing mechanism (for further discussion of this, see Levin and Palmer, 2007; Nakamura and Hamada, 2012; Vandenberg and Levin, 2010).

Nodal and symmetry breaking mechanisms

The cilial flow and ion channel models considered above have been tested in deuterostomes. Asymmetric gene expression follows these symmetry breaking mechanisms, and I would like to reprise that Nodal signalling is highly conserved in this group of animals: there is a great deal of evidence that left-sided *Nodal* expression is conserved in all vertebrates studied so far, and right-sided *Nodal* expression is conserved in all non-chordate deuterostomes studied so far. When considering how symmetry breaking mechanisms evolved in bilaterians, animals outside of the deuterostomes need to be considered, where Nodal is not necessarily conserved in LR asymmetry establishment.

Species without Nodal

As briefly mentioned, the Ecdysozoa lack *Nodal* and other components of the Nodal signalling pathway, and it has been widely thought that it has been lost, although this needs to be confirmed with studies of more lineages within the Ecdysozoa such as priapulids and tardigrades (see Namigai et al., 2014; Appendix 1). So how do these animals establish asymmetry? Evidence has emerged primarily from *Caenorhabditis elegans* and *Drosophila melanogaster*, and points to a mechanism based on the cytoskeleton.

In *C. elegans*, early cell divisions are already asymmetric (Figure 1.5), and LR asymmetry can be reversed by manipulating the direction of early cleavage (Wood, 1991). This cytoskeletal mechanism is thought to be actin and spindle-based, where the inherent polarity of actin filaments at the egg cortex generates cortical asymmetry that is then read into a molecular mechanism together with differential spindle elongation (Pohl and Bao, 2010; Pohl, 2011; Schonegg et al., 2014). In addition to a cytoskeletal mechanism, a molecular mechanism has also been clarified in *C. elegans*, where spindle orientation at the third cleavage is regulated by a G protein which aligns spindles and is necessary for LR asymmetry establishment (Bergmann et al., 2003). Correct spindle localization is also dependent on PAR proteins, which are asymmetrically distributed in the embryo via actin and myosin activity (Munro et al., 2004; reviewed in Nance and Zallen, 2011).

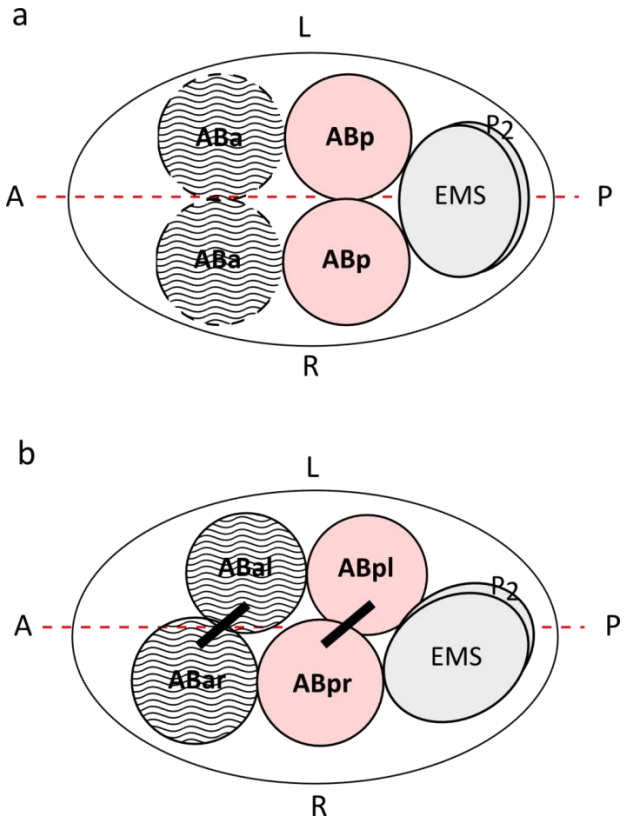


Figure 1.5 Spindle skewing during *C. elegans* LR asymmetry establishment. (A) 6-cell stage prior to spindle skew, (B) spindle skew of ABa and ABp blastomeres (adapted from Pohl and Bao, 2010).

In *D. melanogaster*, a slightly different cytoskeletal mechanism seems to be at work, where mutants of the type ID unconventional Myosin (MyoID) have defaults in asymmetry (Hozumi et al., 2006; Spéder et al., 2006). MyoID is a molecular motor that moves on actin filaments and is thought to interact with adherens junctions through cadherins and beta-catenin (Petzoldt et al., 2012; Spéder et al., 2006). Asymmetry is thought to be mediated by the interactions of the MyoID molecular motors with components such as the atypical cadherin Dachsous (González-Morales et al., 2015), generating transcriptional asymmetry, and polarized cell orientation and asymmetric rotation (Figure 1.6). The Hox gene Abdominal-B functions upstream of this process to regulate MyoID (Coutelis et al., 2013).

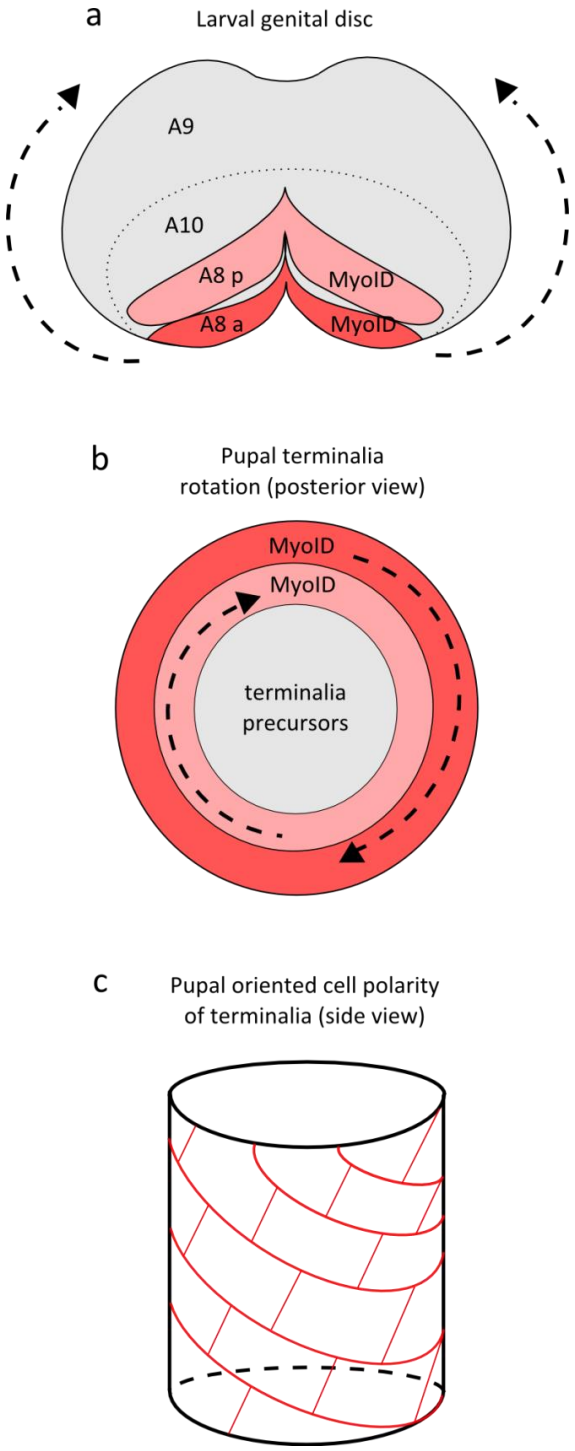


Figure 1.6 LR asymmetry establishment in *D. melanogaster*. (A) Larval genital disc (flat) with MyoID expressed in anterior (A8 a, light pink) and posterior (A8 p, red) segments that later serve as LR organizers, future terminalia (gray), arrows indicate direction of disc eversion during pupation, (B) following disk eversion at pupation A8 segments (light pink, red) fuse dorsally (location of dorsal = MyoID labels in diagram) and undergo terminalia rotation (terminalia precursors in gray, view from posterior), (C) as a result of terminalia rotation pupal terminalia develop with oriented cell polarity (cell polarity represented in red, side view) (adapted from Géminard et al., 2014; González-Morales et al., 2015).

It is important to consider that the two systems that are well-studied in the Ecdysozoa, *C. elegans* and *D. melanogaster*, differ significantly in early development. *C. elegans* undergoes holoblastic cleavage that is asymmetric from the second cleavage onwards, whereas *D. melanogaster* undergoes syncytial development where nuclei multiply in the embryo without cell membrane divisions. This highlights the danger of building hypotheses on these two organisms. We need to be aware that despite being in the same Superphylum, there can be very distinct modes of early development. Before making overarching conclusions concerning the conservation or divergence in the establishment of LR asymmetry, such differences in early development need to be kept in mind.

A unified model for asymmetry establishment

Taken together, the deuterostomes and the ecdysozoans have distinct mechanisms, yet a recurring aspect is the presence of polarized cytoskeletal components that can generate chirality which can then be read into asymmetry establishment relative to the AP and DV axes (Levin and Palmer, 2007; Vandenberg and Levin, 2010). This initial similarity may be a conserved aspect of LR asymmetry establishment, and indeed other cytoskeletal components are also polarized, such as microtubules that are known to differentially transport mRNA in *D. melanogaster* (Roth and Lynch, 2009), and centrosomes that participate in LR asymmetry establishment in *X. laevis* (Lobikin et al., 2012). The chirality of the actin cytoskeleton seen in *C. elegans* has also been seen in amphibian eggs (Danilchik et al., 2006), and in zebrafish LR asymmetry of the

heart is not disrupted in the absence of Nodal, which may be explained by the fact that its development is actin polymerization-dependent (Noël et al., 2013). This patchwork of evidence suggests that the cytoskeleton may generate the first signs of polarity that can then be read into later symmetry breaking events.

In such a model, we begin first with the naïve egg, which already has an animal-vegetal (AV) axis that is roughly correlated with the AP axis (Figure 1.7). The DV axis is likely established at or soon after fertilization, with a possible mechanism being upon sperm entry at fertilization. The inherent chirality of the cytoskeleton now has a two-axis framework upon which it could be read, where the chirality at the cytoskeletal level can now be interpreted using the two axes to break bilateral symmetry. This breaking of symmetry can then be communicated through a plethora of different methods such as ion channel or mRNA localization, Nodal signalling, and others. To understand whether this sequence of events is conserved, we need to understand how widespread such a mechanism is across the Bilateria, especially in organisms that have received comparatively less attention.

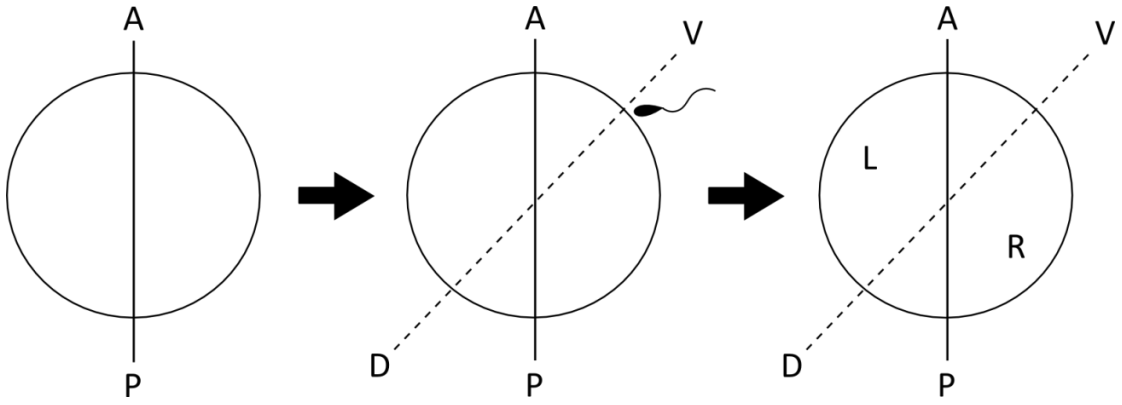


Figure 1.7 Schematic of potential mechanisms of axis establishment in some animal embryos.

Enter the Lophotrochozoa: snails and worms

The Lophotrochozoa is an under-studied Superphylum that includes organisms such as phoronids, bryozoans, rotifers, platyhelminthes, annelids, and molluscs (Figure 1.1 and Figure 1.9). There are many asymmetries in these species, such as the sided position of the operculum in some polychaetes, and most classically the coiling of the mollusc shell. This group of animals is largely united by a mode of cleavage termed spiral cleavage, where sidedness can already be seen at the 8-cell stage through the dextral or sinistral orientation of mitotic spindles (briefly in Figure 1.8, further addressed in Chapter 4). Spiral cleavage is shared by other lophotrochozoan Phyla and may be ancestral for the Lophotrochozoa, although this is somewhat contentious (Giribet, 2008; Hejnol, 2010). Despite the abundance of asymmetries in this Superphylum, we still know relatively little about how LR asymmetry is established in these animals, at least until recently.

Nodal signalling in the Lophotrochozoa

The discovery of *Nodal* in Mollusca (and other lophotrochozoans, see Kenny, Namigai et al., 2014; Appendix 3) renewed the possibility that Nodal signalling could be conserved in the Bilateria. In these studies, both *Nodal* and its downstream effector *pitx* were found to be asymmetrically expressed in molluscs, where sidedness of expression could be correlated with the direction of spiral cleavage at the 8-cell stage in the dextrally cleaving snail *Lottia gigantea* (which does not develop a coiled shell), the sinistrally cleaving snail *Biomphalaria glabrata*, and in the dextral and sinistral morphs of the snail

Lymnaea stagnalis. This correlates with adult shell coiling in the latter two species. Disrupting Nodal function affected *pitx* expression and subsequent morphological asymmetry. However, the conservation of the Nodal pathway was only confirmed in one lineage within the Lophotrochozoa, the molluscs, and requires further investigation using other lophotrochozoan species (further addressed in Chapter 3).

Symmetry breaking mechanisms in Mollusca

Symmetry breaking models prior to asymmetric *Nodal* and *pitx* expression focus on spindle orientation at the 8-cell stage, where the angle of spindle orientation correlates with the direction of morphological asymmetry in molluscs (Figure 1.8; Freeman and Lundelius, 1982; Shibazaki et al., 2004). Evidence of causation was shown by physically manipulating blastomeres in dextral and sinistral snails to alter the direction of spindle orientation (Kuroda et al., 2009). This forced the spiral 8-cell stage to spiral in the opposite direction in the respective snails, and resulted in a corresponding reversal of direction in shell chirality irrespective of the genetic background of the embryo. Later generations of these modified snails reverted back to their genetically-inherited direction of spiralling and shell chirality.

Direction of spiralling at the 8-cell stage is determined by a maternally-inherited nuclear locus, *sinistral*, in molluscs (Asami et al., 2008; Boycott et al., 1930; Diver et al., 1925; Sturtevant, 1923). The transfer of cytoplasm from a dextral egg into a sinistral egg caused dextralization (Freeman and Lundelius, 1982), suggesting that this nuclear locus encodes a specific factor that is

present in the cytoplasm of mollusc eggs that causes dextrality. The identity of this factor is not known, however this *sinistral* locus is being mapped in *Lymnea stagnalis* and *Lymnea peregra* (Liu et al., 2013), which were also used in the embryological studies of this locus. It is likely that this locus will encode a maternal protein (or proteins) that somehow interact with the spindle at the 8-cell stage to direct dextral orientation.

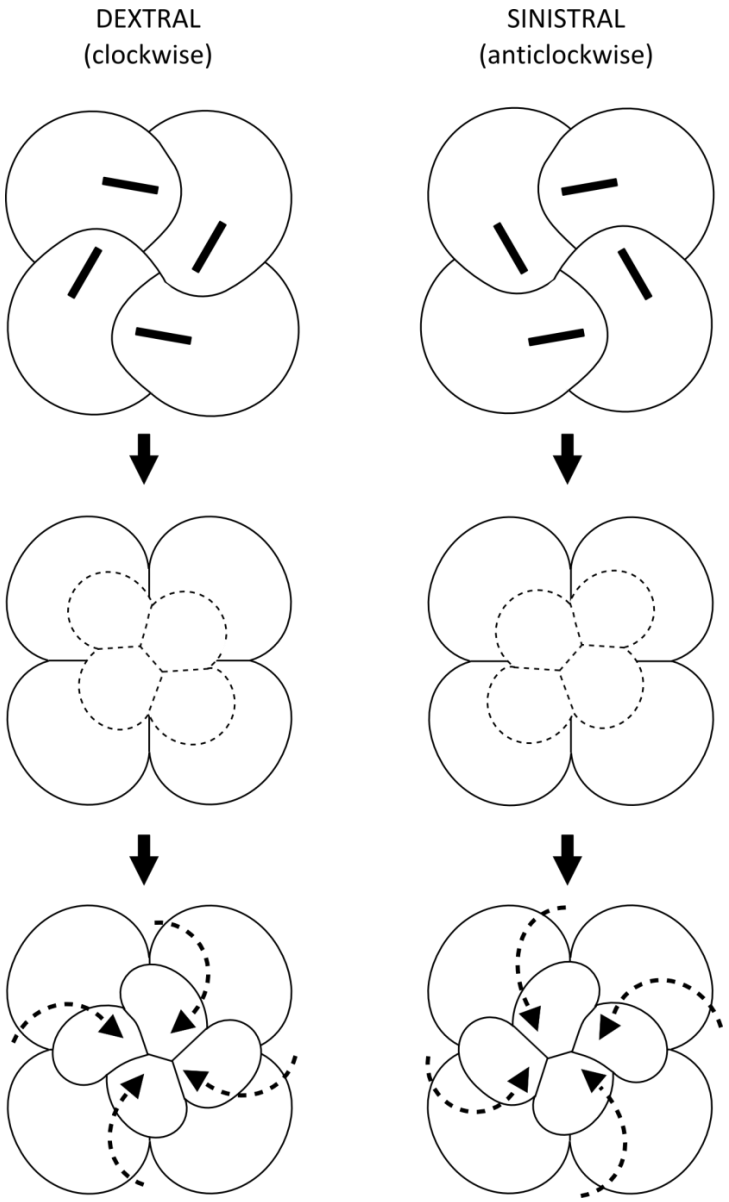


Figure 1.8 Dextral and sinistral coiling at the 8-cell stage in spiralian embryos (adapted from Shibazaki et al., 2004).

Caveats for dextrality and sinistrality in molluscs

The dextral and sinistral forms of the 8-cell stage and the chirality of the adult shell are generally presented as forms that are mirror images of each other, but in reality it is not that straightforward. Although the adult shells of dextral and sinistral molluscs are mirror images of each other, this is not the case for the spiralling of blastomeres at the 8-cell stage. Instead of being mirror images of each other, the dextral and sinistral spiralling of blastomeres are oriented through morphologically and temporally distinct cytoskeletal mechanisms, at least in the species studied so far (Shibazaki et al., 2004). The correlation between the spiralling of blastomeres and the chirality of the adult shell is less linear than currently presented, and there is likely an unappreciated amount of complexity in the cellular mechanism underlying chirality at early cleavage stages.

Symmetry breaking for other axes in molluscs

It is still not known how symmetry is first broken in sinistral and dextral molluscs, however there is further evidence for a cytoskeletal mechanism in establishing asymmetry across other axes in other mollusc species. For example, in the marine gastropod *Ilyanassa obsoleta*, maternally-deposited RNA is involved in specifying the formation of the D-quadrant organizer, a vegetal cell that directs DV axis development (Kingsley et al., 2007; Lambert and Nagy, 2002; Lambert, 2010; Rabinowitz and Lambert, 2010). Similar asymmetric RNA localization is also seen in the gastropod *Crepidula fornicata*

(Henry et al., 2010). It is possible that an asymmetrically deposited maternal mRNA also influences LR asymmetry, but this is not yet clear.

Caveats for LR asymmetry in relation to the DV axis

One important concept to understand is that dextrality and sinistrality at the 8-cell stage (third cleavage) does not necessarily correlate with initial symmetry breaking. This is the first visible sign of chirality, but in some equal-cleaving spiralians (see Chapter 4 for more in modes of cleavage) this precedes DV axis specification, which occurs much later with the formation of the third quartet of micromeres (24-cell stage). As left and right cannot be present without the AP and DV axes, spiral cleavage cannot break LR asymmetry in equal-cleaving spiralians at this point, but likely provides the initial chirality that can be harnessed by a subsequent mechanism that breaks symmetry in relation to the AP and DV axes. In effect, this means that when referring to sinistrality and dextrality at the 8-cell stage, this is not necessarily meant to be taken as the point of symmetry breaking but as the first sign of chirality in the embryo.

LR asymmetry in other lophotrochozoans

Outside of the molluscs, there are no studies showing any similar mechanisms for asymmetry establishment in other lophotrochozoan Phyla. This is mainly because they are under-studied in comparison to molluscs. The spiral cleavage program, however, is seen in a number of species including platyhelminthes, nemerteans, entoprocts, phoronids, annelids, and molluscs

(Lambert, 2010; Merkel et al., 2012; Pennerstorfer and Scholtz, 2012). The conservation of the spiral cleavage program makes it tempting to suggest that the symmetry breaking mechanisms related to spindle orientation in molluscs are also conserved in these lineages. However, there are important differences between molluscs and other lophotrochozoans. For example, sinistral mollusc strains and species occur naturally (if rarely), but the majority of other lineages seem to lack sinistral forms (at least as far as has been documented). It is also possible that similar mechanisms of asymmetry establishment have evolved convergently in spiral-cleaving animals. What is known is largely drawn from studies of the molluscs, and needs to be extended to other lophotrochozoan Phyla to understand the extent to which these genetic, molecular, and cellular mechanisms are conserved in the Lophotrochozoa.

Mechanisms of LR asymmetry establishment across the animal kingdom

Taken together, our understanding of LR asymmetry establishment and its underlying mechanisms is limited to a scattering of model organisms. This limits the phylogenetic distance at which we can generalize these mechanisms to all animals, and it becomes clear that a wider selection of animals needs to be studied to understand conservation and divergence in mechanisms (Figure 1.9).

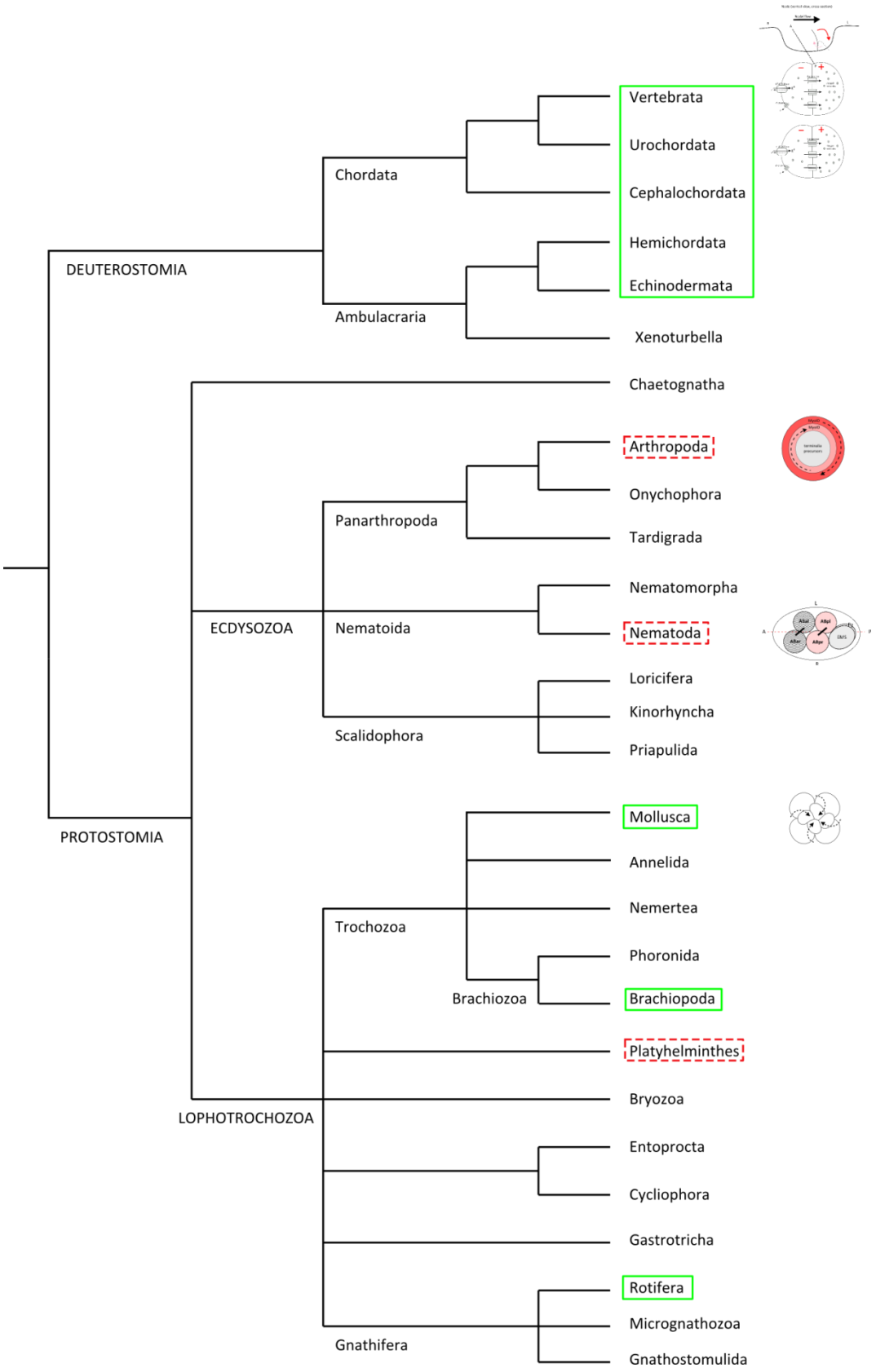


Figure 1.9 *Nodal* presence/absence and mechanisms for LR asymmetry establishment across the animal kingdom. Solid green boxes indicate presence of *Nodal*, dotted red boxes indicate absence of *Nodal*. Mechanism that have been studied run down the right-hand side.

To this effect, I have selected an annelid and a mollusc to study the mechanisms of LR asymmetry establishment: the polychaete annelid *Pomatoceros lamarcki* and the gastropod mollusc *B. glabrata*. An environment to keep *P. lamarcki* in the lab had already been established previously in the lab, and *B. glabrata* were obtained and set-up for culture in the lab (this thesis, see Chapter 4). These two were selected because they were readily available, and more importantly are both equal-cleaving spiralians (see Chapter 4 for more on cleavage modes) with opposite directionality, where *P. lamarcki* is a dextral-spiralling annelid with right-handed placement of the operculum (see Chapter 4), and *B. glabrata* is a sinistral-spiralling mollusc with left-handed chirality of the adult shell (Figure 1.10).

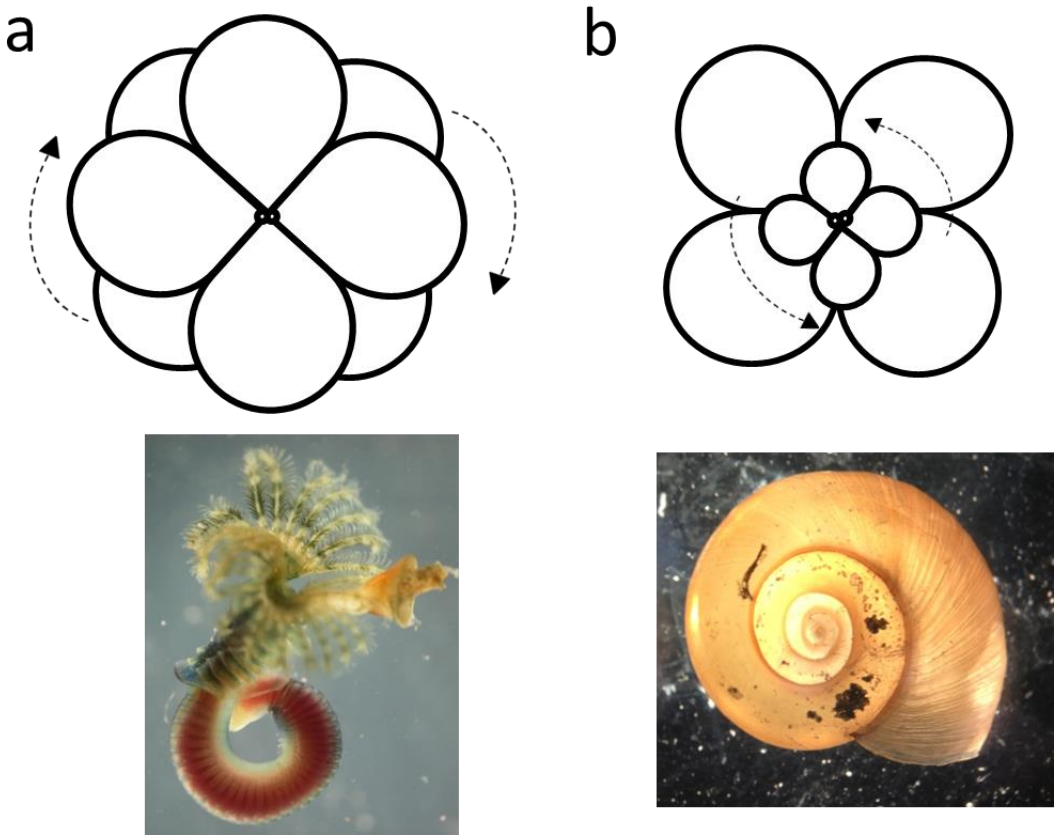


Figure 1.10 Representation of directionality in spiral cleavage in (A) the dextral annelid *P. lamarcki*, and (B) the sinistral mollusc *B. glabrata*.

P. lamarcki

The serpulid annelid *P. lamarcki* (synonymous with *Spirobranchus lamarcki* and *Pomatoceros lamarckii*) is an intertidal polychaete of the Sedentaria (Figure 1.11) and is used in ecotoxicology studies due to its harmful effects on shipping. The Sedentaria clade has recently been classified as a characteristic group within the polychaete annelids, along with errant polychaetes (the Errantia). In the most recent annelid phylogeny, *P. lamarcki* has been placed at the base of the Sedentaria (Struck et al., 2011). This tube-dwelling sedentary annelid undergoes equal spiral development, where the blastomeres are of equal size during early cleavage stages. Among the methods of early spiralian development, the equal cleavage of *P. lamarcki* is a method that has been particularly understudied in comparison to unequal cleavage and cleavage with large polar bodies (see Chapter 4 for more on modes of cleavage).

B. glabrata

B. glabrata is an equal-cleaving, neotropic, freshwater pulmonate gastropod of the family Planorbidae (Figure 1.12), and is best known for being an intermediate host for the disease-carrying trematode *Schistosoma mansoni*, a parasitic flatworm that transmits the human disease schistosomiasis. Most recently, it has been at the center of the LR asymmetry studies that first identified *Nodal* expression within the Lopohotrochozoa (Grande and Patel, 2009; Kuroda et al., 2009).

Figure 1.11 *P. lamarcki* adults. Female (pink/ dark orange body) and male (white/ light orange body).

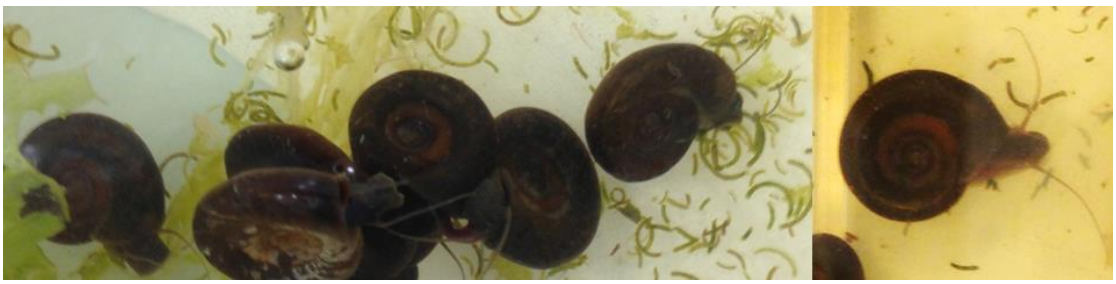


Figure 1.12 *B. glabrata* adults.

Aims of this thesis

This thesis aims to unravel the mechanisms underlying symmetry breaking and asymmetric gene expression during LR asymmetry establishment. Two non-model species were used with the ultimate goal of understanding how LR asymmetry establishment mechanisms may have evolved in bilateral animals (Figure 1.9). The use of non-model species meant that many routine techniques had not been established, and a great deal of technical innovation was done to achieve the work in this thesis. Together, this can be broken down into three aims, highlighting novel techniques that were developed in and for this thesis:

1. To generate a resource to study molecular evolution in a non-model organism. The *P. lamarcki* genome was sequenced and assembled in order to use *P. lamarcki* for studies of molecular evolution (Chapter 2).
2. To see whether *Nodal* is conserved in annelids. *Nodal* was searched for in the genome, and an *in situ* hybridization protocol was developed for *P. lamarcki* (Chapter 3).
3. To observe whether there was directionality at early cleavage stages by analyzing the inner cell dynamics of *P. lamarcki* and *B. glabrata* embryos. A Lightsheet microscope was used to visualize cytoplasmic dynamics. This was challenging as protocols for mounting, visualization, and data analysis were not yet fully developed. Mounting, visualization, and data analysis techniques were developed in this thesis in addition to mounting with drug treatments. Techniques for the dissection of *B. glabrata* eggs were also developed (Chapter 4).

Below are papers that have come out of this thesis:

*joint first author

Appendix 1: Namigai EKO, Kenny NJ, Shimeld SM, 2014. Right across the tree of life: the evolution of left-right asymmetry in the Bilateria. Genesis 52, 458–70.

Appendix 2: Kenny*, Namigai*, Marlétaz, Hui, and Shimeld, 2015. Draft genome assemblies and predicted microRNA complements of the intertidal lophotrochozoans Patella vulgata (Mollusca, Patellogastropoda) and Spirobranchus (Pomatoceros) lamarcki (Annelida, Serpulida). Marine Genomics. *in press*.

Appendix 3: Kenny NJ, Namigai EKO, Dearden PK, Hui JHL, Grande C, Shimeld SM, 2014. The Lophotrochozoan TGF-β signalling cassette - diversification and conservation in a key signalling pathway. Int. J. Dev. Biol. 58, 533-49.

Namigai and Shimeld, 2016. Cytoplasmic flow in a dextral and a sinistral equal-cleaving spiralian is correlated with cell division. *in prep*.

Chapter 2

Molecular resources for studies of annelid development

Introduction

Currently, the available lophotrochozoan genomes include *Helobdella robusta* (leech), *Capitella teleta* (polychaete), *Lottia gigantea* (mollusc), and *Crassostrea gigas* and *Pinctada fucata* (oysters) (Simakov et al., 2013; Takeuchi et al., 2012; Zhang et al., 2012). Given that there are hundreds, if not thousands, of species within each phylum of the Lophotrochozoa, it is an understatement to say that relying on these genomes alone would be a gross misrepresentation. In terms of polychaete annelids, this is especially the case, as the single genome available cannot by itself reflect the diversity of polychaete species.

We have sequenced and assembled the genome of the polychaete *Pomatoceros lamarcki*, a sedentary polychaete annelid. Genome assemblers are constantly evolving and improving, and there is no hard and fast rule in choosing which assembler to use (this is also dependent on the genome being sequenced). In this case, several assemblers were trialled. The genome contains high levels of polymorphism and is a low quality assembly, however we were able to demonstrate that despite the draft-quality of the genome, molecular components such as microRNAs (miRNAs; Kenny, Namigai et al., *in press*; Appendix 2) and TGF-beta components (see Chapter 3; Kenny, Namigai et al., 2014; Appendix 3) can be reliably extracted and studied. The *P. lamarcki* genome has been made public through the publication of a miRNA case study, and this will be presented below.

Methods

Genomic DNA extraction

Sperm from a single adult *P. lamarcki* male was homogenized using an RNase-free polypropylene pellet pestle, washed three times with 2x PHB buffer (0.1M EDTA, 50M Tris, 2.5% SDS, in distilled water), and after addition of 0.015M NaCl and Proteinase K (0.6µg/ml) were digested overnight at 50°C. Genomic DNA (gDNA) was extracted by three phenol-chloroform extractions followed by one chloroform extraction with a 15-30 min. rotation. The samples were washed twice with chloroform before adding 0.1 volumes of sodium acetate and 2.5 volumes of 100% EtOH for incubation at -20°C overnight. gDNA was washed twice with 70% EtOH, air-dried at RT, and resuspended in 100µl of distilled water. gDNA was then prepared for sequencing by the Wellcome Trust Centre for Human Genetics (WTCHG, Oxford, UK).

Genome sequencing and assembly

Two libraries of insert sizes 201bp (base pairs) and 500bp were sequenced using a single lane of Illumina Hiseq2000 with 100bp paired end reads. Jellyfish was used to count k-mers for coverage calculation. The two libraries were then assembled together on a Unix platform using four assemblers: SOAPdenovo, SOAPdenovo2, ABySS, and Velvet (Table 2.3) using a range of k-mer sizes to find the optimal assembly (k-mer sizes 31-71). A blast database was then created in order to search for specific sequences by BLAST. Sequencing was funded by the Elizabeth Hannah Jenkinson Fund awarded to EKO Namigai, JHL Hui, and SM Shimeld.

miRNA identification

miRNA sequences were downloaded from miRbase and searched against the genome using BLASTN and checked for integrity (done in collaboration; see Kenny, Namigai et al., *in press*; Appendix 2). BLASTN was performed with parameters used in other miRNA studies (Quah et al., 2015), with the specifications word size=11, reward=5, penalty=4, gapopen=8 and gapextend=6, and checked for proper miRNA folding by RNAfold. In this thesis, small RNA libraries were not specifically sequenced and identification criteria pertaining to small RNA libraries (Quah et al., 2015; Tarver et al., 2012) were not used here.

miRNA expression

RNA was extracted from *P. lamarcki* whole adults in addition to only the head, abdomen, and trochophores separately, using the miRVana kit (Life Technologies). gDNA was removed by applying a DNase treatment directly to a Qiagen RNeasy column. cDNA was reverse transcribed, and negative controls replaced the reverse transcriptase enzyme with an equal quantity of solute. Primers (Table 2.1) were designed using the Primer3 program, v. 0.4.0 (http://frodo.wi.mit.edu). miRNAs were amplified by PCR and run on 2% agarose gels, and bands were excised and sequenced to confirm the identity of the miRNA. The latter procedure was done for this thesis multiple times with no success, likely due to the low integrity of the RNA extracted. The final successful amplification was done by others using a different sample of RNA (see Kenny, Namigai et al., *in press*; Appendix 2).

Table 2.1 miRNA primers. Red indicates those that were expressed in *P. lamarcki*.

miRNA	product size (bp)	forward primer	reverse primer
5sRNA	50	GCCATACCACGCTGAATACAC	GCTTAACTCCCGTGATTGGA
1957	52	GGGATGTAGCTCAGTGGTAGAG	ACCCCGGGCCTTCC
3533	60	CGAGACCACCTACAACAGCA	GCGTACAAGTCCTTACGGATG
1285	51	GGATAGCACCTGTGAATAGGC	CCAGCTATGTTGGACAGGCTA
7	51	GGAAGACTAGTGATTTTGTTGTTGG	TGATTTCTTGTTAATTTCATTCACG
8	67	GACATCTTCCTTGGCAGCA	GGCATCTTTACCTGACAGTATTACAA
1983	124	CAGCCCCAGTAGCTCAGTCG	AAAACAATTGCCCCAGGTGA
716	59	GGGGATAGTAGGAAATATTCAAACGAG	TGAAACCCTTCCCCACTTCA
3350	51	AAAAACAGATTTCTCATAAGTTGAACA	TGATTACTGAACCCAATTTTGA
1817	68	AATTTGAGGGCATCGAAAACC	TGTTTTCAGGATTATGTATTCATTGTG
1287	73	CGAAGATTCTAGAAAAGTGGTTCGAG	GCACCTATCACTGAATCTGTTGC

Results

Genome sequencing strategy

The *P. lamarcki* genome was previously estimated to be 1.23-1.50pg (Gregory, 2014). As one lane of sequencing yields approximately 30Gbp of sequence per lane, the strategy was to use a single lane for an estimated 20-25 fold coverage. Two fragment lengths of 180bp and 500bp were requested in order to maximize the number of sequences retrieved upon sequencing and assembly.

Genome sequencing and quality analysis

Sequencing was performed using Illumina/Solexa technology (Illumina Hiseq2000) with two libraries of fragments sizes 180bp and 500bp for paired-end sequencing of 100bp reads. Sequence quality of the raw data was assessed using FastQC analysis (Babraham Bioinformatics), which averages the quality of the sequence at each base (Figure 2.1). Quality was deemed sufficient to continue with assembly, with median quality scores being above 30 (considered very good quality) to nearly the 100th base pair. The quality tapers off at the very end of the read.

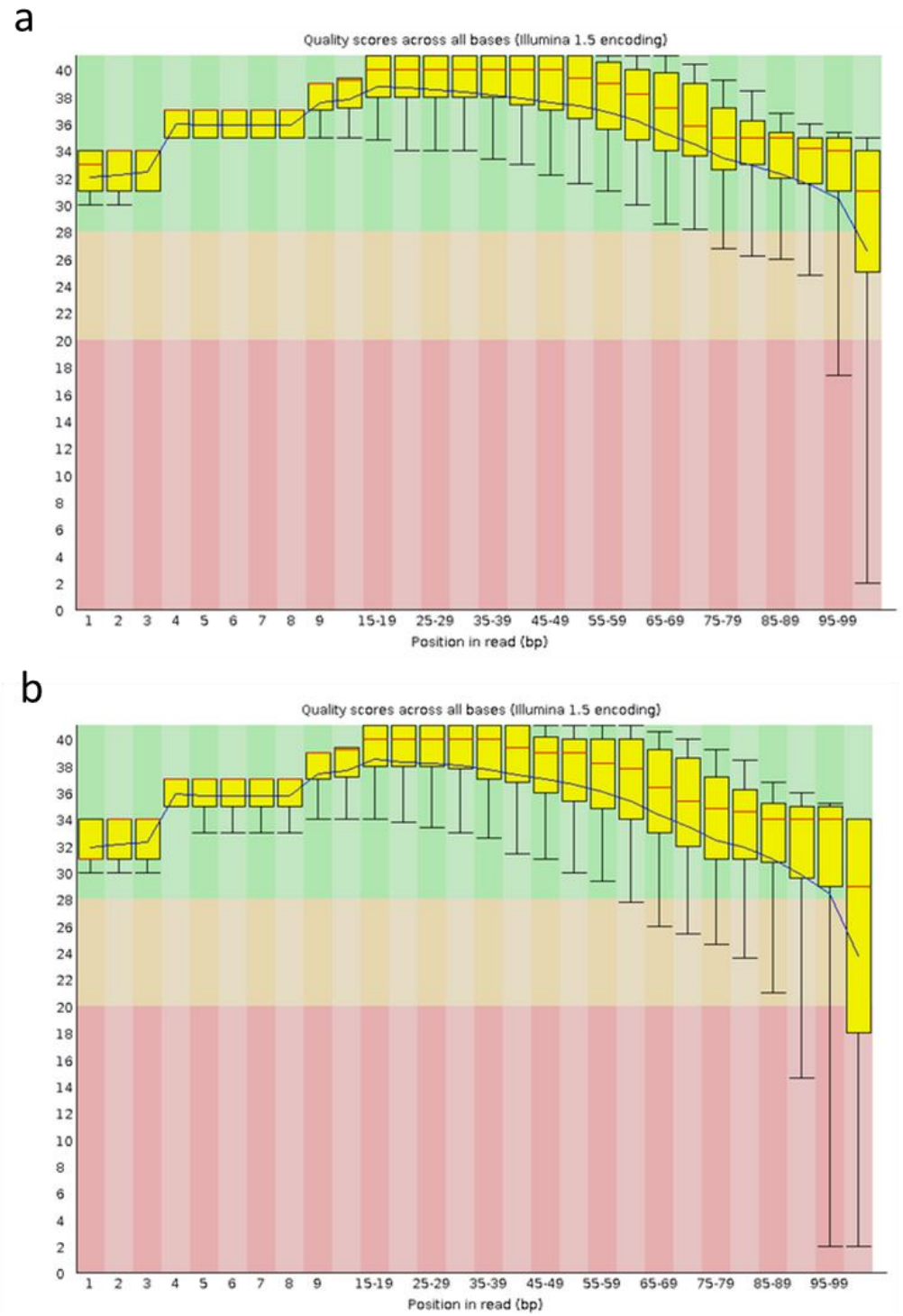


Figure 2.1 FastQC quality analysis of (A) 180bp and (B) 500bp DNA fragment libraries of the *P. lamarcki* genome, with read position (bp) against FastQC quality scores. FastQC scores are represented by: green (28-40) = high, orange (20-28) = intermediate, and red (0-20) = low. Yellow boxes represent the inter-quartile range (25-75%), upper and lower error bars represent the 10% and 90% points respectively, red lines in each yellow box represent the median value, and the single dark blue line running across the yellow boxes represents the mean quality.

Genome coverage and error

Genome coverage represents how much of the genome has been recovered in sequencing. Genome size of the raw reads was estimated by k-mer distribution, where coverage was plotted against k-mers counted using Jellyfish (Marçais and Kingsford, 2011) to find peaks of distribution. This was done first by counting 19-mer fragments (Table 2.2, Figure 2.2), where there was an overall 10x coverage. However a subsequent analysis found that counting 21-mer fragments was more reflective of coverage and genome size for this genome (done in collaboration; see Kenny, Namigai et al., *in press*; Appendix 2). In the 21-mer coverage distribution, there was a double peak (12x and 24x) in the k-mer distribution which when summed reflected a 1.25 Gbp genome size (see Kenny, Namigai et al., *in press*; Appendix 2). This means that 80% of the estimated genome was recovered upon assembly, where the loss of data can be attributed to misassembly of repetitive regions.

Table 2.2 *P. lamarcki* genome sequencing parameters and statistics.

library name	**insert size (bp)**	**read length (bp)**	**number of reads**	**inferred genome size (Gbp)**	**19-mer coverage peak**	**genome coverage**	**overall genome coverage**
274	500	100	67048915	1-2.26*	~4x	4.8x	10x
278	201	100	119500405		~8x	9.7x	

* Method 1: (tot # k-mers represented more than once) /8); 18134670942/8 = 2.26 Gbp
Method 2: total distinct k-mer; 978991876 = ~1

Error was corrected using the Quake program (Kelley et al., 2010). Quake uses the same general method as the EULER assembler (Pevzner et al., 2001). This k-mer distribution method counts the number of oligonucleotides of a certain size k (k-mer) in the reads, and assigns low coverage k-mers as error (about 1-fold coverage) and high coverage k-mers as true sequences. The reads categorized as error are then corrected by editing single bases until all reads are of high coverage (or discarded in the event that this is not possible), with the aim of using the minimum number of single base edits. Quake uses this k-mer distribution method, but incorporates read quality for both the identification of error and for the correct editing of the identified error. When identifying error, a model incorporating quality of reads is used to find the most accurate cut-off distinguishing low coverage reads (error) from high coverage reads. When correcting error, lower quality bases where error is more likely are edited preferentially over higher quality bases. These error-corrected libraries were used to assemble the genome.

Genome assembly

The *P. lamarcki* genome was assembled using two methods: ABySS (Assembled by Short Sequences; Simpson et al., 2009) and Short Oligonucleotide Analysis Package (SOAPdenovo and SOAPdenovo2; Luo et al., 2012). Each assembly program was trialled with different k-mers to find the optimal value, assessed by N50 where the N50 represents the length of scaffold at which 50% of the genome is contained in scaffolds either equal to or greater than that length (Table 2.3). The k-mer size of assembly by SOAPdenovo2

dictates the bp size used to map reads to contigs, and in this case the optimal k-mer size was 51. This optimum is not one that is optimal across genomes, and is largely dependent on the particular genome that is being assembled.

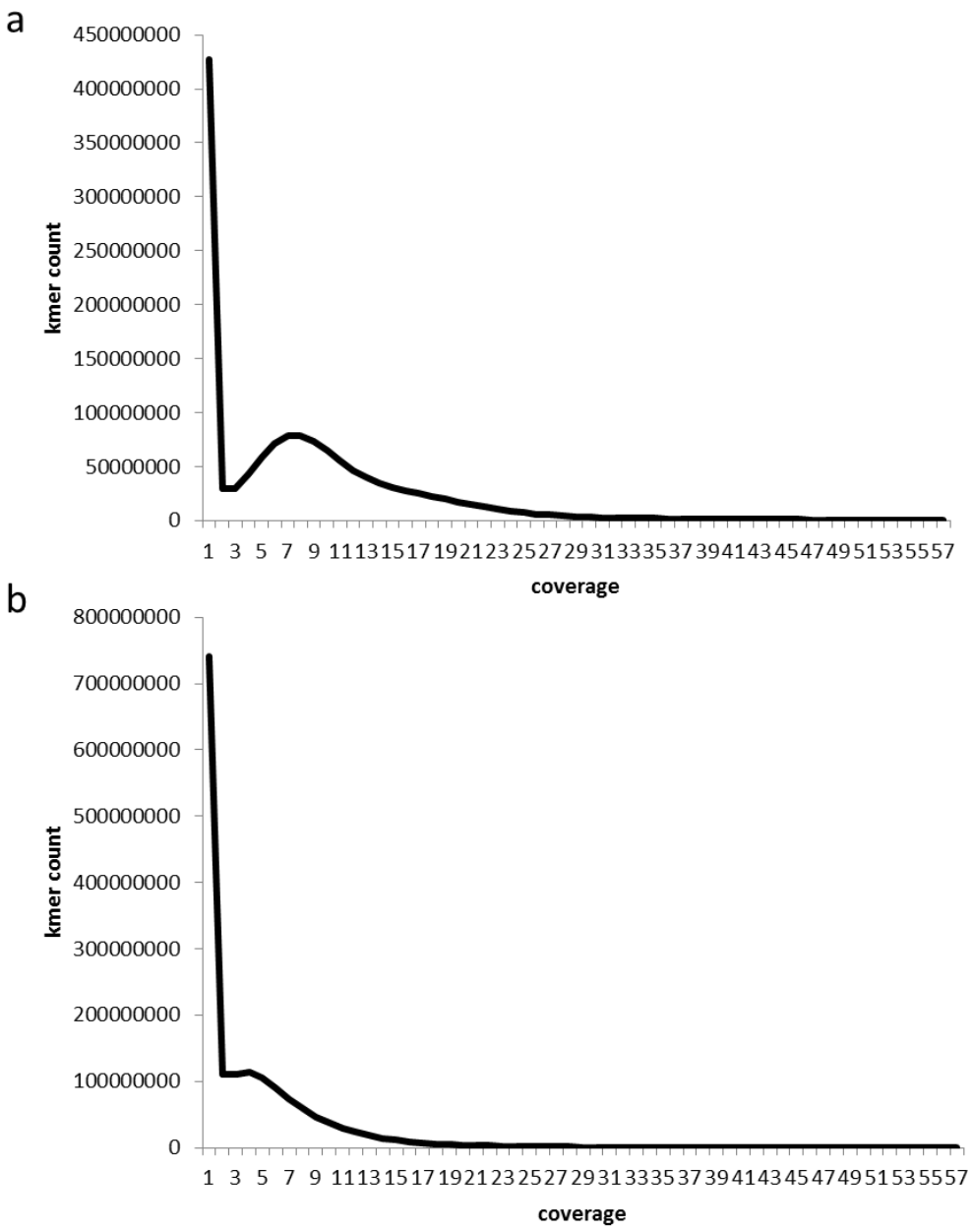


Figure 2.2 19-mer fragment k-mer distribution of *P. lamarcki* genome sequencing for (A) 180 bp and (B) 500 bp libraries. Peaks in the distribution (at approximately 8 in A, and 4.5 in B) indicate genome coverage of the particular library. Error is represented at approximately 1x coverage.

Table 2.3 Assembly parameters and resulting statistics used to find the optimal assembly of the *P. lamarcki* genome using: (A) ABySS (+Quake), (B) SOAPdenovo, and (C) SOAPdenovo2 (+Quake).

a

ABySS								
k-mer size		21	31	41	51	61	quake 21	quake 61
Statistics for contig lengths:	Min contig length:	21	31	41	51	61	21	61
	Max contig length:	12,180	34,118	35,568	42,219	37,876	12,059	42,797
	Mean contig length:	34.62	71.05	105.03	142.04	183.31	34.91	185.19
	Standard deviation of contig length:	30.69	176.32	259.62	335.55	388.5	30.78	396.78
	Median contig length:	26	42	60	79	103	26	104
	N50 contig length:	36	73	118	175	255	36	261
Statistics for numbers of contigs:	Number of contigs:	64,846,330	27,508,157	19,294,214	14,615,040	11,423,672	63,239,626	11,209,097
	Number of contigs >=1kb:	5,525	139,806	201,979	252,559	293,524	5,131	294,231
	Number of contigs in N50:	18,902,934	4,685,090	2,648,614	1,708,328	1,248,872	18,380,371	1,201,949

b

SOAPdenovo							
k-mer size		21	31	41	51	61	71
Statistics for contig lengths:	Min contig length:	100	100	100	100	100	100
	Max contig length:	7,160	14,438	15,565	16,368	15,499	11,148
	Mean contig length:	197.86	225.9	236.8	228.95	253.9	269.06
	Standard deviation of contig length:	146.03	237.87	270.44	248.19	223.82	179.01
	Median contig length:	148	150	152	150	182	234
	N50 contig length:	208	255	271	268	301	301
Statistics for numbers of contigs:	Number of contigs:	2,332,098	3,885,574	5,253,792	7,017,225	6,338,478	4,036,419
	Number of contigs >=1kb:	10,344	60,539	103,843	133,245	100,635	33,376
	Number of contigs in N50:	639,063	857,600	1,091,872	1,443,399	1,474,868	1,166,365

C

SOAPdenovo2							
k-mer size		21	61	quake 21	quake 41	quake 51	quake 61
Statistics for contig lengths:	Min contig length:	100	100	300	300	300	300
	Max contig length:	7,160	11,085	37,223	43,264	37,132	35,887
	Mean contig length:	197.41	247.89	697	646.27	1327.69	751
	Standard deviation of contig length:	145.41	183.79	990.32	1163.76	1439.37	881.79
	Median contig length:	148	200	351	220	898	394
	N50 contig length:	207	281	1,306	1,554	1,939	1,289
Statistics for numbers of contigs:	Number of contigs:	2,333,035	4,888,714	1,037,318	1,619,351	726,277	1,260,243
	Number of contigs >=1kb:	10,262	41,351	208,859	276,510	311,778	288,703
	Number of contigs in N50:	641,031	1,316,688	143,272	164,213	135,692	205,315

Assembly quality

Genome quality was assessed using the N50 as the deciding parameter. The N50 is calculated by ordering contig lengths from largest to smallest, finding the least number of contigs it takes to represent 50% of the genome. The length of the smallest of these contigs is the N50 (Salzberg & York, 2005).

For the purpose of this project, a crude genome assembly is sufficient, and genome quality was not rigorously assessed beyond this metric.

Statistics on optimal assembly

The quake-corrected SOAPdenovo2 assembly (k-mer=51) was selected as the optimal genome assembly based on the N50 statistic (Table 2.4). Sperm from a single animal yielded 964,274,156 bp of sequence upon assembly of the genome.

Table 2.4 Statistics on quake-corrected SOAPdenovo2 k-mer 51 assembly of the *P. lamarcki* genome.

Species/Metric (bp unless stated)	***P. lamarcki***
Number of 2*100bp Paired End Fragments (200bp library)	119,500,405
Number of 2*100bp Paired End Fragments (500bp library)	67,048,915
Min contig length	300
Max contig length	37,132
Mean contig length	1327.69
Std deviation of contig length	1439.37
Median contig length	848
N50 contig length	1,939
Number of contigs	726,277
Number of contigs >=1kb	311,778
Number of contigs in N50	135,692
Number of bases in all contigs	964,274,156
No. of bases in contigs >=1kb	726,465,383
GC Content of contigs (%)	27.97
N content of contigs (%)	0.53485
Genome size (Gbp)	1.25
Overall Coverage	15/30x
***k* mer Coverage (21mer)**	12/24x

Contigs greater than 1kb represent 43% of the total contigs, which suggests low contiguity. However, when looking at the number of bases in contigs greater than 1kb, it becomes apparent that 75.3% of the assembly is contained in these longer contigs. As an appreciable percentage of the assembly is contained in contigs that are 1kb or longer, this means that gene

identification, protein domain content, and other genomic information could plausibly be found in the genome. This is corroborated by the N50 statistic, which, although not ideal, indicates that there is enough contiguity in the genome for this type of analysis. This, of course, is not a conclusive diagnostic, as the integrity of these longer contigs is not known.

The proportion of N's (0.53%) indicates that there are gaps in the genome, where bases are replaced by N's to fill gaps during assembly of contigs into scaffolds. Such gaps are problematic, as this indicates a loss of information. Each genome comes with its own set of difficulties both in sequencing and assembly, making it difficult to pinpoint exactly why there are gaps, but usually they occur due to repetitive regions in the genome, secondary structures, or simply due to low sequence coverage (Tsai et al., 2010). At the sequencing level, some gap-inducing problems may have originated from the choice of DNA fragment size for sequencing. Two DNA fragment sizes (180bp and 500bp) were chosen, providing the possibility to take advantage of the overlap of these two fragments to maximise the sequencing of non-repetitive sequences. It is not clear whether sequencing larger DNA fragments would have prevented gaps, however it does remain a possibility.

The GC content is relatively low (28%, Table 2.4), where GC content above 50% is considered high (Takeuchi et al., 2012). This shows that this genome is AT-rich. What a GC-poor or AT-rich genome can mean is not clear, and has been a point of debate. GC content has been associated with mutation rates of CpG islands (Fryxell & Moon, 2004), fixation bias or GC-fixation bias (Katzman et al., 2011), and a range of other correlations (Romiguier et al., 2010). Low GC content does not mean that this is homogenous across the

genome, and it is thought that GC-poor and GC-rich regions can be concentrated at specific parts of the genome. Conversely, any correlations drawn may be erroneous due to GC bias during sequencing. This is a known problem for Illumina sequencing, where GC bias confounds read mapping, and is a problem that is just beginning to be addressed (Benjamini & Speed, 2012; Ross et al., 2013; Chen et al., 2013).

miRNA case study

miRNAs are short RNA sequences (approximately 50bp) that function to transcriptionally regulate gene expression. They were discovered in animals relatively recently (Pasquinelli et al., 2000), and are potentially useful characters to infer phylogeny as they are thought to be slow-evolving and rarely lost. A problem has been biased sampling across the animal kingdom, and this was addressed using the *P. lamarcki* genome. This was an opportunity to test the utility of the genome to find biologically-relevant information, and as miRNAs are short sequences, the low contiguity of the genome should not be significant. miRNA sequences were searched for in the *P. lamarcki* genome, and 66 putative miRNAs were found (Table 2.5).

A handful were selected for further analyses (Table 2.6) and confirmed for hairpin formation by RNAfold (Gruber et al., 2008). These particular miRNAs were selected to capture miRNAs that were either known to be present in these animals (as a positive control), and miRNAs that had previously been thought to be specific to a different lineage, together with 5SrRNA expression as a control.

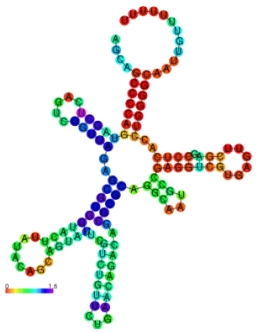
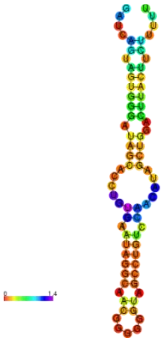
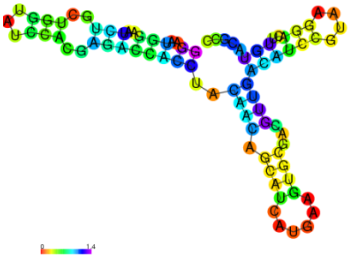
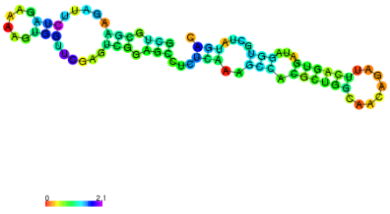
Table 2.5 miRNA sequences found in the *P. lamarcki* genome.

miRNA hit	scaffold	length of hit (bp)	e-value	hit sequence (portion of scaffold hit by BLASTn)
2	scaffold9196	74	7.00E-03	GCTGGTTCTGTCGAAGAGGTTGTGAAATTGCTCTTCCTTTAAC--CATATCACAGCCTGCTTTGATAAGATCAG
7	scaffold120882	66	6.00E-06	GTTTGGAAGACTAGTGATTTTGTTGTTGGACGTGAATGAAATTAACAAGAAATCACTA-TCTTCTT
8	scaffold308814	77	1.00E-05	GGT-TCTGAGGACATCTTCCTTGGCAGCATTACATATGTTGACTGAAGTGTTTGTAATACTGTCAGGTAAAGATGCC
9	scaffold248225	72	2.00E-04	CCATTTTTTCTTTGGTTATCTAGCTGTATGATTAA-AGTTGATGTATCATAAAGCTAGGTTACCAAAGCTAA
29	C1453782	76	3.00E-04	CTTGAGGTCGCTGTTTTCTTTTGG-GGCGTAGATTCGCGTGATTCCCTGTCTAGCACCATTTGAAATCAGTTTTCT
33	scaffold571803	65	0.029	CAGTGGTGCATTGTAGTTGCATTGCATGTATCATAATGAACTCGTGCAGTGCTCCTGCCGGCAAC
71	scaffold552498	61	3.90E-02	TTGAAAGACATGGGTAGTGAGATGTGT---TAAAAAAAGTCACCTTGCTACCATGTCCTTC
87	scaffold515760	71	9.00E-03	TTATCGTGAGTCGCATCTGAATTTTTGTCTCAAGCCTTCAAATAATACGAGGTGAGCAAAGTTTCAGATGT
96	scaffold464026	73	7.00E-04	ACTTGGCACTGGCGGAATAATCACTTATTGTGTTAGCACCAGTGATGTTCCCTTGGTGCCTAATAT-ATCATC
100	scaffold285706	99	2.00E-07	ATATATTCCAAGTTGCACTAACCCGTACAACCGAACTTGTGCCGTTTTATCGACTCGAGAGAACGCAAGCTCGCTTTTACGGGATAGTGTTGCTTTGGG
124	scaffold440006	82	8.00E-04	CAAAATGCTTTGCCTAT--GTGTTCACTTCGTCAGCCTTGGTGGTGGAAAATTGAAACAATTAAGGCACGCGGTGAATGCCA
125	scaffold346750	63	0.092	TCCCTGAGACCCTAACATGTGA--TGACATTCGAGGTTCACACGTTACGATCTCGGGATTAGG
137	C2798015	73	4.00E-09	GCTGTCTTGATTACGGGTATTCTTGGGTTAATAATATACCGGCCAGGATATTATTGCTTGAGAATACACGTAG
153	C6475418	84	5.00E-08	GGAATGGAATCTGCTGGTATCCACGAGACCACCTACAACAGCATCATGAAGTGCGACGTTGACATCCGTAAGGACTTGTACGCC
184	scaffold171685	73	0.014	TCGTTTCCTTTTCACTTCT-TCGCCCCGTCTAC-CGTACTCACGAGCTGGACGGAGGACTGATAAGGGTTGGT
215	scaffold243684	77	0.009	ACCGATCAAAGGACCGACATTAATTTTCAAGTTTGTCTGTC-TGTCTGTCTGACACTCCGTCTGTCCGTCTGT-CTG
219	scaffold216005	137	2.00E-04	AATCAATATCGGGATCTGTCATTACATCATTATTCTATACATACATTTTCAATTGTTTACATTGATCATTGATCAAGATTTATTGATATTATATTCCCTAAATGTGTT--TTCATATCATATGATAGCAAATCAAAC
277	scaffold525653	114	0.009	ACATGCTATTTCTG--TTGTAACCTAGTTACAACGAGTATCCTT-T-ATAGATGGTTTTCATCTGTTTTTGTTTTGTTGTATGTGTGGTCAATGTAATACAAATCAGTTTTGTT
307	scaffold413536	104	8.00E-03	TAATTGATACCAAATAGAATGACATCAAAATATAAATTAGAAAAAAATATCCCA-AAAGACCATGATAATAAAATAACTTATATCAAAATATATATGATTTAAG
315	scaffold558784	59	0.014	AGTTTTGATTGTTGCTCAGAAAGCCGTAAAT-TCAAC---TTGGCTTTCGAGTAATAAT
466	C1177640	74	1.00E-04	TGTGTATATGTGTATATATGTGTGTATGTA--TGTGTGTATTTGTATATATATATATATATATATATACATA
	C2372257	71	0.001	TATTTGTATGTGTGTGTGTGCATATATATATGTGAATATACAGATATGTATATATATATACGCGTATAT
	scaffold119761	55	8.00E-04	TGTTGATATAT-TTATGTATGTATGTGTGTATGTGTGCGTGTTGTGTATACGTAT
467	scaffold593976	72	0.004	GTACGCACATGTATATGCGCACACATTCGAAACACATATATACATTCCCAGGCACACACATACACACACACA
669	scaffold144438	60	0.001	ATGTGTGTATTTGTGTATGTATTTGTGTGTATGTGTATATGTGTATATGAGTATATATAA
	C4033725	88	0.002	TGTGTGTATGTGTGTGTGCGCGTGTGTGTGTGTGCGTGTGTTTTTAAAATACATACACACACACACATACAAAACACACACACACACA
	scaffold588745	69	0.002	TACATATACATACGTGTAAACAACTAAACATAA-ATACACACACATATATAAACCCAAGCACACAAACA

	C3804321	83	0.002	CATACATACATATGTGTGTATGTATGTATGTGTGTGTACAT ACATACACACATACATACACACATACATCCATACACACATA C
705	scaffold340718	82	6.00E-08	GCATGAGGTGCGACGTGGGGTGGTTGCCGAAATCTGAGC TGACATAGCCGACGCGCAGGCGCCCGTTGCTCGCCGCC AGGTC
716	scaffold492884	60	5.00E-05	GGGGATAGTAGGAAATATTCAAACGAGAGTTTTGAAGAAT GAAGTGGGGAAGGGTTTCAT
767	scaffold472580	91	3.00E-03	AAAGATTTCTGCTCATGCACCAAGATT-TCTGGTCATGCACCAAGATA-TCTGGTCATGCACCAAGATAGCTGGCCATGCACCAAGATT TC
982	scaffold407224	107	0.004	GTTAATAAATGCTAAAATCATAAAAGTTTTT--ATATTTA-CAAGTAGATTAATGTTC-ACTTCCATTTATCTGTTATTTCTATTTGTGCGTACTTGTTTC AC-GCT
1285	scaffold330682	76	1.00E-03	GATCAGTAGTGGGATAGCACCTGTGAATAGGCAACGGGG GGTAGCCTGTCCAACATAGCTGGACTTACTTCTTTTT
1287	C1734625	83	1.00E-07	GCTGCGAAGATTCTAGAAAAGTGGTTCGAGTCGGAGCCT CTCAAAGCCACGCTGGCAACAGATTCAGTGATAGGTGCT ATGAC
1497	C2055745	82	0.003	ACCAAATAGTGCCACTATCTTTTAAATCTGCAGTTATCTTC TCAAAAAATAAAAATTAGATAAAGTTTTGAAGAATTACAGG
1817	C5299747	96	0.006	AATTTGAGGGCATCGAA-AACCTCTCTATCATGGGTTTTAAACACAATGAATACATAAT CCTGAAAACAATT--TTAATGTATGTCAGAAAAGTTT
1957	scaffold542862	65	0.059	GGGATGTAGCTCAGTGGTAGAGCGCCCGCTTTGCATGCG GAAGGCCCGGGGTTCGATCCCCCGCA
1983	scaffold327473	130	5.00E-08	AGCA-GCCCCAGTAGCTCAGTCGGTTAGAGCGCGGTACTTATAC AGCAGTATTCGTCTGTTCTGAACAGACAGCGCGCAGGCA ATGCCGAGGTCGTGAGTTCGAGCCTCACCTGGGGCAATT GTTTTTTT
1992	scaffold191985	53	1.00E-02	ACGGTTGCTAGTAGTTTTAAT-TCAACCTATCAGCAGTTGTACCACTGATGTG
1993	scaffold91460	64	7.00E-04	TTTCGTAAATTTCGACATAATGCTGTATAATTGGACCAGTA TTATGCTGTTATTCACGAGATGA
1996	scaffold318934	62	3.00E-03	GTCTCAAGTGAGGTCAGTGCTGCTGTGGCAAGCTAACAA GGGGTGCTGCACTCAGTTGAAAC
2000	scaffold448983	78	3.00E-04	GCATGACATAACTAGAAGTTGGGTAGGCTTTAAGTGAAAT ACCATTTTAAAGTCTTCACTACTTTTAGTTTCGAAATG
2154	C6245532	94	0.006	AGTGAATATTAGAATTTCAGATAGAAAATATTACTTGTGTC TGCTTAAATGATTTT-TTAACTGCATATCTTTAAACATTGTTTATTTATTTAT
2476	scaffold231526	71	0.002	TCCCATGTGGTCTAGTGGCTAGGATTCCGCGCTTTCACCT CGGCGACCGGGGTTCAATTCCCCGCATGGGA
2693	scaffold86750	61	0.004	AGAAATTGTATACTAGTCTAGTTTATTCAAGTGTGTAGCTG CTAGTCTAGTGTACAGTTCA
2702	scaffold275307	76	1.00E-04	GCGCGCTATATTTCTACCACCTGAATG-TCCGCACCGTAACTACTCGTTCAGGTAATAGAAATGTACA CTGCCTGT
2986	scaffold537149	135	2.00E-05	TCATCTTGCGA--ATAAACTGGGAATCATCTTGGGAATCATCTGGGTGTCATC TTGGGATGATCTGGGTATCATCTGGGGATCATCTGGGGA TCATCTGAGTATCATCTGGTGA---TCATCTGGGGATCATCTG
3212	scaffold267616	112	0.003	TTCATGATCATG-ATTTTGTCTTTGAAATAGATTACTTAGAAAACCATAAATTGT GAAATTTTATTGTGATAAATTTAACTATGTTTTCAATAGTTC ACAACAAGTATTATT
3259	scaffold232075	99	0.009	ACAAATTCTAGAATTTAATACTATTAACAACATA--TTCTTGAATTGA-TATATATTGACAAGAAATTCTTGAATTTAGATATATTGACAA CAAATTCT
3350	scaffold122579	128	0.002	AAAGATGAAATAAAAAATAAAAAACAGATTTCTCATAAGTT GAACATGTCA-AAATTG--GGT--TCAGTAATCATAATTTATTACATGAAATCTAATTG--CATTACAAACGTTTATGCAATTTGCA
3362	scaffold388455	96	2.00E-04	AGGCAGTTTGCGATTTTAATGGTAAATTCACCAATATTGTA GCGAAATGGCCTGGTTCAGCGGCTGATTCTAGAATCCTT GATGATTCTTTGTTAA
3367	scaffold558412	86	0.007	AAACTTAATTATATTTTGTCCACTTCCCCTTATCAAAGTAG AAATCATTGAATATGTTAGTAAACAGTTTGTGTAAGGTAAT TCTG

3533	C6475418	84	5.00E-08	GGAATGGAATCTGCTGGTATCCACGAGACCACCTACAAC AGCATCATGAAGTGCGACGTTGACATCCGTAAGGACTTGT ACGCC
3676	scaffold211465	60	3.00E-07	AAAGCGCCTGTCTAGTAAACAGGAGATCCTGGGTTAGAAT CCCAGTAGGGCCTCCTTTTT
3716	scaffold12898	118	0.001	TATTATATTTCTTTGTTAATTATTCTTTCAACAAATTATAATT ATAAGATATTGATATGAATACGTATTAAACAGTTCTAATTTA AATCACATACACATAAAATAAACCAAATAGCTT
3724	C3934081	108	0.007	CTGGAAATTATAAATCAATATATAAATTCACGCTTAAAATG AATTATGAATCATAAAGGAAGAATAGCGAAATGTTAAATCA CAAATGTTTTTCACAATGTTTAATTT
3737	scaffold580568	82	0.004	GATATGTTTGTATGTATAT--ATATGGATAACAATTTTATATAAATGACATGTAAAATATTT-ATTGTTTAGAATATATTTT
3756	scaffold318691	96	3.00E-04	TCTGGTATCTATCA-ACGGA-AATATCATAAATAATATTGCTATATTGTTCCTTTAATTAATA CAATTAAATGTT-AAATGCTAAATATCACAGAA
3757	scaffold112923	130	7.00E-03	ATTTGGATATTAGAAATAGCAAGGCATATCTAAGCTTTAAT ATGTCACCTTATAAAAAAATTGGAATATTGATTAGAAAAAT GTGTACATTAGAATAATTGTTTTGGGGACGTTAATAGCTTA TAAGCAT
3760	scaffold184330	136	1.00E-04	AAACAAAGTTTTTTTTAATTATGTAAAATAATATCAATTATT AGAATACACATGCCACTAAAAATGATAGAATGTAAAATTGA CATTGAAATTTGACTTTCACCATAATGATATTATTGTTAAAC AAAATAAGTAT
3767	C5549604	79	6.00E-04	TGAACTATAAGAATTAATATT--TTATGTTTAAAATGTCCAAATTATTTAAATGTGATATCGAC GGTTTATTTGCTTCT
3901	scaffold101370	115	0.001	CTTTATGATTTTATGAGTTAATAATGACCAATTAAAATGAT GTTTTATATACTTTTTGGTGGCTTTGTGATTAACTTTTTTTA CGATTCATTGATTTGTTTATCAATAACTTTTT
3902	scaffold260348	100	1.00E-05	ATATCAGCATGGTTATATACTAGACATGTATCAGCATGGTT ATATACTAGACGTATATCAGCATGGTTATATACTAGACATG CATCAGCATGGTTATAGA
4125	scaffold423358	80	4.00E-03	ATTTTGCTCA-ACAATTTTTATTTAAAATAAAGTTTTATATTTTTCTATGTTTA AAACATTGTTTTATATTTTTTTTGTT
4426	scaffold21646	58	2.00E-04	GAAGATGGACGCACATTGAGCGACTACAACATCCAGAAG GAGTCTACCCTGCATCTTG
4683	scaffold212053	79	4.00E-04	AAACCCAGGATGAGGCGGGCCTCGAGGTGCACCAATTCT TCCCACTCGTGAAAGTGAATTGCTCGCCGTATTT-ACGTT
4948	scaffold227887	68	7.00E-03	GGCGTAGGCGGGCGCGGGTGTGGCGTGCGTGACGCGC GCTGTCGTCACCAGACCAGTG-GCTTTACCT
4958	scaffold387673	52	5.00E-03	GCTCTGGTTCCGGATCGGGATCTGGATCAGGATCTGGTT CGGGATCAGGATC
5109	scaffold392066	52	0.002	CAACGACCA-ACTTAGAAATGGTGCGGACCAGGGGAATCCGACTGTCTA ATT
bantam	C996355	69	5.00E-05	ATTGGTTTTCATGCTGGTTTCCCAGATATTATACGAAGTTC TGAGATCATTGTGAAAACTAATTATGTG
let-7	scaffold285706	83	1.00E-08	AGGTGAGGTAGTAGGTTGTATAGTTAGAGTTACACCACTA TTTCAGGCGAGCTGTACAACCTTCTAGCTTTCCAGATGTC ACA

Table 2.6 miRNAs expressed in *P. lamarcki*.

miRNA hit	*P. lamarcki* scaffold	length of hit (bp)	e-value	hairpin (RNAfold)
1983	scaffold327473	130	5.00E-08	
1285	scaffold330682	76	1.00E-03	
3533	C6475418	84	5.00E-08	
1287	C1734625	83	1.00E-07	

Expression analyses showed that all of these miRNAs were expressed in *P. lamarcki* adult tissue (Figure 2.3; done in collaboration: see Kenny, Namigai et al., *in press*; Appendix 2). In order to confirm that the miRNAs expressed in *P. lamarcki* were not products of contamination, *P. lamarcki* genomic and transcriptomic sequences were searched for in the human genome, as this was the most likely source of contamination at the WTCHG (Table 2.7). There were no sequences of significant similarity in the human genome, indicating that the miRNAs found were not the result of human DNA contamination. In the case of miR-3533, genomic and transcriptomic sequences hit actin genes. However this particular miRNA is found within an actin-coding region in chicken and cow (see Kenny, Namigai et al., *in press*; Appendix 2), suggesting that this is a real characteristic of this miRNA.

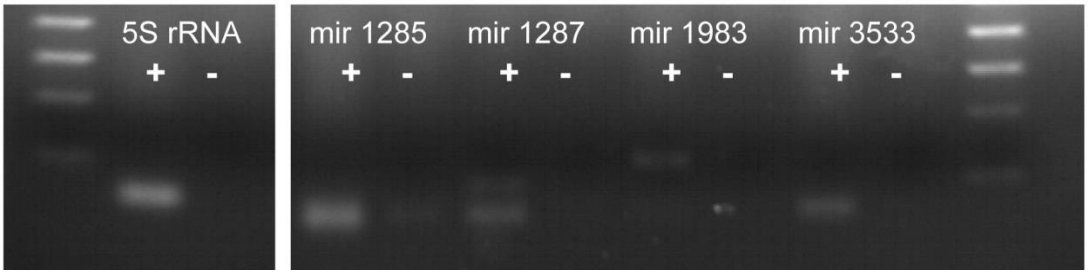


Figure 2.3 miRNA expression in *P. lamarcki*. 5S ribosomal RNA (rRNA) is a positive control, "+" indicates with reverse transcriptase enzyme, "-" indicates negative control without reverse transcriptase enzyme.

Table 2.7 Contamination check for miRNAs studied for expression in *P. lamarcki*.

mirna	scaffold	blastn (genome) against human genome	blastn (transcriptome) against human genome
miR-1285	scaffold330682	no significant similarity found	no significant similarity found
miR-1287	C1734625	no significant similarity found	no significant similarity found
miR-1983	scaffold327473	no significant similarity found	no significant similarity found
miR-3533	C6475418	hits actin	hits actin

miRNAs retrievable but there are caveats

miRNAs were recovered that were found in other lophotrochozoan genomes, miR-1983 and miR-3533, which served as a positive control for the presence of lophotrochozoan miRNAs (Figure 2.3). Interestingly, miRNAs that were previously thought to be vertebrate-specific were found to be present and expressed in *P. lamarcki*, such as those presented here: miR-1285 and miR-1287. These will be further discussed below.

Discussion

Genome size

The Animal Genome Size Database (www.genomesize.com) predicts a genome size of 1.2-1.5 Gbp for *P. lamarcki* and the genome size we have (1.25 Gbp) falls within this range. When compared to other annelid genomes, the *P. lamarcki* genome is of an intermediate size, as annelid genome sizes vary drastically. For example, the polychaete *Dinophilus gyrociliatus* genome is very small at 0.07pg (Soldi et al., 1994) or 0.068 Gbp (at 0.978 Gbp/pg) and the aquatic oligochaete *Spirosperma ferox* genome is very large at 7.64pg (Gregory and Hebert, 2002) or 7.47 Gbp (at 0.978 Gbp/pg). In comparison to other lophotrochozoan genome sequences, the *P. lamarcki* genome is nearly 4 times larger than the two other annelid genomes sequenced so far (Table 2.8). It is possible that the larger size of the *P. lamarcki* genome is correlated with the presence of repetitive regions in the genome, as seems to be the case at least for the *Pinctada fuctata* genome (Takeuchi et al., 2012).

Table 2.8 Summary statistics for select animal genomes. Statistics were taken from NCBI; Aparicio et al., 2002; Consortium, 2001, 2004, 2009, 2006, 2008, 2007; Dehal et al., 2002; Simakov et al., 2013; Takeuchi et al., 2012; Zhang et al., 2012. In cases where there were multiple assemblies, the median value reported at NCBI or the published value was selected. In some cases there are small inconsistencies between sources (usually over-estimation), however these were not rigorously corrected for the purpose of this discussion.

Superphylum	Phylum	Species	genome size (Gbp)	N50 (bp)	GC content (%)
Lophotrochozoa	Mollusca	Lottia gigantea (mollusc)	0.348	1,870,000	33
		Crassostrea gigas (mollusc)	0.559	19,400	35.3
		Pinctada fuctata (mollusc)	1.15	1,600	34
	Annelida	Helobdella robusta (annelid)	0.288	306,000	33
		Capitella teleta (annelid)	0.324	190,000	40
		Pomatoceros lamarcki (annelid, this thesis)	**1.25**	**1,939**	**27.97**
Ecdysozoa	Nematoida	Caenorhabditis elegans (roundworm)	0.097	17,493,829	34
	Panarthropoda	Drosophila melanogaster (fly)	0.18	21,485,538	41.93
		Tribolium castaneum (beetle)	1.52	43,511	33
		Apis mellifera (honeybee)	1.8	45,688	34.1
Deuterostomia	Vertebrata	Mus musculus (mouse)	2.66	32,273,079	41.94
		Gallus gallus (chicken)	1.05	279,750	41.93
		Homo sapiens (human)	3.23	56,413,054	38
		Takifugu rupribes (pufferfish)	0.365	52,883	45.84
	Urochordata	Ciona intestinalis (sea squirt)	0.160	37,096	36.06
	Echinodermata	Stronylocentrotus purpuratus (sea urchin)	0.814	16,785	35

Genome size becomes important when calculating genome coverage. Genome coverage is calculated by taking the total number of bases sequenced and dividing this by the genome size, and is a metric that estimates how much of the genome has been covered by sequencing. This metric is one of the primary statistics used when ascertaining the quality of a sequenced genome, yet it changes as assemblies are improved due to changes to estimates of genome size. For example, the genome of the nematode *Trichinella spiralis* was thought to be 270Mb in size, but upon sequencing is now thought to be

approximately 65Mb (Parra et al., 2009), which would have a large effect on subsequent genome coverage estimates.

Genome quality assessment parameters

Aside from genome coverage, another popular metric is the N50 statistic, where the N50 estimates the contiguity of the genome. The N50 is the length of contig for which half of the genome is assembled in contigs that are of that length or longer. The low N50 of the *P. lamarcki* genome indicates that assembly did not achieve long contigs, which means that the assembly is not very contiguous and may not be fully representative of the genome. This problem is not an isolated case: the lack of long contigs has been a constant underlying problem in *de novo* genome assemblies. The N50 metric is among a number of parameters that are based on contig size, and the heavy dependence on the N50 for quality assessment has been drawn into question (Earl et al., 2011). The N50 only indicates how continuously the genome was assembled, which can be correlated with assembly quality but is not a direct readout of assembly quality. For example, in some cases, the presence of longer contigs may indicate that contigs are being forced together erroneously during assembly. Despite these potential shortfalls, the N50 metric has been largely successful in predicting the success of genome assembly (Table 2.8). The N50 for *P. lamarcki* is low in comparison to other Lophotrochozoa, and indeed most other animals for which a “good” genome has been sequenced (Table 2.8). Interestingly, a genome with a similarly low N50 value, *P. fuctata* (pearl oyster), has a similarly fragmented genome (further addressed in the next

section). It is possible that inaccuracy increases when sequencing efforts are *de novo* with no reference genome for mapping and a lack of multiple sequencing efforts.

The N50 metric is not necessarily the ideal metric to assess genome quality, and does not inform on the completeness of the complement of genes in the sequenced part of the genome. To assess completeness of the genome, a CEGMA (Core Eukaryotic Genes Mapping Approach; Parra et al., 2007) analysis was done (in Kenny, Namigai et al., *in press*; Appendix 2), where the genome assembly was evaluated against a compilation of core eukaryotic proteins. The *P. lamarcki* genome is abnormally incomplete using this method, with only 2.42% (6/248) complete sequences and 12.50% (31/248) partial sequences found in the genome. This very low completeness of the *P. lamarcki* genome is a doubtful result, as the full complement of TGF-beta pathway members had been found in the genome (see Kenny, Namigai et al., 2014; Appendix 3). A BUSCO analysis (Benchmarking Universal Single-Copy Orthologs; Simão et al., 2015), a newer analysis that improves upon CEGMA, was done in order to confirm the CEGMA results. This yields a similar result, with 5% recovery against the *P. lamarcki* genome where of 843 metazoan orthologues, 12 complete, 0 duplicated, 27 fragmented, and 804 missing genes are reported.

It would seem that the CEGMA and BUSCO analyses reflect the low contiguity already reported by the N50 statistic and not the coverage of >10x. However, it could also be the case that the short contig length has inhibited the programs from detecting orthologous sequences in the genome. To see which was the case, the list of orthologous sequences used in BUSCO analyses were

directly searched for in the *P. lamarcki* genome using tBLASTn (Altschul et al., 1990). This markedly improved the statistics, where 84.1% (709/843) of the genes are found in the genome (although this is now an overestimation as there are false positives such as proteins with shared domains). In addition, alignment with previously published transcriptome data for *P. lamarcki* (Kenny and Shimeld, 2012) using the STAR alignment program (Dobin et al., 2013) indicates that 33.66% of the RNA-seq reads map to the genome (see Kenny, Namigai et al., *in press*; Appendix 2). Although still low, these alternative analyses indicate that it is likely low contiguity that has confounded the N50, CEGMA, and BUSCO metrics, and that more of the genome has been successfully recovered than these metrics would suggest.

The problem of metric readout has been approached independently using a variety of methods. For example, in the *P. fuctata* genome, which also has a low N50, the N50 itself became less of a focus. Instead, gene models were built using the gene prediction program AUGUSTUS (Stanke and Morgenstern, 2005) with previously compiled EST (expressed sequence tag) data, which yielded a 70% coincidence between the gene models constructed and the EST sequences (Takeuchi et al., 2012). Although this is not yet an urgent problem, and is not a major problem for genomes that are backed with multiple and deep sequencing efforts, this becomes important for the increasing number of *de novo* assemblies, where there is usually no reference genome available for mapping and where gap-containing scaffolds may erroneously report incomplete genome assemblies when they are just fragmented and actually more complete than appreciated.

Polymorphism in animal genomes

The aforementioned low contiguity of the *P. lamarcki* genome, as reflected by the low N50 number, indicates that there are relatively few long contigs that were successfully assembled. This suggests that the genome is highly fragmented, which in turn likely causes the difficulty in using the CEGMA and BUSCO pipelines. The fragmented state of the genome also suggests that there may be a high level of polymorphism in the *P. lamarcki* genome. This has been encountered before in other marine animals. For example, the *P. fuctata* genome exhibits allelic polymorphism, where there appears a double peak in a plot of coverage against contig length (Takeuchi et al., 2012). This is reminiscent of results in the *P. lamarcki* genome, where a plot of k-mer distribution against coverage has a double peak (Figure 2.4). This indicates that the genome is highly heterozygous (Kajitani et al., 2014).

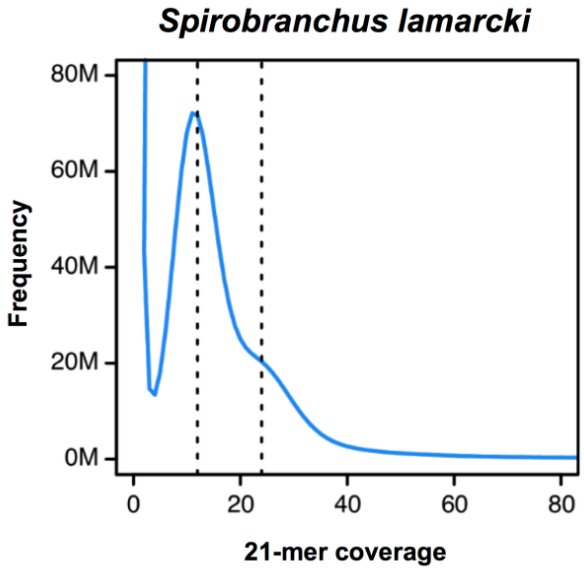


Figure 2.4 21-mer fragment k-mer distribution of *P. lamarcki* genome sequencing (done in collaboration: see Kenny, Namigai et al., *in press*; Appendix 2).

Allelic polymorphism is also seen in a more distantly-related marine animal, the urochordate *Ciona intestinalis*, where allelic polymorphism was also observed in the form of single-nucleotide polymorphisms (SNPs), insertions, and deletions distributed non-uniformly in the genome (Dehal et al., 2002) and the cephalochordate *Branchiostoma floridae* (amphioxus/lancelet) (Putnam et al., 2008). The level of polymorphism in these genomes is considered high, as in most of these cases precautions had been taken to minimize polymorphism by taking sperm from a single individual (as was done for *P. lamarcki*), which minimizes the reflection of maternal and paternal haplotypes. These animals are united by a common mode of fertilization, where fertilization takes place externally with sperm and egg being released freely. The large effective population size that results from this method of breeding has been suggested to account for the high level of allelic variation in these animals (Romiguier et al., 2014). While there is no way to functionally confirm this, observations do seem to support this theory.

AT-rich genomes and their significance in bilateral animals

The *P. lamarcki* genome is AT-rich, and among the animals sequenced so far, has the lowest GC content, and the highest AT content, that has been seen. As previously mentioned, this is probably partly an effect of using Illumina sequencing, where GC bias is known to confound read mapping. However, despite low coverage of the genome, it is possible that the GC content information is relevant as has been seen in the low coverage genomes of two bird species (Card et al., 2014). Low GC content, and consequently high AT

content, has been seen in other organisms to varying degrees (Table 2.8). There are many theories as to what low or high GC content can signify, and genome studies have largely mentioned possibilities such as GC fixation bias (Katzman et al., 2011), an association with mutation rates of CpG islands (Fryxell and Moon, 2005), a correlation between low GC content (high AT) and longer chromosomes (Consortium, 2004), and the association of GC-rich characteristics of intron-exon structures in a number of deuterostome species (Zhu et al., 2009). In addition, genomes have been thought to undergo GC-biased gene conversion, where there is a bias during DNA repair that prefers G and C over A and T in regions with high recombination (Eyre-Walker, 1993). The significance of GC content has been best studied in mammalian genomes, where it has been correlated with chromosome length (Romiguier et al., 2010).

Notably, all of these studies seek to understand a trend of increasing GC content in genomes. Most likely, all of these proposals are relevant to some degree, yet this indirectly implies that GC poor, or AT-rich, genomes do not undergo the aforementioned mechanisms, or to a lesser degree. It is possible that the above mechanisms that have been proposed to explain high GC content may be more relaxed or not functioning to the same degree in GC-poor, or AT-rich, genomes such as *P. lamarcki*. For example, there could be less DNA repair that is GC-biased in AT-rich animals. Alternatively, it is possible that all previous considerations of GC content has focused too much on understanding how GC can be increased, where perhaps more attention should be given to the trend of decreasing AT content. The latter is very possible and merits investigation. Disentangling the functional role of high GC or high AT content would require understanding how this bias is distributed across the

genome, and to understand how they could be functioning. Previous GC content studies have concentrated on animals with moderate to high GC content, and the study of a GC poor, or AT-rich, genome has the potential to further inform on conjectures made for GC evolution.

P. lamarcki *miRNAs*

Despite the abundant problems with assembly and assessment, miRNAs were successfully curated from the *P. lamarcki* genome. Among the miRNAs assessed were ones that, prior to this study, were thought to be vertebrate-specific (for a full discussion of these, see Kenny, Namigai et al., *in press*; Appendix 2). As the discovery of the latter miRNAs in the *P. lamarcki* genome was unexpected, especially given the use of miRNAs in phylogenetic reconstruction, expression analyses were done and confirmed both the expression of miRNAs shared with other lophotrochozoans (miR-1983 and miR-3533), and the expression of miRNAs that were thought to be vertebrate-specific (miR-1285 and miR-1287) (Brameier, 2010; Meunier et al., 2013; Strozzi et al., 2009). The presence of miRNAs that had previously been thought to be lineage-specific to other lineages indicates that there is wider conservation, or an abnormal degree of convergent evolution, of miRNAs in animal genomes.

miRNAs in animal development

The presence of unexpected miRNAs could have multiple meanings in relation to animal development, as miRNAs have been found to participate in or

have at least some relation to almost every important developmental process studied so far. Taking an example that is relevant for the next chapter, miRNAs participate in the process of gastrulation, more specifically in mesendoderm formation, where miR430/427/302 promote mesendoderm formation and supress neuroectoderm formation (Rosa et al., 2009). This is achieved through species-specific targeting of TGF-beta signalling components, where these miRNAs target the ligands Nodal and Lefty (such as in zebrafish), or just Lefty (as in human embryonic stem cells), or Lefty and a subset of Nodal orthologues (as in the frog *Xenopus*) (Ivey and Srivastava, 2015). The promiscuity of miRNA function seems widespread even with the use of a common pathway. Perhaps miRNAs introduce variability and diversity in common mechanisms of differentiation across animals, functioning as an instigator of variation in a process reminiscent of developmental systems drift. From a genomic perspective, this makes it interesting to see whether there are miRNA binding sites in the 3' region of conserved genes, and to see the degree of variability in miRNA control of key developmental processes, and their functional significance. This would involve both looking at miRNA binding sites in the 3' region of genes, testing targets of known miRNAs, and to test the function of these genes with miRNA knockdown.

miRNAs for inferring phylogeny

miRNAs undoubtedly contain information that can be valuable for phylogenetic inferences, however it is more likely that they will inform on the evolution of transcriptional control of development rather than in phylogenetic

reconstruction. This is confirmed by this thesis, where both expected and unexpected miRNAs were found, including those that are likely synapomorphies of annelids and molluscs, and those that were thought to be vertebrate-specific and are likely homoplasies. The unexpected miRNAs found in the *P. lamarcki* genome warn caution in using miRNAs to infer phylogeny, as it remains a possibility that there could be prevalent loss of miRNAs across the animal phylogeny contrary to previous suggestions, or that they could evolve convergently (for further discussion see Kenny, Namigai et al., *in press*; Appendix 2). Which of these two options is the case is not clear, but it does reveal that we do not know enough about the presence or absence of miRNAs in animals to be able justify using them for phylogenetic reconstruction. A cataloguing of miRNA presence and absence needs to be done in a wider range of animals first, in order to understand how miRNAs may have evolved, and later reevaluate how sound it is to use these as phylogenetic markers. It is becoming increasingly possible to survey lesser-known animals in this way, as this thesis has shown that despite the draft quality of this genome, it can still be viable for miRNA analyses and can raise some important biological issues.

Future directions

Three main issues are raised through this molecular resource: genome quality and depth of sequencing, GC content and high heterozygosity or polymorphism, and the capture of developmentally relevant information in the form of small RNA sequences.

To reiterate, it is apparent that this draft genome, while sufficient, can be improved. A higher quality genome would allow us to better compare our data with other sequenced genomes in the Lophotrochozoa (Simakov et al., 2012), such as the recently published annelid genomes (*H. robusta* and *C. teleta*) and would allow us to gain an ever-broader understanding of genome evolution. I am sure that more questions (and answers) will emerge as genome sequencing and assembly technologies continue to improve. With the fast pace of technological advancement, it is probable that the *P. lamarcki* genome can be improved by more and deeper sequencing efforts, and a more honed assembly strategy that builds upon what has been observed in this thesis. However, the current genome is sufficient for the purposes of this thesis and the pursuit of a higher quality genome will be left to others.

The GC-poor and AT-rich characteristic of the genome, in addition to the high levels of heterozygous content, are likely real aspects that need further investigation. The AT/GC balance has been considered in other studies as potential key players in genome evolution, yet we still do not have a good understanding of what it may mean functionally. A deeper sequencing effort could be used to look at areas of GC-rich and GC-poor areas to ascertain whether there are tractable and biologically relevant areas of concentrated GC or AT-rich areas. The high level of polymorphism is also worth analyzing, both qualitatively and quantitatively, in conjunction with whether this is the result of highly repetitive regions. This would help disentangle to what extent the statistics observed at the moment are effects of the sequencing and assembly techniques, and to what extent these are real aspects of the genome.

The miRNA case study warns that more studies are needed to understand the diversity of miRNAs and the diversity of mechanisms by which they function before using them to infer phylogeny. Despite this, these data also show that biologically relevant information is retrievable despite the relatively low quality of the genome, opening the door to using this resource to address other questions, as in the following chapter on Nodal signalling.

Chapter 3

Nodal signalling in annelid development

Introduction

The Nodal pathway is a signalling pathway within the TGF-beta superfamily, a diverse superfamily whose members function in a myriad of developmental contexts (Massagué, 2012, and references therein). Nodal has been the pathway most associated with LR asymmetry establishment, and there is a great deal of evidence that its function and expression are conserved in establishing left-right (LR) asymmetry in deuterostomes. In this chapter, I will present evidence that *Nodal* is present and expressed in the polychaete annelid *P. lamarcki*. Whether this expression is asymmetric is discussed.

Pathways are usually named for a conserved ligand. The Nodal pathway acts in early development to propagate LR asymmetry through a specific suite of antagonists, receptors and transcription factors. For a brief overview of the Nodal pathway, I will take a vertebrate perspective. The ligand Nodal is expressed around the node and in the left lateral plate mesoderm (LPM), and repressed in the right LPM by Lefty2, a feedback inhibitor of Nodal. Lefty1 functions in a similar feedback loop with Nodal, and serves to establish the midline barrier in the embryo, segregating the signals of the left LPM from the right LPM. The Nodal signal is transmitted when Nodal binds to an EGF-CFC protein factor (thought to be Cryptic; Yan et al., 1999), to form a complex that then transduces the Nodal signal by binding to TGF-beta pathway receptors: a type 1 receptor Alk4/5/7, and a type 2 receptor ActRIIa or ActRIIb (Hamada et al., 2002; Schier and Shen, 2000). These receptors are also known as TGF-

beta receptor 1, and TGF-beta receptor 2. This signal is then transduced through the transcription factors Smad2/3, Smad4, and further transmitted through Pitx2, a downstream homeobox transcription factor. Pitx2 regulates other downstream components to mediate asymmetry.

The Nodal-Lefty1/2 feedback loop is far from the only possible interaction in the Nodal pathway. In addition to a suite of other antagonists (such as DAN, Chordin, Gremlin, and BAMBI, among others; Walsh et al., 2010), Nodal is a signalling molecule of the diverse TGF-beta superfamily, sharing receptors and antagonists with the BMP, Activin, and TGF-beta subfamilies. For example, Nodal and other TGF-beta superfamily ligands have been found to complement each other by sharing receptors (Schier, 2003).

Surprisingly, the Lophotrochozoa seems to have been missed in large studies of the TGF-beta superfamily: the evolution of the TGF-beta superfamily has been studied in great detail in the deuterostomes (Massagué, 2012) and the Ecdysozoa (Özüak et al., 2014; Van der Zee et al., 2008), and has been combined with data from earlier branches of the animal tree of life (Adamska et al., 2007; Herpin et al., 2004; Huminiecki et al., 2009; Matus et al., 2006a) to draw insights on the ancestral condition of this pathway in the animal kingdom. In the following, I will present data that has been published in collaboration, where my focus was on the BMP ligands and TGF-beta/BMP family receptors (Figure 3.1). Data on other components of the TGF-beta superfamily were processed primarily by others, where the interesting aspects of promiscuity and crosstalk within the TGF-beta superfamily are further addressed (see Kenny, Namigai et al., 2014; Appendix 3). I have studied the presence of BMP ligands and TGF-beta/BMP family receptors in addition to the expression of *Nodal* and

the downstream effector *pitx* in the annelid *P. lamarcki* in order to ascertain to what extent Nodal signalling is conserved in the Lophotrochozoa.

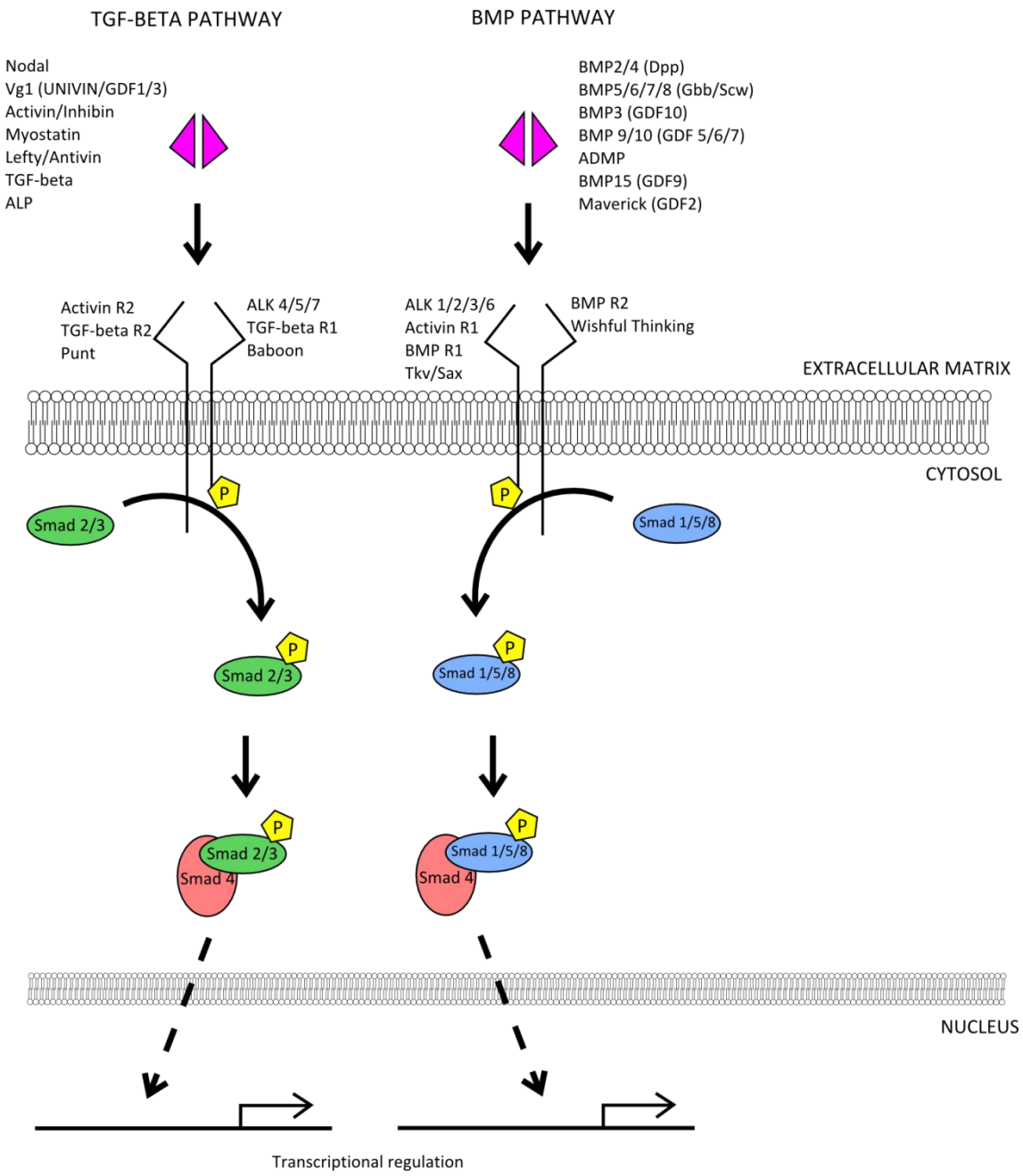


Figure 3.1 The TGF-beta superfamily pathway. The pathway is subdivided into the TGF-beta and BMP pathways. “P” represents phosphorylation, “R” in the names of receptors stands for receptor. Note: the categorization of ligands as functioning through the TGF-beta or BMP pathways does not necessarily correlate with their classification as TGF-beta or BMP ligands.

Methods

Gene candidate extraction

Ligands and receptors were initially identified by Blast2GO as a starting point (Conesa et al., 2005) and confirmed by blasting the sequences to the nr database (NCBI, National Center for Biotechnology Information). Ligands and receptors were also identified by tBLASTn (Altschul et al., 1990) using orthologous sequences against the genome, and confirmed by blasting the sequences to the nr database in NCBI. Sequences were translated into six frames using EMBOSS Transeq (Goujon et al., 2010; Rice et al., 2000) and trimmed down to the TGF-beta domain using SMART (Simple Modular Architecture Research Tool; Schutz et al., 1998; Letunic et al., 2014) to locate the domains. Phylogenetic trees were constructed by aligning these trimmed sequences with known TGF-beta sequences from a variety of species using MAFFT (version 6, Katoh et al., 2002). In cases where there were multiple fragments of a gene, these were aligned separately using MAFFT and visualized using BioEdit to observe whether this was due to assembly of the same gene to different scaffolds, or whether this indicated that there were two separate genes. MAFFT alignments were saved as FASTA files using BioEdit, which were used for the construction of Maximum Likelihood (ML) trees. ML trees were constructed using MEGA6 (Tamura et al., 2013) using the WAG model (Whelan and Goldman, 2001) with 1,000 bootstrap replicates and all other default settings. ML trees were directly exported from MEGA6 as a PDF file and annotated in Inkscape (http://inkscape.org). In the case of Bayesian trees, MAFFT alignments were converted to nexus format using ALTER (Glez-Peña et al., 2010). Bayesian trees were constructed using MrBayes 3.1

(Huelsenbeck and Ronquist, 2001) and after initial testing of various fixed amino acid models at 30,000 generations to test for best fit to the data, the WAG model was chosen (Whelan and Goldman, 2001). For Bayesian analyses, the mcmc (Monte Carlo Markov Chain) search was run for 1,000,000 generations with a burn-in of 25%, where the first 25% of the trees were discarded as burn-in, and all other parameters were left as default. Trees were sampled every 1,000 generations. Final trees were exported in the “simple” format and “allcompat” type, saved in Dendroscope (Huson and Scornavacca, 2012; Huson et al., 2007) as a PDF file, and annotated in Inkscape (http://inkscape.org).

RNA isolation

In order to amplify and clone cDNA from *P. lamarcki* larvae for *in situ* probe synthesis, RNA was isolated from *P. lamarcki* trochophores. Larvae were homogenized with an RNase-free polypropylene pellet pestle. TRIzol reagent was used to isolate the RNA and RNA was extracted with chloroform (to separate RNA from DNA and proteins). The RNA-containing aqueous layer was recovered by precipitation with isopropanol. The resulting RNA pellet was washed with 75% ethanol/DEPC water (0.1% diethylpyrocarbonate), air-dried, and resuspended in DEPC water. In order to ensure the removal of DNA from the isolated RNA sample, PROMEGA RQ1 RNase-Free DNase (Promega) was used according to the manufacturer’s instructions. RNA was once again precipitated with isopropanol. RNA purification was completed by rinsing the pellet with 75% ethanol and dissolving the resulting pellet in DEPC water.

Concentration was measured using the “ThermoScientific NanoDrop 2000” spectrophotometer.

cDNA synthesis

One microgram of the isolated RNA was converted to cDNA using the First Strand cDNA Synthesis Kit (Fermentas) according to the manufacturer’s instructions. This synthesizes cDNA from an RNA template using a genetically engineered Moloney Murine Leukemia Virus (M-MuLV) reverse transcriptase. The oligo(dT) primer supplied selectively anneals to the poly(A) tail of mRNA, allowing the targeted synthesis of cDNA from mRNA rather than other forms of RNA.

Primer design

Primers for *Nodal* and *pitx* were designed using the Primer3 program, v. 0.4.0 (http://frodo.wi.mit.edu, Table 3.1).

Table 3.1 *Nodal* and *pitx* primers.

gene	product size (bp)	forward primer	reverse primer
Nodal	1778	ACGCGGGATGTACAATGCTA	TCTCGAATGCCATAAGTCTTGC
pitx	900	CGGCGAAGAAGGACACACTT	GGGATATCACCCACGCTGAC

Polymerase chain reaction (PCR).

The Promega GoTaq® PCR Core System I was used according to the manufacturer's instructions to amplify gene sequences using the designed primers. The PCR reaction was performed in a thermal cycler with the following setting: initial hold for 2 minutes at 94^0C to melt template cDNA followed by 40 cycles of the following thermal conditions: 30s at 94^0C, 30s at 55^0 C, 1 min. at 72 ^{0}C. Final extension was performed for 5 minutes at 72^0C. The PCR product was analyzed by gel electrophoresis (1.5% agarose, 120V, 30 min.) and visualized with UV transillumination at a wavelength of 360nm. The corresponding band was excised and extracted using the Qiagen© MinElute Extraction Kit according to the manufacturer's instructions.

Cloning and sequencing

The amplified cDNA was cloned into pCR™II-TOPO® vectors (Promega) using TOPO® TA Cloning Kit for Sequencing according to the manufacturer's instructions and transformed into One Shot® TOP10 chemically competent *Escherichia coli*. A single transformed colony was cultured overnight in LB broth and ampicillin and plasmids were purified using the Qiagen QIAprep Spin Miniprep Kit according to the manufacturer's instructions. Cloned DNA sequences were confirmed through sequencing using the ABI PRISM® 3100 Genetic Analyzer according to the manufacturer's instructions and these sequences were confirmed by BLAST analysis.

RNA probe synthesis

For RNA probe synthesis, the specific sequence was amplified from extracted plasmid DNA by PCR using M13 forward and reverse primers, which flank outside of the T7 and Sp6 promoters on the cloning vector. The resulting PCR product was gel purified and resuspended in DEPC water. This PCR product served as a DNA template during probe synthesis, where 0.5µg of cleaned PCR product was added to a 50µl reaction (1X DIG RNA labeling mix (Roche), 1X transcription buffer, 10mM DTT (dithiothreitol), RNase inhibitor (Roche) and T7 or Sp6 polymerase), in order to generate sense (negative control) and antisense probes for each gene sequence. The reaction was incubated for 2.5 hours at 37°C and terminated with a final concentration of 0.1M EDTA (ethylenediaminetetraacetic). Synthesized probes were precipitated with 1/10 volume 8M LiCl and ¼ volume EtOH, overnight at -20°C. Following centrifugation, the pellet was rinsed twice with 75% EtOH, air-dried, and dissolved in 50µl DEPC water. The probes were stored at -80°C until use for *in situ* hybridization.

Fixation

Animals were fixed as previously in the lab (McDougall et al., 2006). Embryos were stepped into ethanol (EtOH) and fixed in 4% paraformaldehyde (PFA) in MOPS overnight at 4°C. Animals were then stored in methanol (MeOH) at -20°C until use for *in situ* hybridization.

in situ *hybridization*

Animals were stepped into PBTw (1 X PBS: 0.137M NaCl, 0.0027M KCl, 0.0101M Na_2HPO_4, 0.0018M KH_2PO_4; 0.1% Tween20) on ice, treated with a 4µg/ml Proteinase K dilution for 5 mins at room temperature (RT), and post-fixed in 4% PFA/MOPS at RT for 45 mins. Animals were washed thrice in PBTw, and transferred to pre-heated *P. lamarcki* hybridization buffer (5X SSC, 5mM EDTA, 50% deionised formamide, 100µg/ml heparin, 0.1% Tween20, 1X Dendhartts, 100µg/ml total yeast RNA) for 3 hours at 65°C. Probes (at 15ng *Nodal* and 7.5ng *pitx*) were denatured at 65°C in hybridization buffer for 10 mins. Embryos were incubated in probe at 65°C for a minimum of 14 hours. Probe was removed from embryos and washed with pre-warmed wash buffer (Washes each with 50% deionised formamide, 0.1% Tween20, and 4x, 2x,or 1x SSC respectively) at 65°C, twice with 4x wash for 10 mins, twice with 2x wash for 10 mins, twice with 1x wash for 10 mins, three times with 1x SSC, 0.1% Tween20 at RT for 20 mins, and twice in PBTw for 5 mins. PBTw was replaced with 1% block (Roche blocking reagent in maleic acid buffer) and incubated at 4°C for 2 hours. A 1:3000 dilution of AP DIG antibody (Roche) was incubated in block on ice for 2 hours. The block on the embryos was replaced with the antibody and incubated at 4°C for a minimum of 14 hours. Embryos were washed four times with PBTw for 15 mins, thrice with 1x APT (0.1M Tris-HCl pH 9.5, 0.1M NaCl, 0.05M $MgCl_2$, 0.1% Tween20). Embryos were incubated in staining solution (50mg/ml NBT-BCIP/APT) and developed at RT. Development was stopped with two washes in PBTw and postfixed in 4% PFA/MOPS overnight at 4°C. Embryos were stepped into MeOH, and stored in 80% glycerol at 4°C or imaged immediately in 80% glycerol.

Results

Gene candidate identification

As a first look at the BMP ligands and TGF-beta/BMP receptors present in the *P. lamarcki* genome, a general BLAST2GO analysis was done (Table 3.2). This analysis recovered six ligands and two receptors. As BLAST2GO is a crude search drawn from the BLAST server, there is no guarantee that the hit sequence is an accurate identifier for these genes. These were therefore further confirmed by blasting against the nr database (NCBI).

Table 3.2 BLAST2GO results for TGF-beta components in the *P. lamarcki* genome.

scaffold	BLAST2GO hit	length of hit	number of hits	e-value
C4403861	tgf-beta family member nodal	197	20	1.12e-19
C7358414	bone morphogenetic protein type 1b	664	20	8.51e-23
scaffold319832	tgf-beta receptor type-1	1385	20	1.43e-17
scaffold389827	bone morphogenetic protein 2	2348	20	4.07e-70
scaffold424608	tgf-beta family member nodal	1210	20	6.79e-13
scaffold555394	vang-like protein 2	2203	20	2.12e-99
scaffold589783	bone morphogenetic protein receptor type-1b	2211	20	9.73e-28
caffold596940	bone morphogenetic protein receptor type-1a	1004	20	7.36e-24

In addition to BLAST2GO, orthologous sequences of BMP ligands and TGF-beta/BMP receptors from other species were used to directly search against the genome and transcriptome by a translated nucleotide query (Table 3.3), and further confirmed by back-blasting this against the nr database (NCBI).

Table 3.3 Scaffolds for BMP ligands and TGF-beta/BMP receptors. Activin receptors were found in collaboration (see Kenny, Namigai et al., 2014; Appendix 3).

	ligand/receptor	**genome scaffold**	**transcriptome scaffold**
BMP ligands	BMP2/4	scaffold389827_14.6	Locus_25644_Transcript_1/1_Confidence _1.000_Length_573
	BMP3	C5521992_14.0	Locus_22445_Transcript_1/2_Confidence _1.000_Length_348
	BMP5/8	scaffold483029_12.8	Locus_15041_Transcript_2/4_Confidence _0.524_Length_202
	Maverick/GDF2	scaffold43629 14.6	Locus_14800_Transcript_3/3_Confidence _0.667_Length_2784
	NODAL	scaffold424608 13.6, C4403861 13.0	Locus_2416_Transcript_2/2_Confidence_ 1.000_Length_2772
	Vg1/UNIVIN/GDF1/3	scaffold578407 16.9	Locus_48251_Transcript_1/1_Confidence _1.000_Length_103
TGF-BETA receptors	ACTIVIN RECEPTOR 1	none	transcriptome
	ACTIVIN RECEPTOR 2	none	transcriptome
	BMP RECEPTOR 1	scaffold589783 15.1, scaffold596940 15.0, C7358414 18.0	Locus_9656_Transcript_1/1_Confidence_ 1.000_Length_2178
	BMP RECEPTOR 2	C4477237_17.0	Locus_98399_Transcript_1/1_Confidence _1.000_Length_118
	TGF-BETA RECEPTOR 1	scaffold319832_7.5	transcriptome

The recovered sequences from both the BLAST2GO analysis and the BLAST results against the genome were translated into protein sequences. The protein sequences were used to search for protein domains (Table 3.4) to find both the correct open reading frame (ORF) and to further confirm that the sequences held conserved TGF-beta domains.

Sequences were trimmed to the portion of the scaffold containing the correct ORF. These trimmed sequences were then used to make alignments (Figure 3.2) and subsequently used for the construction of phylogenetic trees.

Table 3.4 Sequences of BMP ligands and TGF-beta/BMP receptors. Protein domains are highlighted in pink in the "domain sequence" column.

	gene	domain sequence	SMART domain (highlighted region)
BMP ligands	BMP2/4	ETIKGAELRLFRRKANDSFQNLQSN QEYLEMHSSINKAHRIDIYEIIKPSTPT QESITRLIDTRFLKDLSKSKWETFDIH PAVHRWRNSPKSNFGLEVHFTDMK RNKPHLKHVRLRRSTDNSLDSEWNS EQPLLITYSDDGKGETPRRKRRQCQ RRPLYVDFKDVGWSDWIIAPPGYDA YFCLGECPFYLPDHMNATNHAIVQTL VHSVNRQAAPPPCCVPTELSDIAMLF IDENEKVVLRNYQDMTVMACGCR	TGF-beta peptide
	BMP3	IIKYQLFKLMYSIVCSQGGTVSNHASL KSLVQAVGIHGNIPAPCCIPDTMATQ SILYIDHNENIVYKNYRNMAVNSCKC L	TGF-beta peptide
	BMP5/8	GWVDFNITQVTAEWLARSADSLRFVI TIFDSNGLKVKPEHIGIDATQRYEGT RAFMVAFFRAAPLKTQLRQRRSASEI PGLADTQPSDEEYEDESDELEEDEL QFRKPPKHLDFHSRTRRQVTKDDQT NEAEVQNDSPKRGKKKIDKTTQSAD NPRRIIQKSTSGRRNLCHREAFFVRF RELNWQDWIIAPEGYAAYYCTGRCS FPLSSHMNATNHAIVQTLVSLMRPRA VPAPCCAPTKLNHISVLYLDDNSNVIL RKFHNMVVSACGCL	TGF-beta peptide
	MAVERICK/GDF2	IDSIVEDTQPVLEVYTQEQNILRREKR AVDPYDCTQGDGETRCCRYPLEIRF KDIGWDSWIEAPVSFKAYFCDGNCP HAYKLGHNFAQVQAIMHYINADAAPA PCCSASRLSPLTLLHKDEHGKLKVTN YVDMIVEECKCS	TGF-beta peptide
	NODAL	LCAMKDFWVDFATLGWNRWILYPPR YNAQMCGGKCSIPIGWDTNPTNHAIL QSLMRLNDKKVERPCCVPTKLSAITL LYYEGEEIVVRHHEGMVASECGCR	TGF-beta peptide
	Vg1/UNIVIN/GDF1/3	LIPTTYFSFTLYSFLFLYIVLSGNDKDS GMGFIADHRLWDLPIGEAEIVLPQVV MASNVESEIFYQLCSSEVCLVDSSYL TTADGFSSLNITRTIKTLIEGKENITHI HIHFNASGYVNAKNGGATCVIHTAPK SFYFSMASPNPVKLRIRRQKEDNSA GNTGQTGGCMWNHWRFRFKDIGW EDWVIQPKYFDANVCVGVCTLDVLE ENVNMSSHAFLRETYKDASSSDNIQ LVPDAACVPMQMSAINILFVDNDGLS VKRMAEMVAETCGCY	TGF-beta peptide
TGF-BETA receptors	ACTIVIN RECEPTOR 1	VDMATTLQLIHSSAAGLVHLHTEIFGT QGKPAIAHRDIKSKNILVKQNGQCCI ADLGLAVMHSQLTNNLDISENNKVG TRRYMAPELLDGSIKSTCFDSYKQID VYAFSLVMWEIARVCVINGTCEEYRI PYHDCVPSDPSFEEMHKVVCADNR RPAIPIYWSLDERLTKLSQLIRECWN SKAAARLTML	S_TKc domain Serine/Threonine protein kinase catalytic domain
	ACTIVIN RECEPTOR 2	TKMSANSRLSRILGEIMNCSLVIFILT GMTLTNSSPAYEASKQLECMYHNHT CGKDKNCPPTQKCAIPEPGKNVYCH SSWYNDSETGIQFLTQGCWIDNSGD CKDRFECVMDKPAQPGVDANHPDF YYCCCEGDFCNEKVEYKPIYVSSSP APVEPTGPMVITSGSTNALLRTALYSI VPLVGIAIVIILMFWMYRRHKSAYHES LPNADQTPVSEVAPSTPLLHHHYMV QLLEVKARGRFGCVWRAQLMDDTV AVKIFPLQDKASFYTERDFYTLPQVG NHDNILHFIGADRRGEGIGTELWLITE YHDYGSLYDYLKANIVTWSELLKINE GIVSGLAYLHEELPPTKTENYKPAVV HRDLKSKNILLKKDLTACIGDFGLALK	S_TKc domain Serine/Threonine protein kinase catalytic domain

		FESNASVGDTHGQVGTRRYMAPEV LEGAINFNRDAFLRIDMYACALVMWE LVTRCSVQDGPVDEYRLPFEEQAGC YPTLEDMQEHVVQKKERPIIKEHWLK HAGLAVLCNTIEETWDHDPEARLSA GCVQERLHQIMPSLNTPRTPGGHHG GHISHLDTSDSSGGSTSDLEQPLHV VIGHKPNHLAGPDDSSNLDTVQTPLS SSTVTSTGQYNKPTSVVLTMPNIVTP RRTVGDTSSSSDDNNEFIVSRPLSTE RRIHHNTHVPPHRLDVTSPMNTEVA SHSSVTSPIGHDDVTTRLLPSSTETS HDDHDVASVKRESHHVPEVHVLTMS AAGMRNELMNKESNVS	
	BMP RECEPTOR 1	VQRTIAKQITLITSIGKGRYGEVWKGK WRGESVAVKIFFTTEEASWFRETEL YQTVLLRHENILGFIAADIKGTGSWT QLFLITAYHERGSLYDFLHDTVLDTH EALLLCYSASCGLAHLHTEIFGTQGG KPAIAHRDIKSKNILVKNNGTCCIADL GLAVKYLSDSNEVDIFPNTRQGTKRY MAPEVLNETINKYNFDSYRQADMYS FGLVLWEIARRTIVAG	S_TKc domain Serine/Threonine protein kinase catalytic domain
	BMP RECEPTOR 2	RI*YLYLLLSDQKKPAVVHRDLNSRNI LVREDMSLVIVDFGFAMQIDGCRVM RNGNEQTVEQSSLTDVSTLRGX	STYKc domain protein kinase possible dual-specificity Ser/Thr/Tyr kinase
	TGF-BETA RECEPTOR 1	IILINLFLIYCITNISNGLKCRCEPACQN TNHTCETDGVCFSALSRTDEHATPR RIYRCVDQDKLLPPENPLFCQSTSPN VYKLAIGCCRDYDFCNNDLNLTLLSP PKDPMVAKMNEGFFETLDAPKLVILI VGPICVLTILFLSALLIRHKYKIDRGPP AYQRPDPEIQSLMTPGSAQTLREML DDCTNTGSGSGLPLLVQRTIARQIQL VDIIGQGRYGEVWRGRWNDEYVAV KIFSSRDERSWFREAEIYQTVMLRHA NILGFIAADNKDSGTWTQLWLVTDFH ENGSLFDYLNQVTVDVVQMAKLTMS IANGLAHLHMEIVGTQGKPAIAHRDL KSKNILVKRNGYCCIADLGLAVKHDV ETDSVEIPVNPNRVGTKRYMAPEVL DNSINQVHFESFKRADVYSLGLVLW EIVRRCHVQGIKDAYQLPFYDKVPSD PSLEEMKEVVCNQRMRPEIPNRWH DSEALRVMSKLMKECWYANGAARL TALRIKKTIQGMINQEDIKI	S_TKc domain Serine/Threonine protein kinase catalytic domain

BMP ligands in P. lamarcki

BMP ligands are present in *P. lamarcki* (Figure 3.3, Figure 3.4, and Table 3.5), namely BMP2/4, BMP3, BMP5/8, Nodal, and Maverick. In the case of Vg1, the identity of the sequence becomes questionable due to its position in the tree, and the suffix –like has been added in this case. BMP10 and ADMP are missing in *P. lamarcki*, but given that they are present in all other lophotrochozoan genomes, it is likely that the fragmented condition of the genome means that some information such as this was lost in sequencing and assembly. *Adenita*

vaga sequences were included to show that they do not group with other BMP ligands (further discussed in Kenny, Namigai et al., 2014; Appendix 3). The distantly-related GDNF (glial cell line-derived neurotrophic factor) group was used as an outgroup (Saarma, 2000).

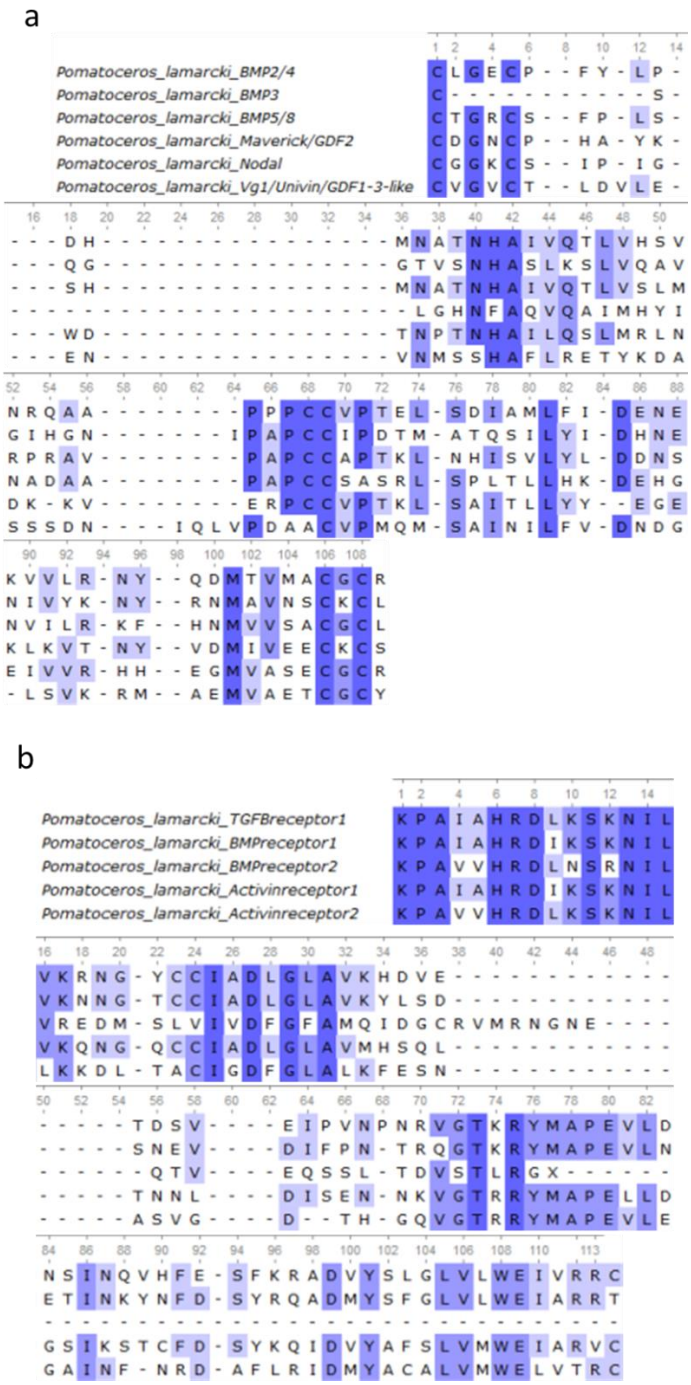


Figure 3.2 Specimen alignment of the conserved protein domain region of (A) BMP ligands and (B) TGF-beta/BMP receptors.

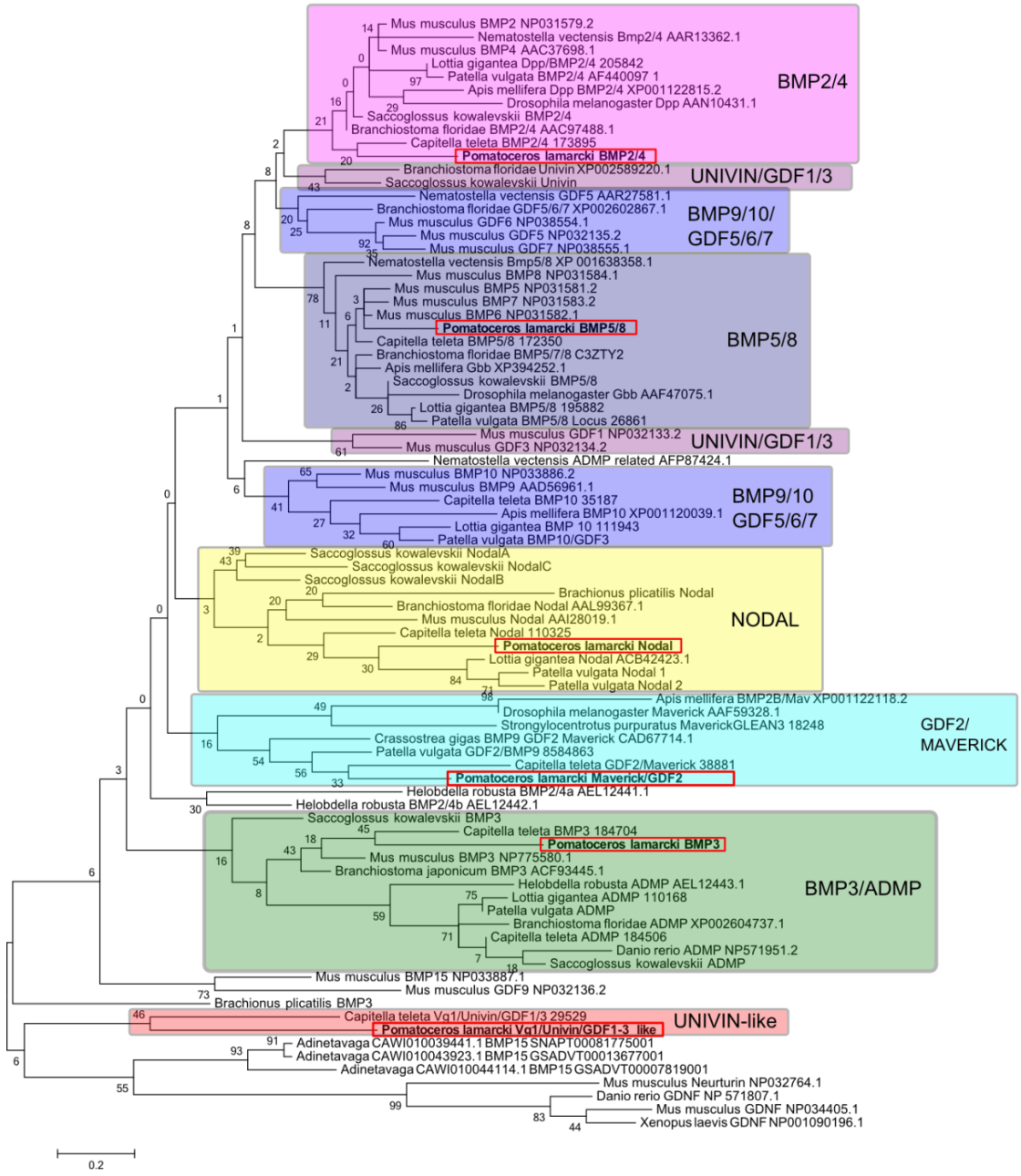


Figure 3.3 ML tree of BMP ligands, based on the WAG model (Whelan and Goldman, 2001). Initial trees constructed by Neighbor-Joining to a matrix of pairwise distances estimated using a JTT model. The tree with the highest log likelihood (-4640.5836) is shown, which involved 84 amino acid sequences. Gaps and missing data were eliminated for a total of 49 positions in the final dataset. Scale bar indicates number of nucleotide substitutions per site, with numbers on branches dependent on branch length.

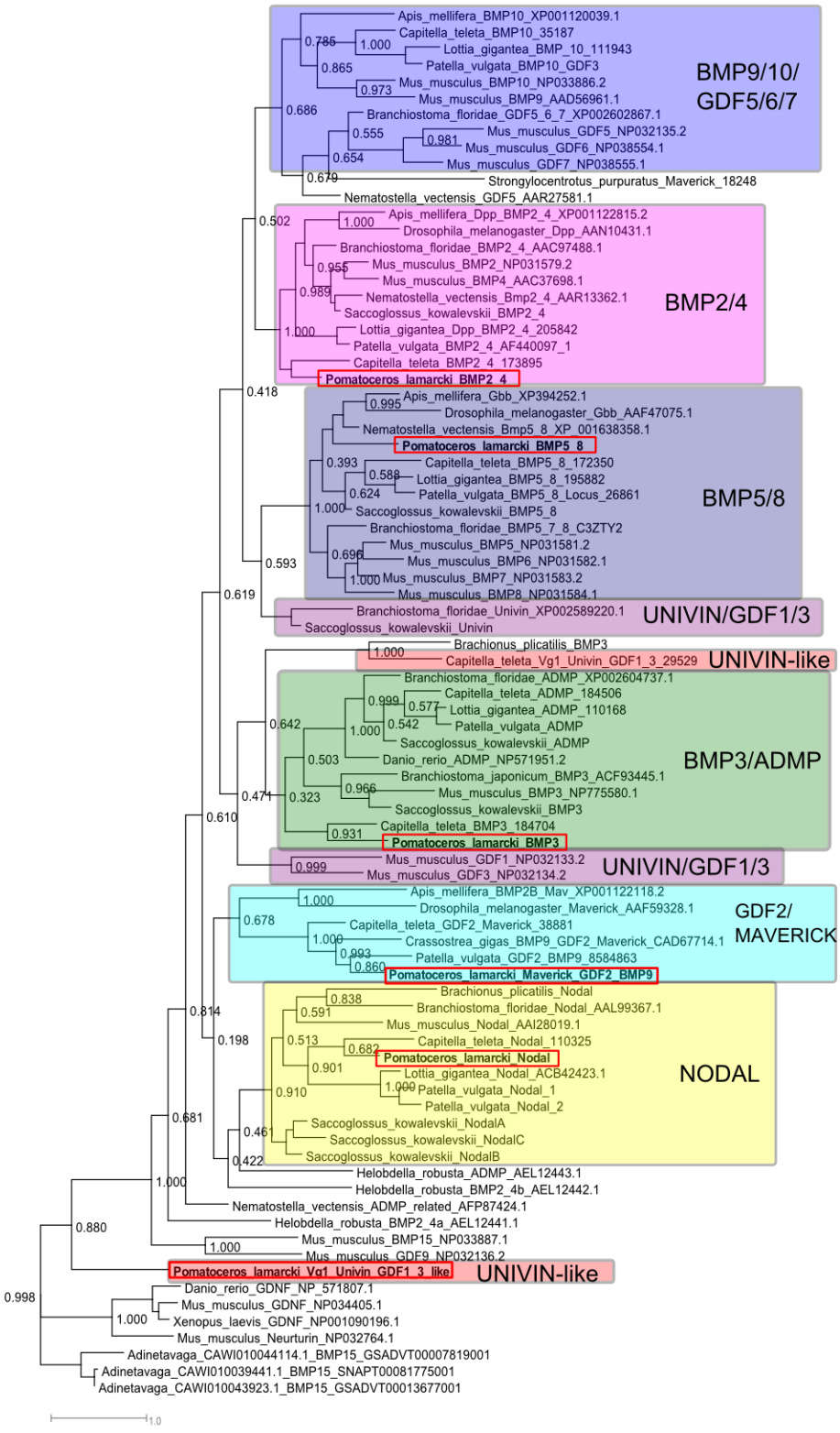


Figure 3.4 Bayesian tree of BMP ligands. Trees were built using the WAG model (Whelan and Goldman, 2001). A Monte Carlo Markov Chain search was run for 1,000,000 generations with 25% burn-in. Scale bar indicates nucleotide substitutions per site, with numbers on branches dependent on branch length.

TGF-beta/BMP receptors in P. lamarcki

Three Type 1 and two Type 2 TGF-beta/BMP receptors are found in *P. lamarcki* (Figure 3.5, Figure 3.6, Table 3.5). These serine/threonine kinase receptors are typically conserved in this ratio, and function by dimerization of Type 1 and Type 2 receptors. Deuterostomes have a higher diversity of receptors, as somewhat reflected in the small sampling of deuterostomes in these trees. Although there were low bootstrap values for the ML trees (Figure 3.5), these were evaluated alongside the Bayesian trees where the posterior probability values were high enough to confidently assign receptors to these groupings (Figure 3.6).

Table 3.5 Presence/absence of BMP ligands and TGF-beta/BMP receptors in *P. lamarcki*.

Family/Pathway	**Ligand/Receptor**	**present/absent**
BMP ligands	Nodal	present
	BMP2/4 (Dpp)	present
	BMP5/6/7/8 (Gbb/Scw)	present
	BMP3 (GDF10)	present
	ADMP	absent
	Maverick (GDF2)	present
	BMP9/10 (GDF5/6/7)	absent
	BMP15 (GDF9)	absent
	Vg1 (Univin,GDF1/3)	present
Type 1 receptor (TGF-beta pathway)	TGF-beta R1 (ALK 4/5/7, Baboon)	present
Type 1 receptor (BMP pathway)	Activin R1 (ALK 1/2/3/6, Tkv/Sax)	present
Type 1 receptor (BMP pathway)	BMP R1 (ALK 1/2/3/6, Tkv/Sax)	present
Type 2 receptor (TGF-beta pathway)	Activin R2 (TGF-beta R2, Punt)	present
Type 2 receptor (TGF-beta pathway)	BMP R2 (Wishful Thinking)	present

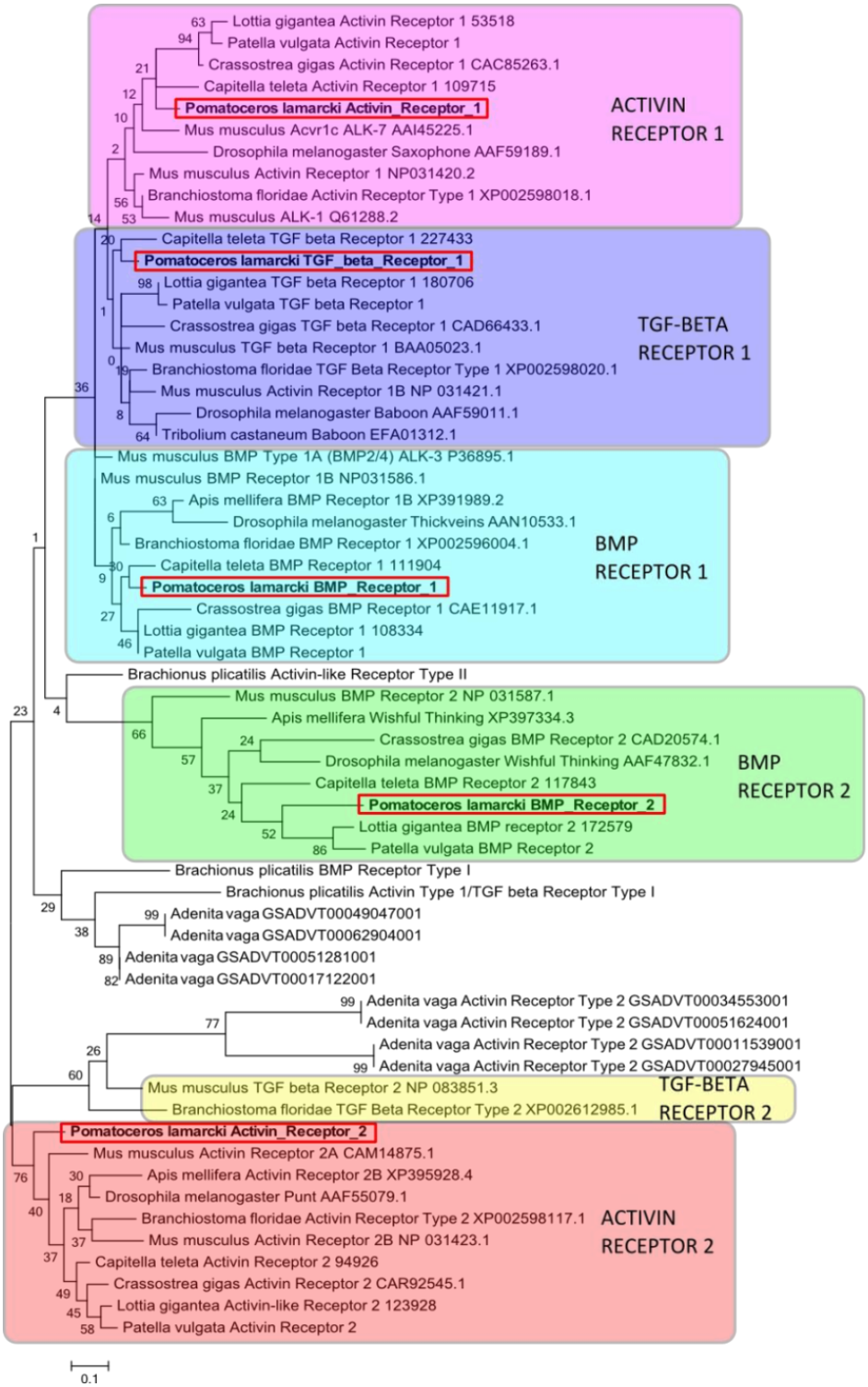


Figure 3.5 ML tree of TGF-beta/BMP receptors, based on the WAG model (Whelan and Goldman, 2001). Initial trees constructed by Neighbor-Joining to a matrix of pairwise distances estimated using a JTT model. The tree with the highest log likelihood (-1953.4744) is shown, which involved 61 amino acid sequences. Gaps and missing data were eliminated for a total of 44 positions in the final dataset. Scale bar indicates number of nucleotide substitutions per site, with numbers on branches dependent on branch length.

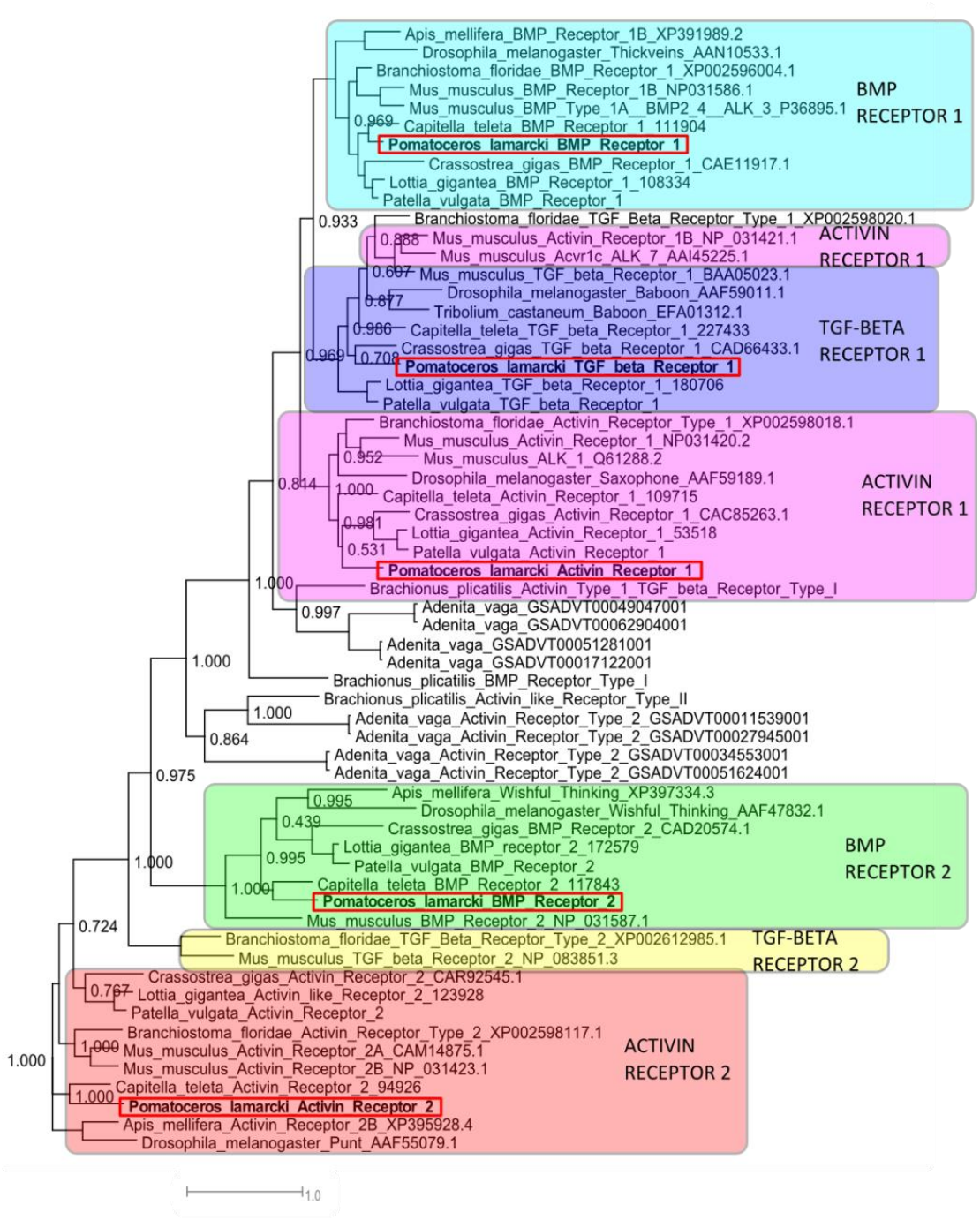


Figure 3.6 Bayesian tree of TGF-beta/BMP receptors. Trees were built using the WAG model (Whelan and Goldman, 2001). A Monte Carlo Markov Chain search was run for 1,000,000 generations with 25% burn-in. Scale bar indicates nucleotide substitutions per site, with numbers on branches dependent on branch length.

Nodal *expression pattern*

Nodal expression was examined by *in situ* hybridization at stages spanning from blastula to trochophore (blastula, gastrula, trochophore). Specific expression was only seen at the trochophore stage, and not earlier at the blastula and gastrula stage. This lack of earlier expression may be real, however it is more likely that this is the result of faults in the protocol as at these stages there was a mixture of complete lack of staining and complete overstaining with the anti-sense probe. *Nodal* is expressed at the trochophore stage in the gut that leads from the stomach to the mouth, and at a concentrated spot just below the stomach, underneath the prototroch area (Figure 3.7). The concentrated spot of expression is likely the mesodermal primordium, which then goes on to form increasingly determined mesodermal tissue.

pitx *expression pattern*

Similar to *Nodal* expression, *pitx* expression was examined from blastula to trochophore, but expression was only seen at the trochophore stage. Again, this is likely the result of a faulty protocol at these stages, as either no stain or overstaining was observed with the anti-sense probe. *Pitx* is expressed over a broader area than *Nodal*, with a similar expression pattern seen along the gut from the stomach to the mouth (Figure 3.8). Taking the mouth as ventral, *pitx* expression is expanded relative to *Nodal* expression on the opposite side of the mouth (presumably the dorsal side). Unlike *Nodal* expression, *pitx* expression

spans the AP axis. Whether this is dorsal expression or sided expression is difficult to tell without a midline marker in these trochophores.

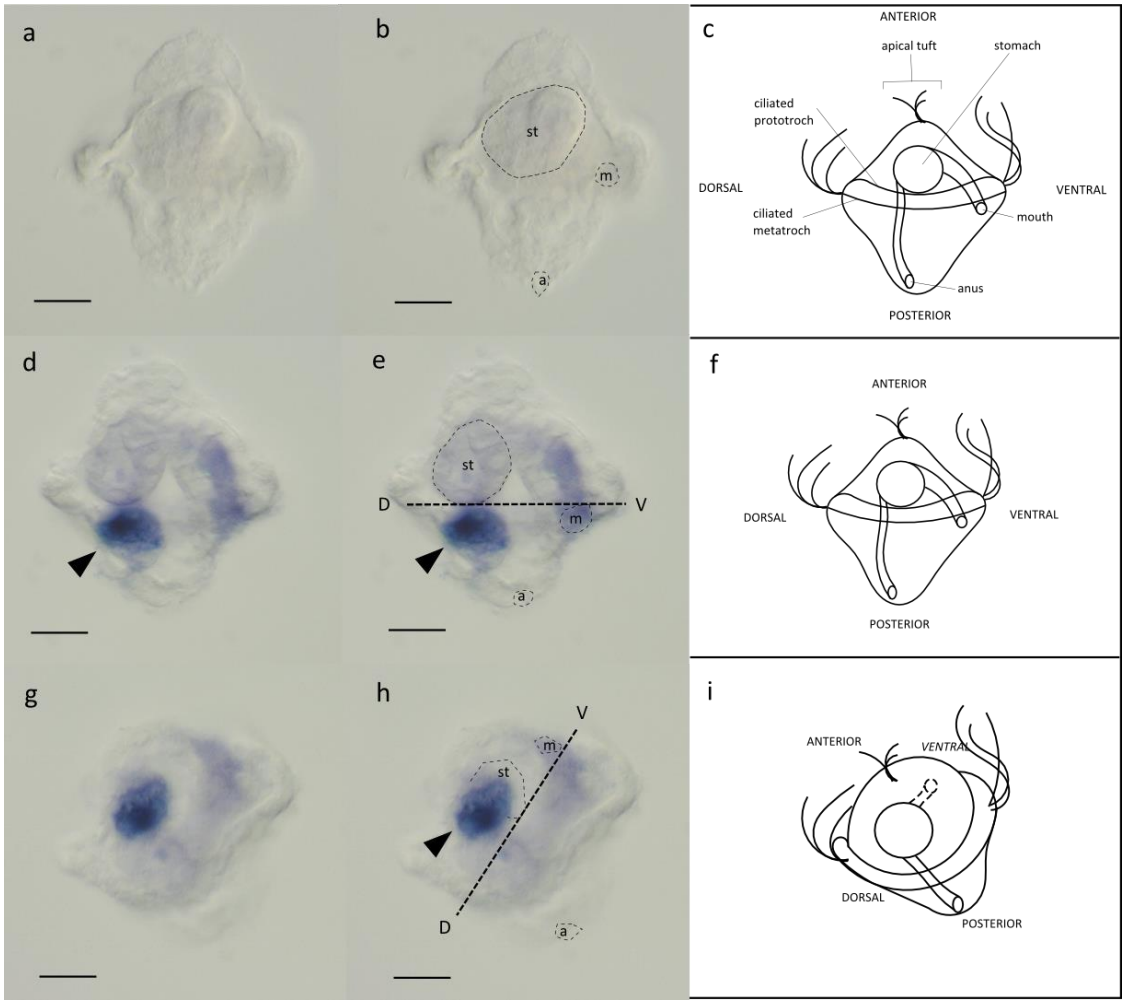


Figure 3.7 (A-C) control sense probe expression with (B) highlighted anatomy and (C) caricature of general anatomy; (D-F) *Nodal* mRNA expression, side view with (E) highlighted anatomy and (F) caricature of general anatomy; (G-I) *Nodal* mRNA expression, dorsal view with (H) highlighted anatomy and (I) caricature of general anatomy. Stomach (st), mouth (m) and anus (a). Black arrowhead indicates concentrated spot of *Nodal* expression below the prototroch. In caricatures, structures drawn with dotted lines and axes labelled in italics indicate that the structure and axis are going into the page along the z-axis. Scale bar is 20µm.

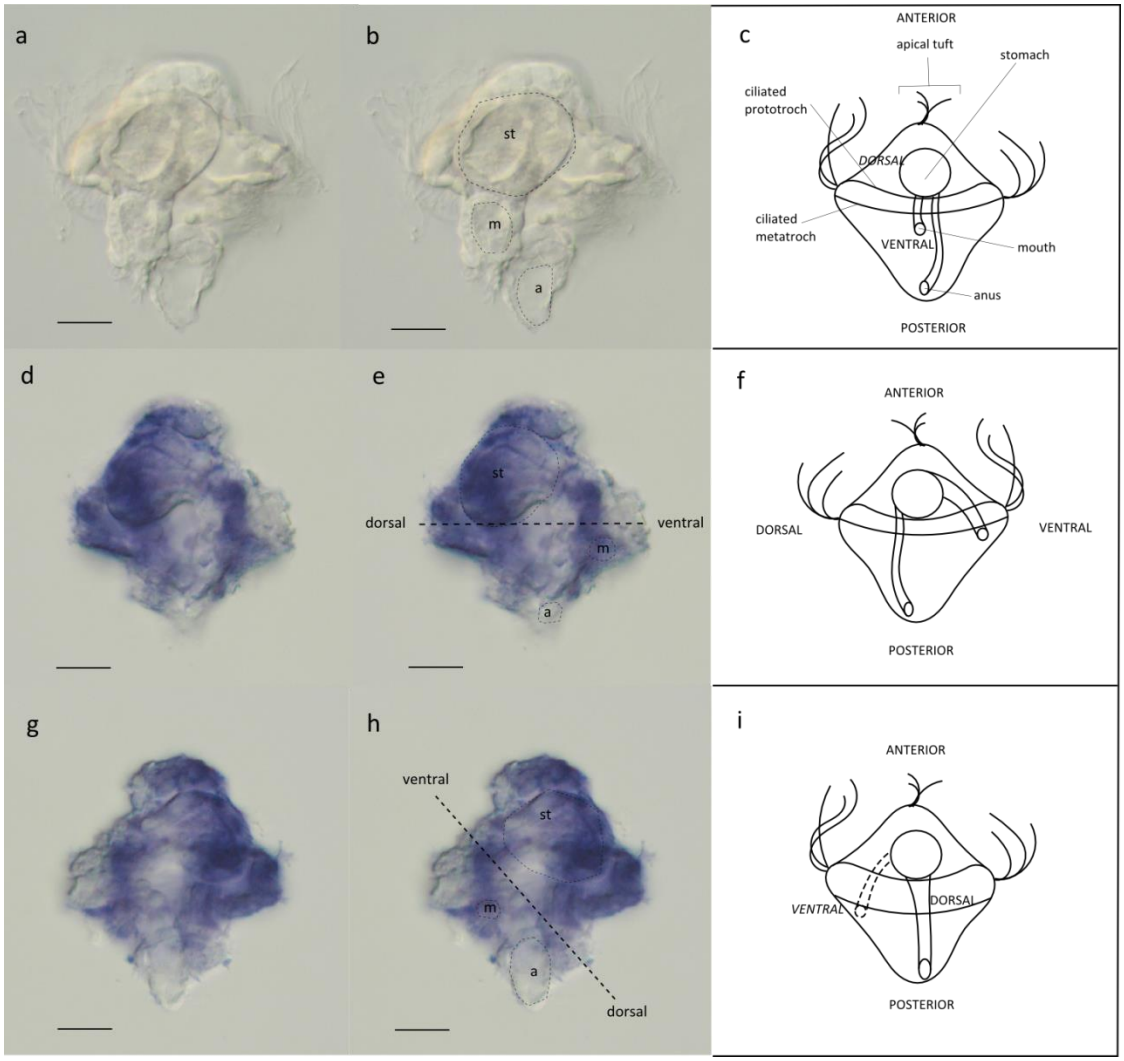


Figure 3.8 (A-C) control sense probe expression with (B) highlighted anatomy and (C) caricature of general anatomy; (D-F) *pitx* mRNA expression, side view with (E) highlighted anatomy and (F) caricature of general anatomy; (G-I) *pitx* mRNA expression, dorsal view with (H) highlighted anatomy and (I) caricature of general anatomy. Stomach (st), mouth (m) and anus (a). Black arrowhead indicates concentrated spot of *Nodal* expression below the prototroch. In caricatures, structures drawn with dotted lines and axes labelled in italics indicate that the structure and axis are going into the page along the z-axis. Scale bar is 20µm.

Discussion

Nodal signaling in bilateral animals

Here, we present the first evidence that *Nodal* and *pitx* are present and expressed in annelids. Combined with data from molluscs, brachiopods, and rotifers, this supports the conservation of *Nodal* and *pitx* in the Lophotrochozoa. Typically, the most parsimonious explanation for the evolution of Nodal signalling is to look at the conservation of *Nodal* and *pitx* in lophotrochozoans and deuterostomes, and to suggest that these components were derived from the last common ancestor of bilaterians, with ecdysozoans undergoing a lineage-specific loss in that superphylum. However, the situation is not as clear-cut as it may seem, and there are key differences in the conservation of the presence of *Nodal*, its expression, and its function in animal development. The significance of each of these aspects will be addressed below, beginning with a discussion of the other components that constitute the TGF-beta superfamily and following with a discussion of Nodal signalling in particular.

TGF-beta receptors

TGF-beta receptors are conserved in *P. lamarcki*, where the expected number is three Type 1 receptors and two Type 2 receptors. These function when Type 1 and Type 2 receptors heterodimerize and are specific to either the TGF-beta pathway or the BMP pathway (Figure 3.1), where TGF-beta receptor 1 and Activin receptor 2 dimerize in the TGF-beta pathway, and BMP receptor 1 and BMP receptor 2 dimerize in the BMP pathway (Massagué, 1998). The

multiple distinct names for the same receptor are not for biological reasons, but because they were independently discovered and named by different groups at different times, resulting in unnecessary confusion. The presence of both Type 1 and Type 2 receptors is conserved across animals with small differences in ratios (Table 3.6). For example, in insects there are usually three Type 1 receptors and two Type 2 receptors. Small differences include animals such as the oyster *Crassostrea gigas*, where there are three Type 1 receptors and one Type 2 receptor (Herpin et al., 2005). In humans and other vertebrates, there is a larger diversity of receptors, most likely due to the genome duplication that is thought to have occurred in vertebrates. Invertebrate deuterostomes also have a slightly higher number of receptors, although this is thought to be an independent duplication event (Huminiecki et al., 2009). Members of Diploblastica (non-Bilateria) such as sponges and ctenophores also have TGF-beta receptors (Huminiecki et al., 2009; Pang et al., 2011; Srivastava et al., 2010). The conservation of core components of the TGF-beta pathway across animals has marked this pathway as an ancient one. There has been considerable diversification both in the number of components and in their functions across animals (Huminiecki et al., 2009). The recovery of five receptors in *P. lamarcki* is similar to that of other Lophotrochozoans, confirming the approximate number observed in this Superphylum and the strong conservation of TGF-beta receptors in animals.

Table 3.6 Presence and absence of TGF-beta superfamily components in a selection of animals (from Kenny, Namigai et al., 2014).

		Ligands	Smads	Receptors	Chordin	Chordin-like	Noggin	Noggin-like	Follistatin	Gremlins	Dans	Dantes	Tsgs	Tolloids	BAMBI	NOMO	SMURFs	References:
Deuterostomes	*Homo sapiens*	33	8	13	1	2	1	0	1	2	1	2	1	3	1	3	2	*Huminiecki et al 2009*
	Ciona intestinalis	10	5	7	1	1	1	0	0	1	0	0	1	1	0	1	1	*Hino et al 2003/Huminiecki et al 2009*
	Strongylocentrotus purpuratus	14	4	6	1	1	0	1	1	1	1	0	1	3	0	2	1	*Lapraz et al 2006*
Ecdysozoans	*Drosophila melanogaster*	7	4	5	1	0	0	0	1	0	0	0	3	2	0	1	1	*van der Zee et al 2008*
	Apis mellifera	7	4	5	1	0	0	0	0	0	1	0	1	1	1	1	1	*van der Zee et al 2008*
	Tribolium castaneum	8	4	5	1	0	0	0	1	1	1	0	1	1	1	1	1	*van der Zee et al 2008*
	Caenorhabditis elegans	5	7	3	0	0	0	0	0	1	0	0	0	1	0	1	0	*Savage-Dunn 2005, Huminiecki et al 2009*
Lophotrochozoans	*Capitella teleta*	16	4	5	0	1	1	1	1	1	0	0	1	2	1	1	1	
	Pomatoceros lamarckii	9	4	5	0	0	1	1	1	1	0	0	1	1	1	1	1	
	Lottia gigantea	10	4	5	1	1	1	1	1	1	0	0	1	1	1	1	1	
	Patella vulgata	12	4	5	1	0	1	1	1	1	0	0	1	1	1	1	1	
	Brachionus plicatilis	4	4	3	0	0	0	0	0	0	0	0	0	0	0	1	0	
	Adenita vaga	10*	10	8	0	0	0	0	2	0	0	0	0	0	0	2	0	*Flot et al 2013*
	Schistosoma mansoni	2	5	5	0	0	1	2	1	0	0	0	0	1	0	1	0	
Diploblasts	*Nematostella vectensis*	6	4	6	1	0	1	1	1	1	1	0	0	1	0	1	0	*Huminiecki et al 2009, Saina and Technau 2009*
	Mnemiopsis leidyi	9	5	4	0	0	0	0	0	0	0	0	0	1	0	1	1	*Pang et al 2011*
	Trichoplax adherens	5	4	4	0	1	1	0	1	0	1	0	0	0	0	1	1	*Huminiecki et al 2009*

TGF-beta signal transducers

SMAD proteins transduce extracellular signals to intracellular signals, and communicate TGF-beta signalling within the cell. This is primarily done by three receptor-regulated SMADs (R-SMADs), SMAD2/3 and SMAD1/5/8, and SMAD4, with a fourth SMAD, SMAD6/7, being an inhibitory SMAD that recruits ubiqutin ligases intracellularly. These four SMADs are the canonical SMADs that function in TGF-beta signalling, and the diversity of these proteins is highly conserved across animals. As SMADs transduce a diversity of signals, it is thought that the pleiotropic effect that would result if one SMAD was lost prevents loss from occurring as often in this family of proteins. *P. lamarcki* has all four SMADs normally found in animals, lending support to the strong conservations seen in other animals (Table 3.6).

TGF-beta ligand regulators

Aside from the inhibitory SMAD6/7, there are a number of other inhibitors of TGF-beta signalling that function as inhibitors at different levels of the signalling pathway. Inhibitors Noggin, Noggin-like, Follistatin, Twisted gastrulation (Tsg) and Gremlin function to inhibit TGF-beta and BMP signals and are all present in *P. lamarcki*. More specifically, Noggin and Noggin-like bind to primarily BMP ligands, Follistatin binds to TGF-beta ligands, and Tsg binds to BMP ligands in addition to some postulated functions in promoting BMP signalling (see Kenny, Namigai et al., 2014). Gremlin falls within the Dan family, which also has a diversity of names in other organisms such as Cerl, Coco, and Cerberus. Dan itself is not found in *P. lamarcki* or in other Lophotrochozoa, however Gremlin is present and is thought to function by sequestering BMP2/4 as proposed by studies in the leech *H. robusta* (Kuo and Weisblat, 2011). It is not clear what this could mean for the evolution of these inhibitory elements, however it seems that Gremlins are more conserved than Dan or Dante across animals (Table 3.6).

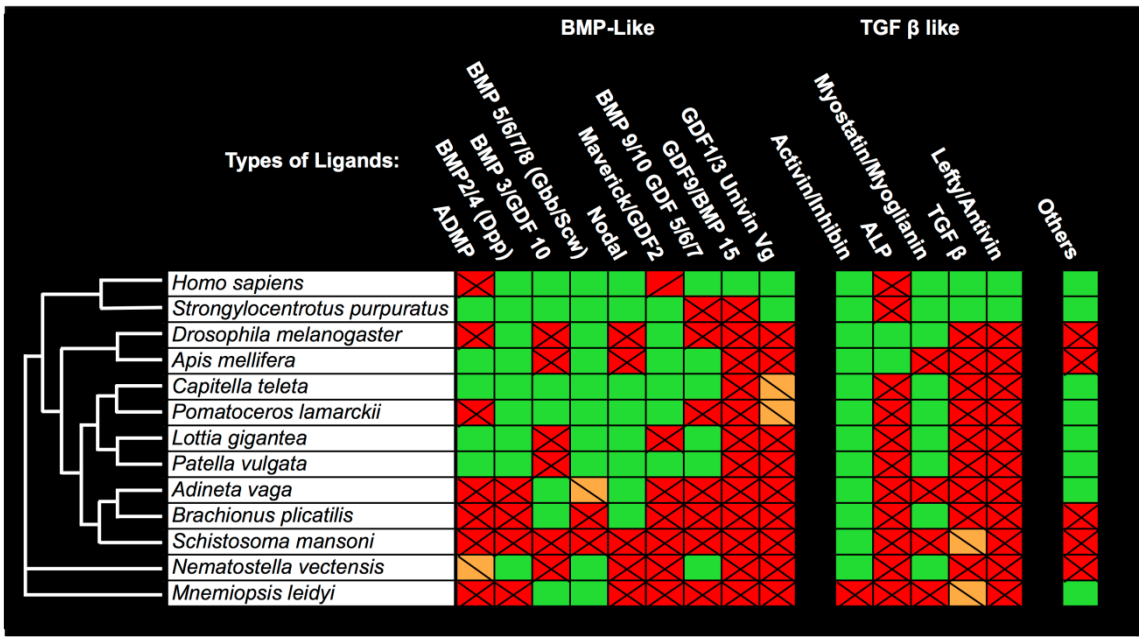
Tolloid, BAMBI, NOMO, and SMURF are all present in *P. lamarcki*, and differ from the aforementioned inhibitors in the way that they function. Tolloids do not function as inhibitors and rather release ligands from inhibitory regulators by cleaving the inhibitory regulator and releasing the ligand. BAMBI (BMP and activin membrane-bound inhibitor) is a pseudoreceptor that competes with Type 2 receptors, NOMO (Nodal Modulator) inhibits Nodal and activin signalling by forming a transmembrane complex at the endoplasmic reticulum, and SMURFs (SMAD specific E3 ubiquitin protein ligase) target R-SMADs for degradation. The fact that *P. lamarcki* has all of these regulators further supports the

conservation of the TGF-beta signalling components. Perhaps the most interesting of these is the presence of NOMO, and its function in regulating mesendoderm formation in zebrafish (Haffner et al., 2004). The fact that this is present in *P. lamarcki* becomes interesting in light of the lack of another Nodal inhibitor, Lefty. This will be addressed in a more detailed discussion of Lefty further below.

TGF-beta and BMP ligands

The majority of the TGF-beta and BMP ligands are present in *P. lamarcki*, with the exception of ALP, TGF-beta, Lefty/Antivin for TGF-beta ligands, and ADMP, BMP9/10 or GDF5/6/7, and GDF9 or BMP15 for BMP ligands. ALP is thus far ecdysozoan-specific, and TGF-beta and Lefty are deuterostome-specific. There have been reports of a homologue of TGF-beta in ctenophores (Pang et al., 2011), however the support for this is not strong. ADMP is also missing in human, fly, rotifer, and ctenophore although found in molluscs and *C. teleta*. BMP9/10 or GDF5/6/7 is also missing in sea urchin, fly, rotifer, and ctenophore although found in molluscs and *C. teleta*. It could be that ADMP and BMP9/10 or GDF5/6/7 were just not recovered in the *P. lamarcki* genome or transcriptome, or that these two ligands were independently lost in *P. lamarcki*.

Table 3.7 Presence and absence of TGF-beta and BMP ligands in a selection of animals (from Kenny, Namigai et al., 2014).



The presence of Nodal signalling components in animals

The one area where there is arguably the least controversy is that *Nodal* and *pitx* are present in most bilateral animals studied so far, with the exception of ecdysozoans and Platyhelminthes. The presence of *Nodal* has even been described outside of bilateral animals to include the cnidarian *Hydra magnipapillata* (Watanabe et al., 2014). *Pitx* is also conserved to even a higher degree than *Nodal*, where this downstream effector is present even in animals that lack *Nodal*. Although evidence needs to be gathered from more early-branching animals, and functional data is needed from cnidarians to solidify this, this evidence suggests that the Nodal-pitx coupling is possibly an ancient one found even in animals outside of the Bilateria.

The story differs from the canonical Nodal pathway when considering other components of the Nodal pathway, notably the Chordin, Chordin-like, Dan,

and Dante family of inhibitors. Lefty is missing in the lophotrochozoan animals studied so far (Kenny, Namigai et al., 2014; Appendix 3). Lefty is a key antagonist of Nodal signalling that functions as a feedback inhibitor of Nodal in deuterostomes, and is a member of the TGF-beta family of ligands. Lefty, Nodal, and Pitx2 are thought to be the three main components of the Nodal pathway that are conserved in both mesendoderm specification and LR asymmetry establishment in vertebrates (Schier, 2009). With Lefty missing in the Lophotrochozoa, this suggests that Lefty is a deuterostome innovation and begs to question how lophotrochozoans achieve sided expression of *Nodal* without this key antagonist. A possibility is that a different TGF-beta family ligand has been co-opted as an antagonist to Nodal. A search for potential candidates could be complicated, as the TGF-beta superfamily functions through a complex combination of pathway crosstalk in different developmental contexts. It also remains a possibility that *Lefty* is present yet has just not been recovered successfully in the genomes studied thus far. However, this is unlikely given the large number of genomes that have been studied.

While these are both possibilities, it is more likely that a different mechanism altogether has evolved in the absence of Lefty. The activation/inhibition feedback loop of Nodal and Lefty is a reaction-diffusion mechanism (Schier, 2009), where there are gradients of ligand expression across large distances that compartmentalize activity of a certain ligand to a certain space in the embryo. From an embryological standpoint, vertebrate embryos have asymmetric *Nodal* expression across the midline, where there are already sheets of cells in place and across which morphogen gradients can plausibly travel. In the Lophotrochozoa, however, there are usually far fewer

cells (hundreds rather than thousands) to be considered during the time that *Nodal* is expressed asymmetrically (at least in molluscs and brachiopods), and it is possible that rather than a gradient mechanism that is less efficient in this smaller context, a different molecular mechanism, such as a short-range signalling mechanism, has been used.

Nodal *expression in annelids and molluscs*

The data gathered in this thesis suggest that *Nodal* and *pitx* are not expressed asymmetrically in *P. lamarcki*. This is initially troubling, as both the presence of *Nodal* and its expression are conserved in other lophotrochozoan animals studied so far. The evidence presented here is far from solid, as expression was only seen at one developmental time point, and *Nodal* expression was not seen at other stages where you would expect it to be (such as earlier at the gastrula stage when the mesoderm is gradually specified). It is also very possible that the stages considered here do not include the stage where *Nodal* is asymmetrically expressed, and that it has just been missed in these analyses, or that in fact the stage considered here (trochophore) is too late to see the first instance of asymmetric expression as organs are already formed at this stage and expression may be associated with the functions of the organs themselves rather than asymmetry establishment. In addition, *pitx* is known to have different isoforms that have different expression patterns in frog and mouse (Schweickert et al., 2000), an aspect that was not studied here in *P. lamarcki* but may explain the lack of asymmetric expression if only one isoform of many was studied here. Finally, in the absence of a good midline marker it is

difficult to ascertain whether expression is symmetric or asymmetric in these animals, especially given the coiled innards of the trochophore larva. If *Nodal* expression is truly not asymmetric in annelids, this would need to be studied in other polychaetes to see whether this lack of asymmetric expression is a common characteristic of polychaetes.

If we believe the lack of asymmetric *Nodal* expression, and take it to mean that the function of Nodal to establish LR asymmetry is not conserved in this polychaete, this could have multiple meanings. For example, it is possible that there was a loss of function of the Nodal ligand, and that a different pathway was co-opted to function in establishing LR asymmetry. As Nodal is known to function both in mesendoderm specification and LR asymmetry establishment in vertebrates, and in other tissues in a range of other animals (Figure 3.9), the expression pattern we see here could reflect a conserved function in mesendoderm specification, but a loss of function in LR specification.

Loss of Nodal function in LR asymmetry establishment would be especially interesting when considering the presence and absence of regulators of Nodal signalling, such as Lefty and NOMO, namely that NOMO is present in *P. lamarcki*, but that Lefty is missing. As NOMO functions in mesendoderm specification but no evidence has emerged that it functions in LR asymmetry establishment, it is possible that the loss of Lefty allowed a loss of Nodal function in LR asymmetry, but that the conservation of NOMO retained Nodal function in mesendoderm specification. This is somewhat supported by the discovery of NOMO in the ctenophore *C. leidyi* despite the absence of a *Nodal* homolog (Pang et al., 2011).

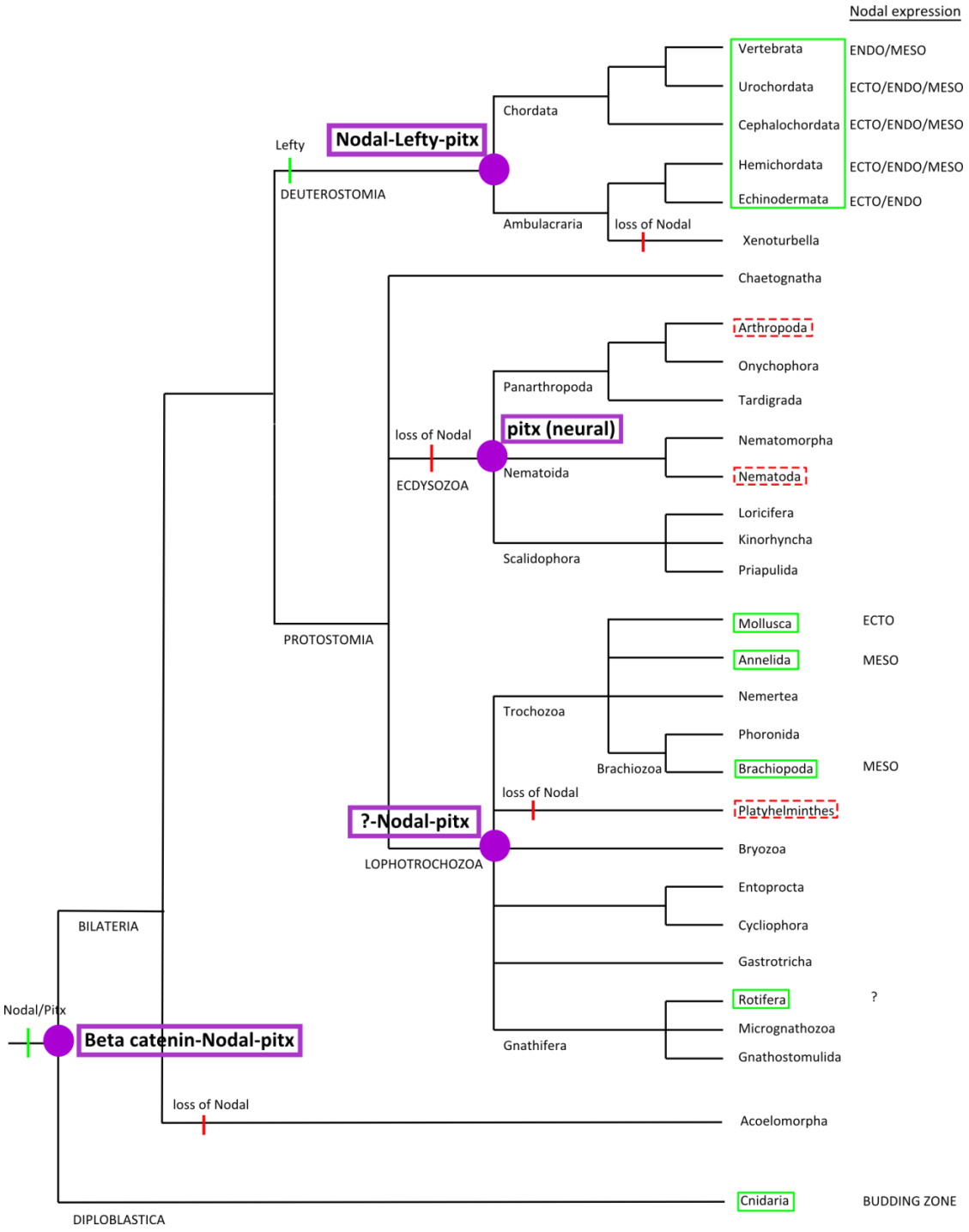


Figure 3.9 Evolutionary hypotheses for Nodal signalling. Green boxes and notches indicate *Nodal* presence or hypothesized gain, red dotted boxes and solid notches indicate *Nodal* absence or hypothesized loss, bold purple boxes and circles indicate hypothesized evolution of molecular machinery at specific nodes. Right-hand column indicates location of *Nodal* expression. ECTO=ectoderm, ENDO=endoderm, MESO=mesoderm.

Despite being highly conserved, there are many variants of the spiral cleavage program

Spiral cleavage is a highly conserved mode of cleavage in the Lophotrochozoa (see Chapter 4), yet there are differences in key molecular mechanisms. For example, there is a shift in timing of MAPK (mitogen-activated protein kinase) signalling activation, a marker of DV axis specification, between molluscs and annelids, where it is activated in the 3D macromere in molluscs and in the 4d micromere in annelids (Henry and Perry, 2008; Koop et al., 2007; Lambert and Nagy, 2003, 2001). In some cases MAPK does not seem involved in DV axis specification, such as in the polychaete annelid *C. teleta* (Amiel et al., 2013). Given the divergence in both timing and function of what was considered a conserved molecular mechanism, it is possible that a similar divergence has taken place in the highly conserved role of Nodal signalling in LR asymmetry establishment.

An important distinction is that most protostome *Nodal* expression studies have been done in gastropod molluscs, which have a distinct mode of developing a very different morphological asymmetry to annelids. In gastropod molluscs, the process of torsion takes place during the development of the adult (and shell), which annelids do not undergo. In addition, the shell is asymmetric by chirality rather than sidedness. It is possible that the inherent differences between molluscs and annelids in the developmental process of generating asymmetry, and the asymmetry itself, contribute to the employment of different pathways in asymmetry establishment.

Development and the evolution of Nodal signalling

Taking a step back, there are notable differences between the way in which bilateral animals develop, and this is reflected in the different tissues in which *Nodal* is expressed in different animals (Figure 3.9). Even within deuterostomes, there are differences despite the conserved molecular feedback loop between Nodal and Lefty. A recent study in *H. magnipapillata* proposes that beta-catenin functions upstream of the Nodal-pitx cascade in cnidarians, and that this linearity may be conserved in bilateral animals (Watanabe et al., 2014). Although this has not been rigorously tested yet across animals, it brings up the possibility that the presence of *Nodal* and *pitx* is very ancient, and that subsequent elaboration on this cascade has taken place in certain lineages. For example, Lefty has taken on a role in deuterostomes to sequester *Nodal* expression to one side, *Nodal* has been lost and pitx has taken on a function in neural development in the Ecdysozoa, and Nodal-pitx function in the Lophotrochozoa to establish LR asymmetry through an as yet unknown upstream mechanism (Figure 3.9).

Deuterostomes first exhibit symmetric *Nodal* expression prior to asymmetric *Nodal* expression with the function of the antagonist Lefty. In contrast, molluscs begin with asymmetric *Nodal* expression from the start and do not have Lefty at all. In this way, Lefty, and the way that this pathway functions, can be thought of as specific to deuterostomes, and the Lophotrochozoa has likely independently evolved a different mechanism to establish LR asymmetry using the same molecular machinery (Grande et al., 2014). Given that LR asymmetry establishment mechanisms vary considerably

across the bilaterian Superphyla, and indeed even within individual Phyla, this is a likely scenario.

Future directions

The presence of *Nodal* supports the conservation of *Nodal* in the Lophotrochozoa, however the lack of sided expression of *Nodal* and *pitx* in *P. lamarcki* raises more questions than it answers. There are a number of experiments that need to be done to understand whether this expression is consistent through all stages, with three main aspects that should be studied moving forward.

It would be important to understand whether Nodal still functions in LR asymmetry establishment, and whether the connectivity to pitx is also maintained. The latter is interesting, as this coupling is hypothesized to be ancient and conserved. These would most easily be tested with a drug that inhibits Nodal function to see whether sided operculum formation takes place upon inhibition, and whether *pitx* expression is affected by Nodal inhibition. If so, this would mean that Nodal still functions in LR asymmetry establishment despite a lack of a sided expression pattern, and that its connectivity to pitx is still present in *P. lamarcki*.

If the former is not true, and that Nodal does not function in LR asymmetry in *P. lamarcki*, It would then be interesting to consider why this loss of function may have occurred. An option is to look at regulators of Nodal, such

as NOMO, and see what this Nodal modulator is doing in the absence of a function in LR asymmetry establishment.

It would be best to eliminate torsion as a confounding factor in understanding the relationship between *Nodal*, chirality, and sidedness. This would require a mollusc species that does not develop a shell in the wild but has directional asymmetry in a similar examination of the presence, expression, and function of Nodal. The aim for this experiment would be to see whether the process of torsion, by which the chiral shell develops in molluscs, is what is employing Nodal rather than primarily as a pathway to establish asymmetry. If a shell-less mollusc that does not form a shell gland and does not undergo torsion has asymmetric expression of *Nodal*, this would indicate that Nodal signalling is not directly correlated with shell formation. Determining whether there are other LR asymmetries in these animals would also be essential and potentially difficult if the asymmetry is purely molecular and not morphological. If a shell-less mollusc does not have asymmetric *Nodal* expression, this would suggest that asymmetric *Nodal* expression may be tightly linked with the presence of a shell gland and the formation of a shell. If that is the case, such a study would be strengthened by a parallel analysis of chiral and non-chiral shells to see whether asymmetric *Nodal* expression is linked not only to shell formation, but also the development of chirality in shells. If asymmetric *Nodal* expression is linked to shell formation, this makes it likely that Nodal function in LR asymmetry is not conserved across deuterostomes and lophotrochozoans, but that Nodal signalling was independently co-opted by molluscs for the development of a chiral shell.

Chapter 4

Cytoplasmic flow in dextral and sinistral spiralians

Abstract

Spiralian development is a conserved mode of cleavage found in many species of the Lophotrochozoa such as molluscs, annelids, and flatworms. Cell fates are specified early in this mode of development, and symmetry breaking has been linked to the direction of spindle orientation at the 8-cell stage. Recently there has been a resurgence of interest in chirality within the embryo to understand symmetry breaking. However, this has predominantly been studied in model organisms such as mouse and *C. elegans*, where cortical flow is chiral. We have studied cytoplasmic flow in two species within the Lophotrochozoa, an annelid and a mollusc. We have analyzed the inner cell dynamics of an equal-cleaving annelid and mollusc at stages prior to the 8-cell stage, and have found that cytoplasmic flow at early cleavage stages correlate with cell division stages, and that this is in part linked to the actin and microtubule cytoskeleton. These results suggest that cytoplasmic flow is not a main player in polarity establishment, and supports cortical flow as a more likely candidate.

Introduction

The early cell divisions of the animal embryo are remarkably conserved across animal phyla, and in some cases have been used as a defining characteristic for subdividing Phyla within the animal kingdom. A key example is the spiral development of the majority of the members of the Lophotrochozoa, which has recently been rechristened the Spiralia (with some controversy; Paps et al. 2009; Giribet 2002; Edgecombe et al. 2011), and consists of a wide range of animals such as annelids, platyhelminthes, nemerteans, and molluscs (Costello and Henley, 1976). Recently, there has been a resurgence of interest in the dynamics of early cleavage in animal development, accompanied by an interest in mechanisms of symmetry breaking in the establishment of the left-right (LR) axis (Ajduk et al., 2014; Naganathan et al., 2014; Schonegg et al., 2014). Most studies have been done in model organisms such as *Caenorhabditis elegans* where the anterior-posterior (AP) and dorsal-ventral (DV) axes are established by the 4-cell stage through a combination of spatial constraints and spindle orientation, (Pohl and Bao, 2010; Pohl, 2011; Schonegg et al., 2014). However, this ecdysozoan does not undergo classical spiral cleavage, and most spiral cleavage studies in lophotrochozoans have focused on the 8-cell stage and later. Recent advances have made techniques increasingly available to organisms that had previously been hampered by the lack of generic technologies and resources, rendering it feasible to study this in non-model species.

Symmetry breaking in spiralian development

Symmetry breaking is an unsolved problem in developmental biology. In spiralian development, the establishment of the LR axis has been connected with the spiral arrangement of blastomeres at the 8-cell stage where sinistrality and dextrality, determined by the angle of spindle orientation at the embryonic stage, is correlated with that of the adult shell in molluscs (Freeman and Lundelius, 1982; Kuroda et al., 2009; Shibazaki et al., 2004). This correlation has exciting implications in that it forms an unprecedented connection between developmentally distant stages. The standing model holds that the spiralling direction of spiral development initiates at the 4-cell to 8-cell stages. What is often over-looked is that this is the first readily visible sign of spiralling, and not necessarily the first event determining dextrality or sinistrality. It remains a possibility that dextrality or sinistrality is present prior to the 8-cell stage, but we know little concerning this in the first few cleavage stages of spiralian development.

Spiralian development

Spiral cleavage is characterized by having a unique spiral arrangement of blastomeres. This spiral forms at the third cleavage, where the top tier of micromeres divides obliquely to the animal-vegetal axis (AV), and alternates this oblique cleavage to generate a spiral arrangement of blastomeres at each subsequent cleavage stage. However, this textbook description of spiral development is a simplification of what is actually a process with diverse modes of development. Rather than one stereotypical mode of spiral cleavage, there

are three (Figure 4.1): unequal cleavage, polar lobe development, and equal cleavage (Boyer and Henry, 1998; Child, 1899; Conklin, 1898; Costello and Henley, 1976; Lambert, 2010; Lillie, 1895; Whitman, 1878; Wilson, 1898, 1892). In unequal and polar lobe development, there is a clear size difference in blastomeres by the first few cleavages, and these size differences can be traced to specific cell fates later in development. This is not the case for equal cleavage, where the blastomeres remain the same size throughout early cleavage.

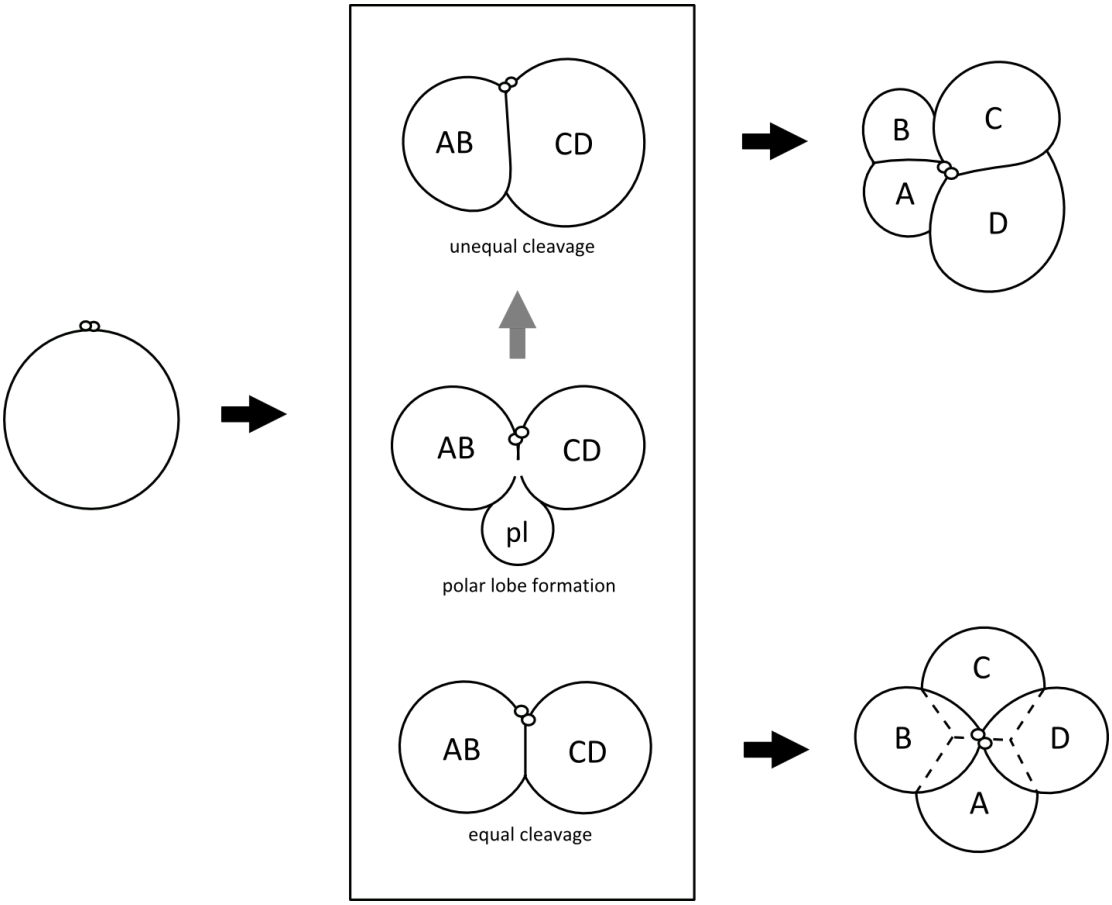


Figure 4.1 Spiral cleavage modes (adapted from Freeman and Lundelius, 1992).

Blastomeres of equal-cleaving spiralians are thought to be equipotent at the 2-cell and 4-cell stages (Arnolds et al., 1983; Martindale and Henry, 1995; Martindale et al., 1985; Morrill et al., 1973). However, there is considerable variation across species, especially in the ability to develop normally in deletion and cell isolation experiments, making it difficult to pinpoint exactly when developmental fates are first specified. When considering the LR axis, which is present by default with the establishment of the AP and DV axes, it becomes a conundrum to regard the LR axis as established at the 8-cell stage when presumably the DV axis may not have been established, at least in equal-cleaving spiralians (Figure 4.2). The DV axis is specified already at the 4-cell stage in unequal-cleaving spiralians, but is thought to be specified through cell-cell interactions at third quartet formation (24-cell stage) in equal-cleaving spiralians (Martindale et al., 1985) determined by the presence of MAPK signalling (Lambert and Nagy, 2001). At cleavage stages prior to third quartet formation, equal-cleaving spiralians develop remarkably normally upon isolation of macromeres at the 2-cell or 4 cell stages by recapitulating normal cleavage patterns (Morrill et al., 1973). This highlights the need for increased taxon sampling in developmental studies of spiralians, especially in equal-cleaving spiralians, as the use of a few reference species may be misleading in the face of the high level of variation seen in the development and life histories of species within the Lophotrochozoa.

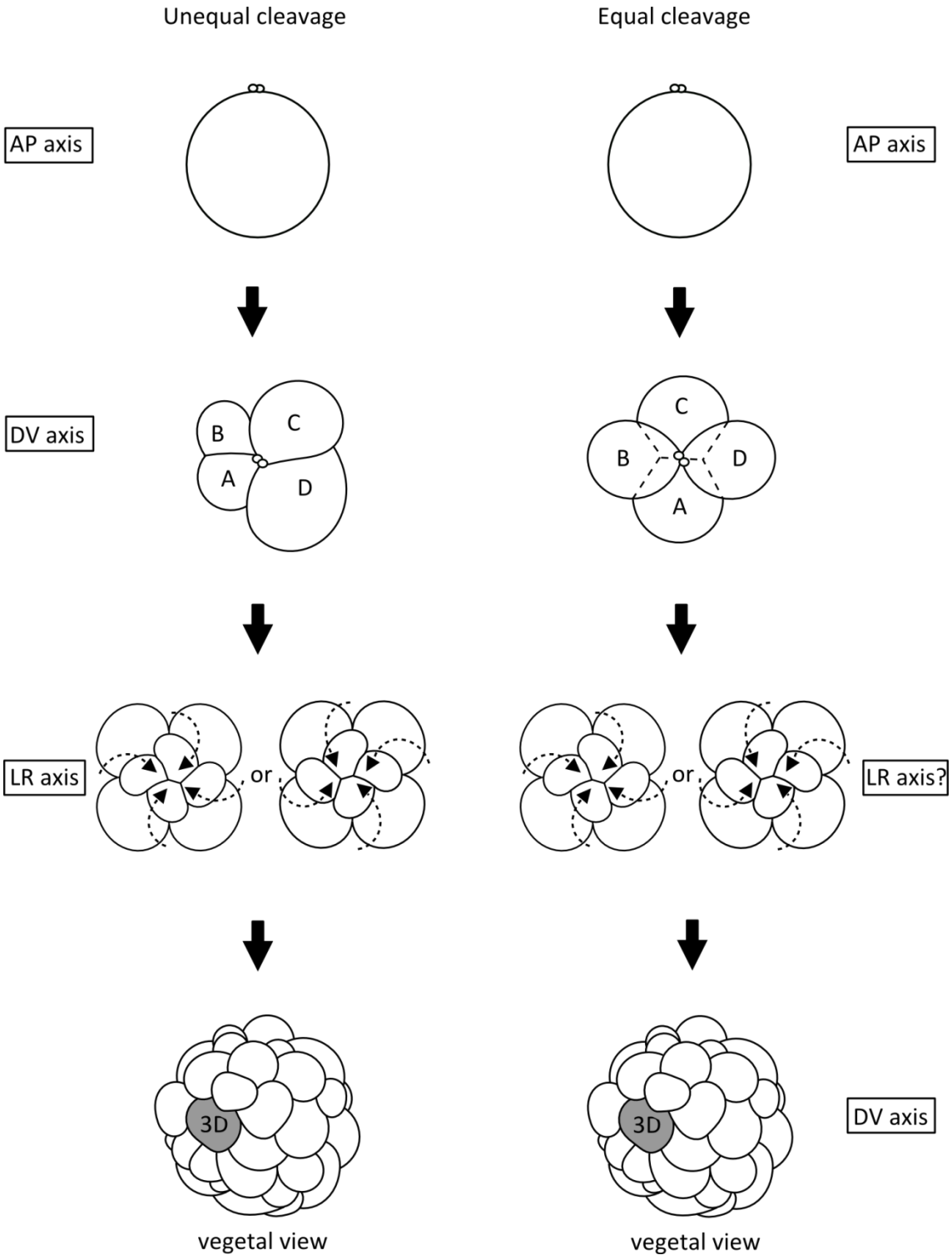


Figure 4.2 Axis specification of unequal and equal cleaving spiral embryos.

In this study, we have categorized the early cleavage stages of two equal-cleaving spiralians and found that there is considerably asynchrony in *Pomatoceros lamarcki* early cleavage, with the addition of a viable 3-cell stage. We also visualized and tracked cytoplasmic flow in *P. lamarcki* and *Biomphalaria glabrata* by fluorescently labelling intracellular vesicles. We found that flow is correlated with cell division at the 2-cell to 4-cell stages, and that these are likely in part dependent on both the actin and microtubule cytoskeletons in both species. Our results demonstrate that the chirality seen in other organisms is likely cortical chirality, and that cytoplasmic chirality is likely not the driving force of polarity establishment.

Methods

Animals collection and spawning

P. lamarcki adults were collected from the coast of Tinside, Plymouth, UK and maintained in recirculating sea water at approximately 12°C. *P. lamarcki* spawn upon being broken out of their tubes (Figure 4.3). Animals were spawned (as previously, McDougall et al., 2006) by breaking the end of the calcareous tube with tongs and breaking open a hole from the posterior end using a blunt probe. The animal was then pushed backwards out of this hole by pushing against the operculum at the anterior end using a blunt probe. Adults were then placed separately in filtered sea water (FSW) and allowed to exude eggs or sperm for 20 mins. The eggs were fertilized with a drop of a mixture of sperm from mixed males and imaged 2 hours thereafter.

B. glabrata adults were donated by Professor Rollinson (Natural History Museum, London, UK) and maintained in shallow containers of standing bottled water at 26°C. Natural breeding was allowed and egg masses (which are fertilized internally) were collected at approximately hourly intervals in order to capture necessary developmental stages (Figure 4.4). Egg masses were developed in the same conditions as the adults.

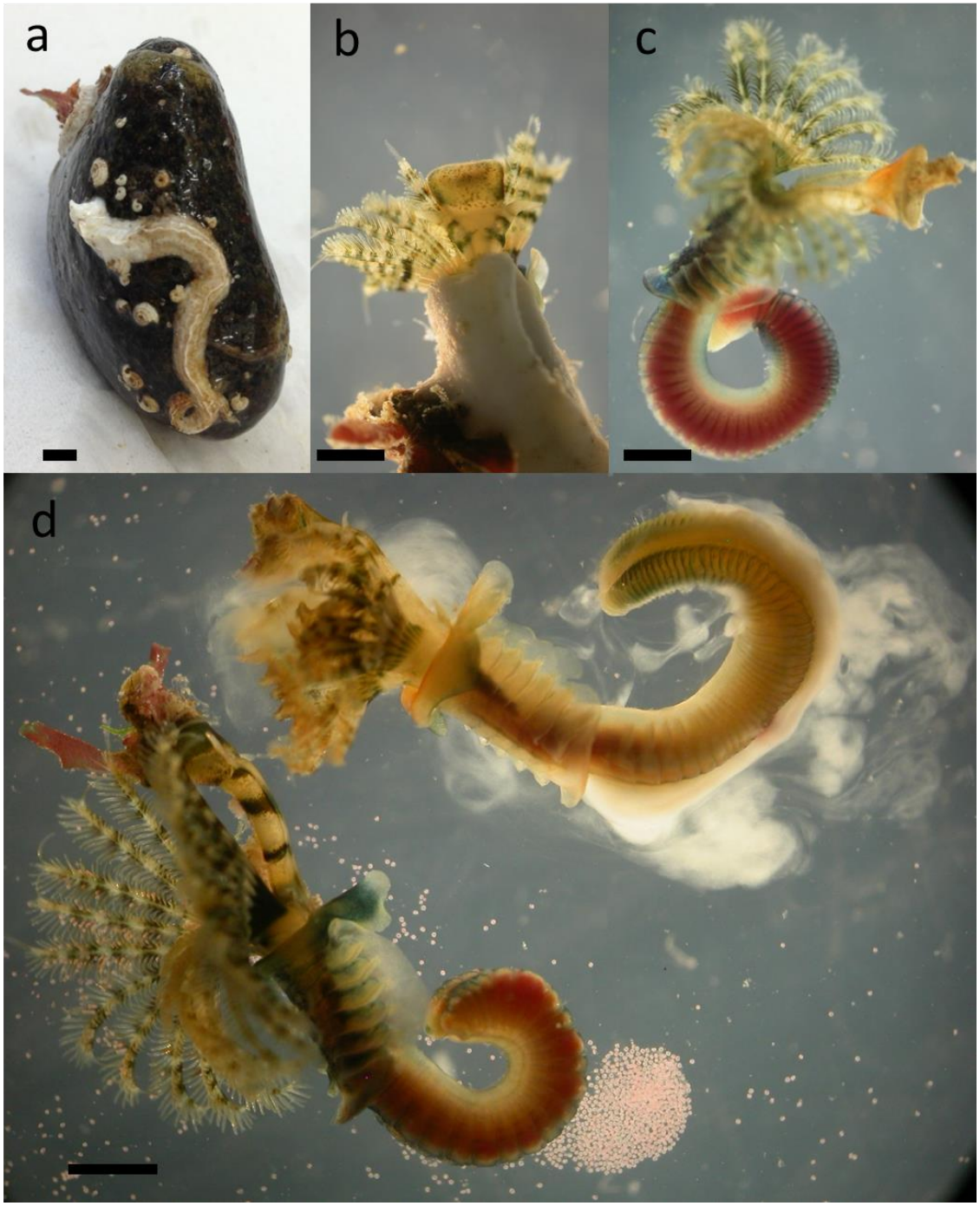


Figure 4.3 Spawning *P. lamarcki* adults. (A) Tube dwelling, (B) adult emerging from tube, (C) female adult, (D) female exuding eggs (bottom) and male exuding sperm (top). Scale bar approximately 1mm.

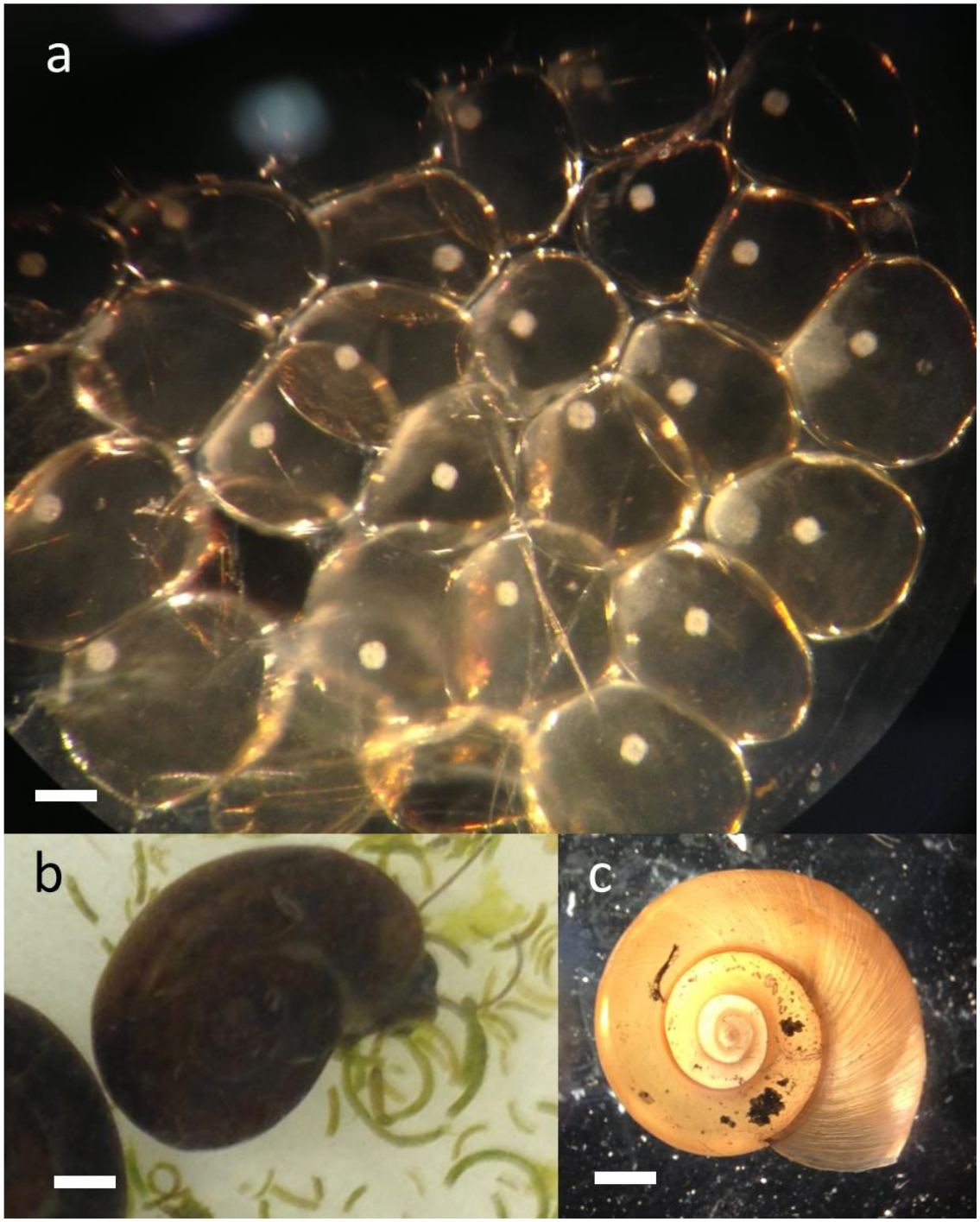


Figure 4.4 *B. glabrata* (A) egg masses, (B) adult, (C) adult shell. Scale bars approximately (A) 0.2mm and (B,C) 2mm.

B. glabrata *egg mass dissection*

Eggs of *B. glabrata* were dissected using fine forceps to break open individual egg capsules under a dissecting microscope. Eggs were gently sucked out of the broken egg capsule using a glass capillary (custom-made by pulling over a bunsen burner and breaking until opening was larger than 200 µm).

Fluorescent staining

Animals were rehydrated into PBTw, washed three times in PBTw at room temperature (RT), and incubated in block-PBTw (2%BSA/PBTw) at 4°C overnight. Block-PBTw was replaced with primary antibody at a 1:1000 dilution of anti-acetylated mono-beta-tubulin (Sigma) in block-PBTw at 4°C overnight. The primary antibody was washed off with four 20 min. washes in block-PBTw at 4°C, and replaced with secondary antibody at a dilution of 1:800 Alexa Fluor 596 (red, Molecular Probes) in block-PBTw at 4°C overnight. The secondary antibody was washed off with four 20 min. washes in block-PBTw at 4°C and embryos were permeabilized in PBTw at RT for one hour. Embryos were incubated in a 1:40 dilution of Oregon Green Phalloidin 488 (Invitrogen Molecular Probes®) in PBTw at RT for one hour. Embryos were washed three times in PBS at RT for 15 mins, immersed in 60% glycerol/PBS and mounted in Vectashield with DAPI (1:1000 dilution in Vectashield). Primary antibody, phalloidin, and DAPI were omitted for respective negative controls.

Confocal microscopy

Confocal images were taken at a magnification of 400x using the Olympus FV1000 Fluoview confocal microscope at RT.

Timelapse imaging

Embryos were imaged at RT under a dissecting scope using a Nikon timelapse recorder (Nikon, Digital Sight DS-L1). See supplementary video (Video 1) in accompanying DVD+R.

B. glabrata *transgenics trials*

A chicken beta-actin GFP vector (Graham et al., 2007) was used for electroporation trials at a concentration of 0.7µg/µl using a NEPA21 CuvetteEP with 70µl mannitol, 3.5 µl DNA and 30µl of solute/embryo. A map of the vector is not publicly available, and would need to be requested from the authors. Optimal parameters were converged upon with trial runs. *B. glabrata* embryos were injected with 1% FastGreen and bottled water using a pulled glass capillary connected to a mouth pipette and a 10-ml syringe prior to electroporation.

Dechorionation and live dye trials

Embryos were trialled for dechorionation and the introduction of live dyes with the below parameters (Table 4.1). All of these methods were done by

incubation of a certain concentration of dechorionating solution or live dye in 500µl of sea water and embryos at specific stages. Mitotracker-stained sperm were drawn up in a 10-ml syringe and added to a Mattek dish with embryos in sea water under the confocal microscope while live recording to try to visualize the sperm entry point.

Table. 4.1 Dechorionation methods and live dyes trialled for *P. lamarcki* embryos.

dechorionation/live dyes	concentration	incubation time	stage
DECHORIONATION			
Sucrose/NaCitrate	1:1	5, 10, 20, 30 secs., 1, 2 mins	0-2 cell
Protease	0.1%	1, 2, 3, 4, 5 mins.	0-2 cell
Trypsin	1x	2, 5, 10, 15, 20 mins	0-2 cell
Sodium Thioglycolate	2%	1, 2, 3, 4, 5 mins.	0-2 cell
2% Sodium Thioglycolate/1% protease	1:1	1, 2, 3, 4, 5 mins.	0-2 cell
20% Pluronic acid	0.1%	1 hour	0-2 cell
Pronase	1x	7 mins	0-2 cell
Pronase-trypsin	1:1	7 mins	0-2 cell
mechanical (pipetting)	n/a	2 mins	0-2 cell
LIVE DYES			
DRAQ5	0.001, 0.002 µl/ml	5 mins	0-2 cell
MitoTracker	50nM	5 mins	sperm
TubulinTracker	0.001, 0.002 µl/ml	1 hour	0-2 cell
Calcium	0.001, 0.002 µl/ml	1 hour	0-2 cell
LysoTracker	0.001, 0.002 µl/ml	1 hour	0-2 cell
FM-64	0.001, 0.002 µl/ml	1 min	0-2 cell

Drug treatment

Embryos were treated with Nocodazole (0.2ug/ml, Sigma) and Cytochalasin B (10ug/ml, Biochemica) at RT for five minutes and imaged in 1% low melting point agarose (Promega, V2111) in sea water or bottled water. Agarose was infused with Nocodazole (0.4ug/ml) or Cytochalasin B (20ug/ml) prior to mounting.

Vesicle staining

Embryos (2 hours post fertilization) were stained with 0.001mM Lysotracker (Red DND-99, Molecular Probes) for 5 mins at RT and diluted into low melting point agarose (Promega, V2111) (Figure 4.5). See supplementary videos (*P. lamarcki* Videos 2-5, *B. glabrata* Videos 6-9) in accompanying DVD+R.

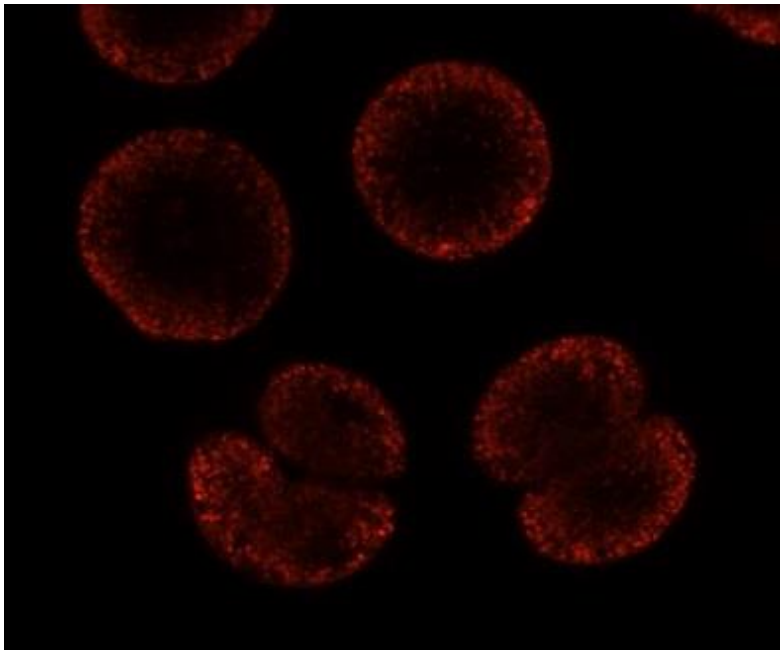


Figure 4.5 LysoTracker-stained *P. lamarcki* embryos.

In-depth methods for Lightsheet microscopy

A brief introduction to Lightsheet microscopy

The technology underlying the Lightsheet microscope was developed as early as the early 1900's, but was only popularized and commercialized recently with the development of Selective/Single-plane Illumination Microscopy (SPIM) in 2004 (Huisken et al., 2004). Since the commercialization of the Lightsheet microscope, it became apparent that the novelty of the data produced, in addition to the enormous size of the image files, demanded significant innovation and troubleshooting before it could be widely used. For the past 10 years, both specimen processing and data analysis have moved forward, but methods are still under development (Reynaud et al., 2014). This prevents the use of this microscope from being a smooth process, and in this way is less advantageous in comparison to well-tried microscopes such as the standard confocal microscope. Its novelty makes it difficult to adapt to different systems. Here, the Lightsheet has been adapted to studies of inner cell dynamics in small marine and freshwater invertebrate embryos (60-120 µm in diameter), and the data have been analyzed for directionality of movement in individual blastomeres at the first few cleavage stages. Below, a general overview of Lightsheet microscopy, and the methods developed in this thesis, will be outlined and discussed.

When analyzing live animals, there are a number of microscopes that could be used that all come with their respective pros and cons. When it comes to imaging very small embryos at early cleavage stages, it becomes important to select a method that inflicts the least amount of harm to the specimen. This is

where Lightsheet microscopy excels, and for this reason it was selected for these experiments, which examine the inner cell dynamics of the first few cleavage stages of spiralian embryos.

Traditional advanced microscopy depends on the general principles of confocal microscopy. In confocal microscopy, the entire specimen is illuminated each time, even though only a portion of this is detected through a pinhole. This means that with each iteration, the fluorescence efficiency decreases, and the animal is subject to harmful laser light. This makes this method phototoxic, prone to photobleaching, and harmful to the animal. Lightsheet microscopy circumvents these problems by illuminating a single plane of the specimen using a sheet of light, rather than illuminating the entire specimen, and by only illuminating the plane that is being detected. This provides intrinsic optical sectioning.

In addition to being gentle, the sheet of light also allows faster data-gathering, as it does not depend on detecting the illumination by scanning pixel by pixel as in confocal microscopy, but detects fluorescence at once in the entire sheet of light. An additional advantage is the ability to change the angle at which the specimen is imaged. Rather than having the specimen in a flat dish or on a slide, that can only be moved using the mounting stage, specimens in the Lightsheet are suspended, and can be rotated to whatever angle is best for the experiment. This also makes it possible to visualize the specimen at all angles, although this was not done here.

Lightsheet data collection

Data were collected on the Z.1 Lightsheet microscope (Zeiss) using a 20x detection objective and 10x illumination objectives. Embryos were embedded in 1% low melting point agarose/sea water or bottled water and imaged in capillaries provided with the microscope at RT for *P. lamarcki* and at 26°C for *B. glabrata*. Details of this method are detailed below.

Animal preparation for Lightsheet imaging

Animals were prepared by spawning and fertilizing in the case of *P. lamarcki*. For *B. glabrata*, which fertilize their embryos internally, the snail environment was checked daily, as often as every two hours, in order to find egg masses that had just been laid and had not undergone cleavage yet. Egg masses were grossly indistinguishable from each other and needed to be checked under a dissecting scope. Once embryos were fertilized, or relevant stages found, there was a 2-hour window during which the sample was prepared for imaging (Figure 4.6). This 2-hour window was needed to wait for the appearance of 2-cell stage embryos (which took approximately two hours to appear after fertilization), in order to make sure that the embryos were viable prior to imaging.

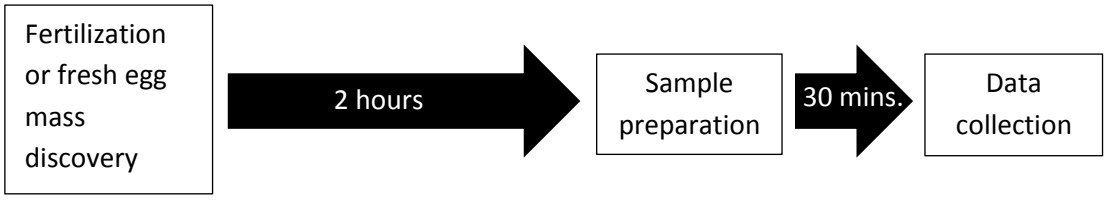


Figure 4.6 Sample preparation timeline for Lightsheet imaging.

2% low melting point agarose was prepared before the day of imaging with the solute that the particular animal lived in (sea water or bottled water), heated to 72°C, and allowed to cool and solidify. On the day of imaging, previously prepared 2% agarose was melted at 72°C just after fertilization of embryos (*P. lamarcki*), or upon the discovery of freshly-laid embryos (*B. glabrata*) to allow ample time for the agarose to fully liquify. Metal plungers and glass capillaries were completely submerged in sea water or bottled water to prevent excessive capillary action from taking place, and relevant dyes or drugs were defrosted.

At the first appearance of a 2-cell stage embryo, there is a 30-minute window during which the specimen is fluorescently labeled, treated with any drugs, embedded in agarose, transported to the microscope, and mounted onto the microscope. In this case, embryos were transferred to a microfuge tube in 200 µl of the respective solute, and in the case of *B. glabrata* embryos were dissected out of egg masses as detailed previously (see *B. glabrata* egg mass dissection). Specimens were treated with LysoTracker by incubation for 5 mins, during which the metal plunger and glass capillary were dried and assembled. Drug treatment followed in the case of treatment.

Embryos and dye were diluted by half with the previously prepared 2% low melting point agarose for a final concentration of 1% low melting point agarose, and mixed gently. In the case of drug-treatment, this agarose was infused with double the concentration of initial drug treatment. The agarose-adding step is very time-dependent, as it is imperative that the agarose does not solidify prior to drawing up the specimen. Taking the specimen to a dissecting scope, animals were gently drawn up the glass capillary by pulling up

the metal plunger, all while preventing bubbles from entering the capillary. This needs to be done gently but swiftly, as the agarose gradually solidifies as it cools. When drawing up the specimen, it is also important that they are drawn up in single file to prevent blocking of the sheet of laser light upon illumination under the microscope. Once the appropriate amount of agarose is drawn up the capillary, just the end of the capillary is submerged in the respective solute to prevent desiccation, and the capillary/embryo/plunger ensemble is transported to the Lightsheet microscope.

Sample chamber preparation and sample mounting

Upon arrival at the microscope, the sample chamber was constructed by assembling three platforms, the sample chamber, and three windows with coverslips and O-rings using a screwdriver and an Allen key. The objective and illumination lenses were mounted into the microscope by hand, and the prepared chamber was screwed into the microscope. The sample chamber was filled with 20-30ml of sea water or bottled water.

A separate mounting adaptor was assembled by connecting a capillary holder, capillary adaptors, microscope adaptor, and capillary stabilizer. The prepared sample ensemble was placed into the mounting adaptor and placed into the microscope.

Data collection

Following mounting of the sample, data collection was done through the Zeiss (ZEN) software, where the sample was lowered into the sample chamber and oriented to be coincident with the sheet of laser light and a biologically appropriate orientation using the software. The microscope was set to relevant timelapse and z-section parameters and images were acquired. These settings were: exposure time 30.0ms, lightsheet thickness 3.09µm, zoom dependent on specimen, acquisition <3ms, laser 561nm at 0.2000%, approximately 70 sections. Single-side illumination was used where the sample was illuminated from one side, causing the other side of the embryo to be less-efficiently illuminated. In this case, only half of the embryo was imaged. This could have been remedied by using dual-side illumination to capture the entire embryo rather than half, however this was not done in order to reduce the file size of the data and to make it manageable to analyze. In order to capture gross inner cell movements, half of the embryo should give us a clue as to whether there is directionality of movement while taking into account that circular movements of larger circumferences will be difficult to infer.

Data analysis

The analysis of the files produced by the Lightsheet is challenging due to the large size of the data (approximately 810GB per file), and the fact that methods of analysis are all under development and untested. An analysis method for vesicle tracking was created here (Figure 4.7) using open-source resources.

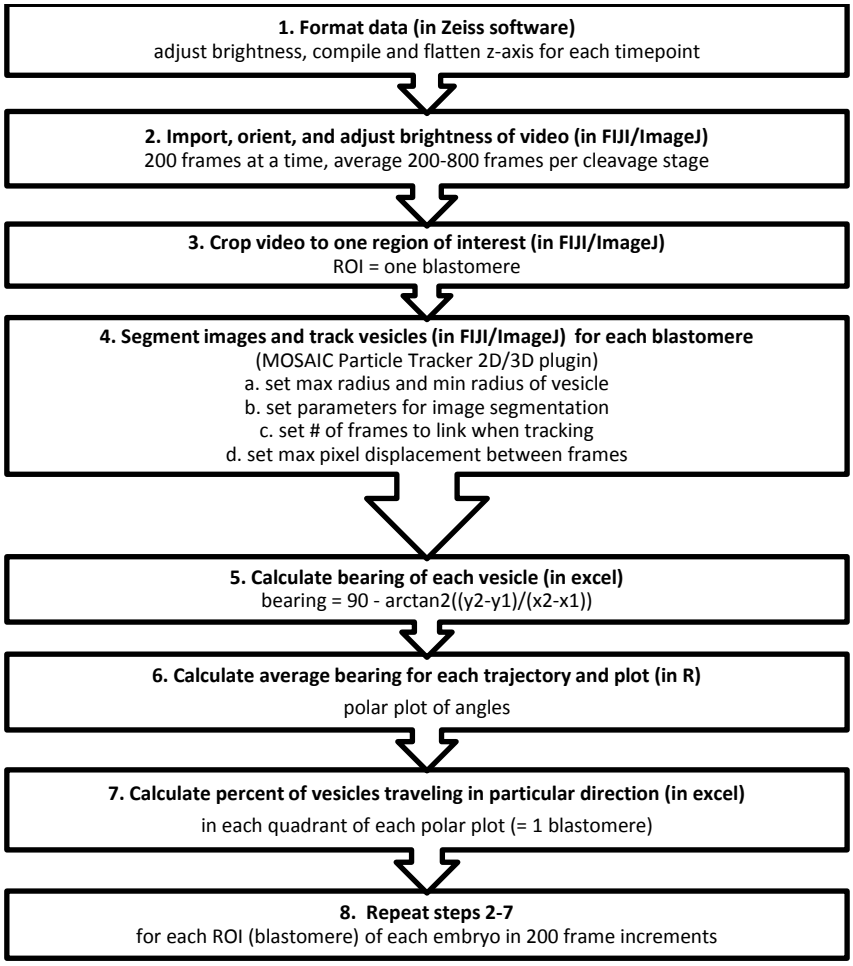


Figure 4.7 Single blastomere vesicle tracking protocol.

Vesicle tracking

Images were compiled using maximum intensity projection to flatten the z-axis (ZEN software) into a timelapse video and saved as an AVI (Audio Video Interleave) file, as the CZI (Carl Zeiss) file format produced by the Lightsheet is incompatible with the processing software used. Videos were too large to be imported directly into FIJI (Fiji Is Just ImageJ; Schindelin et al., 2012). The plug-in BIOFORMATS (Linkert et al., 2010) was therefore used, and videos were imported in 200 frame increments to prevent using up all of the computer RAM. Upon importing, videos were oriented and cropped to a single region of interest

(ROI). Each blastomere was divided into four spatial quadrants, each quadrant corresponding to one ROI (resulting in each blastomere having four ROIs).

Vesicles were tracked in each ROI using MOSAIC Particle Tracker 2D/3D (Helmuth et al., 2010; Shivanandan et al., 2013) in 200 frame increments (Note: the term “vesicle” and the more general term “particle” are used interchangeably in this thesis, with the former in reference to staining and the latter in reference to analysis methods). For *P. lamarcki* at a zoom of 2.5, maximum particle radius (r) was set at 13 pixels as lysosomes are at a range of 0.1-1.2µm (or 0.9257-12.093 pixels), the minimum radius (c) was set at 0.5 pixels to account for smaller vesicles at farther depths, the percentile (p) was set at 1 where particles in the upper first percentile of the image intensity distribution are accepted as a segmented particle, the number of frames (L) used to link optimal trajectories was set at 2 (this sets how many frames can overlap yet still maintain distinct vesicle identities, and the maximum pixel displacement (D) between frames was set at 75. For *B. glabrata* at a zoom of 2.0, the above settings were: r=11, c=0.5, p=1, L=2, D=75, and for zoom 2.5 all were maintained the same except for r=14.

Vesicle tracking data were used to calculate the bearing of each vesicle in excel and saved as a CSV file using the following formula: 90- $arctan2(y2 - y1) \div (x2 - x1)$. This formula calculates bearing (arctan2) relative to the normal (“90-“). An R (Team, 2008) script was written to generate polar plots for all bearings in each ROI (four for each blastomere) using the Rcpp (Eddelbuettel and Fran, 2011), plyr (Wickham, 2011), plotrix (Lemon, 2006), and ggplot2 (Wickham, 2009) packages (Figure 4.8) and compiled to one PDF file using the multiplot function in R.

```
data1 <- read.table("file.csv", header=TRUE, sep=",", na.strings="NA", dec=".", strip.white=TRUE)
                                                        #import data to R
id1 <- data1[,c(1)]                                     #call trajectory data
angle1 <- data1[,c(11)]                                 #call angle data
angleidspatial1 <- data.frame(id=id1, angle=angle1)     #compile into one dataframe
write.csv(angleidspatial1, "file1.csv")                 #write to separate file
p1 <- ggplot(angleidspatial1, aes(x = angle)) +         #make polar plot
        coord_polar(start = 0, direction=1) +           #set 0 as normal
        geom_bar(width = 1, colour = "black", binwidth = 10)+
                                                        #set bar width/color
         scale_x_continuous(breaks=seq(0, 360, by=30), expand=c(0,0), lim=c(0, 360)) +
                                                        #set plot to degrees
        scale_size_area() +                             #set plot area
        theme_bw()                                      #remove default background
print(p1)                                               #print plot to screen
dev.copy(pdf, 'file1.pdf')                              #save plot as pdf
dev.off()                                               #close plot
```

Figure 4.8 R code to generate polar plots. Description of each script follows with a hashtag.

Flow analysis

Particle image velocimetry (PIV) analyses were done using the PIV plugin (Tseng et al., 2012). PIV measures the velocity of labeled particles, termed optic flow, in subregions of an image. This is done for pairs of images and the data are cross-correlated. During this process, the interrogation window size (the subregions of the image) are gradually decreased, which allows for a better PIV resolution. For *P. lamarcki*, two passes of PIV were done, whereas for *B. glabrata*, three passes of PIV were done, both with decreasing window sizes using the basic cross-correlation method.

Results

P. lamarcki *embryos undergo equal cleavage and adults have directional operculum asymmetry*

P. lamarcki embryos are approximately 60 µm in diameter and undergo equal cleavage, where all blastomeres are the same size during equal cleavage (Segrove, 1941) (Figure 4.9). *P. lamarcki* adults have sided operculum asymmetry, where the operculum is consistently found on the right side with no adult (N=2227) displaying mirror-image symmetry (Figure 4.10, Table 4.2).

B. glabrata *undergoes equal cleavage*

B. glabrata embryos are approximately 120 µm in diameter and undergo equal cleavage, where all blastomeres are the same size during equal cleavage (Camey and Verdonk, 1969) (Figure 4.11). Pigmentation is seen in the vegetal portion of the embryo, which is likely associated with yolk in that region.

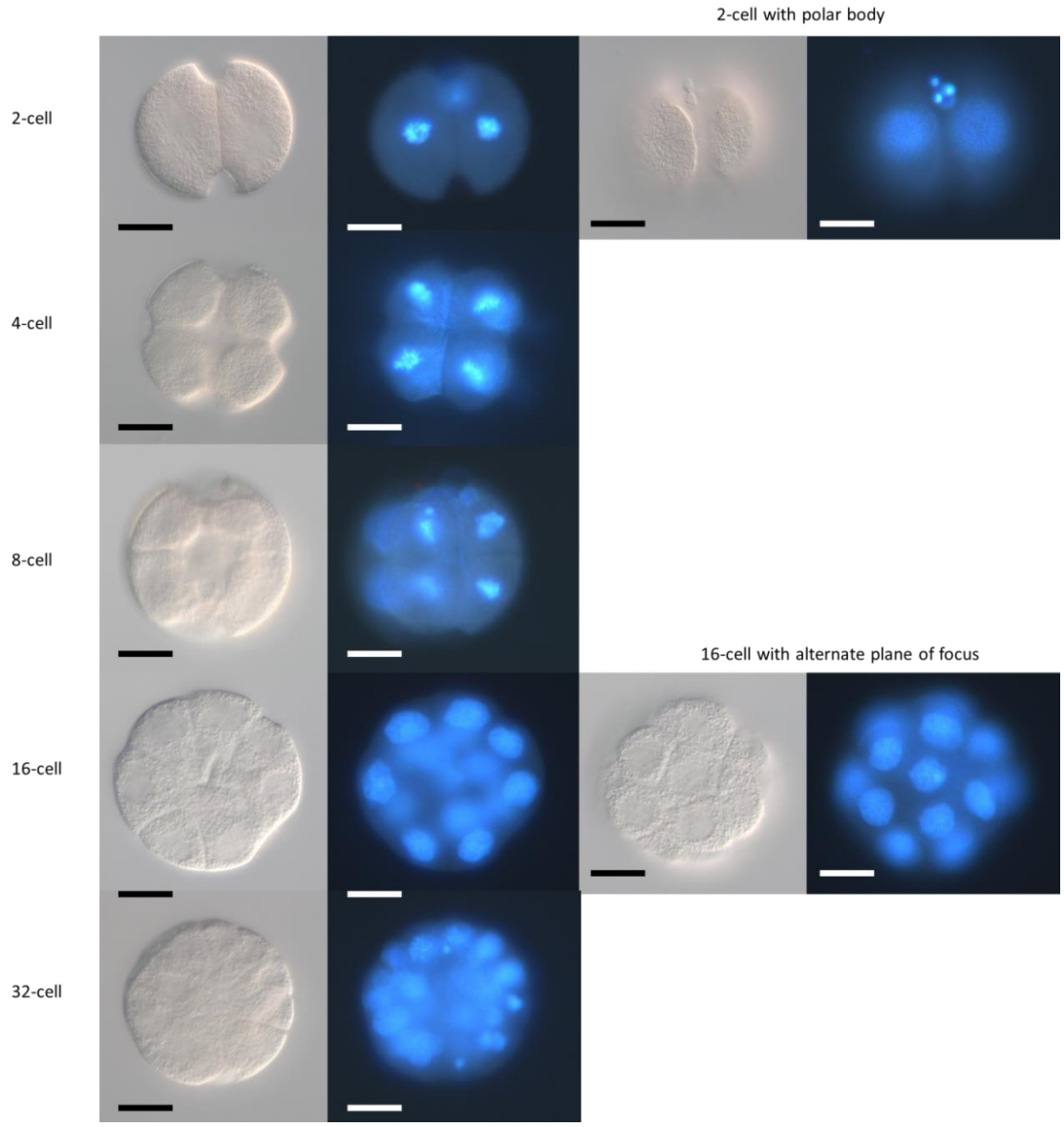


Figure 4.9 *P. lamarcki* early cleavage stages visualized by DIC and DAPI staining (DNA). Right panels show polar bodies at the 2-cell stage and an alternate plane of focus at the 16-cell embryo. 400x magnification, scale bar 20 µm.

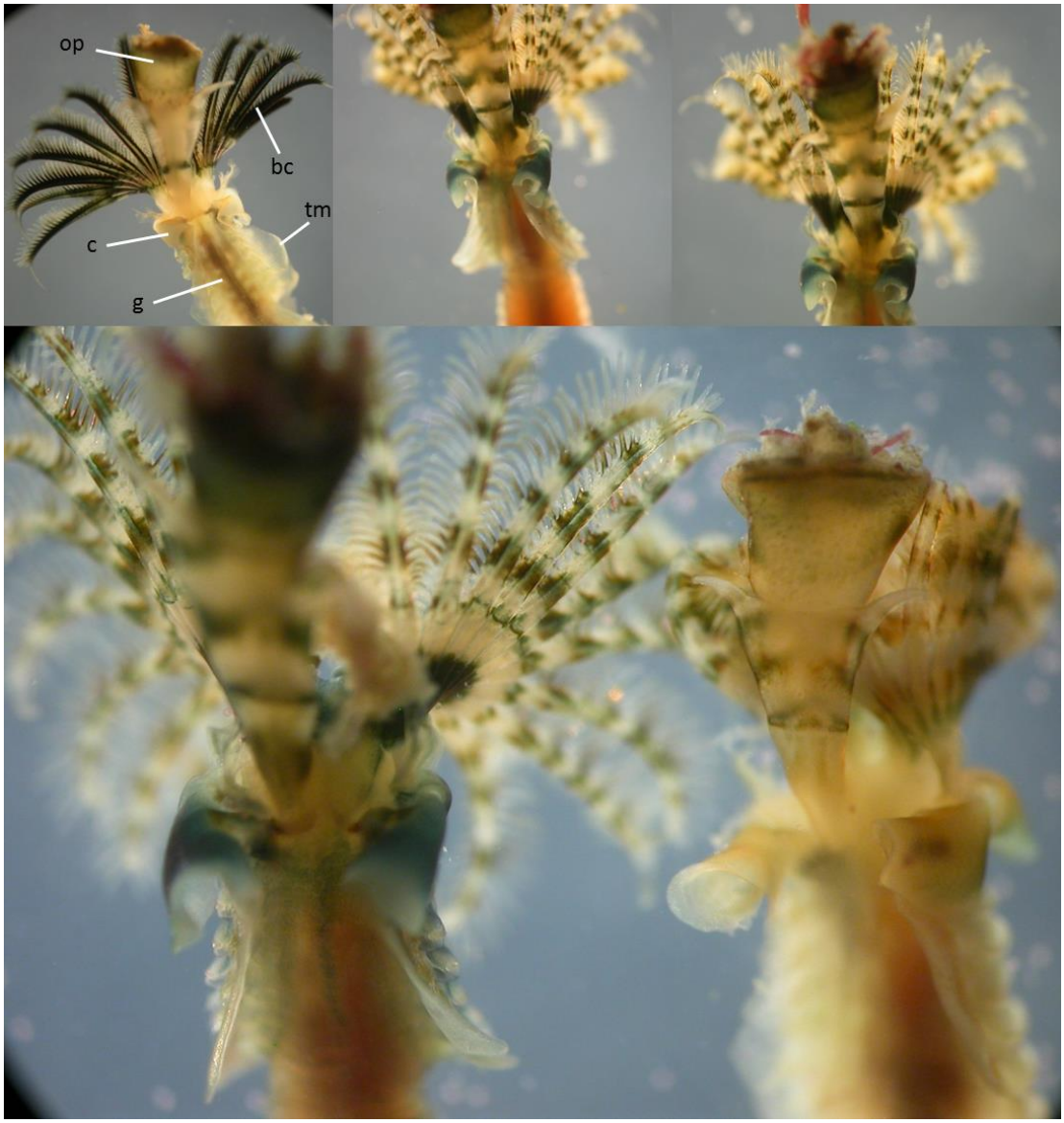


Figure 4.10 *P. lamarcki* operculum sidedness (right side placement). Operculum (op), branchial crown (bc), collar (c), gut (g), thoracic membrane (tm).

Table 4.2 *P. lamarcki* operculum sidedness.

Operculum asymmetry	**male**		**female**		**unknown**	
Side	**L**	**R**	**L**	**R**	**L**	**R**
Total	0	519	0	535	0	1173
Grand Total	**L**	**R**				
	0	2227				

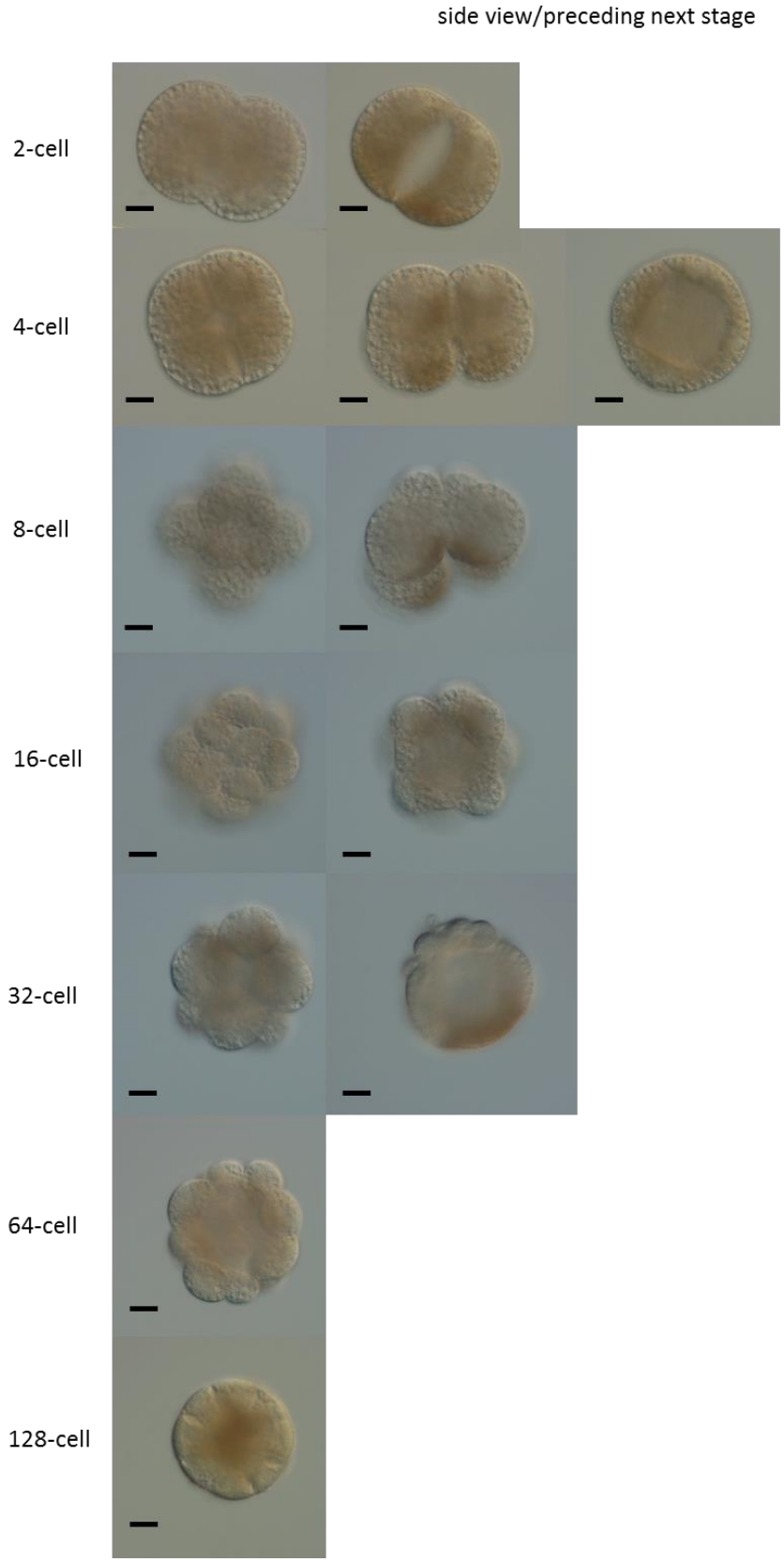


Figure 4.11. *B. glabrata* early cleavage stages visualized by DIC. Right panels show side views of that stage and intermediate stages preceding the next cleavage stage. 400x magnification, scale bar 20 µm.

P. lamarcki *but not* B. glabrata *embryos cleave asynchronously at the 2-cell stage*

Embryos of *P. lamarcki* can cleave asynchronously at the 2-cell stage, where in some embryos there is a transition through a 3-cell stage rather than synchronous cleavage from the 2-cell to the 4-cell stage (Figure 4.12). Embryos that transitioned through a 3-cell stage typically stayed as a 3-cell stage embryo for 30 minutes before proceeding to divide into a 4-cell stage embryo (Figure 4.12, Video 1). In contrast, *B. glabrata* embryos cleave synchronously at the second cleavage stage with no transition through a 3-cell stage (Figure 4.13).

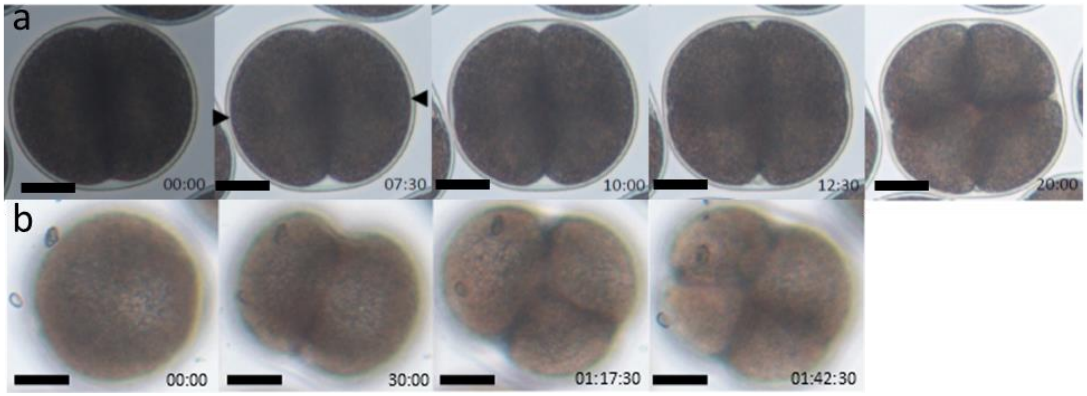


Figure 4.12 *P. lamarcki* early cleavage stages visualized by DIC extracted from timelapse videos. (A) 2-cell to 4-cell stage transition with timing of division extracted from timelapse videos, (B) 3-cell stage with timing of division extracted from timelapse videos. Time is in hours:mins:secs or mins:secs, scale bar is 20 µm.

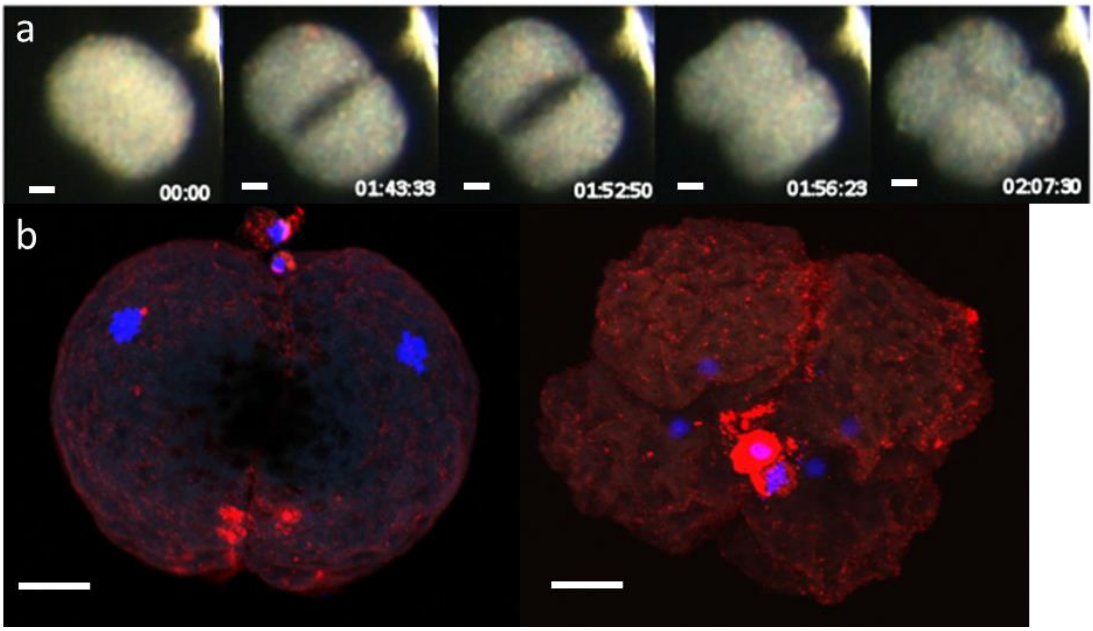


Figure 4.13 *B. glabrata* early cleavage stages visualized by DIC and confocal imaging. (A) 2-4 cell stage transition with timing of division extracted from timelapse videos, (B) 2-4 cell transition with confocal imaging. Confocal images with labelled DNA (blue) and beta-tubulin (red). 400x magnification, scale bar is 20 µm.

Confirmation of 3-cell stage embryos

The 3-cell stage was analyzed further using confocal microscopy in order to ensure that this cleavage stage was not the result of incomplete cytokinesis producing an awkwardly-shaped blastomere at the 2-cell stage without real cell boundaries. Embryos were stained for DNA using DAPI, phalloidin for actin filaments, and monoclonal beta-tubulin antibody to visualize embryos in 3 dimensions (Figure 4.14). In order to capture the 3-cell stage during the approximately 30-minute window of its existence, embryos were fixed at various time points at the 3-cell stage in addition to 4-cell and 8-cell stages for comparison and imaged by confocal microscopy (Figure 4.14).

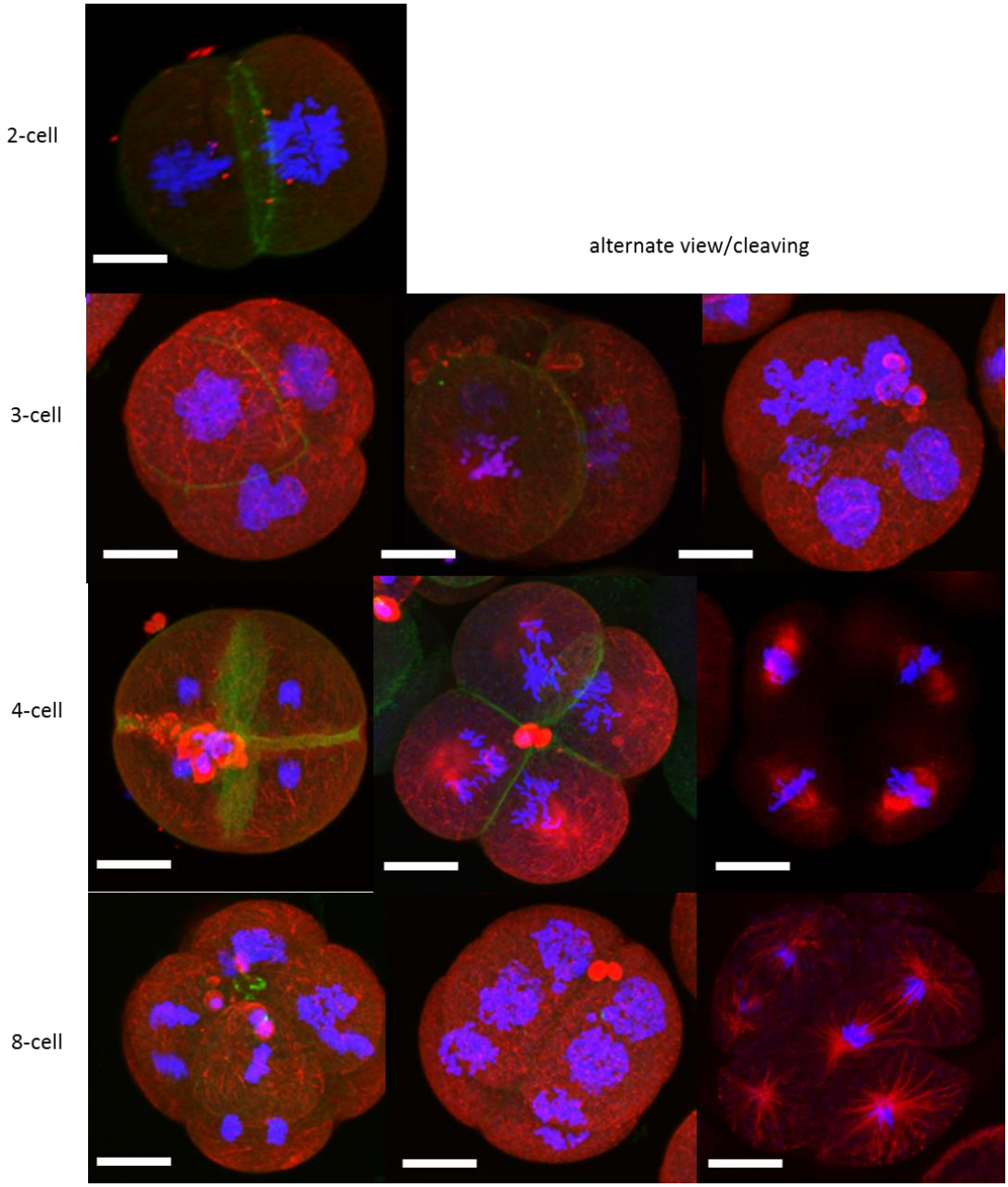


Figure 4.14 *P. lamarcki* early cleavage stages visualized by confocal imaging. Confocal images with labelled DNA (blue), actin (green), beta-tubulin (red). 600x magnification, scale bar is 20 µm.

At the 3-cell stage, embryos clearly had intact cell membranes between all three blastomeres, indicating that there are three bona fide blastomeres at the 3-cell stage rather than an incompletely-cleaving blastomere at the 2-cell stage. As the 3-cell stage transitioned to a 4-cell stage, the appearance of the

third blastomere appears as if one blastomere had simply been delayed in completing cleavage to a 4-cell stage embryo. At the 4-cell stage, embryos at metaphase already exhibited angled behavior associated with the spiral 8-cell stage (Figure 4.14). At the 8-cell stage, the oblique orientation of the blastomeres can best be seen in embryos that are in the midst of cleaving, where it becomes clear that the blastomeres are not aligned linearly.

In order to quantify that the 3-cell stage embryos were able to survive and develop normally, the survival of 3-cell embryos were individually monitored by eye to check for transition through the 3-cell stage. The survival of the embryos to the trochophore larval stage was quantified, where 38.1% (16/42) of embryos that survived transitioned through a 3-cell stage (Table 4.3, $p>0.1$).

Table 4.3 Asynchrony in *P. lamarcki* second cleavage: 3-cell survival quantification (N=96). Pearson's chi-square test indicates no significant difference ($p>0.1$) between survivability of embryos that transition through a 3-cell stage and those that do not transition through a 3-cell stage (χ^2 statistic = 3.526, $p=0.1$ is $\chi^2=2.706$ at df=1, therefore $p>0.1$).

Category	alive	dead	Overall Total
2-cell to 4-cell with 3-cell transition	16	31	
2-cell directly to 4-cell	26	23	
Total	42	54	96

Analysis of cytoplasmic flow in P. lamarcki *and* B. glabrata

Cytoplasmic flow was measured by fluorescently labelling vesicles in fertilized embryos, recording on a Lightsheet microscope, and tracking the direction of movement of each vesicle both in wildtype and drug-treated animals

(Figure 4.15, Videos 2-9). For particle tracking, this was done by dividing each blastomere of each stage into four spatial quadrants, whereas for PIV analysis this was done at once across the entire blastomere to get a sense of overall flow (Figure 4.16).

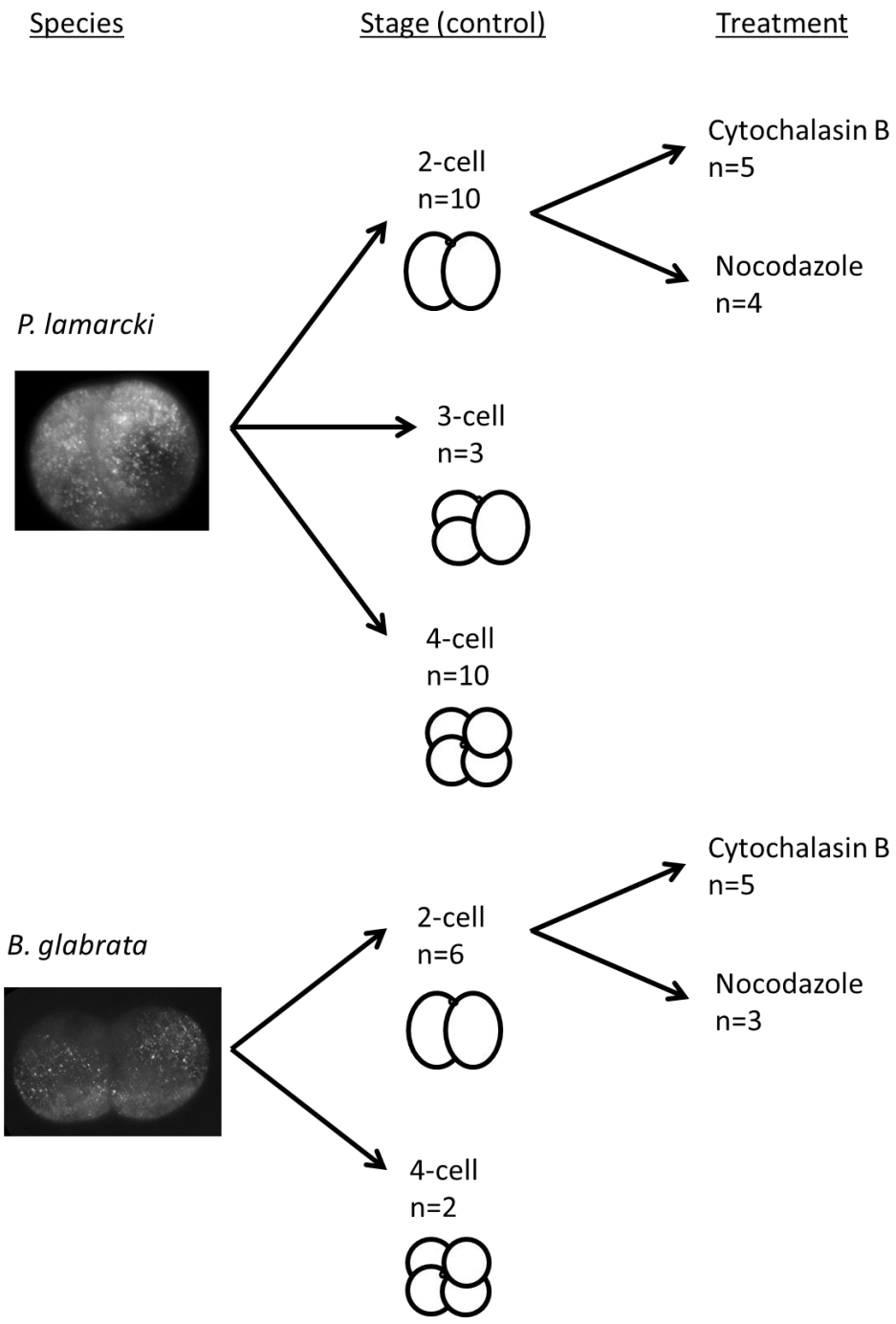


Figure 4.15 Breakdown of cytoplasmic flow analysis of Lightsheet imaging data by species, cell stage, and drug treatment.

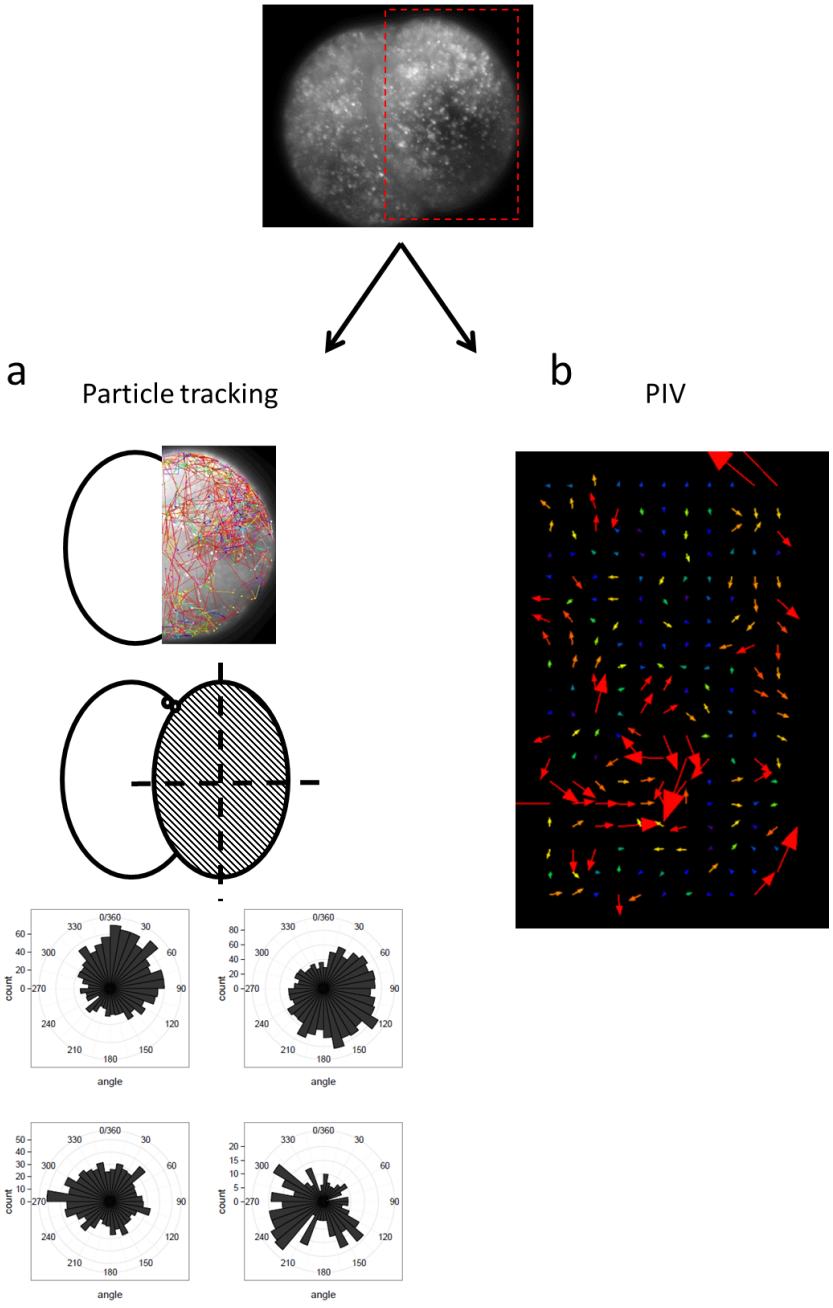


Figure 4.16 *P. lamarcki* particle tracking and PIV analyses of Lightsheet imaging data (A) particle tracking of labelled vesicles by division of one blastomere into four spatial quadrants, (B) PIV analysis of flow for entire blastomere.

For both particle tracking and PIV analyses, this was done in 200 frame increments (Figure 4.17). The bearing of each spatial quadrant of each blastomere was plotted on a 360° plot for particle tracking (Figure 4.16 and Figure 4.17), and relative movement was analyzed in conjunction with the PIV

analysis. The main bearing for each particle tracking plot was surmised by taking the three largest counts for each plot, and together with the spatial PIV plots were summarized to give an overall view of the cell stage (Figure 4.18).

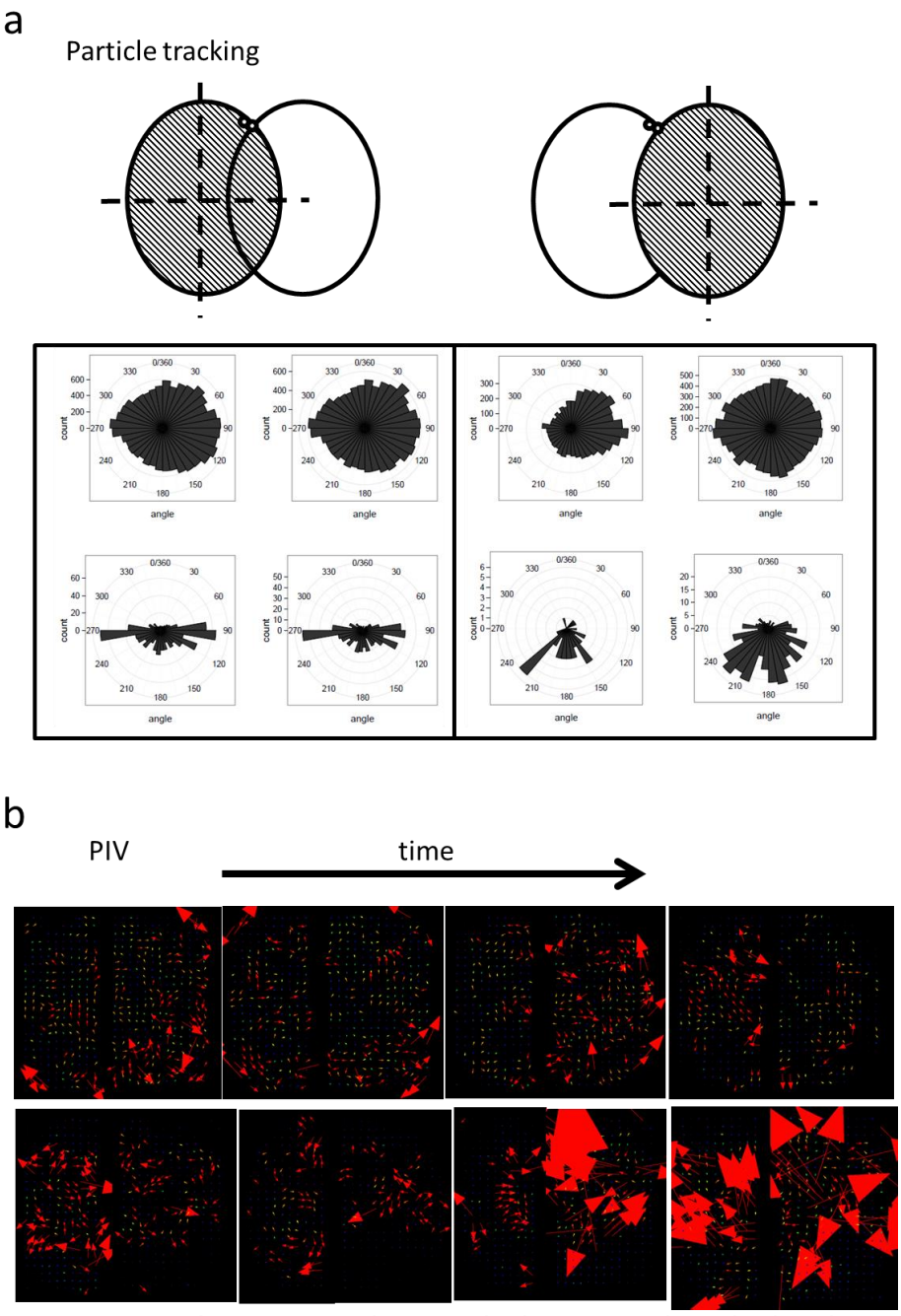


Figure 4.17 An example of a complete analysis of cytoplasmic flow in a 2-cell *P. lamarcki* embryo using Lightsheet data. (A) Particle tracking with analyses of four spatial quadrants (represented by four graphs) in each blastomere of a 2-cell stage embryo. (B) PIV analyses of the same 2-cell embryo with each PIV image being 200 frames at 5 frames/sec, or 40 seconds of recording per PIV image.

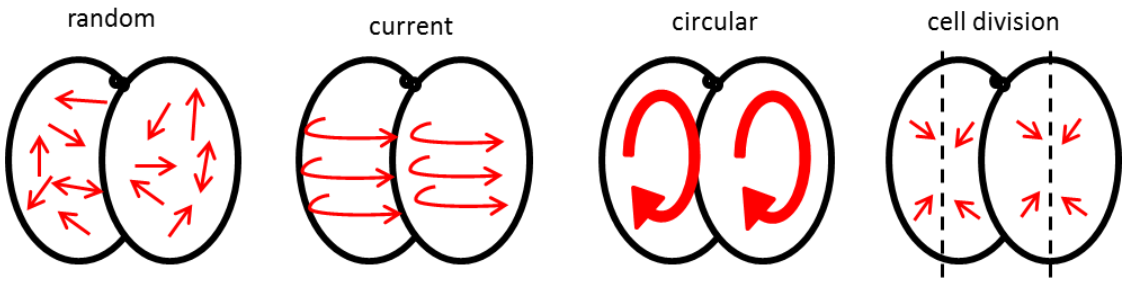


Figure 4.18 Phenotypes observed through analysis of cytoplasmic flow in *P. lamarcki* and *B. glabrata* visualized by Lightsheet microscopy.

Cytoplasmic flow of vesicles is linked to cell division in P. lamarcki *and* B. glabrata

Taken together, four phenotypes of movement were found for each species, where movement was closely linked to stages of cell division (Figure 4.18). The nucleus seemed to be a structure both to flow around and to flow towards depending on how close the embryo was to dividing to the next cleavage stage. There were subtle differences across cleavage stages, however these can be attributed to subtle differences in timing of recording and the angle at which the embryo was suspended at recording (Figure 4.19 and Figure 4.20).

Pomatoceros

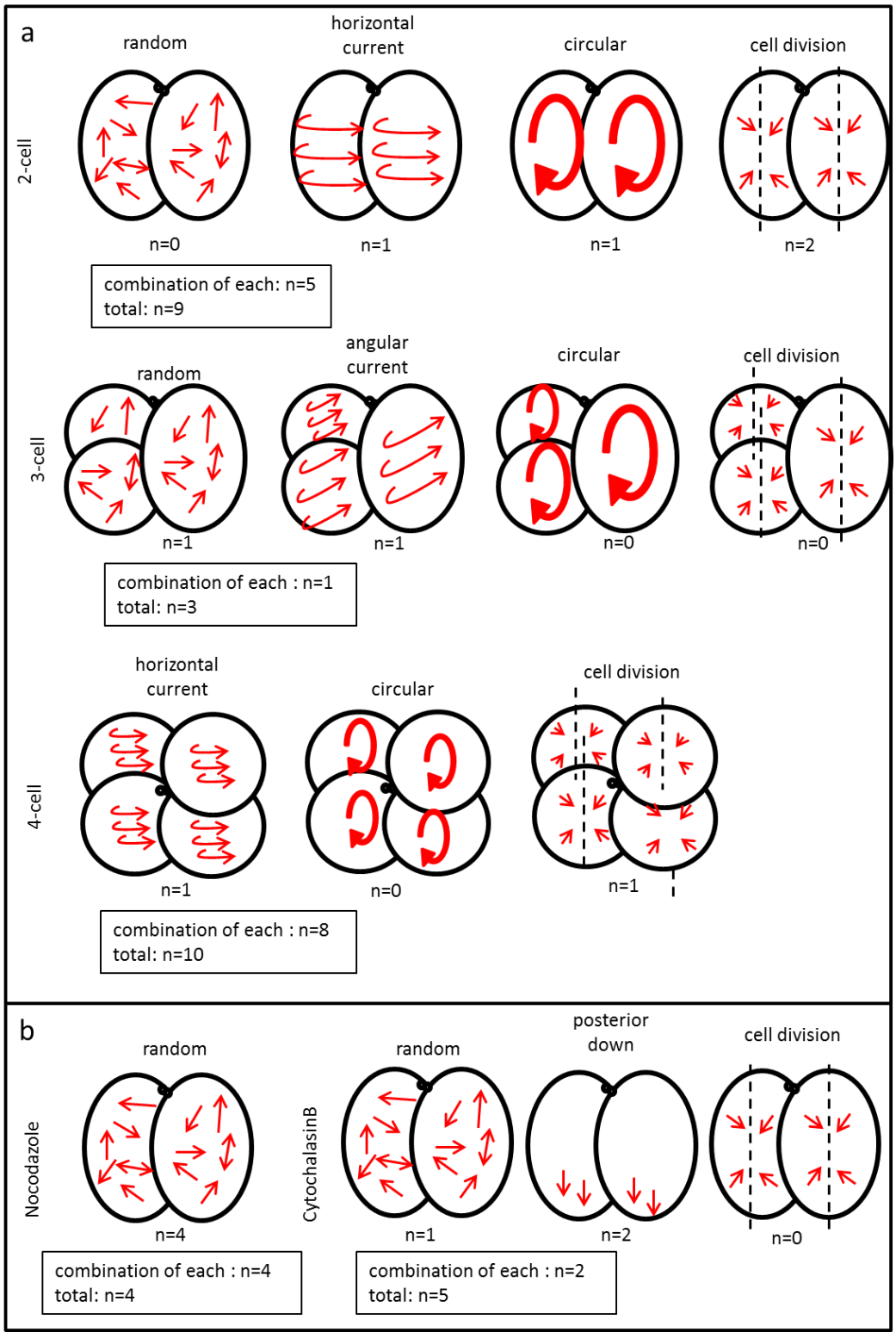


Figure 4.19 Summary of cytoplasmic flow across cleavage stages of *P. lamarcki*. (A) Wildtype embryos, (B) Nocodazole and CytochalasinB treated embryos extrapolated from Lightsheet data.

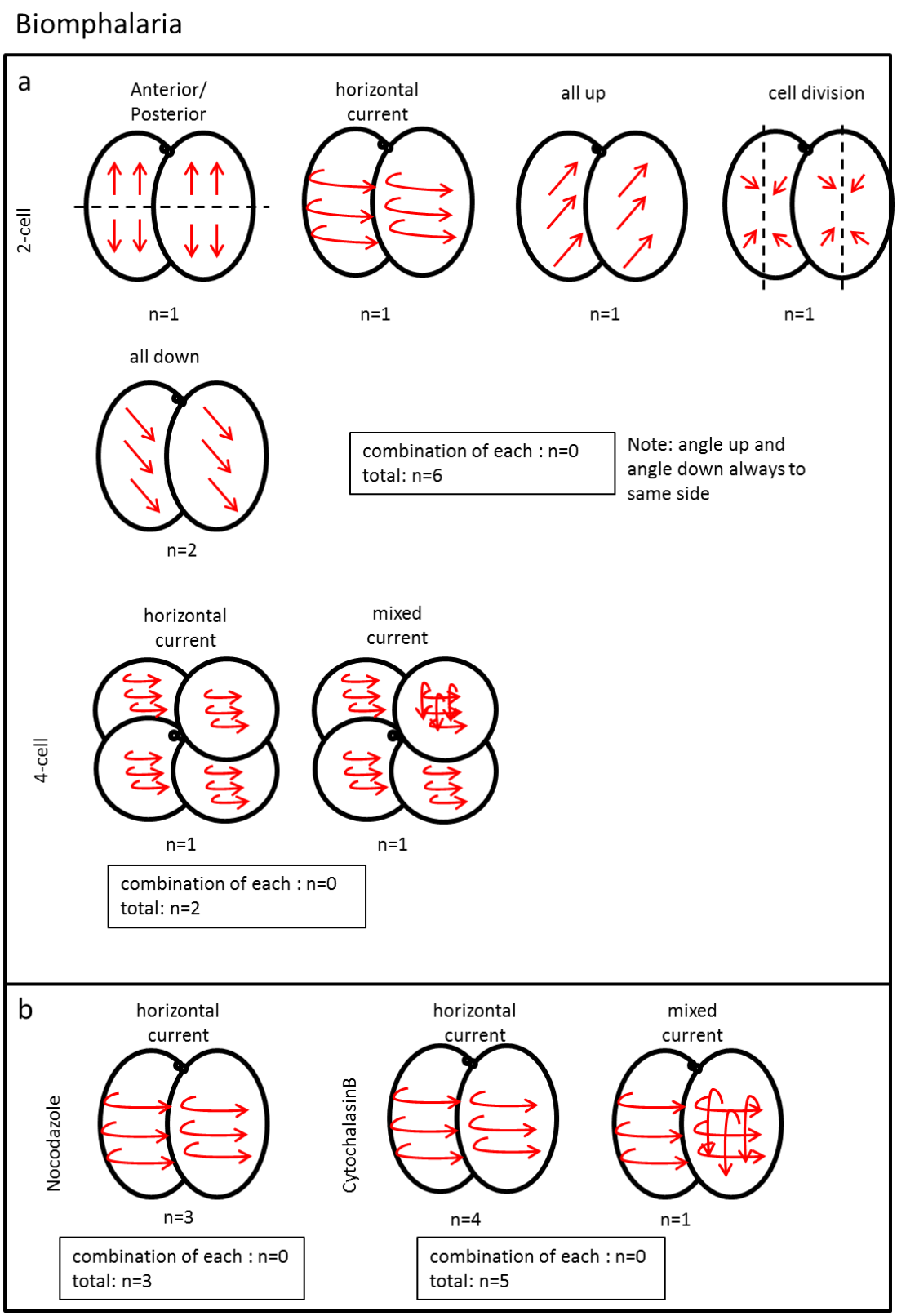


Figure 4.20 Summary of cytoplasmic flow across cleavage stages of *B. glabrata*. (A) Wildtype embryos, (B) Nocodazole and CytochalasinB treated embryos extrapolated from Lightsheet data.

Embryos arrest upon both actin and microtubule inhibition

Embryos with fluorescently labelled vesicles were treated with drugs that inhibit microtubule assembly (Nocodazole) or actin filament assembly (Cytochalasin B) and imaged in a Lightsheet microscope. Flow was not visibly disrupted in comparison to wildtype embryos in treated animals (Figure 4.19, Figure 4.20), however embryos arrested at the treated stage in each case. This may imply that the treatment lost its efficiency during the mounting procedure, or that the treatment was done too late to significantly disrupt the assembly and disassembly of the cytoskeletal components. In addition, embryos were not released from the agarose to see whether they develop further, which would indicate whether the embryos remained viable after recording. It is possible that the drugs were not working as expected, which could be solved by fixing treated and animals, fluorescently staining actin or tubulin, and checking to see whether drug-treated animals had the corresponding cytoskeletal component disrupted. As drug-treatment without subsequent mounting and flow visualization also arrested at the treated cell stage (data not shown), it is also likely that this is a true phenotype.

Dechorionation and live dye experiments are not straightforward

A frustrating limitation to working with *P. lamarcki* and *B. glabrata* is the lack of working protocols for dechorionation and injection for live dye application and imaging. The lack of these techniques means, for example, that ideal experiments such as looking at cortical flow by staining the cortex with a live stain, are not possible. A number of methods were trialled in an effort to develop

a working dechorionation and/or injection protocol, but this was not fruitful in most cases except for the dye used in this thesis to study cytoplasmic dynamics (Table 4.4). With the lack of a cortical live dye or a way to use other live stains such as ones that visualize tubulin, it becomes difficult to analyze obvious alternative sources of chirality generation in these embryos, which is an obvious next step.

Table. 4.4 Results of dechorionation methods and live dyes trialled for *P. lamarcki* embryos.

dechorionation/live dyes	**dead/alive**	**dye penetration to embryo (yes/no)**	**sample size (n)**
DECHORIONATION			
Sucrose/NaCitrate	all dead except 2 mins treatment but still chorion	n/a	approx. 1500 embryos mixed from 2-3 females
Protease	all dead	n/a	
Trypsin	all dead	n/a	
Sodium Thioglycolate	all dead	n/a	
2% Sodium Thioglycolate/1% protease	all dead	n/a	
20% Pluronic acid	dead	n/a	
Pronase	dead	n/a	
Pronase-trypsin	dead	n/a	
mechanical (pipetting)	alive, not dechorionated	n/a	
LIVE DYES			
DRAQ5	dead after 24h	no	approx. 1500 embryos mixed from 2-3 females
MitoTracker	alive	yes	mixed sperm sample from approx. 5 males
TubulinTracker	dead after 24h	no	approx. 1500 embryos mixed from 2-3 females
Calcium	all alive	no	
LysoTracker	all alive	yes	
FM-64	all alive	no	

B. glabrata *construct injection trials*

In the same vein, functional techniques are also lacking in both *P. lamarcki* and *B. glabrata*. For this reason, transgenics were attempted for *B. glabrata* embryos. *P. lamarcki* was not chosen for this, as there is not a working injection or dechorionation protocol to introduce constructs to the embryos. In addition, *P. lamarcki* adults live in calcareous tubes that would be difficult to

manipulate for injections and electroporations (which was the technique used here). *B. glabrata* presents a more feasible option as they develop in egg masses that are formed of individual compartments for each embryo that can easily be injected (Figure 4.21). A beta-actin GFP vector was used, a fusion gene that is likely constructed in a CAG vector. The exact sequence identity of the vector is needed to fully troubleshoot these transgenesis trials, which is something that we did not pursue fully. However, as this is a gene fusion protein, it is possible that beta-actin is just not being expressed in these trials.

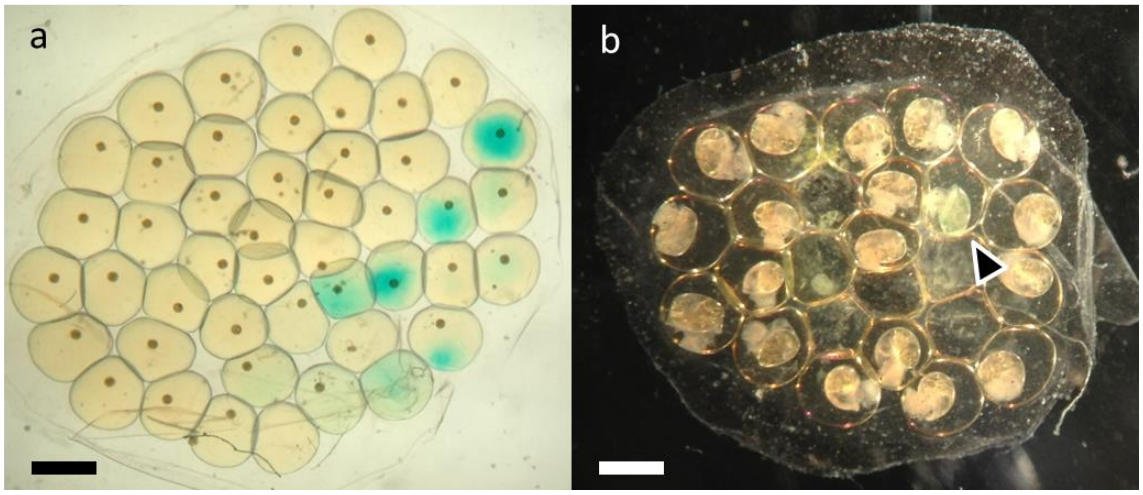


Figure 4.21 (A) Injected *B. glabrata* embryos, (B) after development to juvenile visualized on a dissecting microscope. Black arrow with white border indicates juvenile that retained FastGreen in its shell. Scale bar is approximately 1mm.

Although viable embryos developed, none expressed fluorescent GFP, indicating that the introduction of the constructs had not been successful. In the meantime, functional techniques have begun to emerge from the Lophotrochozoa, with the recent development of an annelid transgenic line of *P. dumerilii* (Backfisch et al., 2014; Zantke et al., 2014) and the development of the CRISPR/CAS9 system for the mollusc *Crepidula fornicata* (Perry and Henry, 2015).

The DV axis in equal-cleaving spiralians

Although this thesis has concentrated on LR asymmetry in preceding chapters, there is no reason for early directionality to be linked to LR asymmetry. Especially in these equal-cleaving spiralian embryos, it can also be the case that any directionality, whether it be cytoplasmic or cortical, can also be associated with the AP and DV axes. As the AP axis is typically correlated with the animal-vegetal axis, which is already marked at the animal pole by the formation of polar bodies, I considered whether the point of sperm entry correlated with the formation of the DV axis, and the later specification of dorsal fates. This was tested by staining *P. lamarcki* sperm with a mitochondrial stain (Table 4.4) and fertilizing embryos with this fluorescently labeled sperm under a confocal microscope (Figure 4.22). Although the staining worked, many sperm attached to the surface of the egg, making it impossible to figure out which one had successfully fused with the egg. It is clear that either an alternative fluorescent label needs to be used, that perhaps labels DNA in the sperm, or an alternative approach needs to be taken. To understand early cell dynamics the DV axis needs to be considered as a viable candidate when analyzing data.

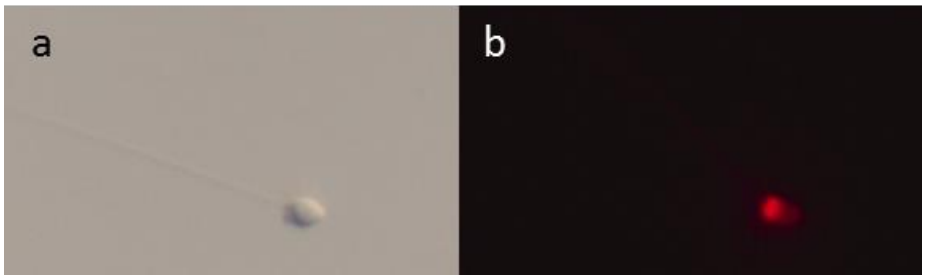


Figure 4.22 MitoTracker-stained *P. lamarcki* sperm visualized by (A) DIC and (B) fluorescence.

Discussion

Cleavage patterns and cytoplasmic flow were analyzed in the dextral *P. lamarcki* and the sinistral *B. glabrata*, both equal-cleaving spiralians with opposite handedness at the 8-cell stage. The main results of flow analysis show that there is little to no chirality in cytoplasmic flow at the first few cleavage stages. In a way, this study almost raises more questions than it answers, both concerning asynchrony in cleavage and cytoplasmic flow. A number of preliminary experiments have been done in an effort to answer some of the most urgent questions, and these feature prominently in the discussion below.

Variation in the timing of cleavage in equal-cleaving annelids

There is asynchrony of cleavage that produces a 3-cell stage in *P. lamarcki.* Asynchrony in spiralian cell cleavage has rarely been documented. There are two publications where asynchrony in cleavage timing has been reported in *Pomatoceros triqueter* embryos. Asynchrony was reported at the 2-cell stage in *P. triqueter* (Dorresteijn and Fischer, 1988; Dorresteijn and Luetjens, 1994), where the nature of this transition was not described, but the transition was said to be less than one minute. This asynchrony was said to be amplified at later stages, and the cause was thought to be correlated with slight size differences between blastomeres. Whether this is related to the DV or LR axes cannot be said with certainty. We have observed that dissimilar cleavage mechanisms can result in similar fates, and confirm observations previously made in molluscs (Martindale et al., 1985).

Yet, fate was really not addressed in this thesis. Although timing of cleavage revealed a 3-cell stage embryo, it is not known whether the precociously cleaving blastomere of the 2-cell embryo already has a specific fate. This is plausible, as in spiralian development developmental fate can be followed and lineaged from early cleavage stages, and fate becomes progressively specified as the embryo cleaves. To see whether blastomeres that cleave precociously to a 3-cell stage have a specific fate, DiI labeling was trialled, where a fluorescent lipophilic dye was used to mark one blastomere in an effort to see whether the precociously cleaving blastomere had a consistent fate later in development. Embryos were successfully labeled with DiI (Figure 4.23), but embryos that were labeled with the dye arrested cell cleavage and did not survive to the larval trochophore stage. This will need to be continued for future researchers in pursuit of understanding any biological significance of the role of precocious cleavage in spiralian development.

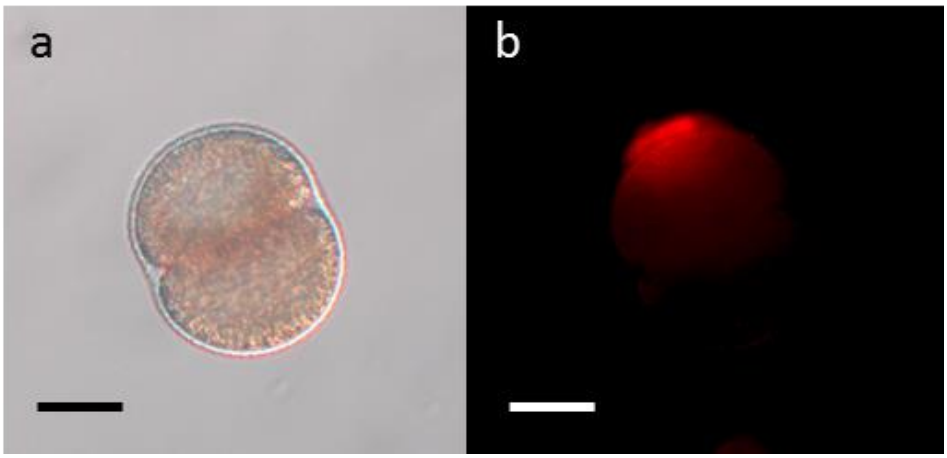


Figure 4.23 DiI labeling of *P. lamarcki* embryos visualized by DIC (A) and fluorescent (B) microscopy.

Asynchrony in cleavage has been seen in entoproct and phoronid embryos, and highlights the irregularity of cleavage as an important aspect of spiralian cleavage (Merkel et al., 2012; Pennerstorfer and Scholtz, 2012). This signals a need for more careful studies such as these, and a closer look at synchrony and asynchrony in cleavage in spriralians. It is possible that this source of variation is not stochastic and may give us new information concerning the dynamics governing cell fate specification. Whether this variability extends to bilateral animals outside of the Lophotrochozoa is not clear, as this type of data is not typically reported. Typically, cleavage variants are ignored as abnormal development and are rarely, if ever, followed to see whether these are really cases of aberrant development. If the same trend of intermediary cleavage stages is seen in other animals, it is possible that these are cases of accentuated timing differences in cleavage cycles that are part of normal animal development. It would be worth understanding just how widespread variability in cleavage stages and timing is in bilateral animals, and it is possible that trends in variability can be informative for understanding the evolution of developmental processes.

Cytoplasmic flow is not chiral prior to the spiral 8-cell stage

The spiral 8-cell stage of spiralian embryos achieves dextral or sinistral spiralling by spindle orientation, but nothing is known about directionality that precedes the orientation of the spindles. An alternative approach to spindles was taken here by looking earlier in development and analyzing cytoplasmic flow within spiralian embryos for signs of chirality that can bias spindle

orientation through a cytoskeletal-based mechanism. This does not seem to be the case. Rather, cytoplasmic flow is associated with stages of cell division (Figure 4.24).

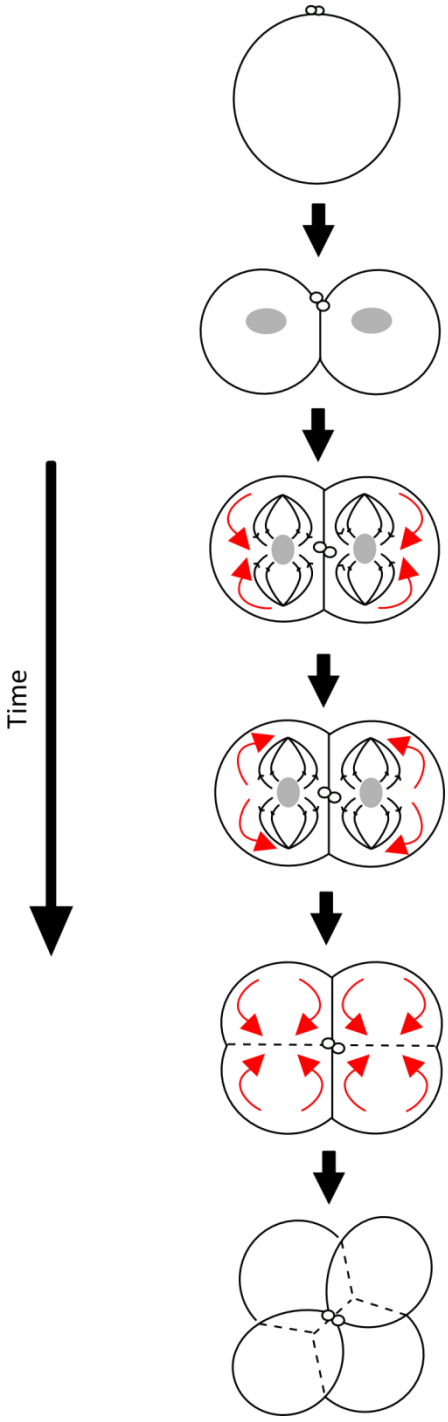


Figure 4.24 Hypothetical model of cytoplasmic dynamics during mitotic cell cycle in *P. lamarcki* and *B. glabrata*.

As the 8-cell stage develops before the presence of presumed organizers in equal-cleaving spiralians (Figure 4.2), it is interesting to consider what kind of determinants could be influencing something as ubiquitous as the cytoskeleton in blastomeres of the early spiralian embryo. Given the intra- and inter-species variation in developmental mechanisms (Seaver, 2014) within polychaetes, it remains a possibility that axis specification is slightly different in timing to other spiralians. The different phenotypes of cytoplasmic flow that we see are likely correlated with cell division (Figure 4.19 and Figure 4.20), and can be a form of developmental noise in relation to axis establishment. However, there is also the possibility that movement is similar to that seen in *C. elegans*, where it is related to the establishment of the DV axis (Singh and Pohl, 2014).

The lack of a strong phenotype upon treatment of embryos with high concentrations of tubulin-disrupting and actin-disrupting compounds may be the result of inefficient disruption, however this is unlikely due to the high concentrations used. As cytoplasmic flow was analyzed rather than cortical flow, it is more likely that nuclear division proceeded normally, as reflected by the lack of a strong difference in cytoplasmic flow, but that subsequent cleavage could not organize the disrupted tubulin and actin components to continue the cell cycle to cytokinesis. This has been seen in similar experiments in the mollusc *Dreissena polymorpha*, where nuclear division was not disrupted upon treatment with actin inhibitors, but subsequent cell cleavage was arrested (Luetjens and Dorresteijn, 1998). However, in the sand dollar, cytoplasmic flow was found to be dependent on the actin cytoskeleton (Rappaport and Rappaport, 1988), making it possible that this would also be the case in *P. lamarcki*.

It is not clear how generalizable cytoplasmic dynamics are across animals, especially given the diversity of early cell cleavage methods such as spiralian cleavage in lophotrochozoans, meroblastic and radial cleavage in deuterostomes, syncytial development in arthropods, among others. In a broad sense, cytoplasmic flow has been correlated with AP axis formation in mouse (Yi et al., 2011) and flow-like communication in the cytoplasm has been studied in the frog (Ishihara et al., 2014). Other than an association with the primary axes of the embryo, the majority of studies on cytoplasmic flow are studied as by-products of examinations of developmental processes involving a great deal of movement such as gastrulation. Yet, the cytoplasm contains a lot of information, and is where gene products are located and communicated. In this way, although cytoplasmic flow is not usually found at the center of research questions, such as it has been in the past in studies of *C. elegans* (Golden, 2000; Road, 1993), it is always assumed to be there and functioning to transport molecules, presumably via the cytoskeleton.

Perhaps a contributing factor for the lack of information concerning cytoplasmic flow in embryos is the emphasis on cell culture and non-animal material to study such minute cell biological details. For example, processes such as cytoplasmic streaming in individual cell movement have been studied in association with cytoplasmic dynamics. There is increasing evidence that individual cells in cell culture generate polar cortical and cytoplasmic flow naturally (Naganathan et al., 2014; Wan et al., 2011), and that this correlates with LR asymmetry establishment. Although this is very different from looking at animal embryos, the cell culture system is often more amenable to study intricate aspects of cell machinery. They may give us a view of the possible

cytoskeletal mechanisms that may be at work in animal embryos, and hopefully these types of considerations will be applied in a more animal-centric way to understand these processes in a developmental context.

Lysosomes and cytoplasmic flow

Interestingly, lysosomes travel along microtubule tracks (Collot et al., 1984; Matteoni and Kreis, 1987). Lysosome movement has also been associated with moving in the direction of microtubule organizing centers in cell culture (Matteoni and Kreis, 1987). In light of our drug treatment experiments that abolish microtubules, it seems likely that lysosomes travel on tracks that are not related to microtubules as well. It also remains a possibility that the drug treatment was not efficient, and microtubules were able to reassemble to transport the lysosomes. Given that we do not know the viscosity of the cytoplasm during this time, it is difficult to ascertain what the movement of the lysosomes would be in the absence of microtubules. Visualizing spindles and microtubules during live imaging in these animals would allow us to ascertain whether this association also exists in spiralians although the techniques are not yet routinely available in these animals.

Cortical flow, not cytoplasmic flow, influences polarity

The polarity seen in *C. elegans* likely points to cortical flow as a better readout for early chirality. Cortical flow has recently received attention, as its polar nature has been neatly correlated with a polar cytoskeleton in *C. elegans*

(Schonegg et al., 2014). The closest anyone has come to understanding such dynamics in early embryos has been in *C. elegans*, where AP axis formation by cortical flow is connected to DV axis formation by the positioning of a small vesicle-like structure called the midbody, which is essential for asymmetric spindle positioning via actin and astral microtubules (Singh and Pohl, 2014). This is similar to more distantly-related animals such as the mouse, where there is increasing evidence for the involvement of the cortex with actin in spindle migration, mediated by the actin nucleators Arp2/3 (actin-related protein) and Fmn2 (formin) (Chaigne et al., 2015; Yi et al., 2013), in the mouse oocyte. Of course, these are distantly-related animals with very distinct modes of early development. However the association between the cortex and asymmetric spindle placement could be a conserved mechanism present in animals, especially those with early cell fate specification such as spiralians (Mogilner and Fogelson, 2015).

The presence of cortical flow in spiralian embryos is varied, with *Platynereis dumerilii* being one species that demonstrates clear flow within the zygote along the AP axis (Dondua et al., 1997; Fischer and Dorresteijn, 2004). Cortical flow has been studied in other spiralians to varied effects and have shown different patterns of movement, such as in a leech (Weisblat and Huang, 2001), a chiton (Huebner and Anderson, 1976), and a range of molluscs including the zebra mussel (Luetjens and Dorresteijn, 1998). Asymmetric cortical movement has been documented at least once in the dextral snails *Lymnea stagnalis* and *Lymnea palustris* and the sinistral snails *Physa acuta* and *Physa fontinalis* (Meshcheryakov and Beloussov, 1975), where cortical movement was correlated with the chirality of spiral cleavage.

Mechanisms for symmetry breaking in the Bilateria

Cytoplasmic chirality was found not to be present in the equal-cleaving animals studied here, pointing to cortical chirality as a likely generator of chirality at these early stages (Figure 4.25). It is likely that the use of cytoskeletal machinery for the establishment of LR asymmetry is widespread in bilateral animals, and that the use of cilia is a derived mechanism found in chordates. This implies that the last common ancestor of the bilaterians (the Urbilaterian) may have used such a mechanism for LR asymmetry establishment. However, given that the cytoskeleton is ubiquitous and used in almost every function, it cannot be ruled out that this may be a case of convergent evolution of cytoskeletal-dependent mechanisms. Further experiments are needed to probe the relationship between symmetry breaking and molecular asymmetry in the form of asymmetric gene expression. With functional techniques becoming amenable to a wider diversity of animals, this should become increasingly possible.

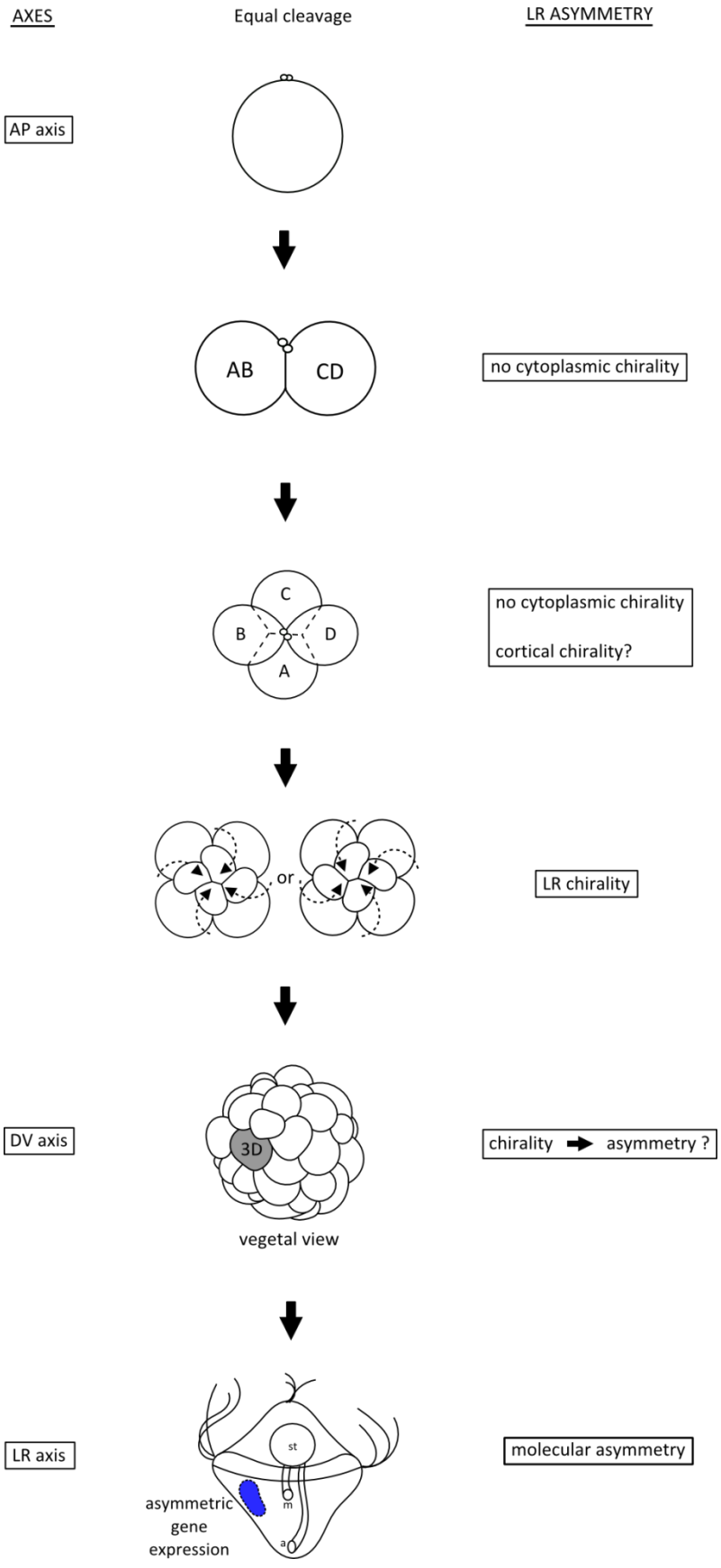


Figure 4.25 Final model to represent thoughts on cytoplasmic chirality in equal-cleaving spiralians.

Future directions

We have shown here that cytoplasmic flow is correlated with cell division. The presence of many different structures and organelles in blastomeres makes it complex to visualize any form of directionality to this flow, and it seems likely that such a stochastic system would not reliably produce polarity. We anticipate studies to build upon this, and to probe at the connection between flow and molecular machinery at the cortical level. This study demonstrates that cytoplasmic polarity is not a likely candidate for symmetry breaking, and pushes future studies to concentrate on cortical flow in these animals.

Future experiments should concentrate on analyzing cortical flow in these embryos to see whether chirality is seen at this stage in spiralian embryos. This is of primary interest, as we still do not know the exact mechanism by which the spiral 8-cell stage becomes spiral. This could only be aided ideally by fusion transgenics, or alternatively by live dyes of other structures such as the nucleus and components of the cytoskeleton such as microtubules and actin filaments in addition to a full 3D analysis of the data contained here and data to be gathered in the future. Drug treatments in conjunction with this would confirm which cytoskeletal component is participating in cytoplasmic and cortical flow in spiralian embryos.

Although equal-cleaving embryos were considered in this thesis, it would be important to expand this study to unequal-cleaving and polar lobe forming spiralian embryos, and to see whether the same cytoskeletal mechanisms govern this conserved mode of early cleavage and fate specification. Sinistrality, although rare, is found in polychaetes as in *Hydroides elegans* (Arenas-Mena,

2007). The position of the operculum is not asymmetric in this animal. How directionality of the cytoplasm or cortex in early embryos may contrast with dextrally-spiralling polychaetes is not known and would help us to understand polarity differences in cytoplasmic or cortical flow within the Annelida.

Chapter 5

General Discussion

A theme for this thesis has become not to trust generalizations and assumptions. Coming back to the question of how left-right (LR) asymmetry is established in spiralian embryos, this thesis demonstrates that previously accepted generalizations both about the molecular machinery to establish LR asymmetry, and aspects of spiralian embryology, have overstepped their broad applicability to other systems and are not necessarily as textbooks would have them be. Rather than a simple linear conservation of mechanisms, the mechanisms underlying LR asymmetry seem to be different in *Pomatoceros lamarcki* relative to molluscs and deuterostomes. However, it is probably more correct to say that molluscs and deuterostomes are different relative to other bilaterians. This will be discussed in more detail below, followed by a discussion of potential cytoskeletal mechanisms for symmetry breaking in spiralian embryos, and concluded with considerations of variability in the light of understanding the evolution of mechanisms in animal development.

Conserved does not necessarily mean functionally identical

Nodal signalling has been touted as a signalling pathway that is conserved in LR asymmetry establishment in bilateral animals, and this claim has thus far been undisputed. The work in this thesis suggests that the presence of the Nodal pathway is conserved in bilateral animals, but that its function in LR asymmetry establishment is not. Although the evidence here is not definitive, taking current evidence under scrutiny, the only evidence of Nodal function in LR asymmetry establishment outside of deuterostomes comes from molluscs, and even here the evidence is not as convincing as it may seem. Although the sided expression of *Nodal* is undisputed, the function of Nodal is less convincing, where the low number of embryos that survive to have a non-coiled shell chirality upon Nodal inhibition (Grande and Patel, 2009) suggests that the ones considered are embryos for which inhibition was less efficient. Although part and parcel in drug treatments, this is a less than ideal piece of evidence on which to base large claims as to the conserved function of Nodal in LR asymmetry establishment across all bilateral animals.

What is possible is that Nodal signalling in LR asymmetry establishment is conserved in deuterostomes, and that other lineages have independently evolved mechanisms for LR asymmetry establishment. Molluscs are likely a case that happened to employ the Nodal signalling pathway for LR asymmetry establishment, perhaps as a by-product of using this pathway for shell formation, rather than this being a conserved function of the Nodal pathway. Given that other BMP ligands, such as BMP2/4 (decapentaplegic, dpp) are involved in shell formation and the development of asymmetry of the mollusc shell

(Hashimoto et al., 2012; Shimizu et al., 2013, 2011), it is not a great leap to suggest that the Nodal pathway has also been co-opted for shell formation, and the intimately associated chirality that comes with forming mollusc shells.

In this way, we propose that Nodal signalling may not be conserved in LR asymmetry establishment in bilaterian animals. Rather, its role in establishing asymmetry is a conserved feature of specifically deuterostome Nodal signalling. Evidence of the role of Nodal in LR asymmetry establishment in some non-deuterostome lineages may indicate independent co-option of Nodal signalling for LR asymmetry establishment rather than conservation (Figure 5.1). Of course, we still cannot rule out the possibility of prevalent loss of the LR asymmetry- establishing function of Nodal across the Bilateria. However, the latter remains an unlikely scenario, and instead we propose that Nodal signalling is conserved in mesendoderm or mesoderm specification and other functions are novel co-options of this pathway.

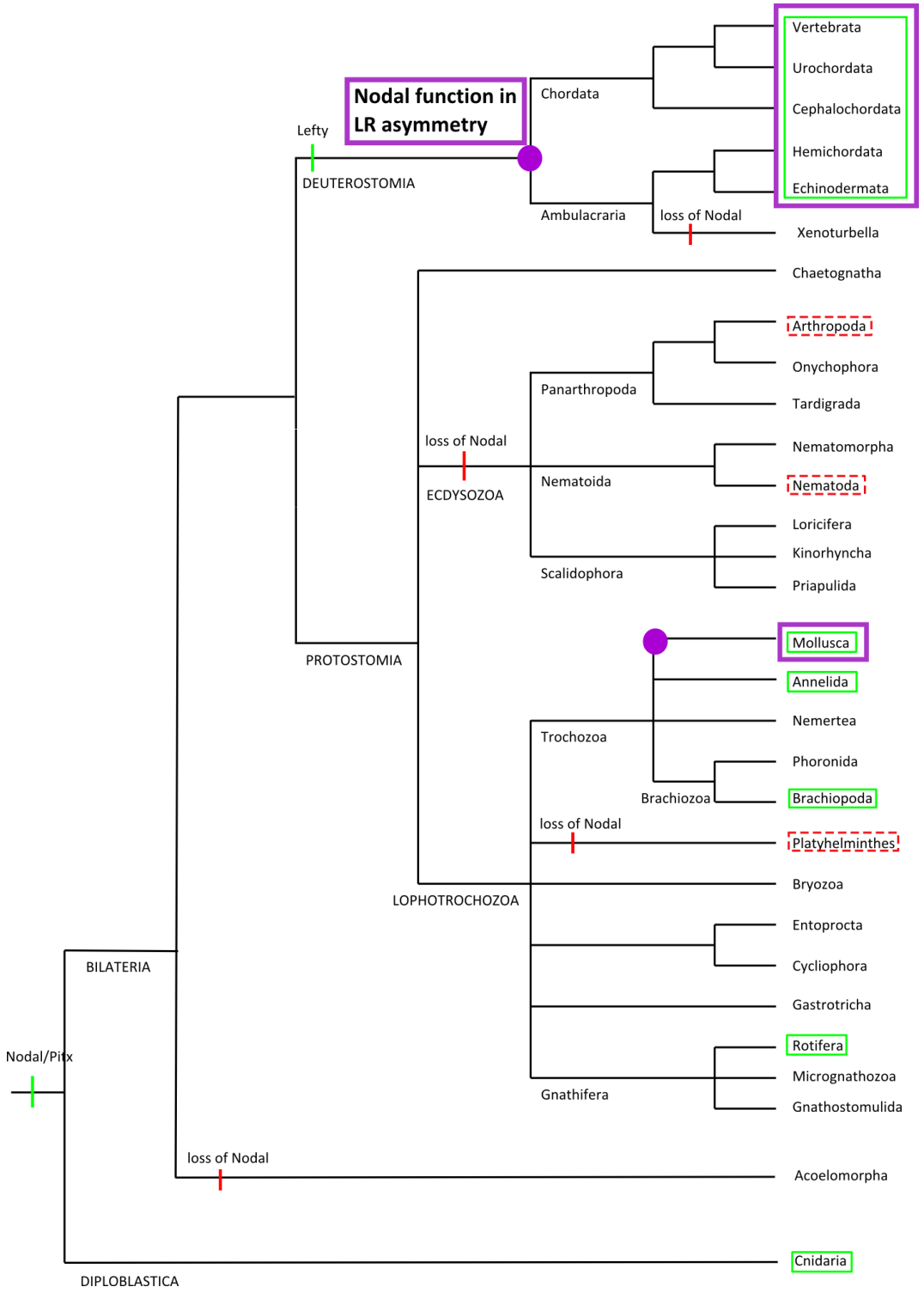


Figure 5.1 The evolution of Nodal function in LR asymmetry establishment. Green boxes and notches indicate *Nodal* presence or hypothesized gain, red dotted boxes and solid notches indicate *Nodal* absence or hypothesized loss, bold purple boxes and circles indicate hypotheses presented in this thesis on the evolution of molecular mechanisms underlying LR asymmetry establishment in bilateral animals.

The window of symmetry breaking is getting smaller

This thesis has not solved the enigma of symmetry breaking. However, it has made smaller the window during which symmetry breaking most likely occurs. Cytoplasmic flow has not generally been associated with LR asymmetry establishment, and my data support this. Although this does point towards cortical flow as the first generator of chirality, the connection between cortical chirality and later asymmetry is still tenuous. The best evidence has been found in *Caenorhabditis elegans*, with the next best source of evidence being the mouse oocyte, and although the evidence in the latter two cases are strong for these two systems, the mechanisms are very specific to the particular stage and particular animal and does not give workable clues for animals like spiralians.

A definitive connection between spindle orientation and prior cytoskeletal asymmetry, if there is one, is still lacking in spiralians. Given what this thesis has found, there is no obvious chirality at the 2-cell and 4-cell stages. As the spindles are oriented at a specific angle as the 4-cell divides into an 8-cell stage, it is somewhere during 8-cell formation that chirality is produced and harnessed for spindle orientation. As this is a fixed asymmetry, and the mysterious component is present in the cytoplasm, it is likely that a maternally deposited sinistral or dextral factor is activated at the 8-cell stage, and that this factor signals to the cytoskeleton to pull spindles in one direction or the other via cytoskeletal architects such as Arp2/3 (actin-related protein) or Fmn2 (formin) (Figure 5.2).

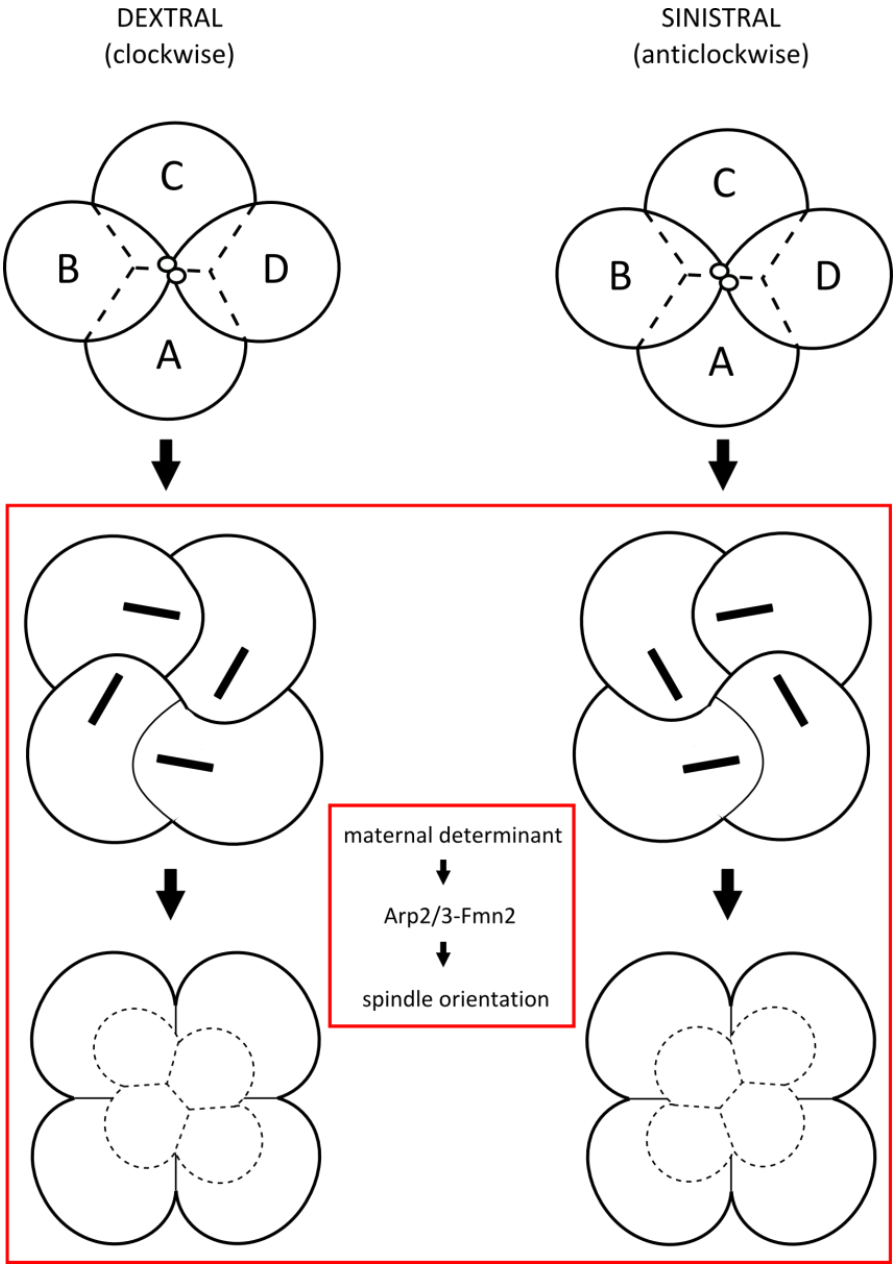


Figure 5.2 Hypothesis for LR symmetry breaking in spiralian embryos. Red boxes indicate hypotheses presented in this thesis, where the large red box indicates the likely stage at which symmetry breaking takes place, and the small inner red box indicates a hypothesized cytoskeletal mechanism.

In this way, the tools that can be used for symmetry breaking are limited at early stages of development, and it is not surprising that the cytoskeleton features prominently in all animals. Despite the diversity of modes of early

development across animals, the cytoskeleton is really the only constant factor. Rather than a common mechanism for symmetry breaking it is more likely that diverse animals have harnessed the cytoskeleton in a myriad of ways to achieve symmetry breaking, and that this is highly dependent on whether this asymmetry is heritable or not, and on the effects that the chosen mode of symmetry breaking would have on subsequent development of that animal. This would mean that the quest to find the mechanism underlying symmetry breaking would have to be undertaken independently for different animals, and each animal may have a slightly different take on how it is done.

Taking a further step back, perhaps the reason why symmetry breaking has been such an impossible enigma to solve is that there is no generalizable mechanism. Researchers have shied away from diversity in favor of high impact mechanisms that can be as generalizable as possible, yet it may be this favoritism that is blinding researchers from seeing that there is no single mechanism for symmetry breaking.

Generalizing animal development can do more harm than good

What has become apparent through this thesis is that there is a considerable amount of variation in the development of bilateral animals. The over-simplification of developmental processes can be detrimental in this respect, as variation is an important part of animal development. The level of variation in mechanisms of development, from cell cleavage modes and inner cell dynamics to general modes of body plan establishment, is usually

disregarded for the sake of simplicity and the search for common mechanisms that unite animal development. Although generalizable processes are typically the most lauded when it comes to scientific merit and recognition, there is a danger of cherry-picking and skimming over real variations, which themselves can be informative for understanding the evolution of development.

The extant animals that we see today evolved from long-gone ancestors, and although evolution can be very fast, it is likely that animals evolved by using the abundant variation in their development. By only looking at what is not variable in extant animals, we are ignoring important tools that are used in the evolution of developmental mechanisms. Variation fits into the picture somehow, whether through developmental systems drift or a maintenance of robustness. Underappreciating the level of variation is misleading and prevents us from understanding the true evolution of the mechanisms underlying animal development.

References

Adamska M, Degnan SM, Green KM, Adamski M, Craigie A, Larroux C, Degnan BM, 2007. Wnt and TGF β expression in the sponge Amphimedon queenslandica and the origin of metazoan embryonic patterning. PLoS One 2, e1031–e1031.

Ajduk A, Biswas Shivhare S, Zernicka-Goetz M, 2014. The basal position of nuclei is one pre-requisite for asymmetric cell divisions in the early mouse embryo. Dev. Biol. 392, 133–140.

Altschul S, Gish W, Miller W, Myers E, Lipman D, 1990. Basic local alignment search tool. J Mol Biol 215, 403–410.

Amiel AR, Henry JQ, Seaver EC, 2013. An organizing activity is required for head patterning and cell fate specification in the polychaete annelid Capitella teleta: new insights into cell-cell signaling in Lophotrochozoa. Dev. Biol. 379, 107–22.

Aparicio S, Chapman J, Stupka E, Putnam N, Chia J, Dehal P, Christoffels A, Rash S, Hoon S, Smit A, Gelpke MDS, Roach J, Oh T, Ho IY, Wong M, Detter C, Verhoef F, Predki P, Tay A, Lucas S, Richardson P, Smith SF, Clark MS, Edwards YJK, Doggett N, Zharkikh A, Tavtigian S V, Pruss D, Barnstead M, Evans C, Baden H, Powell J, Glusman G, Rowen L, Hood L, Tan YH, Elgar G, Hawkins T, Venkatesh B, Rokhsar D, Brenner S, 2002. Whole-Genome Shotgun Assembly and Analysis of the Genome of Fugu rubripes. Science 297, 1301–1310.

Arenas-Mena C, 2007. Sinistral equal-size spiral cleavage of the indirectly developing polychaete Hydroides elegans. Dev Dyn 236, 1611–1622.

Arnolds WJA, van den Biggelaar JAM, Verdonk NH, 1983. Spatial aspects of cell interactions involved in the determination of dorsoventral polarity in equally cleaving gastropods and regulative abilities of their embryos, as studied by micromere deletions in Lymnaea and Patella. Dev. Biol. 192, 75–85.

Asami T, Gittenberger E, Falkner G, 2008. Whole-Body Enantiomorphy and Maternal Inheritance of Chiral Reversal in the Pond Snail Lymnaea stagnalis. J Hered 99, 552–557.

Backfisch B, Kozin V, Kirchmaier S, Tessmar-Raible K, Raible F, 2014. Tools for gene-regulatory analyses in the marine annelid Platynereis dumerilii. PLoS One 9, e93076.

Beck S, Le Good JA, Guzman M, Ben Haim N, Roy K, Beermann F, Constam DB, 2002. Extraembryonic proteases regulate Nodal signalling during gastrulation. Nat Cell Biol 4, 981–985.

Bergmann DC, Lee M, Robertson B, Tsou MF, Rose LS, Wood WB, 2003. Embryonic handedness choice in C. elegans involves the Galpha protein GPA-16. Development 130, 5731–5740.

Beyer T, Danilchik M, Thumberger T, Vick P, Tisler M, Schneider I, Bogusch S, Andre P, Ulmer B, Walentek P, Niesler B, Blum M, Schweickert A, 2012. Serotonin Signaling Is Required for Wnt-Dependent GRP Specification and Leftward Flow in Xenopus. Curr. Biol. 22, 33–39.

Blum M, Feistel K, Thumberger T, Schweickert A, 2014. The evolution and conservation of left-right patterning mechanisms. Development 141, 1603–13.

Boorman CJ, Shimeld SM, 2002. Pitx homeobox genes in Ciona and amphioxus show left-right asymmetry is a conserved chordate character and define the ascidian adenohypophysis. Evol Dev 4, 354–365.

Boycott A, Diver C, Garstang S, Turner F, 1930. The inheritance of sinistrality in Limnaea peregra. Philos Trans R Soc L. B Biol Sci 219, 51–130.

Boyer BC, Henry JQ, 1998. Evolutionary Modifications of the Spiralian Developmental Program. Amer Zool 38, 621–633.

Brameier M, 2010. Genome-wide comparative analysis of microRNAs in three non-human primates. BMC Res. Notes 3, 1–6.

Brown NA, Wolpert L, 1990. The development of handedness in left/right asymmetry. Development 109, 1–9.

Camey T, Verdonk NH, 1969. The Early Development of the Snail Biomphalaria glabrata and the Origin of the Head Organs. Netherlands J. Zool. 20, 93–121.

Capdevila J, Vogan KJ, Tabin CJ, Izpisúa Belmonte JC, 2000. Mechanisms of left-right determination in vertebrates. Cell 101, 9–21.

Card DC, Schield DR, Reyes-velasco J, Fujita MK, Andrew AL, Oyler-mccance SJ, Fike JA, Tomback DF, Ruggiero RP, Castoe TA, 2014. Two Low Coverage Bird Genomes and a Comparison of Reference-Guided versus De Novo Genome Assemblies. PLoS One 9, 1–13.

Chaigne A, Campillo C, Gov N, Voituriez R, Sykes C, Verlhac M, Terret M, 2015. A narrow window of cortical tension guides asymmetric spindle positioning in the mouse oocyte. Nat. Commun. 6, 1–10.

Child C, 1899. The Significance of the Spiral Type of Cleavage and its Relation to the Process of Differentiation. Biol. Lect. Deliv. Mar. Biol. Lab. 231–266.

Collot M, Louvard D, Singer S, 1984. Lysosomes are associated with microtubules and not with intermediate filaments in cultured fibroblasts. Proc. Natl. Acad. Sci. U. S. A. 81, 788–792.

Conklin EG, 1898. Cleavage and Differentiation. Biol. Lect. Deliv. Mar. Biol. Lab. 17–43.

Consortium HG, 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–921.

Consortium ICGS, 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432, 695–777.

Consortium MGS, 2009. Lineage-Specific Biology Revealed by a Finished Genome Assembly of the Mouse. PLoS Biol. 7, e1000112.

Consortium SUGS, 2006. The Genome of the Sea Urchin Strongylocentrotus purpuratus. Science 314, 941–952.

Consortium TGS, 2008. The genome of the model beetle and pest Tribolium castaneum. Nature 452, 949–55.

Consortium THGS, 2007. Insights into social insects from the genome of the honeybee Apis mellifera. Nature 443, 931–949.

Costello DP, Henley C, 1976. Spiralian Development: A Perspective. Am. Zool. 16, 277–291.

Coutelis J-B, Géminard C, Spéder P, Suzanne M, Petzoldt AG, Noselli S, 2013. Drosophila left/right asymmetry establishment is controlled by the Hox gene abdominal-B. Dev. Cell 24, 89–97.

Cui C, Little CD, Rongish BJ, 2009. Rotation of organizer tissue contributes to left-right asymmetry. Anat Rec 292, 557 – 61.

Danilchik M, Brown E, Riegert K, 2006. Intrinsic chiral properties of the Xenopus egg cortex: an early indicator of left-right asymmetry? Development 133, 4517–4526.

Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, De Tomaso A, Davidson B, Di Gregorio A, Gelpke M, Goodstein DM, 2002. The draft genome of Ciona intestinalis: insights into chordate and vertebrate origins. Science (80-.). 298, 2157–2167.

Diver C, Boycott A, Garstang S, 1925. The inheritance of inverse symmetry in Lymnaea peregra. J Genet 15, 113–200.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR, 2013. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21.

Dondua A, Kostyuchenko R, Fedorova Z, 1997. Effects of some cytoskeleton inhibitors on ooplasmic segregation in the Nereis virens egg. Int. J. Dev. Biol. 41, 853–858.

Dorresteijn A, Fischer A, 1988. The process of early development, in: Westheide, W., Hermans, C.O. (Eds.), The Ultrastructure of Polychaeta. pp. 335–352.

Dorresteijn A, Luetjens C, 1994. Morphometric analysis of cellular specification in Platynereis and Pomatoceros embryogenesis (Annelida, Polychaeta), in: Dauvin, J.-C., Laubier, L., Reish, D. (Eds.), Actes de La 4ième Conférence Internationale Des Polychètes. Mém. Mus. Natn. Hist. Nat., Paris, pp. 45–50.

Duboc V, Rottinger E, Lapraz F, Besnardeau L, Lepage T, 2005. Left-right asymmetry in the sea urchin embryo is regulated by nodal signaling on the right side. Dev Cell 9, 147–158.

Dunn CW, Giribet G, Edgecombe GD, Hejnol A, 2014. Animal Phylogeny and Its Evolutionary Implications. Annu. Rev. Ecol. Evol. Syst. 45, 371–395.

Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, Smith SA, Seaver E, Rouse GW, Obst M, Edgecombe GD, Sorensen M V, Haddock SH, Schmidt-Rhaesa A, Okusu A, Kristensen RM, Wheeler WC, Martindale MQ, Giribet G, 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452, 745–749.

Earl D, Bradnam K, John JS, Darling A, Lin D, Fass J, On H, Yu K, Buffalo V, Zerbino DR, Diekhans M, Nguyen N, Ariyaratne PN, Sung W, Ning Z, Haimel M, Simpson JT, Fonseca NA, Docking TR, Ho IY, Rokhsar DS, Chikhi R, Lavenier D, Chapuis G, Naquin D, Koren S, Yang S, Wu W, Chou W, Srivastava A, Shaw TI, Ruby JG, Skewes-cox P, Betegon M, Dimon MT, Solovyev V, Seledtsov I, Kosarev P, Vorobyev D, Ramirez-gonzalez R, Leggett R, Maclean D, Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Yin S, Sharpe T, Hall G, Kersey PJ, Durbin R, Jackman SD, Chapman JA, Huang X, Derisi JL, Caccamo M, Li Y, Jaffe DB, Green RE, Haussler D, Korf I, Paten B, 2011. Assemblathon 1 : A competitive assessment of de novo short read assembly methods. Genome Res. 21, 2224–2241.

Eddelbuettel D, Fran R, 2011. Rcpp : Seamless R and C ++ Integration 40.

Edgecombe GD, Giribet G, Dunn CW, Hejnol A, Kristensen RM, Neves RC, Rouse GW, Worsaae K, Sørensen M V., 2011. Higher-level metazoan relationships: recent progress and remaining questions. Org. Divers. Evol. 11, 151–172.

Eyre-Walker A, 1993. Recombination and mammalian genome evolution. Proc R Soc L. B 252, 237–243.

Fischer A, Dorresteijn A, 2004. The polychaete Platynereis dumerilii (Annelida): a laboratory animal with spiralian cleavage, lifelong segment proliferation and a mixed benthic/pelagic life cycle. Bioessays 26, 314–325.

Freeman G, Lundelius J, 1982. The Developmental Genetics of Dextrality and Sinistrality in the Gastropod Lymnaea peregra. Roux s Arch. Dev. Biol. 191, 69–83.

Freeman G, Lundelius W, 1992. Evolutionary specification implications of the mode of D quadrant in coelomates with spiral cleavage. J. Evol. Biol. 5, 205–247.

Fryxell KJ, Moon W-J, 2005. CpG mutation rates in the human genome are highly dependent on local GC content. Mol. Biol. Evol. 22, 650–8.

Fukumoto T, Kema IP, Levin M, 2005. Serotonin signalling is a very early step in patterning of the left-right axis in chick and frog embryos. Curr. Biol. 15, 794–803.

Géminard C, González-Morales N, Coutelis J-B, Noselli S, 2014. The myosin ID pathway and left-right asymmetry in Drosophila. Genesis 52, 471–80.

Gerhart J, 2000. Inversion of the chordate body axis: are there alternatives? Proc. Natl. Acad. Sci. U. S. A. 97, 4445–8.

Giribet G, 2002. Current advances in the phylogenetic reconstruction of metazoan evolution . A new paradigm for the Cambrian explosion? 24, 345–357.

Giribet G, 2008. Assembling the Lophotrochozoan (spiralian) tree of life. Philos Trans R Soc L. B Biol Sci 363, 1513–1522.

Glez-Peña D, Gómez-Blanco D, Reboiro-Jato M, Fdez-Riverola F, Posada D, 2010. ALTER: program-oriented conversion of DNA and protein alignments. Nucleic Acids Res. 38, W14–W18.

Golden A, 2000. Cytoplasmic flow and the establishment of polarity in C . elegans 1-cell embryos. Curr. Opin. Genet. Dev. 10, 414–420.

González-Morales N, Géminard C, Lebreton G, Cerezo D, Coutelis J-B, Noselli S, 2015. The Atypical Cadherin Dachsous Controls Left-Right Asymmetry in Drosophila. Dev. Cell 33, 675–89.

Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R, 2010. A new bioinformatics analysis tools framework at EMBL-EBI. Nucleic Acids Res. 38, W695–699.

Graham A, Blentic A, Duque S, Begbie J, 2007. Delamination of cells from neurogenic placodes does not involve an epithelial-to-mesenchymal transition. Development 134, 4141–5.

Grande C, Martín-Durán JM, Kenny NJ, Truchado-García M, Hejnol A, 2014. Evolution, divergence and loss of the Nodal signalling pathway: new data and a synthesis across the Bilateria. Int. J. Dev. Biol. 58, 521–32.

Grande C, Patel NH, 2009. Nodal signalling is involved in left-right asymmetry in snails. Nature 457, 1007–1011.

Gregory TR, 2014. Animal Genome Size Database [WWW Document]. URL http://www.genomesize.com

Gregory TR, Hebert PD., 2002. Genome size estimates for some oligochaete annelids. Can. J. Zool. 80, 1485–1489.

Gros J, Feistel K, Viebahn C, Blum M, Tabin C, 2009. Cell Movements at Hensen ' s Node Establish Left/Right Asymmetric Gene Expression in the Chick. Science 324, 941–944.

Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL, 2008. The Vienna RNA websuite. Nucleic Acids Res. 36, W70–74.

Haffner C, Frauli M, Topp S, Irmler M, Hofmann K, Regula JT, Bally-Cuif L, Haass C, 2004. Nicalin and its binding partner {N}omo are novel {N}odal signaling antagonists. EMBO J. 23, 3041–3050.

Hamada H, Meno C, Watanabe D, Saijoh Y, 2002. Establishment of vertebrate left-right asymmetry. Nat. Rev. Genet. 3, 103–113.

Hashimoto N, Kurita Y, Wada H, 2012. Developmental role of dpp in the gastropod shell plate and co-option of the dpp signaling pathway in the evolution of the operculum. Dev. Biol. 366, 367–73.

Hejnol A, 2010. A twist in time--the evolution of spiral cleavage in the light of animal phylogeny. Integr Comp Biol 50, 695–706.

Helmuth J a, Paul G, Sbalzarini IF, 2010. Beyond co-localization: inferring spatial interactions between sub-cellular structures from microscopy images. BMC Bioinformatics 11, 372.

Henry JJ, Perry KJ, 2008. MAPK activation and the specification of the D quadrant in the gastropod mollusc, Crepidula fornicata. Dev. Biol. 313, 181–95.

Henry JJ, Perry KJ, Fukui L, Alvi N, 2010. Differential localization of mRNAs during early development in the mollusc, Crepidula fornicata. Integr Comp Biol 50, 720–33.

Herpin A, Lelong C, Becker T, Rosa F, Favrel P, Cunningham C, 2005. Structural and functional evidence for a singular repertoire of BMP receptor signal transducing proteins in the lophotrochozoan Crassostrea gigas suggests a shared ancestral BMP/activin pathway. FEBS J. 272, 3424–3440.

Herpin A, Lelong C, Favrel P, 2004. Transforming growth factor-β-related proteins: an ancestral and widespread superfamily of cytokines in metazoans. Dev. Comp. Immunol. 28, 461–485.

Hirokawa N, Tanaka Y, Okada Y, Takeda S, 2006. Nodal Flow and the Generation of Left-Right Asymmetry. Cell 125, 33–45.

Hozumi S, Maeda R, Taniguchi K, Kanai M, Shirakabe S, Sasamura T, Speder P, Noselli S, Aigaki T, Murakami R, Matsuno K, 2006. An unconventional myosin in Drosophila reverses the default handedness in visceral organs. Nature 440, 798–802.

Huebner E, Anderson E, 1976. Comparative Spiralian Oogenesis- Structural Aspects: An Overview. Amer Zool 16, 315–343.

Huelsenbeck JP, Ronquist F, 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755.

Huisken J, Swoger J, Del Bene F, Wittbrodt J, Stelzer E, 2004. Live Embryos by Selective Plane Illumination Microscopy. Science 305, 13–16.

Huminiecki L, Goldovsky L, Freilich S, Moustakas A, Ouzounis C, Heldin C-H, 2009. Emergence, development and diversification of the TGF-beta signalling pathway within the animal kingdom. BMC Evol. Biol. 9, 28.

Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, Rupp R, 2007. Dendroscope: An interactive viewer for large phylogenetic trees. BMC Bioinformatics 8, 1–6.

Huson DH, Scornavacca C, 2012. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. Syst. Biol. 61, 1061–7.

Ishihara K, Nguyen PA, Wu M, Groen AC, Field CM, Mitchison TJ, 2014. Organization of early frog embryos by chemical waves emanating from centrosomes. Philos. Trans. R. Soc. B 369, 1–9.

Ivey K, Srivastava D, 2015. microRNAs as Developmental Regulators. Cold Spring Harb. Perspect. Biol. 7, 1–10.

Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, Yabana M, Harada M, Nagayasu E, Maruyama H, Kohara Y, Fujiyama A, Hayashi T, Itoh T, 2014. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome Res. 24, 1384–1395.

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

Katzman S, Capra JA, Haussler D, Pollard KS, 2011. Ongoing GC-biased evolution is widespread in the human genome and enriched near recombination hotspots. Genome Biol. Evol. 1–17.

Kenny NJ, Namigai EKO, Dearden PK, Hui JHL, Grande C, Shimeld SM, 2014. The Lophotrochozoan TGF-β signalling cassette - diversification and conservation in a key signalling pathway. Int. J. Dev. Biol. 58, 533–49.

Kenny NJ, Shimeld SM, 2012. Additive multiple k-mer transcriptome of the keelworm Pomatoceros lamarckii (Annelida; Serpulidae) reveals annelid trochophore transcription factor cassette. Dev. Genes Evol. 222, 325–339.

Kingsley E, Chan X, Duan Y, Lambert J, 2007. Widespread RNA segregation in a spiralian embryo. Evol Dev 9, 527–39.

Koop D, Richards GS, Wanninger A, Gunter HM, Degnan BM, 2007. The role of MAPK signaling in patterning and establishing axial symmetry in the gastropod Haliotis asinina. Dev. Biol. 311, 200–212.

Kuo D-H, Weisblat DA, 2011. A New Molecular Logic for BMP-Mediated Dorsoventral Patterning in the Leech Helobdella 21, 1282–1288 ST – A New Molecular Logic for BMP–Medi.

Kuroda R, Endo B, Abe M, Shimizu M, 2009. Chiral blastomere arrangement dictates zygotic left/right asymmetry pathway in snails. Nature 462, 790–794.

Lambert JD, 2010. Developmental patterns in spiralian embryos. Curr Biol 20, R72–77.

Lambert JD, Nagy LM, 2001. MAPK signaling by the D quadrant embryonic organizer of the mollusc Ilyanassa obsoleta. Development 128, 45–56.

Lambert JD, Nagy LM, 2002. Asymmetric inheritance of centrosomally localized mRNAs during embryonic cleavages. Nature 420, 682–686.

Lambert JD, Nagy LM, 2003. The MAPK cascade in equally cleaving spiralian embryos. Dev. Biol. 263, 231–241.

Lemon J, 2006. Plotrix: a package in the red light district of R. R-News 6, 8–12.

Levin M, 2005. Left-right asymmetry in embryonic development: a comprehensive review. Mech. Dev. 122, 3–25.

Levin M, Johnson R, Stern C, Kuehn M, Tabin C, 1995. A molecular pathway determining left-right asymmetry in chick embryogenesis. Cell 82, 803–814.

Levin M, Nascone N, 1997. Two molecular models of initial left-right asymmetry generation. Med. Hypotheses 49, 429–435.

Levin M, Palmer AR, 2007. Left/right patterning from the inside out: widespread evidence for intracellular control. BioEssays 29, 271–287.

Levin M, Thorlin T, Robinson KR, Nogi T, Mercola M, 2002. Asymmetries in H+/K+-ATPase and Cell Membrane Potentials Comprise a Very Early Step in Left-Right Patterning. Cell 111, 77–89.

Lillie FR, 1895. The Embryology of the Unionidae. J. Morphol. 10, 1–101.

Linkert M, Rueden CT, Allan C, Burel J-M, Moore W, Patterson A, Loranger B, Moore J, Neves C, Macdonald D, Tarkowska A, Sticco C, Hill E, Rossner M, Eliceiri KW, Swedlow JR, 2010. Metadata matters: access to image data in the real world. J. Cell Biol. 189, 777–782.

Liu M, Davey J, Banerjee R, Han J, Yang F, Aboobaker A, Blaxter M, Davison A, 2013. Fine Mapping of the Pond Snail Left-Right Asymmetry (Chirality) Locus Using RAD-Seq and Fibre-FISH. PLoS One 8, e71067–e71067.

Lobikin M, Wang G, Xu J, Hsieh Y-W, Chuang C-F, Lemire JM, Levin M, 2012. Early, nonciliary role for microtubule proteins in left-right patterning is conserved across kingdoms. Proc. Natl. Acad. Sci. U. S. A. 109, 12586–12591.

Luetjens C, Dorresteijn A, 1998. The site of fertilisation determines dorsoventral polarity but not chirality in the zebra mussel embryo. Zygote 6, 125–135.

Marçais G, Kingsford C, 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27, 764–70.

Martindale MQ, Doe CQ, Morrill JB, 1985. The role of animal-vegetal interaction with respect to the determination of dorsoventral polarity in the equal-cleaving spiralian , Lymnaea palustris. Roux's Arch. Dev. Biol. 281–295.

Martindale MQ, Henry JQ, 1995. Modifications of cell fate specification in equal-cleaving nemertean embryos : alternate patterns of spiralian development. Development 121, 3175–3185.

Massagué J, 1998. TGF- β Signal Transduction. Annu. Rev. Biochem. 67, 753–791.

Massagué J, 2012. TGFβ signalling in context. Nat. Rev. Mol. Cell Biol. 13, 616–30.

Matteoni R, Kreis T, 1987. Translocation and Clustering of Endosomes and Lysosomes Depends on Microtubules. J. Cell Biol. 105, 1253–1265.

Matus DQ, Pang K, Marlow H, Dunn CW, Thomsen GH, Martindale MQ, 2006a. Molecular evidence for deep evolutionary roots of bilaterality in animal development. Proc. Natl. Acad. Sci. U. S. A. 103, 11195–11200.

Matus DQ, Thomsen GH, Martindale MQ, 2006b. Dorso/ventral genes are asymmetrically expressed and involved in germ-layer demarcation during cnidarian gastrulation. Curr Biol 16, 499–505.

McDougall C, Chen W-C, Shimeld S, Ferrier D, 2006. The development of the larval nervous system, musculature and ciliary bands of Pomatoceros lamarckii (Annelida): heterochrony in polychaetes. Front. Zool. 3, 16.

Merkel J, Wollesen T, Lieb B, Wanninger A, 2012. Spiral cleavage and early embryology of a loxosomatid entoproct and the usefulness of spiralian apical cross patterns for phylogenetic inferences. BMC Dev Biol 12, 1–11.

Meshcheryakov VN, Beloussov L V, 1975. Asymmetrical rotations of blastomeres in early cleavage of gastropoda. Dev. Genes Evol. 177, 193–203.

Meunier J, Lemoine F, Soumillon M, Weier M, Guschanski K, Hu H, Khaitovich P, Kaessmann H, 2013. Birth and expression evolution of mammalian microRNA genes. Genome Res. 23, 34–45.

Mogilner A, Fogelson B, 2015. Cytoskeletal Chirality: Swirling Cells Tell Left from Right. Curr. Biol. 25, R490–R514.

Morrill JB, Blair CA, Larsen WJ, 1973. Regulative Development in the Pulmonate Gastropod, Lymnaea palustris, as Determined by Blastomere Deletion Experiments. J. Exp. Zool. 183, 47–56.

Munro E, Nance J, Priess JR, 2004. Cortical Flows Powered by Asymmetrical Contraction Transport PAR Proteins to Establish and Maintain Anterior-Posterior Polarity in the Early C. elegans Embryo. Dev Cell 7, 413–424.

Naganathan SR, Fürthauer S, Nishikawa M, Jülicher F, Grill SW, 2014. Active torque generation by the actomyosin cell cortex drives left–right symmetry breaking. Elife 3, 1–16.

Nakamura T, Hamada H, 2012. Left-right patterning: conserved and divergent mechanisms. Development 139, 3257–62.

Namigai EKO, Kenny NJ, Shimeld SM, 2014. Right across the tree of life: the evolution of left-right asymmetry in the Bilateria. Genesis 52, 458–70.

Nance J, Zallen JA, 2011. Elaborating polarity: PAR proteins and the cytoskeleton. Development 138, 799–809.

Nishide K, Mugitani M, Kumano G, Nishida H, 2012. Neurula rotation determines left-right asymmetry in ascidian tadpole larvae. Development 139, 1467–75.

Noël ES, Verhoeven M, Lagendijk AK, Tessadori F, Smith K, Choorapoikayil S, den Hertog J, Bakkers J, 2013. A Nodal-independent and tissue-intrinsic mechanism controls heart-looping chirality. Nat. Commun. 4, 2754.

Nonaka S, Tanaka Y, Okada Y, Takeda S, Harada A, Kanai Y, Kido M, Hirokawa N, 1998. Randomization of left-right asymmetry due to loss of nodal cilia generating leftward flow of extraembryonic fluid in mice lacking KIF3B motor protein. Cell 95, 829–837.

Özüak O, Buchta T, Roth S, Lynch J, 2014. Ancient and diverged TGF-β signaling components in Nasonia vitripennis. Dev. Genes Evol. 224, 223–33.

Palmer AR, 2004. Symmetry breaking and the evolution of development. Science 306, 828–833.

Pang K, Ryan JF, Baxevanis AD, Martindale MQ, 2011. Evolution of the TGF-β signaling pathway and its potential role in the ctenophore, Mnemiopsis leidyi. PLoS One 6, e24152.

Paps J, Baguñà J, Riutort M, 2009. Lophotrochozoa internal phylogeny: new insights from an up-to-date analysis of nuclear ribosomal genes. Proc. Biol. Sci. 276, 1245–54.

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

Parra G, Bradnam K, Ning Z, Keane T, Korf I, 2009. Assessing the gene space in draft genomes. Nucleic Acids Res. 37, 289–97.

Pasquinelli AE, Reinhart BJ, Slack F, Martindale MQ, Kuroda MI, Maller B, Hayward DC, Ball EE, Degnan B, Mu P, Srinivasan A, Fishman M, Finnerty J, Corbo J, Levine M, Leahy P, Davidson E, Ruvkun G, 2000. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. Nature 408, 86–89.

Pennerstorfer M, Scholtz G, 2012. Early cleavage in Phoronis muelleri (Phoronida) displays spiral features. Evol Dev 14, 484–500.

Perry KJ, Henry JQ, 2015. CRISPR/Cas9-mediated genome modification in the mollusc, Crepidula fornicata. Genesis 53, 237–244.

Petzoldt AG, Coutelis J-B, Géminard C, Spéder P, Suzanne M, Cerezo D, Noselli S, 2012. DE-Cadherin regulates unconventional Myosin ID and Myosin IC in Drosophila left-right asymmetry establishment. Development 139, 1874–1884.

Pevzner PA, Tang H, Waterman MS, 2001. An eulerian path approach to DNA fragment assembly. Proc. Natl. Acad. Sci. U. S. A. 98, 9748–9753.

Pick KS, Philippe H, Schreiber F, Erpenbeck D, Jackson DJ, Wrede P, Wiens M, Alié A, Morgenstern B, Manuel M, Wörheide G, 2010. Improved Phylogenomic Taxon Sampling Noticeably Affects Nonbilaterian Relationships. Mol Biol Evol 27, 1983–1987.

Pohl C, 2011. Left-right patterning in the C. elegans embryo: Unique mechanisms and common principles. Commun Integr Biol 4, 34–40.

Pohl C, Bao Z, 2010. Chiral Forces Organize Left-Right Patterning in C. elegans by Uncoupling Midline and Anteroposterior Axis. Dev Cell 19, 402–412.

Putnam NH, Butts T, Ferrier DEK, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu J-K, Benito-Gutiérrez EL, Dubchak I, Garcia-Fernàndez J, Gibson-Brown JJ, Grigoriev I V, Horton AC, de Jong PJ, Jurka J, Kapitonov V V, Kohara Y, Kuroki Y, Lindquist E, Lucas S, Osoegawa K, Pennacchio LA, Salamov AA, Satou Y, Sauka-Spengler T, Schmutz J, Shin-I T, Toyoda A, Bronner-Fraser M, Fujiyama A, Holland LZ, Holland PWH, Satoh N, Rokhsar DS, 2008. The amphioxus genome and the evolution of the chordate karyotype. Nature 453, 1064–1071.

Quah S, Hui JHL, Holland PWH, 2015. A Burst of miRNA Innovation in the Early Evolution of Butterflies and Moths. Mol. Biol. Evol. 32, 1161–1174.

Rabinowitz JS, Lambert JD, 2010. Spiralian quartet developmental potential is regulated by specific localization elements that mediate asymmetric RNA segregation. Development 137, 4039–49.

Rappaport R, Rappaport B, 1988. Reversing Cytoplasmic Flow in Nucleated , Constricted Sand Dollar Eggs. J. Exp. Zool. 247, 92–98.

Raya A, Izpisúa Belmonte JC, 2006. Left-right asymmetry in the vertebrate embryo: from early information to higher-level integration. Nat Rev Genet 7, 283–93.

Reynaud EG, Peychl J, Huisken J, Tomancak P, 2014. Guide to light-sheet microscopy for adventurous biologists. Nat. Methods 12, 30–34.

Rice P, Longden I, Bleasby A, 2000. The European Molecular Biology Open Software Suite EMBOSS : The European Molecular Biology Open Software Suite. Trends Genet. 16, 276–277.

Road H, 1993. Cortical and Cytoplasmic Flow Polarity in Early Embryonic Cells of Caenorhabditis elegans 121, 1343–1355.

Romiguier J, Gayral P, Ballenghien M, Bernard A, Cahais V, Chenuil A, Chiari Y, Dernat R, Duret L, Faivre N, Loire E, Lourenco J, Nabholz B, Roux C,

Tsagkogeorga G, Weber A-T, Weinert L, Belkhir K, Bierne N, Glémin S, Galtier N, 2014. Comparative population genomics in animals uncovers the determinants of genetic diversity. Nature 515, 261–263.

Romiguier J, Ranwez V, Douzery E, Galtier N, 2010. Contrasting GC-content dynamics across 33 mammalian genomes : Relationship with life-history traits and chromosome sizes. Genome Res. 20, 1001–1009.

Rosa A, Spagnoli F, Brivanlou A, 2009. The miR-430/427/302 family controls mesendodermal fate specification via species-specific target selection. Dev. Cell 16, 517–527.

Roth S, Lynch JA, 2009. Symmetry Breaking During Drosophila Oogenesis. Cold Spring Harb Perspect Biol 1, 1–22.

Saarma M, 2000. GDNF - a stranger in the TGF-beta superfamily ? Eur. J. Biochem. 267, 6968–6971.

Schier A, 2003. Nodal signaling in vertebrate development. Annu. Rev. Cell Dev. Biol. 19, 589–621.

Schier A, Shen M, 2000. Nodal signalling in vertebrate development. Nature 403, 385–9.

Schier AF, 2009. Nodal Morphogens. Cold Spring Harb. Perspect. Biol. 1, 1–20.

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez J-Y, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A, 2012. Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676–682.

Schonegg S, Hyman A, Wood WB, 2014. Timing and mechanism of the initial cue establishing handed left-right asymmetry in Caenorhabditis elegans embryos. Genesis 52, 572–580.

Schweickert A, Campione M, Steinbeisser H, Blum M, 2000. Pitx2 isoforms: Involvement of Pitx2c but not Pitx2a or Pitx2b in vertebrate left-right asymmetry. Mech. Dev. 90, 41–51.

Seaver EC, 2014. Variation in spiralian development: insights from polychaetes. Int. J. Dev. Biol. 58, 457–67.

Segrove F, 1941. The Development of the Serpulid Pomatoceros triqueter (L.). Q. J. Microsc. Sci. 82, 467–540.

Shibazaki Y, Shimizu M, Kuroda R, 2004. Body Handedness Is Directed by Genetically Determined Cytoskeletal Dynamics in the Early Embryo. Curr. Biol. 14, 1462–1467.

Shimeld SM, Levin M, 2006. Evidence for the regulation of left-right asymmetry in Ciona intestinalis by ion flux. Dev Dyn 235, 1543–53.

Shimizu K, Iijima M, Setiamarga D, Sarashina I, Kudoh T, Asami T, Gittenberger E, Endo K, 2013. Left-right asymmetric expression of dpp in the mantle of gastropods correlates with asymmetric shell coiling. Evodevo 4, 1–7.

Shimizu K, Sarashina I, Kagi H, Endo K, 2011. Possible functions of Dpp in gastropod shell formation and shell coiling. Dev Genes Evol 221, 59–68.

Shivanandan A, Radenovic A, Sbalzarini IF, 2013. MosaicIA : an ImageJ / Fiji plugin for spatial pattern and interaction analysis. BMC Bioinformatics 14, 1–10.

Simakov O, Marlétaz F, Cho S-J, Edsinger-Gonzales E, Havlak P, Hellsten U, Kuo D-H, Larsson T, Lv J, Arendt D, Savage R, Osoegawa K, de Jong P, Grimwood J, Chapman JA, Shapiro H, Aerts A, Otillar RP, Terry AY, Boore JL, Grigoriev I V, Lindberg DR, Seaver EC, Weisblat DA, Putnam NH, Rokhsar DS, 2013. Insights into bilaterian evolution from three spiralian genomes. Nature 493, 526–531.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V, Zdobnov EM, 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 1–3.

Singh D, Pohl C, 2014. Coupling of rotational cortical flow, asymmetric midbody positioning, and spindle rotation mediates dorsoventral axis formation in C. elegans. Dev. Cell 28, 253–267.

Soldi R, Ramella L, Gambi M, Sordino P, Sella G, 1994. Genome size in polychaetes: relationship with body length and life habit, in: Dauvin, J.-C., Laubier, L., DJ, R. (Eds.), Actes de La 4ième Conférence Internationale Des Polychètes. Mém. Mus. Natn. Hist. Nat., Paris, pp. 129–135.

Spéder P, Adam G, Noselli S, 2006. Type ID unconventional myosin controls left/right asymmetry in Drosophila. Nature 440, 803–807.

Srivastava M, Simakov O, Chapman J, Fahey B, Gauthier MEA, Mitros T, Richards GS, Conaco C, Dacre M, Hellsten U, Larroux C, Putnam NH, Stanke M, Adamska M, Darling A, Degnan SM, Oakley TH, Plachetzki DC, Zhai Y, Adamski M, Calcino A, Cummins SF, Goodstein DM, Harris C, Jackson DJ, Leys SP, Shu S, Woodcroft BJ, Vervoort M, Kosik KS, Manning G, Degnan BM, Rokhsar DS, 2010. The Amphimedon queenslandica genome and the evolution of animal complexity. Nature 466, 720–726.

Stanke M, Morgenstern B, 2005. AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 33, W465–W467.

Strozzi F, Mazza R, Malinverni R, Williams JL, 2009. Annotation of 390 bovine miRNA genes by sequence similarity with other species. Anim. Genet. 40, 125.

Struck TH, Paul C, Hill N, Hartmann S, Hösel C, Kube M, Lieb B, Meyer A, Tiedemann R, Purschke G, Bleidorn C, 2011. Phylogenomic analyses unravel annelid evolution. Nature 471, 95–98.

Sturtevant A, 1923. Inheritance of direction of coiling in Limnaea. Science 58, 269–270.

Tabin C, 2006. The Key to Left-Right Asymmetry. Cell 127, 27–32.

Tabin CJ, Vogan KJ, 2003. A two-cilia model for vertebrate left-right axis specification. Genes Dev 17, 1–6.

Takeuchi T, Kawashima T, Koyanagi R, Gyoja F, Tanaka M, Ikuta T, Shoguchi E, Fujiwara M, Shinzato C, Hisata K, Fujie M, Usami T, Nagai K, Maeyama K, Okamoto K, Aoki H, Ishikawa T, Masaoka T, Fujiwara A, Endo K, Endo H, Nagasawa H, Kinoshita S, Asakawa S, Watabe S, Satoh N, 2012. Draft Genome of the Pearl Oyster Pinctada fucata: A Platform for Understanding Bivalve Biology. DNA Res. 19, 117–130.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S, 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol. Biol. Evol. 30, 2725–9.

Tarver JE, Donoghue PCJ, Peterson KJ, 2012. Do miRNAs have a deep evolutionary history? BioEssays 34, 857–866.

Team RDC, 2008. R: A language and environment for statistical computing.

Thompson H, Shaw MK, Dawe HR, Shimeld SM, 2012. The formation and positioning of cilia in Ciona intestinalis embryos in relation to the generation and evolution of chordate left-right asymmetry. Dev Biol 364, 214–23.

Tsai I, Otto T, Berriman M, 2010. Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps. Genome Biol. 11, 1–9.

Tseng Q, Duchemin-Pelletier E, Deshiere A, Balland M, Guillou H, Filhol O, Théry M, 2012. Spatial organization of the extracellular matrix regulates cell-cell junction positioning. Proc. Natl. Acad. Sci. U. S. A. 109, 1506–1511.

Van der Zee M, da Fonseca RN, Roth S, 2008. TGF beta signaling in Tribolium: vertebrate-like components in a beetle. Dev. Genes Evol. 218, 203–13.

Vandenberg LN, Levin M, 2010. Far from solved: a perspective on what we know about early mechanisms of left-right asymmetry. Dev. Dyn. 239, 3131–3146.

Vandenberg LN, Levin M, 2012. Polarity proteins are required for left-right axis orientation and twin-twin instruction. Genesis 50, 219–34.

Vandenberg LN, Levin M, 2013. A unified model for left-right asymmetry? Comparison and synthesis of molecular models of embryonic laterality. Dev. Biol. 379, 1–15.

Walsh DW, Godson C, Brazil DP, Martin F, 2010. Extracellular BMP-antagonist regulation in development and disease: tied up in knots. Trends Cell Biol. 20, 244–56.

Wan LQ, Ronaldson K, Park M, Taylor G, Zhang Y, Gimble JM, Vunjak-Novakovic G, 2011. Micropatterned mammalian cells exhibit phenotype-specific left-right asymmetry. Proc. Natl. Acad. Sci. U. S. A. 108, 12295–12300.

Watanabe H, Schmidt H a, Kuhn A, Höger SK, Kocagöz Y, Laumann-Lipp N, Ozbek S, Holstein TW, 2014. Nodal signalling determines biradial asymmetry in Hydra. Nature 515, 112–115.

Weisblat D, Huang F, 2001. An overview of glossiphoniid leech development. Can. J. Zool. 79, 218–232.

Whelan S, Goldman N, 2001. A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach. Mol. Biol. Evol. 18, 691–699.

Whitman C, 1878. The Embryology of Clepsine. Q. J. Microsc. Sci. 71, 215–315.

Wickham H, 2009. ggplot2: elegant graphics for data analysis. Springer New York.

Wickham H, 2011. The Split-Apply-Combine Strategy for Data 40.

Wilson EB, 1892. The Cell-lineage of Nereis. J. Morphol. 6, 361–478.

Wilson EB, 1898. Considerations on Cell-lineage and Ancestral Reminiscence: a re-examination of some points in the early development of annelids and polyclades. Ann. N. Y. Acad. Sci. 11, 1–27.

Wood WB, 1991. Evidence from reversal of handedness in C. elegans embryos for early cell interactions determining cell fates. Nature 349, 536–538.

Yan YT, Gritsman K, Ding J, Burdine RD, Corrales JD, Price SM, Talbot WS, Schier a F, Shen MM, 1999. Conserved requirement for EGF-CFC genes in vertebrate left-right axis formation. Genes Dev. 13, 2527–37.

Yi K, Rubinstein B, Li R, 2013. Symmetry breaking and polarity establishment during mouse oocyte maturation. Philos. Trans. R. Soc. B 368, 1–7.

Yi K, Unruh JR, Deng M, Slaughter BD, Rubinstein B, Li R, 2011. Dynamic maintenance of asymmetric meiotic spindle position through Arp2/3-complex-driven cytoplasmic streaming in mouse oocytes. Nat. Cell Biol. 13, 1252–1258.

Yu J-K, Holland LZ, Holland ND, 2002. An amphioxus nodal gene (AmphiNodal) with early symmetrical expression in the organizer and mesoderm and later asymmetrical expression associated with left–right axis formation. Evol Dev 4, 418–425.

Zantke J, Bannister S, Rajan VBV, Raible F, Tessmar-Raible K, 2014. Genetic and genomic tools for the marine annelid Platynereis dumerilii. Genetics 197, 19–31.

Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, Yang P, Zhang L, Wang X, Qi H, Xiong Z, Que H, Xie Y, Holland PWH, Paps J, Zhu Y, Wu F, Chen Y, Wang J, Peng C, Meng J, Yang L, Wang J, 2012. The oyster genome reveals stress adaptation and complexity of shell formation. Nature 490, 49–54.

Zhou X, Sasaki H, Lowe L, Hogan BL, Kuehn MR, 1993. Nodal is a novel TGF-beta-like gene expressed in the mouse node during gastrulation. Nature 361, 543–547.

Zhu L, Zhang Y, Zhang W, Yang S, Chen J, Tian D, 2009. Patterns of exon-intron architecture variation of genes in eukaryotic genomes. BMC Genomics 10, 1–12.

Appendix 1

Namigai EKO, Kenny NJ, Shimeld SM, 2014. Right across the tree of life: the evolution of left-right asymmetry in the Bilateria. Genesis 52, 458–70.

genesis 52:458–470 (2014)

REVIEW

Right Across the Tree of Life: The Evolution of Left–Right Asymmetry in the Bilateria

Erica K.O. Namigai, Nathan J. Kenny, and Sebastian M. Shimeld*

Department of Zoology, University of Oxford, South Parks Road, Oxford, United Kingdom

Received 9 December 2013; Revised 27 January 2014; Accepted 29 January 2014

Abstract: Directional left/right (LR) asymmetries, in which there are consistent, heritable differences in morphology between the left and right sides of bilaterally symmetrical organisms, are found in animals across the Bilateria. For many years, we have lacked evidence for shared mechanisms underlying their development. This led to the supposition that the mechanisms driving establishment of LR asymmetries, and consequently the asymmetries themselves, had evolved separately in the three major Superphyla that constitute the Bilateria. The recent discovery that the transforming growth factor-beta (TGF-β) ligand Nodal plays a role in the regulation of LR asymmetry in both Deuterostomia and Lophotrochozoa has reignited debate in this field, as it suggests that at least this aspect of the development of the LR axis is conserved. In this review, we discuss evidence for shared mechanisms of LR asymmetry establishment across the bilaterian tree of life and consider how these mechanisms might have diverged across the Metazoa over the last 500 million years or so of evolution. As well as the likelihood that Nodal is an ancestral mechanism for regulating LR asymmetry, we reemphasize cytoskeletal architecture as a potential shared mechanism underlying symmetry breaking. However, convergent evolution remains a distinct possibility and study of a wider diversity of species will be needed to distinguish between conserved and lineage-specific mechanisms. genesis 52:458–470, 2014. © 2014 Wiley Periodicals, Inc.

Key words: asymmetry; Lophotrochozoa; Ecdysozoa; LR; Bilateria

BILATERAL SYMMETRY AND ASYMMETRY

Most living animals are classified within the Bilateria (Fig. 1) and have a body plan that is, for at least part of their life cycle, fundamentally bilaterally symmetrical. However, morphological asymmetries across the left/right (LR) axis are present in many such species. Classic examples include the directional coiling of snail shells, the looping of the vertebrate gut, and the left-sided placement of the human heart (Palmer, 2004). It should be noted that the phrase "LR axis" could be regarded as something of a misnomer, as from a mechanistic developmental perspective there is in effect one axis on either side of the midline (as discussed in Palmer, 2004). However, for the purposes of this review, we will use the term LR axis to indicate both the left and right sides of the midline, while keeping in mind that this is not a bona fide continuous axis with a single patterning mechanism spanning from one side to the other.

The Bilateria are divided into three major Superphyla, the Ecdysozoa, Lophotrochozoa, and Deuterostomia, with the Acoelomorpha posited as the earliest-diverging bilaterian lineage and predating their radiation (Fig. 1, and references therein). The bilateral symmetry seen in these clades contrasts with the animal lineages historically known as the "Radiata" such as cnidarians and ctenophores. These are typically described as showing some form of radial symmetry, although there is some evidence for bilaterality at the molecular level in at least one cnidarian species (Matus *et al.*, 2006a,b). The relative placement of these lineages in the tree of life leads

Abbreviations: AP, anterior–posterior; DV, dorsal–ventral; LR, left/right; TGF-β, transforming growth factor-beta.

*Correspondence to: Sebastian M. Shimeld, Evolution and Development Research Group, Department of Zoology, University of Oxford, South Parks Road, Oxford, United Kingdom.
E-mail: sebastian.shimeld@zoo.ox.ac.uk

Published online 9 February 2014 in
Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/dvg.22748

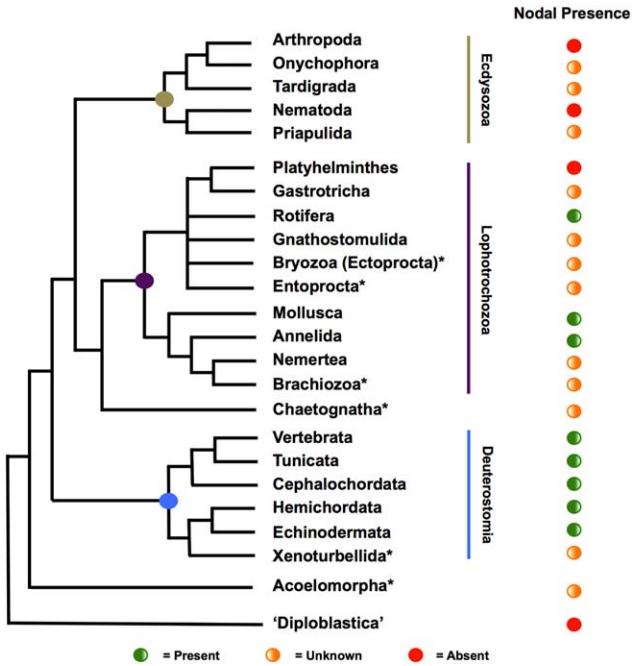


FIG. 1. Consensus cladogram of Metazoan inter-relationships (based primarily on Dunn *et al*., 2008; Pick *et al*., 2010). Major divisions of the Bilateria into the three superphyla Deuterostomia, Lophotrochozoa, and Ecdysozoa represented with colored bars and circles at clade nodes. Note that there is continued disagreement in the literature about some of the nodes shown in this tree, for example, the precise arrangement of the Tardigrada, Nematoda, and Priapulida, and the relative position of the Platyhelminthes. Some other taxa whose placement in the cladogram is contested are indicated with an asterisk. All non-Bilateria are incorporated into the paraphyletic group "Diploblastica"—for discussion of the possible inter-relationships in this grouping, we refer the interested reader to Hejnol *et al*. (2009) or Pick *et al*. (2010). At the right, we indicate the known presence of the *Nodal* gene as determined by prior publications (Duboc *et al*., 2005; Grande and Patel, 2009; Kuroda *et al*., 2009; Morokuma *et al*., 2002; Yu *et al*., 2002) or our analyses (Kenny *et al*., manuscript submitted) with a green dot, known absence of *Nodal* (when complete genomes available) with a red dot, and currently undetermined presence or absence with an orange dot.

to the inference that bilateral symmetry evolved only once and that the common ancestor of the Bilateria, as the name suggests, was bilaterally symmetrical.

During embryonic development, bilateral symmetry is apparent once the anterior–posterior (AP) and dorsal–ventral (DV) axes are established. True symmetry is, however, generally short lived in development, as most bilaterians quickly establish asymmetries across their LR axis. These can be limited to the results of developmental noise causing subtle differences in growth and pattern on the left and right sides (known as fluctuating asymmetry) or to the random asymmetric placement of structures in which all individuals in a population are asymmetric but sidedness is stochastic (known as antisymmetry). To many, however, the most interesting asymmetries are directional asymmetries, which are fixed in a certain direction in a population, and under genetic control (Palmer, 2004).

Directional asymmetries are found across the Bilateria, but how they are established and patterned has been the cause of recent debate. Much of our understanding of these processes has come from studies in vertebrates. A critical question is, what actually breaks symmetry in the first place, allowing left and right sides to follow different developmental programs? Studies in model vertebrates have focused on two main mechanisms: the nodal flow hypothesis, which proposes that the chirality of cilia coupled with their arrangement at the node generates LR directional fluid flow and hence drives LR asymmetry (reviewed by Capdevila *et al*., 2000; Nonaka *et al*., 1998; Tabin, 2006), and the ion flow hypothesis, which proposes that asymmetry is established by differential transport of charged molecules in response to the different voltages across the LR axis of the embryo, propagated by the action of molecules like H^+/K^+ ATPases (Levin *et al*., 2002). Both models have caveats when considering their general applicability to all vertebrate embryos, and we do not have the space in this review to fully explore these. Instead, the reader is referred to recent reviews that discuss these issues (Hirokawa *et al*., 2006; Levin and Palmer, 2007; Nakamura and Hamada, 2012; Vandenberg and Levin, 2010). Importantly, as discussed below in detail, neither model is a strong candidate for explaining the breaking of symmetry across the Bilateria. This, coupled with a general lack of data from most invertebrate lineages, led to a general supposition that the mechanisms for establishment of directional asymmetry evolved separately in the three bilaterian Superphyla (Palmer, 1996).

NODAL SIGNALING AND LR ASYMMETRY

A second critical question has been how broken symmetry can be translated into asymmetric gene expression, developmental programs, and hence morphology. In vertebrate studies, there is some consensus here in that in all species so far examined, it involves the left-sided expression of the transforming growth factor-beta (TGF-β) ligand *Nodal* and the homeobox transcription factor *Pitx2*. *Nodal* was first described in 1993 (Zhou *et al*., 1993), and along with performing a key role in early embryogenesis with respect to mesendoderm induction, its later left-sided expression adjacent to the node induces additional left-sided *Nodal*, left-sided expression of *Pitx2*, and hence left-sided morphogenesis (Beck *et al*., 2002; Levin *et al*., 1995). Mouse and chicken have single copies of the *Nodal* gene, whereas *Xenopus* possesses five paralogs (*Xnr1*, *2*, *4*, *5*, and *6*) and zebrafish three (*cyclops*, *squint*, and *southpaw*). As can be seen at the right of Figure 1, *Nodal* is so far conspicuous by its absence in the Ecdysozoa, despite the

widespread presence of LR asymmetries in this Superphylum. Initially, this reinforced the idea that mechanisms regulating LR asymmetry were not conserved between the bilaterian Superphyla. However, more recent study has shown that at least one copy of *Nodal* is present in all invertebrate deuterostomes thus far studied, as well as being identifiable in the genomes of most sequenced members of the Lophotrochozoa (Duboc *et al*., 2005; Grande and Patel, 2009; Kuroda *et al*., 2009; Yu *et al*., 2002; Kenny *et al*., manuscript submitted).

These key studies have shown that the asymmetric expression of both *Nodal* and *Pitx* (the gene family to which Pitx2 belongs) are also found in both invertebrate deuterostomes and lophotrochozoans. Furthermore, disrupting Nodal function can affect the expression of *Pitx* and the development of asymmetric morphology (Grande and Patel, 2009; Kuroda *et al*., 2009). This has led to re-examination of the question of whether directional LR asymmetry evolved independently in each of the three bilaterian Superphyla or whether at least some of the mechanisms establishing LR asymmetry are ancestral and hence conserved. It also seems that *Nodal* is expressed on the right-hand side of most invertebrates (Duboc and Lepage, 2008; Grande and Patel, 2009; Kuroda *et al*., 2009) and plays its role in patterning asymmetry on this side of the body, whereas in the chordates, it is found on the left-hand side (Boorman and Shimeld, 2002a,b; Chea *et al*., 2005; Morokuma *et al*., 2002; Yu *et al*., 2002). This mirror-image distribution of roles is likely related to the so-called DV inversion event, suggesting that bilateral asymmetry (or at least one-sided expression of *Nodal*) predates this.

In this review, we focus on the evolutionary ancestry of mechanisms for establishing and propagating directional asymmetries, exploring studies across all three Superphyla. Focusing first on the invertebrate deuterostomes, then the lophotrochozoans, and finally the ecdysozoans, we will summarize current understanding, evaluate mechanistic similarities between lineages, and propose evolutionary explanations for the similarities and differences we see in the mechanisms between living bilaterian lineages.

ASYMMETRY IN INVERTEBRATE DEUTEROSTOMES

As immediate outgroups to the well-studied Vertebrata, invertebrate deuterostomes are key systems for understanding the ancestry of chordate LR asymmetry mechanisms. There are probably only three living lineages of invertebrate deuterostomes (Fig. 1): Tunicata (sea squirts and allies), Cephalochordata (amphioxus and allies), and Ambulacraria (sea urchins and hemichordates; the placement of Xenoturbella is controversial, and as we currently know little of its development, it will not be considered further here). All have been studied to some extent, with most data coming from the tunicates, and we will summarize these findings in turn.

Tunicates

The Tunicata are the invertebrate lineage most closely related to the vertebrates (Fig. 2). Tunicate embryos develop into a tadpole larva with dorsal, hollow neural tube and underlying notochord. Larvae typically show asymmetry in the sensory vesicle sited in the anterior brain (Boorman and Shimeld, 2002a). The larvae of ascidians, which comprise the majority of tunicates, metamorphose into a sessile filter feeding adult, which also shows asymmetry, most obviously in the positioning and coiling of the gut and other internal organs (Boorman and Shimeld, 2002a).

The regulation of asymmetry has been primarily studied in two ascidian species, *Ciona intestinalis* and *Halocynthia roretzi*. These species belong to widely divergent ascidian lineages; however, their early development is remarkably similar with each showing essentially the same stereotypical pattern of cleavage and cell lineage (Lemaire, 2009). In both species, the nodal–Pitx signaling system is deployed on the left-hand side of the body, as it is in vertebrates (Boorman and Shimeld, 2002b; Morokuma *et al*., 2002). However, the primary tissue differs, with both *H. roretzi* and *C. intestinalis* genes being broadly expressed in the left-side ectoderm at the neurula and tailbud stages, as opposed to around the node and in lateral plate mesoderm in vertebrates (Collignon *et al*., 1996; Lowe *et al*., 1996). One possible explanation for this difference is that tunicates are thought to have lost the node-based body axis patterning used by vertebrates (and amphioxus; see below), replacing it with lineage and local signaling-based mechanisms (Kourakis and Smith, 2005; Lemaire, 2009). Conservation of the process extends to the regulation of *Pitx* by nodal (Morokuma *et al*., 2002; Yoshida and Saiga, 2008) and to the deployment of lefty: although the precise regulatory role of lefty is yet to be dissected, overexpression experiments produce similar phenotypes to treatment with the nodal receptor antagonist SB431542, suggesting that it also functions to inhibit nodal signaling (Mita and Fujiwara, 2007).

A small number of studies have also begun to investigate the mechanism underlying how left-sided expression of these genes is initially triggered, although with sometimes conflicting results. Shimeld and Levin (2006) investigated whether manipulation of ion channels or pumps by pharmacology affected asymmetry in C. *intestinalis*, finding that compounds which in vertebrates interfered with H^+K^+ ATPase activity could affect asymmetry as revealed by ectopic right-sided expression of *Pitx* in treated embryos. With respect to

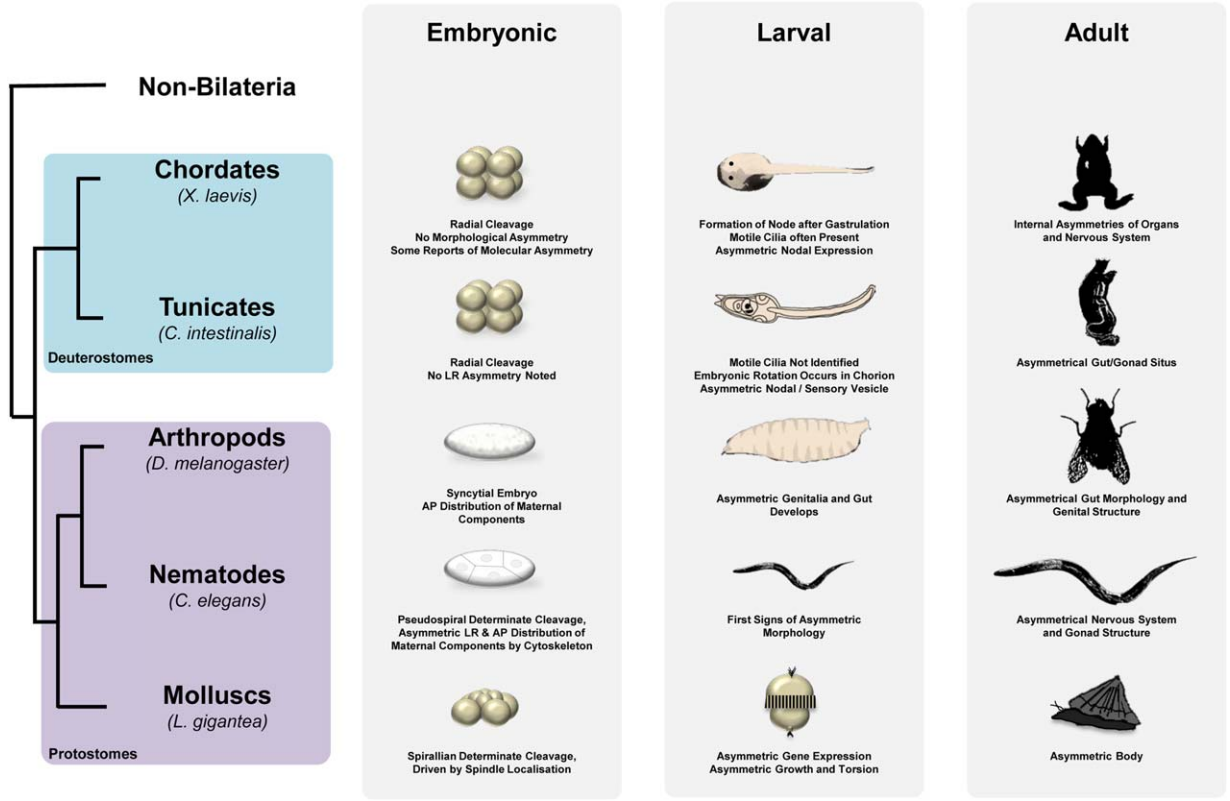


FIG. 2. Cladogram with illustrations of key stages in the development of some of the clades, referred to in text, at early embryonic, larval, and adult stages, illustrated with generally representative or well-researched organisms. Some of the key findings related to asymmetry establishment or presence are noted in text under illustrations. Please see the main text for relevant references.

the potential for cilia to be involved in symmetry-breaking in tunicates, classical descriptions of ascidian development have failed to identify cilia prior to the larval stage, when asymmetry is already well established (for review, see Satoh, 1994). However, this question has been recently revisited using fixation protocols designed to preserve primary cilia (Thompson *et al*., 2012). This study identified small (1–2 μm long) transient cilia on all ectoderm cells (other than those at the midline) at the late neurula and early tailbud stages, coincidental with the activation by these cells of *nodal* expression. The cilia are posteriorly localized on each cell; however, analysis of their structure by TEM found collapsed axonemes and a lack of dynein arms, consistent with primary cilia structure but inconsistent with motility. This led the authors to postulate that the cilia were not involved in a nodal flow-like mechanism in ascidians but were more likely to be sensory in nature. However, a parallel study in *H. roretzi* showed that during this period of development, embryos rotated gently inside their chorions and always came to rest on their left side (Nishide *et al*., 2012). Furthermore, they experimentally demonstrated that contact between the side of the embryo and chorion wall was sufficient to induce *nodal* gene expression. These authors also presented evidence that other ascidian embryos rotated inside their chorions, including those of *C. intestinalis*, although the degree of rotation varied and embryos did not always end up lying in the same orientation. Cilia, although not proven to be involved, are obviously likely candidates for driving such rotation.

These conflicting data have yet to be reconciled. One possibility is that ascidian embryos develop different types of cilia at different places and/or times such that TEM analysis missed those responsible for embryo motility. Alternatively other forces may drive embryo rotation, for example, the *C. intestinalis* chorion is lined internally with a dense layer of maternally supplied test cells that are only a few micrometers away from the embryo, in contrast to the large space inside a *H. roretzi* chorion. Nonmotile cilia might also be involved in asymmetry via a sensory function, as contact between embryo and chorion substratum appears necessary for symmetry breaking in *H. roretzi*. Mechanistic diversity between ascidians is also possible; as discussed above, although morphologically similar, *C. intestinalis* and *H. roretzi* are genetically very divergent. Furthermore, removal of the chorion has different effects on the two species, leading to bilateral expression of left-sided genes in *C. intestinalis* (Shimeld and

Levin, 2006; Yoshida and Saiga, 2008), but loss of left-sided expression of *nodal* in *H. roretzi* (Nishide *et al*., 2012). This suggests that if an embryo–substrate contact mechanism is deployed in these species to break symmetry, it is working in opposite ways in the two species (repressing left-sidedness in *C. intestinalis*, activating it in *H. roretzi*).

In summary, although it is clear that ascidians share left-sided *nodal* and *Pitx* expression with vertebrates and that the regulation of *Pitx* by nodal is also conserved, it is less clear whether upstream mechanisms of symmetry breaking are also shared ancestrally. An intriguing possibility is that a node cilia-based symmetry breaking mechanism was deployed by the common ancestor and that ascidians have adapted this to the use of cilia (be they sensory, motile, or both) over the whole ectoderm as a corollary of losing the node. Understanding symmetry breaking and *nodal* regulation in amphioxus would help resolve this.

Cephalochordates

Amphioxus and allies form the Cephalochordata, the earliest diverging living chordate lineage (Fig. 1). The amphioxus larva is one of the most asymmetric bilaterians described. It possesses a canonical chordate body plan, with dorsal neural tube, notochord, and somite muscle blocks (Conklin, 1932). However, the somites are asymmetrically aligned, the mouth opens on the left side of the head, the gill slits extend up to the right side of the head, and the anterior head cavities undergo very different morphogenesis on left and right sides (reviewed by Boorman and Shimeld, 2002a). Some symmetry is restored at morphogenesis; however, asymmetry clearly persists in brain, gut, and associated organs.

Both *nodal* and *Pitx* are left sided in amphioxus (Boorman and Shimeld, 2002b; Yasui *et al*., 2000; Yu *et al*., 2002). Their expression more closely resembles that of vertebrates than ascidians, despite the latter two being more closely related; for example, *nodal* expression initiates symmetrically before resolving into asymmetry, and both genes are mesendodermally expressed (Yasui *et al*., 2000; Yu *et al*., 2002). This may be a corollary of organizer-based patterning in amphioxus and of the involvement of nodal in mesoderm induction (Onai *et al*., 2010; Yasui *et al*., 2000).

To date, the mechanism that breaks symmetry in amphioxus and leads to asymmetric *nodal* expression is unknown. Early amphioxus embryos are heavily ciliated and rotate inside their chorion prior to hatching, powered by these cilia (Conklin, 1932). Gastrulation internalizes ciliated cells such that the archenteron of the early neurula is lined with cells with long cilia, although whether these remain actively beating has not been described. Although this raises the possibility that cilia flow could account for symmetry breaking in amphioxus, hard evidence is lacking, and further experiments are needed to investigate this, for example, direct imaging of cilia activity, measurement of fluid flow within the amphioxus archenteron, and/or inhibition of cilia activity.

Echinoderms and Hemichordates

Echinoderms and hemichordates form a monophyletic grouping termed the Ambulacraria (Fig. 1). Most hemichordates are long, thin, and worm-shaped animals, although one hemichordate lineage, the pterobranchs, are tube dwellers (Sato and Holland, 2008). At least some adult hemichordates have subtle asymmetries, for example, in the location of an anterior pore (Kaul-Strehlow and Stach, 2013). Whether pterobranch hemichordates are asymmetric is less clear; classical descriptions of these animals as having an asymmetric gonad have recently been shown to be antisymmetric rather than directionally asymmetric (Sato and Holland, 2008). As of yet, the regulation of hemichordate asymmetry has not been described.

Echinoderms are a diverse Phylum with a variety of adult body plans including sea urchins, starfish, brittle stars, sea cucumbers, and sea lilies, all united by underlying pentaradial symmetry in the adult form. However, larvae are bilaterally symmetrical, and in sea urchins, the adult forms from a rudiment that develops on the left side of the larva (Fox *et al*., 2004). The positioning of this rudiment is under the control of nodal–Pitx signaling across the LR axis, demonstrating mechanistic conservation with chordates. The regulatory interactions of these genes are quite well dissected but will not be discussed further here; instead, the reader is referred to these recent studies (Duboc and Lepage, 2008; Luo and Su, 2012; Warner *et al*., 2012). In contrast to chordates, however, *nodal* and *Pitx* are deployed on the right-hand side of the embryo. The implication of this difference will be discussed below in detail, when we consider lophotrochozoan asymmetries.

ASYMMETRY AND ITS REGULATION IN INVERTEBRATE DEUTEROSTOMES

Chordates clearly share an ancient conserved role for *nodal* and *Pitx* on the left side of the body. Although this may have been considerably modified in ascidians, fundamentally they are left-expressed and likely drive left-sided morphogenesis. This is also shared with ambulacrarians; however, in these lineages, the pathway appears to operate on the right side (with the caveat that most echinoderm lineages have yet to be studied and the diversity of both adult and larval forms in this clade is considerable). Less clear is whether mechanisms of breaking symmetry and initiating asymmetric *nodal* expression are also conserved. We currently lack sufficient data from invertebrate deuterostome lineages to assess this and full

understanding will require integration of imaging and gene manipulation studies, both challenging approaches in the small embryos of nonmodel species.

ASYMMETRY IN THE LOPHOTROCHOZOA

The Lophotrochozoa is a Superphylum, including diverse organisms such as phoronids, bryozoans, rotifer, platyhelminthes, annelids, and molluscs (Fig. 1). Asymmetries can be found in many of these organisms, such as the sided operculum of some polychaete worms, the localization of ovary structures in rotifer, asymmetries in larval behavior, and, most classically, in the coiling of mollusc shells. Many undergo spiral cleavage (discussed further below), in which mitotic spindle orientation early in development directs sidedness.

Despite this insight, and the abundance of directional asymmetries within this Superphylum, we have still had relatively little insight into the molecular mechanisms underlying asymmetry in this clade until recently. This changed in 2009, with studies showing *nodal* and *pitx* to be asymmetrically expressed in snails (Grande and Patel, 2009; Kuroda *et al*., 2009), where their sidedness of expression correlated with the direction of spiral cleavage in the dextrally cleaving snail *Lottia gigantea* (which does not develop a coiled shell), the sinistrally cleaving snail *Biomphalaria glabrata*, and in dextral and sinistral morphs of the snail *Lymnaea stagnalis*. The subsequent direction of shell coiling in later development seen in *B. glabrata* and *L. stagnalis* morphs also correlated with sidedness of gene expression. It has also been shown that asymmetric growth and torsion in larval snails is driven by TGF-β signaling (Kurita and Wada, 2011), with the TGF-β superfamily ligand Dpp (BMP2/4 or Decapentaplegic) suggested to integrate asymmetric cues in postembryonic shell morphogenesis through a gradient of Dpp in the shell gland (Shimizu *et al*., 2011, 2013). Despite these advances, much remains unknown, with most studies only considering gastropod molluscs. In the next sections of this review, we will summarize current knowledge of lophotrochozoan asymmetry and suggest key questions still to be addressed.

The Cytoskeleton and LR Asymmetry in Molluscs

Mollusc embryos primitively undergo spiral cleavage (Fig. 2). This is apparent from the four- to eight-cell stage, when at the third cleavage, the tier of four animal micromeres is slightly offset from the four vegetal macromeres, creating a spiral form that either twists sinistrally or dextrally. Successive cleavages follow this offset pattern through to gastrulation. While in larval development dextrally cleaving embryos show dextral shell coiling, sinistrally cleaving embryos show sinistral shell coiling. Spiral cleavage is shared by many other lophotrochozoan Phyla and may be ancestral for the Lophotrochozoa, although this is somewhat contentious (Giribet, 2008; Hejnol, 2010). Although not described outside the Lophotrochozoa, spiral cleavage may hold clues as to the ways in which asymmetry is established in other clades.

There is excellent experimental evidence that the angle of spindle orientation during the third cleavage is both coincident with and determines later asymmetries (Freeman and Lundelius, 1982; Shibazaki *et al*., 2004). A particularly elegant study by Kuroda *et al*. (2009) showed that the direction of spiral rotation during third cleavage is directly connected to the orientation of shell coiling in later development. This was shown by physically manipulating blastomeres in dextral and sinistral snails, changing the arrangement of spindles and hence forcing spiral cleavage in the opposite orientation. This resulted in the concordant alteration of chirality in later development, with shell coiling corresponding to the direction of blastomere manipulation, regardless of the genetic background of the embryo. Subsequent generations reverted back to their genetically inherited direction of cleavage and coiling.

Early genetic studies indicated that the chirality of at least some mollusc embryos depends on a maternally inherited nuclear locus, termed *sinistral* (Asami *et al*., 2008; Boycott *et al*., 1930; Diver *et al*., 1925; Sturtevant, 1923). An experimental embryological study has further connected this genetic basis to a molecular mechanism, with the transfer of cytoplasm of a dextral egg into a sinistral egg causing subsequent dextralization (Freeman and Lundelius, 1982). This shows that there is a specific factor in the cytoplasm that communicates dextrality; however, the nature of this factor is still unknown. The pond snails *L. stagnalis* and *Lymnaea peregra* have historically been used for such studies, as they can be naturally found and cultured in both chiral forms (for review, see Schilthuizen and Davison, 2005). Mapping of the locus responsible has been undertaken (Liu *et al*., 2013), and although the nature of the gene or genes is not yet clear, this can only be a matter of time now. A likely scenario is that the locus will encode a maternally supplied protein or proteins that organize dextral orientation of the spindle at the third cleavage.

Interestingly (and somewhat counterintuitively), sinistral and dextral snails, although mirror images at the gross level of adult morphology and shell coiling, are not mirror images of each other at the cellular level during cleavage. Instead, they show morphologically and temporally distinct mechanisms of orientated cell division involving the cytoskeleton, at least for the small number of species studied to date (Shibazaki *et al*., 2004). Future studies are needed to better understand how these mechanisms operate, how phylogenetically widespread they are, and how they could be related to the maternally inherited *sinistral* locus.

These studies, while promising clear insight in the near future into the nature of sinistral versus dextral development in molluscs, do not yet tell us how symmetry is initially broken in mollusc embryos. It is worth noting that there is evidence for a cytoskeletal role in establishing asymmetry across other axes in the development of other mollusc species. Subcellular- and blastomere-specific RNA localization has been noted in the eggs and early cleavage stages of the marine gastropod mollusc *Ilyanassa* (Kingsley *et al*., 2007; Lambert and Nagy, 2002; Rabinowitz and Lambert, 2010), with maternally deposited RNA playing a role in specifying the DV axis via formation of the D-quadrant organizer, a specific vegetal cell that directs subsequent DV development (Lambert, 2010). Whether this extends to LR asymmetry of maternally deposited mRNA is not clear; however, we speculate that somewhere between the first cell cleavage and the specification of the D-quadrant organizer, asymmetric cues could be communicated to specify the LR axis. Asymmetric RNA localization has also been identified in the gastropod *Crepidula fornicata* (Henry *et al*., 2010) and would benefit from studies in other mollusc lineages and other lophotrochozoan Phyla to understand whether this is an ancestrally conserved mechanism or one that was convergently evolved with a shared cleavage pattern.

Other Lophotrochozoans

The Lophotrochozoa remain under-investigated in the developmental literature, and outside the mollusc examples discussed above, there have been no studies indicating that similar mechanisms for establishing asymmetry are at work in other Phyla within this clade. Spiral cleavage, however, is widespread, being found in at least annelids, platyhelminthes, nemerteans, entoprocts, molluscs, and phoronids (Lambert, 2010; Merkel *et al*., 2012; Pennerstorfer and Scholtz, 2012). This suggests that underlying mechanisms of symmetry breaking and cleavage orientation may be equally as ancient, although this idea needs to be treated with some caution as there are some differences between molluscs and other spirally cleaving embryos; for example, although sinistral mollusc strains and species are naturally (if rather rarely) found, the majority of other lineages seem to lack such sinistral forms. Furthermore, there is still the possibility that despite a conserved method of cleavage, similar mechanisms of asymmetry establishment may have evolved convergently in spiraling cleaving animals. Recent advances in genome sequencing show that the *nodal* gene is broadly conserved in lophotrochozoans (Kenny *et al*., manuscript submitted), although whether it is involved in LR asymmetry remains to be determined. Given the modifications of spiralian cleavage that can be found within this Superphylum, from equal cleavage to unequal cleavage and polar body formation, this provides interesting material to help us understand the relationship between cleavage patterns and molecular mechanisms in the establishment of LR asymmetry.

Summary of the Lophotrochozoa: An Emerging Picture of Cytoskeletal Architecture Linking to Asymmetric Nodal Expression

In summary, investigation into genetic, molecular, and cellular mechanisms underlying LR asymmetry in the Lophotrochozoa indicate that one aspect of this process—the role of the Nodal pathway in directing asymmetric morphogenesis—is conserved with deuterostomes. This would make it ancestral for these two lineages, and hence a character of the bilaterian common ancestor (if we exclude the acoelomorphs; Fig. 1). Axis establishment in the Lophotrochozoa is closely linked to the chirality of early cell cleavage, and although the exact mechanism for establishing polarity remains unclear, this is likely via cytoskeletal organization, certainly through spindle orientation, and possibly through asymmetric RNA localization (although direct evidence here is lacking). This model makes some sense when we consider deuterostomes and ecdysozoans (as discussed below). However, it should be recognized that it is effectively the tip of the iceberg drawn from a small number of studies in a few species primarily within the Gastropoda, that is, just one Class of one Phylum, the Mollusca. We still do not know whether these mechanisms are conserved in other molluscs, let alone other lophotrochozoan Phyla.

A further consideration that influences our understanding of the establishment of LR asymmetry in these taxa is that spiral cleavage is visible at the third cleavage of spiralian embryos. This is before DV axis specification, which occurs after the third micromere quartet forms and is mediated by the D-quadrant macromere (for review, see Lambert, 2010). Thus, early spirality correlates with LR asymmetry but precedes DV specification. As left and right sides cannot in theory be defined until both AP and DV are established, spiral cleavage does not break LR symmetry per se, although may provide underlying chirality that can be "read" into LR asymmetry after the DV axis is set. Further studies are needed that probe the interaction between the spiral third cleavage, the D-quadrant macromere, and LR asymmetry. Much remains to be determined, and given the diversity of the Lophotrochozoa, additional surprises are possible.

LR ASYMMETRY IN THE ECDYSOZOA

Examples of asymmetry in living Ecdysozoa include the coiling of the gut and genital apparatus in *Drosophila melanogaster*, cell lineage in *Caenorhabditis elegans* embryos, and neural organization in a range of species.

Further examples have been well catalogued and reviewed previously (Coutelis *et al*., 2008; Frasnelli *et al*., 2012; Palmer, 2009; Spéder *et al*., 2007) and therefore will not be detailed here. Instead, we will focus on mechanisms drawing attention to how some ecdysozoan methods for establishing LR asymmetry may be shared across the Bilateria.

The absence of *Nodal* in the Ecdysozoa suggests one of two hypotheses; that *Nodal* was lost in the Ecdysozoa and independently lost in the Platyhelminthes, or that *Nodal* has been independently gained by the Lophotrochozoa (apart from Platyhelminthes) and the Deuterostomia. Given that loss of ancient genes controlling development is a frequent phenomenon in animal evolution (Paps *et al*., 2012), and convergent gain of such genes in such widely divergent lineages has never, to our knowledge, been satisfactorily identified, the most plausible scenario is that Nodal was present in the common ancestor of the Lophotrochozoa and Deuterostomia. As this common ancestor was also shared with the Ecdysozoa (Fig. 1), we can conclude that ecdysozoans primitively had *Nodal* (and hence a role in LR asymmetry); however, this was either lost early in ecdysozoan evolution or has been lost independently in several ecdysozoan lineages. If we accept, as discussed above, that *Nodal* played a role in LR asymmetry in the bilaterian common ancestor, how and why loss occurred raises questions as to the development of (and indeed the presence of) asymmetry in stem group Ecdysozoa. We can speculate that the role of *Nodal* could have been superseded by lineage-specific mechanisms and that asymmetry could have been lost before being re-evolved or that developmentally earlier-acting mechanisms for establishing asymmetry could be operating without the need for *Nodal* to further propagate this signal. The ultimate reasons behind such evolutionary change will be difficult to establish without evidence from currently understudied ecdysozoan Phyla such as priapulids and tardigrades. However, the mechanisms now used by ecdysozoans for establishing LR asymmetries seem to be based on cytoskeletal cues, and these may at least help in understanding the ancestry of symmetry breaking.

Cytoskeletal Cues in Ecdysozoan LR Asymmetry

Asymmetric cell divisions play a key role in differentially segregating cellular determinants in a wide variety of species (e.g., Betschinger and Knoblich, 2004) and have the potential to initiate the development of morphological asymmetries, although hard evidence linking such cell level processes directly to organismal LR asymmetry is generally lacking. In *C. elegans*, early cell divisions are asymmetric, and pioneering work by Wood (1991) showed that physical reversal of cleavage asymmetry by micromanipulation led to reversal of subsequent LR asymmetry of the cell lineage. This experiment is analogous to that performed by Kuroda *et al*. (2009) as described above, in which lophotrochozoan spiral cleavage was physically reversed with similar results. In *C. elegans* a chiral actin cytoskeleton is present in the egg (Pohl and Bao, 2010). The daughter cells that result from this polarity have markedly different cytoplasmic constitution and size. It is hypothesized in this species that asymmetries may be formed by differential forces in spindle elongation due to chiral cortical and cytoskeletal components, which would generate asymmetric cell divisions (Pohl, 2011; Pohl and Bao, 2010). This hypothesis needs to be tested in other Ecdysozoa with asymmetric cell divisions in the early embryo, such as the tardigrades, to understand whether this could be an ancestrally shared mechanism within the Ecdysozoa. A critical question remains as to what subsequent mechanism then generates morphological asymmetry, although one recent study has shown that in *C. elegans*, this may in part be based on the regulation of cortical contractility and cleavage furrow formation (Pohl and Bao, 2010).

Furthermore, it is known that at the third cell cleavage, spindle orientation by a G protein is crucial for the correct establishment of LR asymmetry (Bergmann *et al*., 2003). These G proteins have also been shown to have a role in the alignment of spindles in *D. melanogaster*, by binding to atypical protein kinase C and par proteins (reviewed by Ahringer, 2003). The *par* genes, originally identified in *C. elegans* but now known to be broadly conserved, are known to disrupt this asymmetric division process when mutated, affecting spindle localization (reviewed by Nance and Zallen, 2011). Par proteins are themselves asymmetrically distributed by the action of actin and myosin in the cell, which is generally said to follow from the organization of the cytoskeleton around the point of sperm entry (Munro *et al*., 2004), and have also been identified in an annelid (as suggested by Weisblat, 2007) and a mollusc, with at least one mollusc Par protein able to interact with microtubules and localizing with spindles during early cleavage (Hozumi *et al*., 2006).

In concert with cytoskeletal components, molecular motors could be the proximate cause of LR asymmetry by asymmetric localization of cargoes such as mRNA. A good example is the type ID unconventional Myosin (MyoID), which has been shown to cause situs inversus when mutated in *D. melanogaster* (Hozumi *et al*., 2006; Spéder *et al*., 2006). Such molecular motors move along F-actin in the cytoskeleton, and any asymmetry in the actin cytoskeleton could then be translated into transcriptional asymmetry. MyoID can also interact with adherens junctions via β-catenin (Petzoldt *et al*., 2012; Spéder *et al*., 2006), and these junctions could also act as mediators of LR asymmetry signals. Adherens junctions have been shown to be involved with LR

asymmetry in concert with cadherins in *D. melanogaster* (Petzoldt *et al*., 2012) and the chick (Burdine and Caspary, 2013). An ancestrally shared role for adherens junctions could also explain the links seen between the Planar Cell Polarity (PCP) pathway (reviewed by Aw and Levin, 2009), inversins (Petzoldt *et al*., 2012; Watanabe *et al*., 2003), and LR asymmetry in a diverse phylogenetic grouping of species. However, this should remain a tentative conclusion, as it is easy to imagine that convergently evolved mechanisms of asymmetry involving epithelia would independently co-opt the adherens junction–cadherin machinery typical of such cells.

Conserved Cytoskeletal Mechanisms in Ecdysozoans?

Chirality of the actin microfilament cytoskeleton could be the root of all downstream differences in directionality, with additional cytoskeletal components acting on this, effectively representing the hypothetical "F" molecule, proposed by Brown and Wolpert (1990) as an asymmetric molecule that could integrate AP and DV polarity to specify left from right. Both *C. elegans* and *D. melanogaster* seem to use such a system. However, extrapolating this to conclude conservation (and hence assuming it is ancestral for other ecdysozoans) has caveats. First, the timing of action is quite different in *C. elegans* and *D. melanogaster*, occurring at early cleavage in the former and much later in development in the latter. This may be a consequence of the unusual mode of early development of *D. melanogaster*, in which both AP and DV axes are maternally defined, a derived condition for a subgroup of the insects. Without study of other arthropods and other ecdysozoan Phyla, we cannot test this. Second, the actin cytoskeleton is such a fundamental component of eukaryotic cells that convergent evolution is a clear possibility. Identification of additional similarities in underlying mechanisms would help to support the case for conservation rather than convergence.

A Conserved Role for Other Mechanisms Between Ecdysozoa and Other Bilateria?

Some of the genes implicated in modulation of asymmetrical signals in deuterostomes have also been observed playing a similar role in ecdysozoans. Notch signaling and calcium concentration, for example, play a role in establishing asymmetry in vertebrate embryos (Lopes *et al*., 2010; Raya *et al*., 2003, 2004). The specification of two neurons in *C. elegans*—ASE left and ASE right—are driven by the bilaterally asymmetric activation of T-box transcription factors under the control of calcium levels and the Notch pathway (Bertrand *et al*., 2011; Poole and Hobert, 2006; Schumacher *et al*., 2012). As the areas where Notch is expressed in vertebrates and *C. elegans* have no obvious comparison, this is likely to represent co-option of widely used signaling pathways, rather than an ancestrally conserved mechanism. micro-RNAs (miRNAs) have also been shown to be involved in the establishment of asymmetry in *C. elegans* (Johnston and Hobert, 2003); however, although miRNAs have also been shown to regulate asymmetry via the regulation of *Nodal* in a range of contexts in vertebrates, there is currently no evidence for homology of this process between the Ecdysozoa and vertebrates (Barroso del Jesus *et al*., 2011; Choi *et al*., 2007; Martello *et al*., 2007).

SUMMARY AND CONCLUSIONS: AN EMERGING PICTURE OF LR ASYMMETRY IN THE BILATERIA

The discovery of a role for *Nodal* in regulating LR asymmetry in deuterostomes and molluscs has led to the conclusion that this was ancestral for the Bilateria. Although this needs confirmation from additional lineages, especially other lophotrochozoans, it suggests that the common ancestor, the "Urbilaterian," was asymmetric, and hence that this method of regulating asymmetry was lost by some and possibly all ecdysozoans. In chordates, the sidedness has flipped, with right-sided deployment of nodal ancestral and chordates evolving left-sided expression instead. This very likely relates to the DV axis inversion thought to have occurred at some point in deuterostome evolution, which would by its nature flip our interpretation of left and right.

Asymmetric *Nodal* expression is a consequence, not a cause, of symmetry breaking. Although studies on various animal lineages point to rather different mechanisms of symmetry breaking in different species (for example as discussed above; cilia in mice, spiral cleavage in molluscs, and microfilament chirality in *C. elegans*), a common theme is the use of polarized cytoskeletal structures to generate chirality that can be organized relative to the AP and DV axes and hence read into LR asymmetry (as previously discussed by Levin and Palmer, 2007; Vandenberg and Levin, 2010). Several additional lines of evidence support a general role for chirality of microfilaments or microtubules in LR asymmetry. Like microfilaments, microtubules possess an inherent polarity, and in many species, this organization is the basis for differential mRNA transport (e.g., in *D. melanogaster*; Roth and Lynch, 2009). Microtubule organizing centers (both centrioles and basal bodies) have been implicated in LR asymmetry in a range of contexts in several species across the Bilateria (Beisson and Jerka-Dziadosz, 1999; Lobikin *et al*., 2012). Some studies have provided evidence for a role for cytoskeletal cues lying upstream of LR asymmetry, including actin cytoskeletal chirality in amphibian eggs (Danilchik *et al*., 2006) and microtubule spindle organization in molluscs (Shibazaki *et al*., 2004). This might

also help to explain the continued development of LR asymmetric heart morphology in the apparent absence of Nodal signaling in the zebrafish, which is also actin polymerization-dependent (Noel *et al*., 2013).

We speculate that these are indeed aspects of an ancestrally shared framework underlying the breaking of LR asymmetry in the Bilateria and should be renewed as a focus of future studies to understand symmetry breaking during LR asymmetry establishment. Primitively, it is likely that bilaterian oocytes were formed with an axis, the animal–vegetal axis, laid down maternally (Martindale and Hejnol, 2009). The DV axis is then likely to have been determined at fertilization via sperm entry, generating a zygote with two axes that could be transformed into organismal AP and DV pattern. Cytoskeletal chirality in the oocyte has the potential to interact with these axes due to the polarized nature of microfilaments and/or microtubules to break symmetry. Downstream interpretation of broken symmetry could take many forms, H^+K^+ATPase localization, mRNA localization, directional cell movement, *Nodal* activation, and so forth, with the latter the best candidate for an ancestral mechanism. This would imply that the use of cilia to break symmetry is a chordate or vertebrate innovation.

However, as yet, we still have too little data from most animal lineages to reliably conclude either cytoskeletal activity or Nodal expression was not convergently evolved. Arriving at the party late, invertebrates still have much to teach us about the development of LR asymmetry. As we learn more about how these disparate animals establish their own LR axis, our perspective on those mechanisms we see in the vertebrates and their evolutionary origin can only improve, paving the way toward a true understanding on the fundamentals of this process across the Metazoa.

ACKNOWLEDGMENTS

The authors thank the members of the Shimeld and Holland laboratories and attendees at the ISDB Satellite Symposium on LR asymmetry for discussion. N.J. Kenny was supported by a Clarendon Scholarship.

LITERATURE CITED

Ahringer J. 2003. Control of cell polarity and mitotic spindle positioning in animal cells. Curr Opin Cell Biol 15:73–81.

Asami T, Gittenberger E, Falkner G. 2008. Whole-body enantiomorphy and maternal inheritance of chiral reversal in the pond snail *Lymnaea stagnalis*. J Hered 99:552–557.

Aw S, Levin M. 2009. Is left–right asymmetry a form of planar cell polarity? Development 136:355–366.

Barroso del Jesus A, Lucena-Aguilar G, Sanchez L, Ligero G, Gutierrez-Aranda I, Menendez P. 2011. The Nodal inhibitor Lefty is negatively modulated by the microRNA miR-302 in human embryonic stem cells. FASEB J 25:1497–1508.

Beck S, Le Good JA, Guzman M, Ben Haim N, Roy K, Beermann F, Constam DB. 2002. Extraembryonic proteases regulate Nodal signalling during gastrulation. Nat Cell Biol 4:981–985.

Beisson J, Jerka-Dziadosz M. 1999. Polarities of the centriolar structure: Morphogenetic consequences. Biol Cell 91:367–378.

Bergmann DC, Lee M, Robertson B, Tsou MF, Rose LS, Wood WB. 2003. Embryonic handedness choice in *C. elegans* involves the Gα protein GPA-16. Development 130:5731–5740.

Bertrand V, Bisso P, Poole RJ, Hobert O. 2011. Notch-dependent induction of left/right asymmetry in *C. elegans* interneurons and motoneurons. Curr Biol 21:1225–1231.

Betschinger J, Knoblich JA. 2004. Dare to be different: Asymmetric cell division in *Drosophila*, *C. elegans* and vertebrates. Curr Biol 14:R674–R685.

Boorman CJ, Shimeld SM. 2002a. The evolution of left–right asymmetry in chordates. Bioessays 24:1004–1011.

Boorman CJ, Shimeld SM. 2002b. Pitx homeobox genes in *Ciona* and amphioxus show left–right asymmetry is a conserved chordate character and define the ascidian adenohypophysis. Evol Dev 4:354–365.

Boycott A, Diver C, Garstang S, Turner F. 1930. The inheritance of sinistrality in *Limnaea peregra*. Philos Trans R Soc Lond B Biol Sci 219:51–130.

Brown NA, Wolpert L. 1990. The development of handedness in left/right asymmetry. Development 109: 1–9.

Burdine RD, Caspary T. 2013. Left–right asymmetry: Lessons from Cancun. Development 140:4465–4470.

Capdevila J, Vogan KJ, Tabin CJ, Izpisúa Belmonte JC. 2000. Mechanisms of left–right determination in vertebrates. Cell 101:9–21.

Chea HK, Wright CV, Swalla BJ. 2005. Nodal signaling and the evolution of deuterostome gastrulation. Dev Dyn 234:269–278.

Choi W-Y, Giraldez AJ, Schier AF. 2007. Target protectors reveal dampening and balancing of Nodal agonist and antagonist by miR-430. Science 318:271–274.

Collignon J, Varlet I, Robertson EJ. 1996. Relationship between asymmetric nodal expression and the direction of embryonic turning. Nature 381:155–158.

Conklin EG. 1932. The embryology of amphioxus. J Morphol 54:69–151.

Coutelis JB, Petzoldt AG, Spéder P, Suzanne M, Noselli S. 2008. Left–right asymmetry in *Drosophila*. Semin Cell Dev Biol 19:252–262.

Danilchik M, Brown E, Riegert K. 2006. Intrinsic chiral properties of the Xenopus egg cortex: An early indicator of left–right asymmetry? Development 133: 4517–4526.

Diver C, Boycott A, Garstang S. 1925. The inheritance of inverse symmetry in *Lymnaea peregra*. J Genet 15:113–200.

Duboc V, Lepage T. 2008. A conserved role for the Nodal signaling pathway in the establishment of dorso-ventral and left–right axes in deuterostomes. J Exp Biol 310:41–53.

Duboc V, Rottinger E, Lapraz F, Besnardeau L, Lepage T. 2005. Left–right asymmetry in the sea urchin embryo is regulated by nodal signaling on the right side. Dev Cell 9:147–158.

Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, Smith SA, Seaver E, Rouse GW, Obst M, Edgecombe GD, Sorensen MV, Haddock SHD, Schmidt-Rhaesa A, Okusu A, Kristensen RM, Wheeler WC, Martindale MQ, Giribet G. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452:745–749.

Ruppert EE, Fox R, Barnes RD. 2004. Invertebrate zoology: A functional evolutionary approach. Thompson-Brooks/Cole. Belmont, CA, USA.

Frasnelli E, Vallortigara G, Rogers LJ. 2012. Left–right asymmetries of behaviour and nervous system in invertebrates. Neurosci Biobehav Rev 36:1273–1291.

Freeman G, Lundelius J. 1982. The developmental genetics of dextrality and sinistrality in the gastropod *Lymnaea peregra*. Roux Arch Dev Biol 191:69–83.

Giribet G. 2008. Assembling the Lophotrochozoan (spiralian) tree of life. Philos Trans R Soc Lond B Biol Sci 363:1513–1522.

Grande C, Patel NH. 2009. Nodal signalling is involved in left–right asymmetry in snails. Nature 457:1007–1011.

Hejnol A. 2010. A twist in time—The evolution of spiral cleavage in the light of animal phylogeny. Integr Comp Biol 50:695–706.

Hejnol A, Obst M, Stamatakis A, Ott M, Rouse GW, Edgecombe GD, Martinez P, Baguna J, Bailly X, Jondelius U, Wiens M, Muller WE, Seaver E, Wheeler WC, Martindale MQ, Giribet G, Dunn CW. 2009. Assessing the root of bilaterian animals with scalable phylogenomic methods. Proc Biol Sci 276: 4261–4270.

Henry JJ, Perry KJ, Fukui L, Alvi N. 2010. Differential localization of mRNAs during early development in the mollusc, *Crepidula fornicata*. Integr Comp Biol 50:720–733.

Hirokawa N, Tanaka Y, Okada Y, Takeda S. 2006. Nodal flow and the generation of left–right asymmetry. Cell 125:33–45.

Hozumi S, Maeda R, Taniguchi K, Kanai M, Shirakabe S, Sasamura T, Speder P, Noselli S, Aigaki T, Murakami R, Matsuno K. 2006. An unconventional myosin in *Drosophila* reverses the default handedness in visceral organs. Nature 440:798–802.

Johnston RJ, Hobert O. 2003. A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. Nature 426:845–849.

Kaul-Strehlow S, Stach, T. 2013. A detailed description of the development of the hemichordate *Saccoglossus kowalevskii* using SEM, TEM, histology and 3D-reconstructions. Front Zool 10: 53.

Kingsley E, Chan X, Duan Y, Lambert J. 2007. Widespread RNA segregation in a spiralian embryo. Evol Dev 9:527–539.

Kourakis M, Smith W. 2005. Did the first chordates organize without the organizer? Trends Genet 21: 506–510.

Kurita Y, Wada H. 2011. Evidence that gastropod torsion is driven by asymmetric cell proliferation activated by TGF-β signalling. Biol Lett 7:759–762.

Kuroda R, Endo B, Abe M, Shimizu M. 2009. Chiral blastomere arrangement dictates zygotic left–right asymmetry pathway in snails. Nature 462:790–794.

Lambert JD. 2010. Developmental patterns in spiralian embryos. Curr Biol 20:R72–R77.

Lambert JD, Nagy LM. 2002. Asymmetric inheritance of centrosomally localized mRNAs during embryonic cleavages. Nature 420:682–686.

Lemaire P. 2009. Unfolding a chordate developmental program, one cell at a time: Invariant cell lineages, short-range inductions and evolutionary plasticity in ascidians. Dev Biol 332:48–60.

Levin M, Johnson R, Stern C, Kuehn M, Tabin C. 1995. A molecular pathway determining left–right asymmetry in chick embryogenesis. Cell 82:803–814.

Levin M, Palmer AR. 2007. Left–right patterning from the inside out: Widespread evidence for intracellular control. Bioessays 29:271–287.

Levin M, Thorlin T, Robinson K, Nogi T, Mercola M. 2002. Asymmetries in H+/K+-ATPase and cell membrane potentials comprise a very early step in left–right patterning. Cell 111:77–89.

Liu M, Davey J, Banerjee R, Han J, Yang F, Aboobaker A, Blaxter M, Davison A. 2013. Fine mapping of the pond snail left–right asymmetry (chirality) locus using RAD-Seq and Fibre-FISH. PLoS One 8:e71067–e71067.

Lobikin M, Wang G, Xu J, Hsieh Y-W, Chuang C-F, Lemire JM, Levin M. 2012. Early, nonciliary role for microtubule proteins in left-right patterning is conserved across kingdoms. Proc Natl Acad Sci USA 109: 12586–12591.

Lopes SS, Lourenço R, Pacheco LS, Moreno N, Kreiling J, Saúde L. 2010. Notch signalling regulates left–right asymmetry through ciliary length control. Development 137:3625–3632.

Lowe LA, Supp DM, Sampath K, Yokoyama T, Wright CV, Potter SS, Overbeek P, Kuehn MR. 1996. Conserved left–right asymmetry of nodal expression and alterations in murine situs inversus. Nature 381: 158–161.

Luo Y-J, Su Y-H. 2012. Opposing Nodal and BMP signals regulate left–right asymmetry in the sea urchin larva. PLoS Biol 10:e1001402.

Martello G, Zacchigna L, Inui M, Montagner M, Adorno M, Mamidi A, Morsut L, Soligo S, Tran U, Dupont S, Cordenonsi M, Wessely O, Piccolo S. 2007. MicroRNA control of Nodal signalling. Nature 449:183–188.

Martindale MQ, Hejnol A. 2009. A developmental perspective: Changes in the position of the blastopore during bilaterian evolution. Dev Cell 17:162–174.

Matus DQ, Pang K, Marlow H, Dunn CW, Thomsen GH, Martindale MQ. 2006a. Molecular evidence for deep evolutionary roots of bilaterality in animal development. Proc Natl Acad Sci USA 103:11195–11200.

Matus DQ, Thomsen GH, Martindale MQ. 2006b. Dorso/ventral genes are asymmetrically expressed and involved in germ-layer demarcation during cnidarian gastrulation. Curr Biol 16:499–505.

Merkel J, Wollesen T, Lieb B, Wanninger A. 2012. Spiral cleavage and early embryology of a loxosomatid entoproct and the usefulness of spiralian apical cross patterns for phylogenetic inferences. BMC Dev Biol 12:11.

Mita K, Fujiwara S. 2007. Nodal regulates neural tube formation in the *Ciona intestinalis* embryo. Dev Genes Evol 217:593–601.

Morokuma J, Ueno M, Kawanishi H, Saiga H, Nishida H. 2002. *HrNodal*, the ascidian nodal-related gene, is expressed in the left side of the epidermis, and lies upstream of *HrPitx*. Dev Genes Evol 212:439–446.

Munro E, Nance J, Priess JR. 2004. Cortical flows powered by asymmetrical contraction transport PAR proteins to establish and maintain anterior–posterior polarity in the early *C. elegans* embryo. Dev Cell 7:413–424.

Nakamura T, Hamada H. 2012. Left–right patterning: Conserved and divergent mechanisms. Development 139:3257–3262.

Nance J, Zallen JA. 2011. Elaborating polarity: PAR proteins and the cytoskeleton. Development 138:799–809.

Nishide K, Mugitani M, Kumano G, Nishida H. 2012. Neurula rotation determines left–right asymmetry in ascidian tadpole larvae. Development 139:1467–1475.

Noel ES, Verhoeven M, Lagendijk AK, Tessadori F, Smith K, Choorapoikayil S, den Hertog J, Bakkers J. 2013. A Nodal-independent and tissue-intrinsic mechanism controls heart-looping chirality. Nat Commun 4:2754.

Nonaka S, Tanaka Y, Okada Y, Takeda S, Harada A, Kanai Y, Kido M, Hirokawa N. 1998. Randomization of left–right asymmetry due to loss of nodal cilia generating leftward flow of extraembryonic fluid in mice lacking KIF3B motor protein. Cell 95:829–837.

Onai T, Yu J-K, Blitz IL, Cho KWY, Holland LZ. 2010. Opposing Nodal/Vg1 and BMP signals mediate axial patterning in embryos of the basal chordate amphioxus. Dev Biol 344:377–389.

Palmer AR. 1996. From symmetry to asymmetry: Phylogenetic patterns of asymmetry variation in animals and their evolutionary significance. Proc Natl Acad Sci USA 93:14279–14286.

Palmer AR. 2004. Symmetry breaking and the evolution of development. Science 306:828–833.

Palmer AR. 2009. Animal asymmetry. Curr Biol 19:473–477.

Paps J, Holland PW, Shimeld SM. 2012. A genome-wide view of transcription factor gene diversity in chordate evolution: Less gene loss in amphioxus? Brief Funct Genomics 11:177–186.

Pennerstorfer M, Scholtz G. 2012. Early cleavage in *Phoronis muelleri* (Phoronida) displays spiral features. Evol Dev 14:484–500.

Petzoldt AG, Coutelis J-B, Géminard C, Spéder P, Suzanne M, Cerezo D, Noselli S. 2012. DE-Cadherin regulates unconventional Myosin ID and Myosin IC in Drosophila left–right asymmetry establishment. Development 139:1874–1884.

Pick KS, Philippe H, Schreiber F, Erpenbeck D, Jackson DJ, Wrede P, Wiens M, Alié A, Morgenstern B, Manuel M, Wörheide G. 2010. Improved phylogenomic taxon sampling noticeably affects nonbilaterian relationships. Mol Biol Evol 27:1983–1987.

Pohl C. 2011. Left–right patterning in the *C. elegans* embryo: Unique mechanisms and common principles. Commun Integr Biol 4:34–40.

Pohl C, Bao Z. 2010. Chiral forces organize left–right patterning in *C. elegans* by uncoupling midline and anteroposterior axis. Dev Cell 19:402–412.

Poole RJ, Hobert O. 2006. Early embryonic programming of neuronal left/right asymmetry in *C. elegans.* Curr Biol 16:2279–2292.

Rabinowitz JS, Lambert JD. 2010. Spiralian quartet developmental potential is regulated by specific localization elements that mediate asymmetric RNA segregation. Development 137:4039–4049.

Raya A, Kawakami Y, Rodriguez-Esteban C, Buscher D, Koth CM, Itoh T, Morita M, Raya RM, Dubova I, Bessa JG, de la Pompa JL, Izpisua Belmonte JC. 2003. Notch activity induces Nodal expression and mediates the establishment of left–right asymmetry in vertebrate embryos. Genes Dev 17:1213–1218.

Raya A, Kawakami Y, Rodriguez-Esteban C, Ibanes M, Rasskin-Gutman D, Rodriguez-Leon J, Buscher D, Feijo JA, Izpisua Belmonte JC. 2004. Notch activity

acts as a sensor for extracellular calcium during vertebrate left–right determination. Nature 427:121–128.

Roth S, Lynch JA. 2009. Symmetry breaking during Drosophila oogenesis. Cold Spring Harb Perspect Biol 1:a001891.

Sato A, Holland PW. 2008. Asymmetry in a pterobranch hemichordate and the evolution of left–right patterning. Dev Dyn 237:3634–3639.

Satoh N. 1994. Developmental biology of ascidians, Cambridge University Press, London and New York.

Schilthuizen M, Davison A. 2005. The convoluted evolution of snail chirality. Naturwissenschaften 92:504–515.

Schumacher JA, Hsieh Y-W, Chen S, Pirri JK, Alkema MJ, Li W-H, Chang C, Chuang C-F. 2012. Intercellular calcium signaling in a gap junction-coupled cell network establishes asymmetric neuronal fates in *C. elegans*. Development 139:4191–4201.

Shibazaki Y, Shimizu M, Kuroda R, Project KC, Agency T, Building P. 2004. Body handedness is directed by genetically determined cytoskeletal dynamics in the early embryo. Curr Biol 14:1462–1467.

Shimeld SM, Levin M. 2006. Evidence for the regulation of left–right asymmetry in *Ciona intestinalis* by ion flux. Dev Dyn 235:1543–1553.

Shimizu K, Iijima M, Setiamarga D, Sarashina I, Kudoh T, Asami T, Gittenberger E, Endo K. 2013. Left–right asymmetric expression of dpp in the mantle of gastropods correlates with asymmetric shell coiling. EvoDevo 4:15.

Shimizu K, Sarashina I, Kagi H, Endo K. 2011. Possible functions of Dpp in gastropod shell formation and shell coiling. Dev Genes Evol 221:59–68.

Spéder P, Adam G, Noselli S. 2006. Type ID unconventional myosin controls left/right asymmetry in *Drosophila*. Nature 440:803–807.

Spéder P, Petzoldt A, Suzanne M, Noselli S. 2007. Strategies to establish left/right asymmetry in vertebrates and invertebrates. Curr Opin Genet Dev 17: 351–358.

Sturtevant AH. 1923. Inheritance of direction of coiling in *Limnaea*. Science 58:269–270.

Tabin C. 2006. The key to left–right asymmetry. Cell 127:27–32.

Thompson H, Shaw MK, Dawe HR, Shimeld SM. 2012. The formation and positioning of cilia in *Ciona intestinalis* embryos in relation to the generation and evolution of chordate left–right asymmetry. Dev Biol 364:214–223.

Vandenberg LN, Levin M. 2010. Far from solved: A perspective on what we know about early mechanisms of left–right asymmetry. Dev Dyn 239: 3131–3146.

Warner JF, Lyons DC, McClay DR. 2012. Left–right asymmetry in the sea urchin embryo: BMP and the asymmetrical origins of the adult. PLoS Biol 10: e1001404.

Watanabe D, Saijoh Y, Nonaka S, Sasaki G, Ikawa Y, Yokoyama T, Hamada H. 2003. The left–right determinant Inversin is a component of node monocilia and other 9+0 cilia. Development 130:1725–1734.

Weisblat DA. 2007. Asymmetric cell divisions in the early embryo of the leech *Helobdella robusta*. Prog Mol Subcell Biol 45:79–85.

Wood WB. 1991. Evidence from reversal of handedness in *C. elegans* embryos for early cell interactions determining cell fates. Nature 349:536–538.

Yasui K, Zhang S, Uemura M, Saiga H. 2000. Left–right asymmetric expression of *BbPtx*, a Ptx-related gene, in a lancelet species and the developmental left-sidedness in deuterostomes. Development 127: 187–195.

Yoshida K, Saiga H. 2008. Left–right asymmetric expression of Pitx is regulated by the asymmetric Nodal signaling through an intronic enhancer in *Ciona intestinalis*. Dev Genes Evol 218:353–360.

Yu J-K, Holland LZ, Holland ND. 2002. An amphioxus nodal gene (*AmphiNodal*) with early symmetrical expression in the organizer and mesoderm and later asymmetrical expression associated with left–right axis formation. Evol Dev 4:418–425.

Zhou X, Sasaki H, Lowe L, Hogan BL, Kuehn MR. 1993. Nodal is a novel TGF-β-like gene expressed in the mouse node during gastrulation. Nature 361:543–547.

Appendix 2

Kenny*, Namigai*, Marlétaz, Hui, and Shimeld, 2015. Draft genome assemblies and predicted microRNA complements of the intertidal lophotrochozoans Patella vulgata (Mollusca, Patellogastropoda) and Spirobranchus (Pomatoceros) lamarcki (Annelida, Serpulida). Marine Genomics. *in press*.

*joint first author

Marine Genomics xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Marine Genomics

Draft genome assemblies and predicted microRNA complements of the intertidal lophotrochozoans *Patella vulgata* (Mollusca, Patellogastropoda) and *Spirobranchus* (*Pomatoceros*) *lamarcki* (Annelida, Serpulida)

Nathan J. Kenny [a,b,1], Erica K.O. Namigai [b,1], Ferdinand Marlétaz [b], Jerome H.L. Hui [a,*], Sebastian M. Shimeld [b,*]

[a] Simon F.S. Li Marine Science Laboratory of School of Life Sciences and Center for Soybean Research of the State Key Laboratory of Agrobiotechnology, The Chinese University of Hong Kong, Shatin, Hong Kong
[b] Department of Zoology, University of Oxford, Oxford OX1 3PS, UK

ARTICLE INFO

Article history:
Received 20 March 2015
Received in revised form 19 June 2015
Accepted 14 July 2015
Available online xxxx

Keywords:
Spiralian
Lophotrochozoan
Annelid
Mollusc
Genome
MicroRNA
Phylogeny

ABSTRACT

MicroRNAs (miRNA) are small non-coding RNAs that act post-transcriptionally to regulate gene expression levels. Some studies have indicated that microRNAs may have low homoplasy, and as a consequence the phylogenetic distribution of microRNA families has been used to study animal evolutionary relationships. Limited levels of lineage sampling, however, may distort such analyses. Lophotrochozoa is an under-sampled taxon that includes molluscs, annelids and nemerteans, among other phyla. Here, we present two novel draft genomes, those of the limpet *Patella vulgata* and polychaete *Spirobranchus* (*Pomatoceros*) *lamarcki*. Surveying these genomes for known microRNAs identifies numerous potential orthologues, including a number that have been considered to be confined to other lineages. RT-PCR demonstrates that some of these (*miR-1285*, *miR-1287*, *miR-1957*, *miR-1983* and *miR-3533*), previously thought to be found only in vertebrates, are expressed. This study provides genomic resources for two lophotrochozoans and reveals patterns of microRNA evolution that could be hidden by more restricted sampling.

1. Introduction

Originally discovered in the nematode *Caenorhabditis elegans* in 1993 (Lee et al., 1993), microRNAs (miRNAs) did not attract much attention until the discovery of the first conserved miRNAs in animals in 2000 (Pasquinelli et al., 2000). miRNAs have important and widespread roles in many aspects of the biology of animals, plants and even some viruses, with some playing evolutionarily-ancient roles (Bartel, 2004; Axtell and Bartel, 2005; Plaisance-Bonstaff and Renne, 2011). For example, it is likely that the majority of mammalian mRNAs are regulated by miRNAs (Friedman et al., 2009), while miR-1 appears to have an ancient role in muscle development (Kloosterman and Plasterk, 2006). As a consequence, over the last decade the study of miRNAs has become a rapidly moving field in a range of contexts, most commonly in study of the post-transcriptional regulation of gene expression (see Bartel, 2009; Hui et al., 2013a,b) but also in the field of phylogenetic reconstruction (Tarver et al., 2013; Kenny et al., 2015).

One reason that miRNAs have been utilized in phylogenetic reconstruction was an initially reported low rate of homoplasy (Tarver et al., 2013). Initial investigation suggested that once miRNAs were incorporated into genomes, they would seldom be lost (Sempere et al., 2006). Further studies suggested that they might be used as slow-evolving genomic characters, such that mapping their gain across a cladogram would allow the derivation of evolutionary relationships (for examples, see Wheeler et al., 2009; Tarver et al., 2013). This approach has been used to shed light on several recalcitrant cases in animal phylogeny (e.g., Rota-Stabelli et al., 2010; Campbell et al., 2011; Philippe et al., 2011). More recently, however, it has been noted that heterogeneous rates of gain and loss of miRNA loci, as well as their secondary loss, may be more common than previously suspected, especially in some lineages, and also that sampling error has affected some historic analyses (Fromm et al., 2013; Thomson et al., 2014; Quah et al., 2015). Attempts have been made to correct some of these problems via the re-analysis of previously published datasets (Field et al., 2014). While work remains to be done in this regard, if these problems can be addressed miRNA remain potentially useful for the reconstruction of phylogeny, and their flanking sequences have also been shown to contain useful phylogenetic signal at the intra-ordinal and -familial levels (Kenny et al., 2015).

* Corresponding authors.
E-mail addresses: jeromehui@cuhk.edu.hk (J.H.L. Hui), sebastian.shimeld@zoo.ox.ac.uk (S.M. Shimeld).
[1] Contributed equally.

http://dx.doi.org/10.1016/j.margen.2015.07.004
1874-7787/

Uneven genome sampling across animal phylogeny, however, remains a limitation. As shown in Fig. 1A, three major clades make up the bilaterians (Halanych et al., 1995; Aguinaldo et al., 1997). Deuterostomia, and particularly Chordata, are relatively well sampled genomically, and Ecdysozoa are also well represented (Kenny et al., 2013). However the third assemblage, variously named Lophotrochozoa or Spiralia, is relatively poorly represented in both genomic and miRNA databases. This assemblage includes a number of phyla including Mollusca, Annelida, Brachiopoda, Nemertea and Platyhelminthes, with some authors using Lophotrochozoa and Spiralia synonymously, while others reserve Lophotrochozoa for a subgroup of these phyla (for example see Struck et al., 2014). The uneven distribution of characters, such as stereotypical spiral cleavage, the occurrence of a trochophore larval stage and the presence of a lophophore feeding organ, make it difficult to decide which synapomorphy best represents this clade, which was originally based on molecular phylogeny (Halanych et al., 1995). We use Lophotrochozoa as inclusive of all these taxa, though whichever nomenclature is adopted their poor sampling has a range of ramifications for phylogeny reconstruction using miRNA data, and in particular the inference of gains of 'novel' miRNA sequences in single clades.

Here, we present the draft genomic sequences of two marine lophotrochozoans— the gastropod mollusc *Patella vulgata* and the serpulid annelid *Spirobranchus* (*Pomatoceros*) *lamarcki*. These species are members of diverse and ecologically vital phyla, and to our knowledge are only the fourth mollusc and third annelid genome resources to be published (after Takeuchi et al., 2012; Zhang et al., 2012; Simakov et al., 2013). These genomes will therefore be useful for a range of investigations.

The common European limpet *P. vulgata* (Fig. 1B) is a univalve gastropod and typical true limpet of the family Patellidae. It is distributed throughout Europe, as far north as the Arctic Circle and as far south as Portugal. It is found attached to firm substrates from the high shore to the edge of the sublittoral zone, although it predominates in areas of wave action. The order Patellogastropoda, to which *P. vulgata* belongs, can be found worldwide, and is well described with members widely used as models in studies of ecology, development and evolution (Lindberg et al., 1998; Nakano and Sasaki, 2011). A mantle-derived transcriptome also exists for this species (Werner et al., 2013).

S. lamarcki (Fig. 1C) is a tube-building serpulid worm which is widespread in intertidal and sublittoral zones around the United Kingdom and northern Europe. They attach to firm substrates and are noted for their detrimental effect on shipping, earning them the common name 'keelworm' (Hamer et al., 2001). *S. lamarcki* is also a useful model for embryological work, as it provides a readily-accessible source of embryonic and larval material (McDougall et al., 2006) and both a *S. lamarcki* EST dataset (Takahashi et al., 2009) and embryonic transcriptome (Kenny and Shimeld, 2012) are available. *S. lamarcki* is a member of the newly redefined Sedentaria class, as is the only other available polychaete genome *Capitella teleta*, however it is relatively phylogenetically distant from both *C. teleta* and *Helobdella robusta*, the other published annelid genome (Struck et al., 2011; Simakov et al., 2013). We note that *S. lamarcki* has recently been the subject of taxonomic revision and the name is synonymous with *Pomatoceros lamarcki* and *Pomatoceros lamarckii* (which are widely used in the literature) (ten Hove, 2015).

Our draft genome assemblies recover 578,961,269 and 964,274,156 bp of sequence for *P. vulgata* and *S. lamarcki* respectively. Using the known catalogue of metazoan miRNAs as the queries for BLAST searches for initial assignation of identity, several unexpected miRNA candidate loci (*miR-1285*, *miR-1287*, *miR-1957*, *miR-1983* and *miR-3533*) were found to be present in these lophotrochozoans. This study provides new genomic resources for an undersampled clade, and suggests that broader sampling will be useful for revealing the evolutionary history of miRNAs.

2. Materials and methods

2.1. P. vulgata DNA extraction and genome sequencing

Adult *P. vulgata* were collected from Tinside, Plymouth, UK. Gonads were dissected from a single male and left in a petri dish in filtered seawater to allow sperm to disassociate from somatic tissue. Large fragments of somatic tissue were removed from the petri dish, and the liquid including sperm transferred to a 15 mL tube. This was then spun at 4000 RPM at 4 °C for 5 min. The supernatant was then removed, and the pellet washed thrice in 3 times its volume of 1× PHB (0.1 M EDTA, 50 M Tris, 2.5% SDS, in distilled water), and spun at 4000 RPM at 4 °C to pellet following each wash step. 1 mL PHB containing 3 μL of 5 M NaCl and 60 μL of 10 mg μL^{-1} Proteinase K was then added to the pellet, which was gently pipetted. This was then left overnight at 50 °C. After digestion, the solution was phenol/chloroform extracted, a

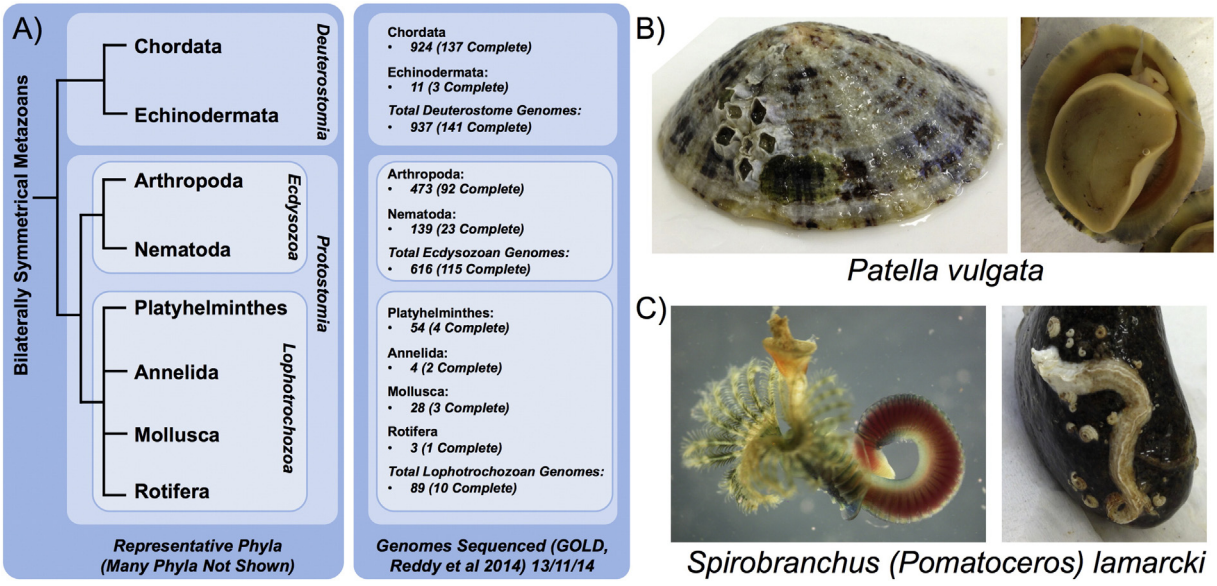


Fig. 1. A representative phylogeny of bilaterally symmetrical animals, sampling at the genomic level across the Bilateria, and images of *Patella vulgata* and *Spirobranchus* (*Pomatoceros*) *lamarcki*. A. Phylogeny is representative and based on a consensus of recent studies. Numbers cited in figure taken from GOLD (Reddy et al., 2014). 'Complete' refers to the GOLD nomenclature differentiating finalized projects from those still underway. B. Oblique (left) and ventral (right) view of adult *P. vulgata*. C. Adult *S. lamarcki* (left), and calcareous tube in situ on a rock gathered from Plymouth seafront (right).

Please cite this article as: Kenny, N.J., et al., Draft genome assemblies and predicted microRNA complements of the intertidal lophotrochozoans *Patella vulgata* (Mollusca, Patelloga..., Mar. Genomics (2015), http://dx.doi.org/10.1016/j.margen.2015.07.004

process which was repeated three times (at which point no identifiable protein layer was observed). The DNA pellet was then ethanol precipitated. The washed pellet was then left to air dry at room temperature, and resuspended in 100 μL milliQ filtered water. DNA concentration was determined using a Nanodrop 1000 spectrophotometer.

A sample of genomic DNA was prepared for sequencing by the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics (Oxford, UK) with nominal fragment size libraries of 200 bp and 500 bp. Genomic DNA from the same sample has been stored for future sequencing. A single lane of Illumina HiSeq 2000 was generated, with 100 bp paired end read length. Initial assessments of the quality of the genomic data were performed using FastQC (Andrews, 2010). The NCBI SRA has been used as the long term repository for raw read data, which are available under the accession number SRP055157.

2.2. S. lamarcki DNA extraction and genome sequencing

S. lamarcki adults were collected from the coast of Tinside, Plymouth, UK and maintained in an aquarium at 12 °C. Genomic DNA (gDNA) was extracted from the sperm of a single adult worm. Sperm were homogenized using an RNase-free polypropylene pellet pestle, washed three times with 2× PHB buffer, and digested overnight in 0.015 M NaCl and Proteinase K (0.6 μg ml^{-1}) at 50 °C. gDNA was extracted by three phenol-chloroform extractions followed by one chloroform extraction with a 15–30 min rotation. Samples were extracted twice with chloroform and incubated in 0.1 volumes of 5 M Sodium acetate and 2.5 volumes of 100% ethanol at −20 °C overnight. gDNA was washed twice with 70% ethanol, air-dried at room temperature, and resuspended in 100 μl of distilled water. Sequencing was performed by the Wellcome Trust Centre for Human Genetics (Oxford, UK) using a single lane on the Illumina HiSeq 2000 with 100 bp paired-end reads, multiplexing two libraries (nominally 201 bp and 500 bp fragment library sizes, including read length). Some gDNA was retained for future sequencing, Quality was assessed using FastQC (Andrews, 2010). Raw reads have been uploaded to the NCBI SRA, and are available under the accession number SRP055158.

2.3. Genome assembly and coverage

Several assembly programmes (Velvet, ABySS, SOAPdenovo) were trialled at a variety of k-mer lengths and the 'best' assembly determined empirically from contig sizes and genome coverage as estimated from total number of base pairs in long (>1 kb) contigs. Genome sizes were estimated for both species using *k*-mer spectrum approach by counting all occurrences of 21-mers in sequenced data with Jellyfish (Fig. 2), where *k* coverage is translated into actual sequencing coverage using the equation $C_k = C \times (L - k + 1) / L$, where C is real coverage, L is read length, k is k-mer size and C_k is k-mer coverage. Read cleaning and error correction was attempted for *P. vulgata*, but was found to decrease quality of final assembly as assayed by the metrics used above. Error correction was performed with Quake (Kelley et al., 2010) using a 19-mer for *S. lamarcki*. For *P. vulgata* ABySS 1.3.4 (Simpson et al., 2009) at a k-mer length of 57 (abyss-pe driver script and all default settings) was used for further analyses, while for *S. lamarcki* SOAPdenovo2 (Luo et al., 2012) at a k-mer length of 51 was used. Final metrics relating to genome assemblies presented here can be seen in Table 1. Genome assemblies themselves are available from the Oxford Research Archive under DOI: 10.5287/bodleian:xp68kh25x.

To assess the quality of genome assemblies, we first assayed the recovered fraction of sets of conserved eukaryotic orthologues in those assemblies using CEGMA (Parra et al., 2007) and BUSCO Version 1.1b (Simão et al., 2015). We then determined the mapping rates of RNA-seq data using the STAR splice-aware aligner to map previously published transcriptomic data (Method: Dobin et al., 2013. Transcriptomes: Werner et al., 2013; Kenny and Shimeld, 2012, available under accession SRA055301).

2.4. miRNA searches and identification

Genomes were compared to all known metazoan miRNA sequences, as downloaded from miRbase (Griffiths-Jones et al., 2008) on the 6th of February 2013 using BLASTN with the following settings: −word_size 11 -reward 5 -penalty −4 -gapopen 8 -gapextend 6. Putative miRNA sequences obtained by blast were checked to confirm that both arms of the putative miRNA were present, for general homogeneity of 5′ (seed) sequence, for robust hairpin structures and for a lack of similarity to known protein, tRNA or rRNA sequences. These criteria are similar to those found in Tarver et al. (2012) and Quah et al. (2015), although as small RNA libraries were not sequenced as a part of our investigation, criteria related to processing and overhang listed in Tarver et al. (2012) were not included in our process. We also performed BLASTN comparison of identified contigs to the NCBI nr database to exclude the possibility that contamination with human DNA (or DNA from other species represented in this database) underlay the identification of candidate miR loci.

2.5. Transcriptional validation of predicted miRNAs

RNA was extracted from *P. vulgata* (head, mantle and foot samples) and *S. lamarcki* (whole adult) samples using a miRVana kit (Life Technologies) and residual gDNA removed using a Qiagen RNeasy kit with on-column DNase treatment according to the manufacturer's

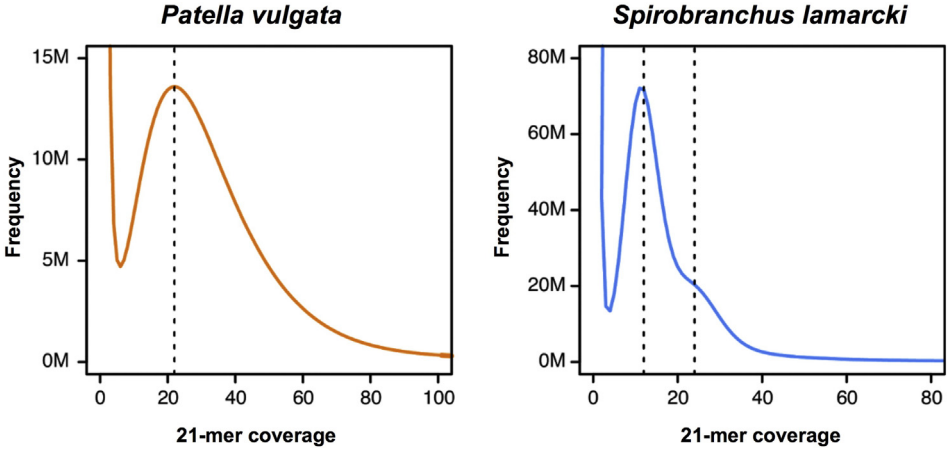


Fig. 2. Coverage distributions of sequence read data. 21-mer distribution showing *k* mer coverage peak (22× and 12/24× in *P. vulgata* and *S. lamarcki* respectively) that are translated into actual sequencing coverages using the equation $C_k = C \times (L - k + 1) / L$, where C is real coverage, L is read length, k is *k* mer size and C_k is *k* mer coverage.

Please cite this article as: Kenny, N.J., et al., Draft genome assemblies and predicted microRNA complements of the intertidal lophotrochozoans *Patella vulgata* (Mollusca, Patelloga..., Mar. Genomics (2015), http://dx.doi.org/10.1016/j.margen.2015.07.004

Table 1
Sequencing and assembly statistics for *Patella vulgata* and *Spirobranchus* (*Pomatoceros*) *lamarcki* genomes. Note: the word 'contig' is used as no long mate pair libraries were used for scaffolding our assemblies.

Species/Metric (bp unless stated)	*Patella vulgata*	*Spirobranchus lamarcki*
Number of 2*100 bp Paired End Fragments (200 bp library)	118,329,948	119,500,405
Number of 2*100 bp Paired End Fragments (500 bp library)	69,107,905	67,048,915
Min contig length	300	300
Max contig length	47,432	37,132
Mean contig length	1960.27	1327.69
Std deviation of contig length	2199.21	1439.37
Median contig length	1224	848
N50 contig length	3160	1939
Number of contigs	295,348	726,277
Number of contigs > = 1 kb	170,706	311,778
Number of contigs in N50	51,138	135,692
Number of bases in all contigs	578,961,269	964,274,156
No. of bases in contigs > = 1 kb	506,356,838	726,465,383
GC Content of contigs (%)	35.18	27.97
N content of contigs (%)	0.00062	0.53485
Genome size (Gbp)	1.46 (perl)/0.95 (*k* mer)	1.25
Overall Coverage	25.60× (perl)/28.2× (*k* mer)	15/30×
k mer Coverage (21mer)	22×	12/24×

protocol. No correction to the protocol was made for small RNA size, so some small RNAs may have been lost at this step. cDNA was generated from RNA using a Takara Primescript Reverse Transcriptase (RT) kit, with negative RT controls made using equal quantities of RNA but no Primescript 5X solution. Primers designed using Primer3Plus were used to perform PCR at the following settings: 94 °C 3 min, 35× (94 °C 30 s, 55 °C 30 s, 72 °C 30 s), 72 °C 7 min. Samples were run in 2% *w*/*v* agarose gels alongside 1 kb + (Invitrogen) ladder. Primers used to perform PCR, alongside details of band sizes and miRNA identity can be seen in Table 2. Gel bands were excised and DNA extracted using a Qiagen Qiaquick gel extraction kit. Bands were cloned into pMD18T vector, transformed into DL-1 *E. coli* and plasmid DNA miniprepped with a Qiagen Qiaprep spin miniprep kit after blue/white selection. Sequencing was performed by Techdragon (Hong Kong) using M13-47 primer on an Applied Biosystems 3730xl DNA Analyser.

Table 2
miRNAs amplified from RNA samples to confirm expression, with primer sequences and predicted band size.

miRNA:	Primers:	Expected Band Size:
P. vulgata 5 s rRNA (control)	F: ACCACGTTGAAAACACCAGTTC R: CGGTCACCCATCCAAGTACTAA	81 bp
S. lamarcki 5 s rRNA (control)	F: GCCATACCACGCTGAATACAC R: GCTTAACTCCCGTGATTGGA	50 bp
S. lamarcki miR-1285	F: GGATAGCACCTGTGAATAGGC R: CCAGCTATGTTGGACAGGCTA	51 bp
S. lamarcki miR-1287	F: CGAAGATTCTAGAAAAGTGGTTCGAG R: GCACCTATCACTGAATCTGTTGC	73 bp
P. vulgata miR-1957	F:GGGATGTAGCTCAGTGGTAGAG R: GAACCCGGGGCCTTTCAC	54 bp
P. vulgata miR-1983	F: AGCGCGCCGTACTTATAGACAG R: GTGAGGCTCGAACTCACAACCT	83 bp
S. lamarcki miR-1983	F: CAGCCCCAGTAGCTCAGTCG R: AAAACAATTGCCCCAGGTGA	124 bp
P. vulgata miR-3533	F: GGCATGGAATCTTCTGGTATTC R: CAAGTCTTTACGGATATCAACATCAC	78 bp
S. lamarcki miR-3533	F: CGAGACCACCTACAACAGCA R: GCGTACAAGTCCTTACGGATG	60 bp

3. Results

3.1. Statistics on assemblies

FastQC assessment of read quality ascertained raw read data to be excellent, with median PHRED scores above 30 through to the 100th base for both read directions in all libraries in both species. Despite this, QUAKE-treated *S. lamarcki* read genome assemblies were empirically found to be better than those constructed from raw reads alone, and were thus used for further analysis. The genomic datasets presented here comprise 578,961,269 and 964,274,156 bp of sequence for *P. vulgata* and *S. lamarcki* respectively. Genome sizes were independently estimated from raw read data using the *k*-mer spectrum approach, which assesses coverage peak based on *k*-mer count (Fig. 2). We found an approximate genome size of 1.46 Gbp (25.7× base pair coverage) in *P. vulgata* using the estimate_genome_size.pl script (Ryan, 2013), and an estimate of 0.95 Gbp (22× *k*-mer coverage, 28.2× base pair coverage) using the *k*-mer spectrum approach, with a single peak of *k*-mer coverage. Conversely, we observed a double peak (12× and 24×) in the *k*-mer distribution of *S. lamarcki* which is the hallmark of high heterozygosity content (Kajitani et al., 2014), and when summed is consistent with a genome size of 1.25 Gb. We therefore recover approximately 50% and 80% of the estimated genomes of these species. This shortfall is likely explained by the misassembly of the repetitive fraction in these genomes, and the moderate coverage and short fragment sizes libraries used for assembly will make proper recovery of these portions of the genome challenging.

Our genome size estimates are consistent with previous experimental measures for *P. vulgata* and *S. lamarcki*, which indicate respective sizes of 1 Gbp and 1.2–1.5 Gbp according to the Animal Genome Size Database (www.genomesize.com/). The C-value (in pg) of the ten species of Patellogastropoda as extracted from this database varies between 0.43 (*Lottia gigantea*) and 0.94 (*Acmaea mitra*). The *P. vulgata* genome is therefore larger than that of all other patellogastropods, but we note that our wild-collected individual is likely highly heterozygous, which may artificially inflate our estimate. Annelid genome sizes vary widely, from the exceptionally small (*Dinophilus gyrociliatus*, 0.06 pg (Soldi et al., 1994)) to the very large (*Spirosperma ferox*, 7.64 pg; Gregory and Hebert (2002)), and *S. lamarcki* is therefore within the normal range for annelid genome size.

The *P. vulgata* GC percentage, 35.18%, is similar to that found in a previous transcriptomic study (Werner et al., 2013: 33.56%), as well as resembling that seen in other molluscs. However, at 27.97%, the GC content of the *S. lamarcki* genomic assembly is exceptionally low. It is markedly lower than that seen in previous EST (42.42%, Takahashi et al., 2009) and transcriptome (43.33%, Kenny and Shimeld, 2012) analyses. It is also lower than that recorded in previously published annelid datasets (*Capitella teleta* 40%, *Helobdella robusta* 33%; Simakov et al., 2013). Whether this reflects the biology of these organisms or is the result of bias in our genomic sequencing and assembly remains to be confirmed.

Contiguity is generally low, although a sizeable percentage of the assemblies are contained in contigs longer than 1 kb in length (87.5% of *P. vulgata* and 75.3% of *S. lamarcki*). At this size, contigs can easily be assayed for protein domain content, allowing identification of genes as well as some information about intronic and non-coding regions. The N50 results (*P. vulgata*: 3160, *S. lamarcki*: 1939) are also of sufficient length to ensure many regions of the genome have adequate contiguity for establishing whole gene sequence.

3.2. Assessment of assembly quality

Our genome assemblies contain 68/248 (27.42%) complete, 144/248 (58.06%) partial (*P. vulgata*) and 6/248 (2.42%) complete, 31/248 (12.50%) partial recovery (*S. lamarcki*) of the core eukaryotic gene mapping dataset as assessed by CEGMA (Parra et al., 2007). To confirm these statistics, we utilized BUSCO (Simão et al., 2015), which returned similar

results to CEGMA. Of the BUSCO set of 843 metazoan orthologues, *P. vulgata* possessed 177 complete, 7 duplicated, 165 fragmented and 501 missing genes, for a total recovery of 41% of the dataset. *S. lamarcki*'s assembly contained 12 complete, 0 duplicated, 27 fragmented and 804 missing genes (5% recovery).

The poor recovery in *S. lamarcki* is probably a consequence of the low contiguity of the genome assembly. To test whether these genes were present but unrecovered by CEGMA or BUSCO due to low contig length, we ran TBLASTN against the genome assemblies using the 843 BUSCO orthologs as queries, with an *E* value cut-off of 1e-6. Of these, 798 (94.7%) of these had at least one hit in *P. vulgata*, and 709 (84.1%) in *S. lamarcki*. This is an overestimation of their recovery, as there will be shared domains in some of these proteins capable of generating a hit above threshold, but indicates the assemblies possess more of the coding fraction of the genomes than raw CEGMA or BUSCO output suggests. There was also only a 10.6% difference between *S. lamarcki* and *P. vulgata* using this approach, compared to 25% to 45.56% for CEGMA and BUSCO, indicating that the greater CEGMA/BUSCO recovery from *P. vulgata* may be a consequence of its increased average contig length. Finally, comparison of published transcriptome data to the assemblies using the STAR splice-aware aligner found higher levels of recovery, with 77.07% (*P. vulgata*) and 33.66% (*S. lamarcki*) of RNA-seq reads mapped. We conclude that while contiguity is limited (especially for *S. lamarcki*), we recover the majority of *P. vulgata* coding sequence and a considerable proportion *S. lamarcki* coding sequence in the assemblies. Further, low contiguity is less likely to affect the detection of miRNA loci, which are very short in length.

3.3. In silico identification of miRNA genes

Using the complete miRNA dataset contained in miRbase we recovered 45 (*P. vulgata*) and 54 (*S. lamarcki*) putative miRNAs (Supplementary File A). This is less than the number of miRNAs catalogued for *Lottia gigantea* (59) and *Capitella teleta* (129) (mirBase v21), but higher than the number reported in other lophotrochozoan species such as *Platynereis dumerilii* (34; Christodoulou et al., 2010). Nineteen of these putative miRNAs were found in both *P. vulgata* and *S. lamarcki*, increasing confidence that they are bona fide miRNAs.

Of the canonical miRNA families, we recovered a large number of well-conserved examples (as described in Wheeler et al., 2009) from both genomes. Eumetazoan miRNA families such as *miR-100*, and bilaterian families such as *miR-7*, *miR-8* and miR-*9* (among many others) were found in both genomes. We were also able to find miRNAs only previously described in molluscs or annelids (Wheeler et al., 2009; Tarver et al., 2013) in *P. vulgata* and *S. lamarcki* respectively, such as *miR-1985*, *miR-1986* and *miR-1988* in *P. vulgata* and *miR-1996* and *miR-2000* in *S. lamarcki*. These findings reinforce the categorisation of these particular miRNAs as gastropod and annelid synapomorphies respectively (e.g. Tarver et al., 2013).

Surprisingly, we also noted the presence of a range of miRNA families that had been previously described as confined to lineages other than Lophotrochozoa. For example, *miR-1285*, *miR-1287*, *miR-1957*, *miR-1983* and *miR-3533* (undescribed outside Vertebrata), *miR-3350* (described only in silkworm; Cai et al. (2010)) and *miR-494* and *miR-767* (only reported in eutherians; Tarver et al. (2013)), were found in either *S. lamarcki*, *P. vulgata*, or both.

To ascertain whether these genes are present more broadly across the Lophotrochozoa, we first checked them against published lophotrochozoan miRNA sequences that have not been accessioned into miRbase (Xue et al., 2008; Christodoulou et al., 2010; Bao et al., 2014), but none of these mIRNAs have previously been noted. To check whether this was through oversight rather than true absence, we searched the published genomes of the molluscs *P. fucata*, *C.gigas* and *L. gigantea* and the annelids *H. robusta and C. teleta* for these sequences. miRNAs *miR-494*, *miR-767*, *miR-1287* and *miR-3350*, were not identified in the other lophotrochozoan species examined, a finding that supports reports of heterogeneous rates of gain and loss (Fromm et al., 2013; Thomson et al., 2014). However candidate *miR-1957*, *mir-1983* and *mir-3533* sequences were found in all lophotrochozoan genomes examined, suggesting that these species possess an as-yet uncharacterised diversity of miRNA genes (Supplementary File B). Some miRNAs, such as the putative *L. gigantea miR-1957*, are 100% identical to their *P. vulgata* counterpart. These three miRNAs could therefore have been present in the common ancestor of molluscs and annelids, but be as yet unidentified in living descendant by small RNA sequencing approaches.

3.4. Verification of transcription

In order to discern whether these surprising putative miRNAs were in fact transcribed, we extracted short fragment RNA from head, mantle and foot tissue of *P. vulgata* and from whole adult *S. lamarcki*. We then used RT-PCR to assay for the presence of several miRNA, using primers as described in Table 2. Not all miRNA sequences identified in our analysis were found to be transcribed in these samples. At present, we could not differentiate whether this could indicate the predicted miRNAs are artefacts, or whether they are expressed in tissues or stages we did not examine. Positive results are shown in Fig. 3C. In both *P. vulgata* and *S. lamarcki*, evidence for the transcription of the *miR-1983* and *miR-3533* loci was obtained (Fig. 3C): these miRs have been previously only described in mice (Babiarz et al., 2008) and/or other vertebrates (Wang et al., 2009; Vegh et al., 2013). In addition, we also noted the transcription of *miR-1957* (mouse specific; Kuchenbauer et al. (2008)) in *P. vulgata* and *miR-1285* and *miR-1287* (vertebrate specific; Strozzi et al. (2009), Brameier (2010), Meunier et al. (2013)) in *S. lamarcki*. Together this provides evidence for the transcription of several of these miRNA loci. It should however be noted that miRbase shows *miR-3533* as so far only identified in chicken and cow, and in both cases as mapping to a region of the genome that also encodes an actin gene. This co-localisation with actin genes is also found in our lophotrochozoan sequences, and may both contribute to the level of homology as shown in Fig. 3A and underlie its transcription.

4. Discussion and conclusions

The two genome assemblies presented here will be useful for a range of investigations, given the currently sparse sampling of lophotrochozoans. Our preliminary assemblies and *k*-mer based genome size estimates also provide a basis for establishing appropriate strategies to improve genome assembly in these and related species. As with some other recently-sequenced marine animals, such as the pacific oyster *Crassostrea gigas* (Zhang et al., 2012) and the cephalochordate *Branchiostoma belcheri* (Huang et al., 2014), these two genomes are highly polymorphic. This is likely explained by very large effective population sizes deriving from planktonic larval dispersal (Romiguier et al., 2014). High levels of polymorphism represent a challenge for assembly and annotation. Nevertheless, while contiguity is low, assembly is still sufficient to carry out the miRNA analysis described here.

miRNAs represent potentially useful, slow-evolving characters for the inference of phylogeny, and have been used to suggest solutions to a variety of recalcitrant problems in metazoan phylogeny. However, given the sparse and biased nature of our sampling of genomic diversity across the tree of life, it would perhaps be better to adopt caution when making claims as to the phylogenetic inter-relationships of taxa, especially in cases where only a small number of miRNAs are used as the basis for such claims, or where only single datasets are utilized as the raw material for such inference. Here, we show that an informative survey of miRNAs is possible through moderate coverage genome sequencing and assembly, and we suggest a similar approach might be applied more broadly across animal diversity. We found many (45 *P. vulgata* and 54 *S. lamarcki*) putative miRNAs in our datasets, representing a considerable proportion of the expected complements of lophotrochozoans when compared to the well-annotated *L. gigantea* and *C. teleta* genomes. The discovery of potential mollusc- and annelid-specific miRNAs in

Please cite this article as: Kenny, N.J., et al., Draft genome assemblies and predicted microRNA complements of the intertidal lophotrochozoans *Patella vulgata* (Mollusca, Patelloga..., Mar. Genomics (2015), http://dx.doi.org/10.1016/j.margen.2015.07.004

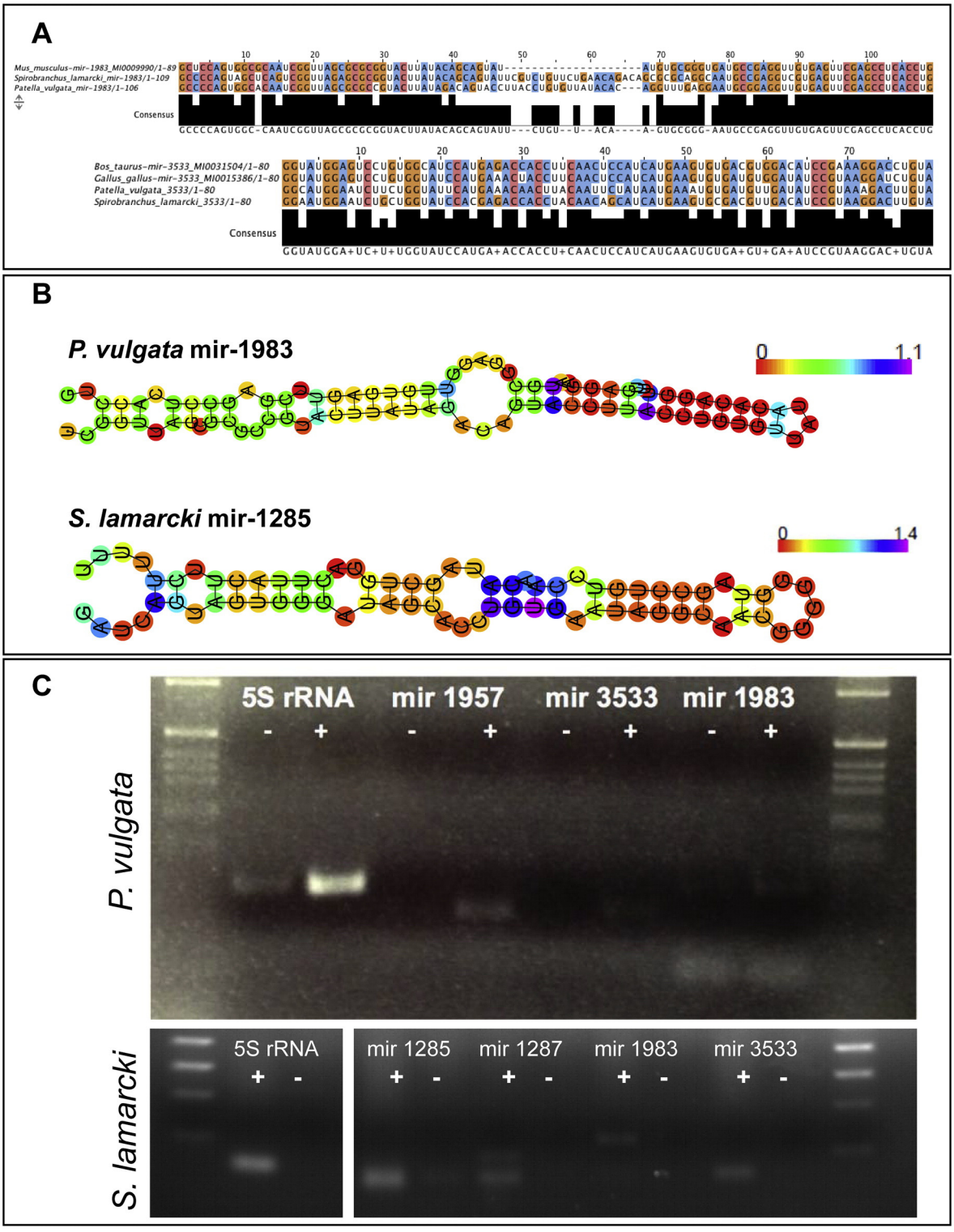


Fig. 3. Evidence for the Expression of "Vertebrate like" miRNA in *Patella vulgata* and *Spirobranchus* (*Pomatoceros*) *lamarcki*. A. Example alignments of novel miRNA candidates with known orthologues. B. Hairpin structures of example miRNAs, *S. lamarcki* miR-1285 and *P. vulgata* miR-1983, as displayed by RNAfold (Gruber et al., 2008) with positional entropy values displayed. C. RT-PCR results showing the expression of miRNAs in both species examined. + and − indicate analyses conducted with and without reverse transcription (RT). Note the presence of a weak band indicating amplification of small amounts of residual genomic DNA in the 5 s RNA band for *P. vulgata* −RT sample (due to the high copy number of rRNA genes in the genome). We did not observe bands indicating miRNA sequence amplification in other -RT controls. Smears at the base of miR-1983 in *P. vulgata* are primer dimers, a weak band can be seen above in +RT lane.

P. vulgata and *S. lamarcki* respectively also reinforces the likelihood that those particular genes are synapomorphies of these clades, rather than limited to the (often single species) samples where they were first described. This demonstrates the utility of extra datasets when addressing the still under-researched field of miRNA evolution.

While *miR-1957*, *miR-1983* and *miR-3533* (noting the caveat for mir-3533 described above) are found in both the species examined here and other sequenced lophotrochozoan species, some of the more surprising findings are limited to only one of the two assemblies described here, as can be seen in Fig. 4. When coupled with their absence from other

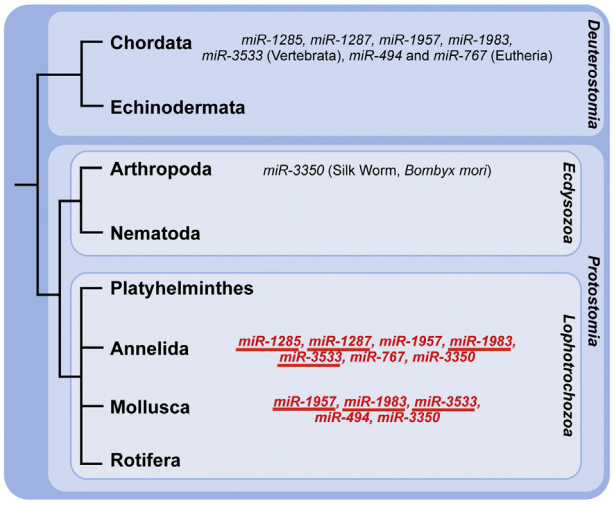


Fig. 4. Distribution of phylogenetically distant miRNA gene apparent homologues. Phylogenetic distribution of miRNA genes of apparent homology discussed in this paper. Those shown in black have been described previously, while those shown in red, bolded text are described for the first time in this paper. Those underlined have further been shown to be transcribed in *S. lamarcki* (Annelida) or *P. vulgata* (Mollusca).

sequenced lophotrochozoan genomes, this raises the possibility of homoplasy, notably for the putative *miR-494*, *miR-767*, *miR-1287* and *miR-3350*. However sampling of animal, and particularly lophotrochozoan, phylogeny remains too sparse to assert with full confidence whether the sequences found, particularly those of *miR-494*, *miR-767*, *miR-1287* and *miR-3350*, could have arisen by convergent evolution or are the remnants of prevalent loss across metazoan phylogeny. These competing hypotheses will be better tested when a broader range of genomes are available, drawn from the widest possible range of bilaterian species. Datasets such as the ones presented here represent key comparison points in the interim. Either way, however, the identification of unexpected miRNA sequences in these species suggests that current assumptions concerning the utility of miRNAs in reconstructing phylogeny need qualification: either more loss has occurred across metazoan phylogeny than postulated previously, or convergent evolution across wide evolutionary distances is possible. The complex evolutionary history of miRNAs is interesting in light of the important role that they play in animal biology, and understanding the true nature of their gain and loss will be vital for an understanding of how miRNAs influence and are influenced by genomic evolution.

Without detailed cataloguing of miRNA family sequences in a more diverse range of animals, some of the key assumptions made about miRNAs for use in phylogeny will remain untested. Only greater density of sampling across the metazoan tree of life will reveal further the true nature of miRNA conservation and loss. The two novel lophotrochozoan genomes presented here, as well as being useful for a wide range of investigations, are another step in this process.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.margen.2015.07.004.

Statement of competing Interests

The authors declare no conflict of interests. The funding source had no input into the decision to publish or in preparation of this manuscript.

Acknowledgments

The authors thank the Elizabeth Hannah Jenkinson Fund for grants supporting the sequencing of the genomes listed here (*P. vulgata*: grant to NJK and JHLH, *S. lamarcki*: grant to EKON and JHLH). For sequencing we thank the High-Throughput Genomics unit at the Wellcome Trust Centre for Human Genetics, Oxford. NJK was supported by a Clarendon Scholarship for work on this project. We also thank the members of our laboratories for their many helpful comments and support.

References

Aguinaldo, A.M.A., Turbeville, J.M., Linford, L.S., Rivera, M.C., Garey, J.R., Raff, R.A., Lake, J.A., 1997. Evidence for a clade of nematodes, arthropods and other moulting animals. Nature 387, 489–493. http://dx.doi.org/10.1038/387489a0.

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Axtell, M.J., Bartel, D.P., 2005. Antiquity of microRNAs and their targets in land plants. Plant Cell 17, 1658–1673. http://dx.doi.org/10.1105/tpc.105.032185.

Babiarz, J.E., Ruby, J.G., Wang, Y., Bartel, D.P., Blelloch, R., 2008. Mouse ES cells express endogenous shRNAs, siRNAs, and other microprocessor-independent, Dicer-dependent small RNAs. Genes Dev. 22, 2773–2785. http://dx.doi.org/10.1101/gad.1705308.

Bao, Y., Zhang, L., Dong, Y., Lin, Z., 2014. Identification and comparative analysis of the *Tegillarca granosa* haemocytes MicroRNA transcriptome in response to Cd using a deep sequencing approach. PLoS One 9, e93619. http://dx.doi.org/10.1371/journal.pone.0004034.

Bartel, D.P., 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116, 281–297. http://dx.doi.org/10.1016/S0092-8674(04)00045-5.

Bartel, D.P., 2009. MicroRNAs: target recognition and regulatory functions. Cell 136, 215–233. http://dx.doi.org/10.1016/j.cell.2009.01.002.

Brameier, M., 2010. Genome-wide comparative analysis of microRNAs in three non-human primates. BMC research notes 3, 64. http://dx.doi.org/10.1186/1756-0500-3-64.

Cai, Y., Yu, X., Zhou, Q., Yu, C., Hu, H., Liu, J., Lin, H., Yang, J., Zhang, B., Cui, P., Hu, S., Yu, J., 2010. Novel microRNAs in silkworm (*Bombyx mori*). Funct. Integr. Genomics 10, 405–415. http://dx.doi.org/10.1007/s10142-010-0162-7.

Campbell, L.I., Rota-Stabellli, O., Marchioro, T., Longhorn, S.J., Edgecombe, G.D., Telford, M.J., Philippe, H., Rebecchi, L., Peterson, K.J., Pisani, D., 2011. MicroRNAs and phylogenomics resolve the relationships of tardigrada and suggest the velvet worms are the sister group of arthropoda. Proc. Natl. Acad. Sci. U. S. A. 108, 15920–15924. http://dx.doi.org/10.1073/pnas.1105499108.

Christodoulou, F., Raible, F., Tomer, R., Simakov, O., Trachana, K., Klaus, S., et al., 2010. Ancient animal microRNAs and the evolution of tissue identity. Nature 463, 1084–1088. http://dx.doi.org/10.1038/nature08744.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. http://dx.doi.org/10.1093/bioinformatics/bts635.

Field, D.J., Gauthier, J.A., King, B.L., Pisani, D., Lyson, T.R., Peterson, K.J., 2014. Toward consilience in reptile phylogeny: miRNAs support an archosaur, not lepidosaur, affinity for turtles. Evol. Dev. 16, 189–196. http://dx.doi.org/10.1111/ede.12081.

Friedman, R.C., Farh, K.K.-H., Burge, C.B., Bartel, D.P., 2009. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. 19, 92–105. http://dx.doi.org/10.1101/gr.082701.108.

Fromm, B., Worren, M.M., Hahn, C., Hovig, E., Bachmann, L., 2013. Substantial loss of conserved and gain of novel microRNA families in flatworms. Mol. Biol. Evol. 30, 2619–2628. http://dx.doi.org/10.1093/molbev/mst155.

Gregory, T.R., Hebert, P.D.N., 2002. Genome size estimates for some oligochaete annelids. Can. J. Zool. 80, 1485–1489. http://dx.doi.org/10.1139/z02-145.

Griffiths-Jones, S., Saini, H.K., van Dongen, S., Enright, A.J., 2008. MiRBase: tools for microRNA genomics. Nucleic Acids Res. 36, D154–D158. http://dx.doi.org/10.1093/nar/gkm952.

Gruber, A.R., Lorenz, R., Bernhart, S.H., Neubock, R., Hofacker, I.L., 2008. The Vienna RNA website. Nucleic Acids Res. 36, W70–W74. http://dx.doi.org/10.1093/nar/gkn188.

Halanych, K.M., Bacheller, J., Liva, S., Aguinaldo, A.A., Hillis, D.M., Lake, J.A., 1995. 18S rDNA evidence that the lophophorates are protostome animals. Science 267, 1641–1643. http://dx.doi.org/10.1126/science.7886451.

Hamer, J., Walker, G., Latchford, J., 2001. Settlement of *Pomatoceros lamarkii* (Serpulidae) larvae on biofilmed surfaces and the effect of aerial drying. J. Exp. Mar. Biol. Ecol. 260, 113–132. http://dx.doi.org/10.1016/S0022-0981(01)00247-7.

Huang, S., Chen, Z., Yan, X., Yu, T., Huang, G., Yan, Q., et al., 2014. Decelerated genome evolution in modern vertebrates revealed by analysis of multiple lancelet genomes. Nat. Commun. 5. http://dx.doi.org/10.1038/ncomms6896.

Hui, J.H.L., Marco, A., Hunt, S., Melling, J., Griffiths-Jones, S., Ronshaugen, M., 2013a. Structure, evolution and function of the bi-directionally transcribed iab-4 microRNA locus in insects. Nucleic Acids Res. 41, 3352–3361. http://dx.doi.org/10.1093/nar/gks1445.

Hui, J.H.L., Bendena, W.G., Tobe, S.S., 2013b. Future perspectives on the research of juvenile hormones and sesquiterpenoids in arthropod endocrinology and ecotoxicology. Modelling Biological Effects and Environmental Fate. CRC Press. In Juvenile hormones and juvenoids, pp. 15–30.

Kajitani, R., Toshimoto, K., Noguchi, H., Toyoda, A., Ogura, Y., Okuno, M., Yabana, M., Harada, M., Nagayasu, E., Maruyama, H., Kohara, Y., Fujiyama, A., Hayashi, T., Itoh, T., 2014. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome Res. 24, 1384–1395. http://dx.doi.org/10.1101/gr.170720.113.

Kelley, D.R., Schatz, M.C., Salzberg, S.L., 2010. Quake: quality-aware detection and correction of sequencing errors. Genome Biol. 11, R116. http://dx.doi.org/10.1186/gb-2010-11-11-r116.

Kenny, N.J., Shimeld, S.M., 2012. Additive multiple k-mer transcriptome of the keelworm *Pomatoceros lamarckii* (Annelida; Serpulidae) reveals annelid trochophore

transcription factor cassette. Dev. Genes Evol. 222, 325–339. http://dx.doi.org/10.1007/s00427-012-0416-6.
Kenny, N.J., Quah, S., Holland, P.W.H., Tobe, S.S., Hui, J.H.L., 2013. How do comparative genomics and microRNAs change our views on arthropod endocrinology and their adaptations to the environment? Gen. Comp. Endocrinol. 188, 16–22. http://dx.doi.org/10.1016/j.ygcen.2013.02.013.
Kenny, N.J., Sin, Y.W., Hayward, A., Paps, J., Chu, K.H., Hui, J.H.L., 2015. The phylogenetic utility of functional constraint on microRNA sequence evolution. Proc. R. Soc. B http://dx.doi.org/10.1098/rspb.2014.2983.
Kloosterman, W.P., Plasterk, R.H.A., 2006. The diverse functions of microRNAs in animal development and disease. Dev. Cell 11, 441–450. http://dx.doi.org/10.1016/j.devcel.2006.09.009.
Kuchenbauer, F., Morin, R.D., Argiropoulos, B., Petriv, I., Griffith, M., Heuser, M., Yung, E., Piper, J., Delaney, A., Prabhu, A.L., Zhao, Y., McDonald, H., Zeng, T., Hirst, M., Hansen, C.L., Marra, M.A., Humphries, R.K., 2008. In-depth characterization of the microRNA transcriptome in a leukemia progression model. Genome Res. 18, 1787–1797. http://dx.doi.org/10.1101/gr.077578.108.
Lee, R.C., Feinbaum, R.L., Ambros, V., 1993. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. Cell 75, 843–854. http://dx.doi.org/10.1016/0092-8674(93)90529-Y.
Lindberg, D.R., Estes, J.A., Warheit, K.I., 1998. Human influences on trophic cascades along rocky shores. Ecol. Appl. 8, 880–890.
Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D.W., Yiu, S., Peng, S., Xiaoqian, Z., Liu, G., Liao, X., Li, Y., Yang, H., Wang, J., Lam, T., Wang, J., 2012. SOAPdenovo2 : an empirically improved memory-efficient short-read de novo assembler. Gigascience 1, 1–6. http://dx.doi.org/10.1186/2047-217X-1-18.
McDougall, C., Chen, W.-C., Shimeld, S.M., Ferrier, D.E.K., 2006. The development of the larval nervous system, musculature and ciliary bands of *Pomatoceros lamarckii* (Annelida): heterochrony in polychaetes. Front. Zool. 3, 16.
Meunier, J., Lemoine, F., Soumillon, M., Liechti, A., Weier, M., Guschanski, K., Hu, H., Khaitovich, P., Kaessmann, H., 2013. Birth and expression evolution of mammalian microRNA genes. Genome Res. 23, 34–45. http://dx.doi.org/10.1101/gr.140269.112.
Nakano, T., Sasaki, T., 2011. Recent advances in molecular phylogeny, systematics and evolution of patellagastropod limpets. J. Molluscan Stud. 77, 203–217.
Parra, G., Bradnam, K., Korf, I., 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23, 1061–1067. http://dx.doi.org/10.1093/bioinformatics/btm071.
Pasquinelli, A.E., Reinhart, B.J., Slack, F., Martindale, M.Q., Kuroda, M.I., Maller, B., Hayward, D.C., Ball, E.E., Degnan, B., Müller, P., Spring, J., Srinivasan, A., Fishman, M., Finnerty, J., Corbo, J., Levine, M., Leahy, P., Davidson, E., Ruvkun, G., 2000. Conservation of the sequence and temporal expression of *let-7* heterochronic regulatory RNA. Nature 408, 86–89. http://dx.doi.org/10.1038/35040556.
Philippe, H., Brinkmann, H., Copley, R.R., Moroz, L.L., Nakano, H., Poustka, A.J., Wallberg, A., Peterson, K.J., Telford, M.J., 2011. Acoelomorph flatworms are deuterostomes related to Xenoturbella. Nature 470, 255–258. http://dx.doi.org/10.1038/nature09676.
Plaisance-Bonstaff, K., Renne, R., 2011. Viral miRNAs. Antiviral RNAi. Humana Press, pp. 43–66 http://dx.doi.org/10.1007/978–1–61779–037-9_3.
Quah, S., Hui, J.H., Holland, P.W., 2015. A burst of miRNA innovation in the early evolution of butterflies and moths. Mol. Biol. Evol. 32, 1161–1174. http://dx.doi.org/10.1093/molbev/msv004.
Reddy, T.B.K., Thomas, A.D., Stamatis, D., Bertsch, J., Isbandi, M., Jansson, J., Mallajosyula, J., Pagani, I., Lobos, E.A., Kyrpides, N.C., 2014. The Genomes OnLine Database (GOLD) v. 5: a metadata management system based on a four level (meta) genome project classification. Nucleic Acids Res. gkv455v1-gkv455. http://dx.doi.org/10.1093/nar/gku950.
Romiguier, J., Gayral, P., Ballenghien, M., Bernard, A., Cahais, V., Chenuil, A., Chiari, Y., Dernat, R., Duret, L., Faivre, N., Loire, E., Lourenco, J.M., Nabholz, B., Roux, C., Tsagkogeorga, G., Weber, A.A.-T., Weinert, L.A., Belkhir, K., Bierne, N., Glémin, S., Galtier, N., 2014. Comparative population genomics in animals uncovers the determinants of genetic diversity. Nature 515, 261–263. http://dx.doi.org/10.1038/nature13685.
Rota-Stabelli, O., Campbell, L., Brinkmann, H., Edgecombe, G.D., Longhorn, S.J., Peterson, K.J., Pisani, D., Philippe, H., Telford, M., 2010. A congruent solution to arthropod phylogeny: phylogenomics, microRNAs and morphology support monophyletic mandibulata. Proc. R. Soc. Lond. B Biol. Sci. 278, 298–306. http://dx.doi.org/10.1098/rspb.2010.0590.
Ryan, J.F., 2013. estimate_genome_size.pl (Version 0.03) [Computer Software]. Sars International Centre for Marine Molecular Biology, Bergen, Norway (Retrieved from http://josephryan.github.com/estimate_genome_size.pl/).
Sempere, L.F., Cole, C.N., McPeek, M.A., Peterson, K.J., 2006. The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. J. Exp. Zool. B Mol. Dev. Evol. 306, 575–588. http://dx.doi.org/10.1002/jez.b.21118.
Simakov, O., Marletaz, F., Cho, S.J., Edsinger-Gonzales, E., Havlak, P., Hellsten, U., Kuo, D.H., Larsson, T., Lv, J., Arendt, D., Savage, R., Osoegawa, K., de Jong, P., Grimwood, J., Chapman, J.A., Shapiro, H., Aerts, A., Otillar, R.P., Terry, A.Y., Boore, J.L., Grigoriev, I.V., Lindberg, D.R., Seaver, E.C., Weisblat, D.A., Putnam, N.H., Rokhsar, D.S., 2013. Insights into bilaterian evolution from three spiralian genomes. Nature 493, 526–531. http://dx.doi.org/10.1038/nature11696.
Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. Assessing genome assembly and annotation completeness with benchmarking universal single-copy orthologs. http://busco.ezlab.org (Version 1.1b, May 2015).
Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J., Birol, I., 2009. ABySS: a parallel assembler for short read sequence data. Genome Res. 19, 1117–1123. http://dx.doi.org/10.1101/gr.089532.108.
Soldi, R., Ramella, L., Gambi, M.C., Sordino, P., Sella, G., 1994. Genome size in polychaetes: relationship with body length and life habit. In: Dauvin, J.-C., Laubier, L., Reish, D.J. (Eds.), Actes de la 4ième Conférence internationale des Polychètes, Mém. Mus. Natn. Hist. Nat. 162. Editions du Muséum, Paris, France, pp. 129–135.
Strozzi, F., Mazza, R., Malinverni, R., Williams, J.L., 2009. Annotation of 390 bovine miRNA genes by sequence similarity with other species. Anim. Genet. 40, 125. http://dx.doi.org/10.1111/j.1365-2052.2008.01780.x.
Struck, T.H., Paul, C., Hill, N., Hartmann, S., Hösel, C., Kube, M., Lieb, B., Meyer, A., Tiedemann, R., Purschke, G., Bleidorn, C., 2011. Phylogenomic analyses unravel annelid evolution. Nature 470, 95–98. http://dx.doi.org/10.1038/nature09864.
Struck, T.H., Wey-Fabrizius, A.R., Golombek, A., Hering, L., Weigert, A., Bleidorn, C., Klebow, S., Iakovenko, N., Hausdorf, B., Petersen, M., Kuck, P., Herlyn, H., Hankeln, T., 2014. Platyzoan paraphyly based on phylogenomic data supports a noncoelomate ancestry of spiralia. Mol. Biol. Evol. 31, 1833–1849. http://dx.doi.org/10.1093/molbev/msu143.
Takahashi, T., McDougall, C., Troscianko, J., Chen, W.C., Jayaraman-Nagarajan, A., Shimeld, S.M., Ferrier, D.E., 2009. An EST screen from the annelid *Pomatoceros lamarckii* reveals patterns of gene loss and gain in animals. BMC Evol. Biol. 9, 240. http://dx.doi.org/10.1186/1471-2148-9-240.
Takeuchi, T., Kawashima, T., Koyanagi, R., Gyoja, F., Tanaka, M., Ikuta, T., Shoguchi, E., Fujiwara, M., Shinzato, C., Hisata, K., Fujie, M., Usami, T., Nagai, K., Maeyama, K., Okamoto, K., Aoki, H., Ishikawa, T., Masaoka, T., Fujiwara, A., Endo, K., Endo, H., Nagasawa, H., Kinoshita, S., Asakawa, S., Watabe, S., Satoh, N., 2012. Draft genome of the pearl oyster *Pinctada fucata*: a platform for understanding bivalve biology. DNA Res. 19, 117–130. http://dx.doi.org/10.1093/dnares/dss005.
Tarver, J.E., Donoghue, P.C., Peterson, K.J., 2012. Do miRNAs have a deep evolutionary history? BioEssays 34, 857–866. http://dx.doi.org/10.1002/bies.201200055.
Tarver, J.E., Sperling, E.A., Nailor, A., Heimberg, A.M., Robinson, J.M., King, B.L., Pisani, D., Donoghue, P.C.J., Peterson, K.J., 2013. MiRNAs: small genes with big potential in metazoan phylogenetics. Mol. Biol. Evol. 30, 2369–2382. http://dx.doi.org/10.1093/molbev/mst133.
ten Hove, H., 2015. *Spirobranchus lamarcki* (Quatrefages, 1866). In: Read, G., Fauchald, K. (Eds.), (2015) World Polychaeta database (Accessed through: World Register of Marine Species at http://www.marinespecies.org/aphia.php?p=taxdetails&id=560033 on 2015–05–29).
Thomson, R.C., Plachetzki, D.C., Mahler, D.L., Moore, B.R., 2014. A critical appraisal of the use of microRNA data in phylogenetics. Proc. Natl. Acad. Sci. U. S. A. 111, E3659–E3668. http://dx.doi.org/10.1073/pnas.1407207111.
Vegh, P., Foroushani, A.B., Magee, D.A., McCabe, M.S., Browne, J.A., Nalpas, N.C., Conlon, K.M., Gordon, S.V., Bradley, D.G., MacHugh, D.E., Lynn, D.J., 2013. Profiling microRNA expression in bovine alveolar macrophages using RNA-seq. Vet. Immunol. Immunopathol. 155, 238–244. http://dx.doi.org/10.1016/j.vetimm.2013.08.004.
Wang, Y., Brahmakshatriya, V., Zhu, H., Lupiani, B., Reddy, S.M., Yoon, B.J., Gunaratne, P.H., Kim, J.H., Chen, R., Wang, J., Zhou, H., 2009. Identification of differentially expressed miRNAs in chicken lung and trachea with avian influenza virus infection by a deep sequencing approach. BMC Genomics 10, 512. http://dx.doi.org/10.1186/1471-2164-10-512.
Werner, G.D., Gemmell, P., Grosser, S., Hamer, R., Shimeld, S.M., 2013. Analysis of a deep transcriptome from the mantle tissue of *Patella vulgata* Linnaeus (Mollusca: gastropoda: patellidae) reveals candidate biomineralising genes. Mar. Biotechnol. 15, 230–243. http://dx.doi.org/10.1007/s10126-012-9481-0.
Wheeler, B.M., Heimberg, A.M., Moy, V.N., Sperling, E.A., Holstein, T.W., Heber, S., Peterson, K.J., 2009. The deep evolution of metazoan microRNAs. Evol. Dev. 11, 50–68.
Xue, X., Sun, J., Zhang, Q., Wang, Z., Huang, Y., Pan, W., 2008. Identification and characterization of novel microRNAs from *Schistosoma japonicum*. PLoS One 3, e4034. http://dx.doi.org/10.1371/journal.pone.0004034.
Zhang, G., Fang, X., Guo, X., Li, L., Luo, R., et al., 2012. The oyster genome reveals stress adaptation and complexity of shell formation. Nature 490, 49–54. http://dx.doi.org/10.1038/nature11413.

Please cite this article as: Kenny, N.J., et al., Draft genome assemblies and predicted microRNA complements of the intertidal lophotrochozoans *Patella vulgata* (Mollusca, Patelloga..., Mar. Genomics (2015), http://dx.doi.org/10.1016/j.margen.2015.07.004

Appendix 3

Kenny NJ, Namigai EKO, Dearden PK, Hui JHL, Grande C, Shimeld SM, 2014. The Lophotrochozoan TGF-β signalling cassette - diversification and conservation in a key signalling pathway. Int. J. Dev. Biol. 58, 533-49.

Int. J. Dev. Biol. 58: 533-549 (2014)
doi: 10.1387/ijdb.140080nk

THE INTERNATIONAL JOURNAL OF DEVELOPMENTAL BIOLOGY
www.intjdevbiol.com

The Lophotrochozoan TGF-β signalling cassette - diversification and conservation in a key signalling pathway

NATHAN J. KENNY[1,3], ERICA K.O. NAMIGAI[1], PETER K. DEARDEN[2], JEROME H.L. HUI[3], CRISTINA GRANDE[4] and SEBASTIAN M. SHIMELD*[,1]

[1]Evolution and Development Research Group, Department of Zoology, University of Oxford, UK, [2]Laboratory for Evolution and Development, Genetics Otago and Gravida, The National Centre for Growth and Development, Biochemistry Department, University of Otago, Aotearoa, New Zealand, [3]School of Life Sciences, Chinese University of Hong Kong, Shatin, Hong Kong and [4]Departamento de Biología Molecular and Centro de Biología Molecular "Severo Ochoa", CSIC-Universidad Autónoma de Madrid, Madrid, Spain

ABSTRACT TGF-β signalling plays a key role in the patterning of metazoan body plans and growth. It is widely regarded as a 'module' capable of co-option into novel functions. The TGF-β pathway arose in the Metazoan lineage, and while it is generally regarded as well conserved across evolutionary time, its components have been largely studied in the Ecdysozoa and Deuterostomia. The recent discovery of the *Nodal* molecule in molluscs has underlined the necessity of untangling this signalling network in lophotrochozoans in order to truly comprehend the evolution, conservation and diversification of this key pathway. Three novel genome resources, the mollusc *Patella vulgata,* annelid *Pomatoceros lamarcki* and rotifer *Brachionus plicatilis,* along with other publicly available data, were searched for the presence of TGF-β pathway genes. Bayesian and Maximum Likelihood analyses, along with some consideration of conserved domain structure, was used to confirm gene identity. Analysis revealed conservation of key components within the canonical pathway, allied with extensive diversification of TGF-β ligands and partial loss of genes encoding pathway inhibitors in some lophotrochozoan lineages. We fully describe the TGF-β signalling cassette of a range of lophotrochozoans, allowing firm inference to be drawn as to the ancestral state of this pathway in this Superphylum. The TGF-β signalling cascade's reputation as being highly conserved across the Metazoa is reinforced. Diversification within the activin-like complement, as well as potential wide loss of regulatory steps in some Phyla, hint at specific evolutionary implications for aspects of this cascade's functionality in this Superphylum.

KEY WORDS: *TGF-β, Lophotrochozoa, BMP, Activin, signalling*

Introduction

The TGF-β signalling pathway (Fig. 1A) has been well studied in a wide variety of traditional model systems, and is regarded as a 'module' (Wagner 1996) capable of regulating homeostasis, growth, and differentiation in a range of contexts (Derynck and Miyazono 2008, Moustakas and Heldin 2009, Massagué 2012). While the TGF-β pathway most likely arose in its canonical form in the metazoan common ancestor (Pang *et al.,* 2011), its divergence across the Metazoa and the ancestral roles played by its components are still in many ways unknown.

Attempts have been made to categorise the ancestral metazoan TGF-β cassette (Herpin *et al.,* 2004, Matus *et al.,* 2006, Adamska *et al.,* 2007, Huminiecki *et al.,* 2009). According to such analysis it is known that the original Metazoan repertoire consisted of at least four TGF-β receptors and four Smads (Suga *et al.,* 1999, Huminiecki *et al.,* 2009), although the extent of the original ligand cassette and regulatory repertoire remains uncatalogued. While

Abbreviations used in this paper: BAMBI, BMP and activin membrane-bound inhibitor; TGF, transforming growth factor; BMP, bone morphogenic protein; JTT, Jones-Taylor-Thornton; ML, maximum likelihood; R-Smads, receptor-regulated Smads; Tsg, twisted gastrulation; WAG, Whelan and Goldman.

***Address correspondence to:** Sebastian M. Shimeld. Evolution and Development Research Group, Department of Zoology, University of Oxford, South Parks Road, OX1 3PS, UK. Tel: +44-(0)-1865-281994. E-mail: sebastian.shimeld@zoo.ox.ac.uk - web: http://zoo-shimeld.zoo.ox.ac.uk/home

Supplementary Material (one table and one figure) for this paper is available at: http://dx.doi.org/10.1387/ijdb.140080nk

Accepted: 20 June 2014.

ISSN: Online 1696-3547, Print 0214-6282

fully-fledged TGF-β signalling components have not yet been found outside the Metazoa, the choanoflagellate *Monosiga brevicollis* has been shown to possess a gene with an MH2 domain similar to that of a Smad class protein (Srivastava *et al.,* 2010, Pang *et al.,* 2011). Both the placozoan *Trichoplax adhaerens* and the ctenophore *Mnemiopsis leidyi* have a fairly complete central signalling pathway complement (Humineicki *et al.,* 2009, Pang *et al.,* 2011), and sponges possess at least the basic receptor and Smad complements found in the Bilateria (Suga *et al.,* 1999). The evolution of some elements of the cassette is less well understood, particularly those components that act to modulate signalling.

The TGF-β signalling pathway has been well studied in traditional ecdysozoan (van der Zee *et al.,* 2008) and deuterostome (Massagué *et al.,* 2000) model systems such as *Drosophila melanogaster* and *Mus musculus*. The few attempts that have been made at categorizing elements of the lophotrochozoan TGF-β cassette have typically been limited to individual elements and/or single species (for example, Herpin *et al.,* 2005, Freitas *et al.,* 2007, Kuo and Weisblat 2011). Discoveries such as that of Nodal in the Mollusca (Grande and Patel 2009) have led to further interest in the true pattern of conservation and diversification of genes in this pathway. With the genomic resources now on-hand we should be able to trace its ancestral form and function, at least across the Bilateria.

In essence while the TGF-β pathway regulates a number of highly complex processes in animal tissues its core mode of action is simple, and can be seen schematically in Fig. 1A. TGF-β ligands form dimers and bind sequentially to Type II and Type I receptors, which form a complex and are phosphorylated. Upon activation of the Type I receptor within the signalling complex, receptor-regulated Smads (R-Smads) are recruited from the cell membrane with the aid of proteins such as Smad anchor for receptor activation (SARA) (Itoh and ten Dijke 2007). R-Smads are then phosphorylated and activated by the receptor complex (Massagué *et al.,* 2005). R-Smads can be divided into two families, depending on the ligands to which they respond– Mad/Smad 1/5/8 responds to BMP signalling, and Smox/Smad 2/3 to TGF-β, Activin, and Nodal signalling (Heldin and Moustakas 2012). Once activated, R-Smads bind to a co-Smad (Medea/Smad 4) to form a complex that mediates transcription in the nucleus, resulting in up- and down-regulation of target genes (Ross and Hill 2008). Inhibitory Smads (known as Dad or Smad 6/7), compete with R-Smads for activation by receptor complexes, thus regulating the pathway (Ross and Hill 2008).

Complexity is introduced to the TGF-β signalling cascade by the diversity of regulatory mechanisms which modulate it both extra- and intra-cellularly, and which are perhaps more free to vary than the core signalling cascade. TGF-β ligands can be removed from the extracellular environment by ligand traps, such as the Chordins, Noggins and DANs (Balemans and Van Hul 2002). These are vital for many aspects of development, including the correct specification of dorsoventral polarity, and have been catalogued in ecdysozoans and deuterostomes (Holley *et al.,* 1995). It has been observed that ecdysozoans have less diversity in these protein classes than vertebrate models (van der Zee *et al.,* 2008). Tolloid, a zinc metalloprotease, is capable of cleaving Chordin, hence re-releasing the trapped ligand, as well as cleaving other potential repressors of TGF-β signalling, such as proteoglycans (Scott *et al.,* 1999). This plays a key role in establishing the body plan of early embryos, including those of lophotrochozoans (Herpin *et al.,* 2007). TGF-β ligand binding to receptor serine/threonine kinases can also be up- or down - regulated at the cell surface by membrane anchored co-receptors and receptors (Shi and Massagué 2003). Some co-receptors, such as Cripto and the EGF-CFC class of genes, allow active ligand-receptor complexes to be formed by acting as cofactors and are vital for the function of some ligands (Cheng *et al.,* 2003, Shen and Schier 2000). Down-regulation can be performed by pseudoreceptors such as BMP and membrane bound inhibitor (BAMBI, also known as Nma), which compete with functional Type I receptors for ligand binding (Onichtchouk *et al.,* 1998). The existence of these regulatory mechanisms has been noted in protostomes previously (van der Zee *et al.,* 2008), but the degree to which these are conserved across the Bilateria is unknown, and the possibility that these regulatory mechanisms have diversified, changed in function or have been lost in some lineages is yet to be explored.

Further intracellular regulation of TGF-β signalling also occurs. FKBP12, Dad/SMAD7 recruited E3 ubiquitin ligases and SMAD ubiquitination regulatory factors (SMURFs) can up- and down-regulate signalling within the cell (Shi and Massagué 2003, Itoh and ten Dijke 2007). These, and other regulatory mechanisms, often participate in other signalling cascades. The full repertoire of regulatory interactions with the TGF-β cascade is beyond the scope of this manuscript, and we refer the interested reader to the detailed reviews available on this topic (e.g. Shi and Massagué 2003, Moustakas and Heldin 2009, Al-Salihi *et al.,* 2012).

For all its importance and ubiquity, the TGF-β pathway within the Lophotrochozoa has yet to be satisfactorily documented. Some studies have found evidence of diversification at the ligand level within the leech *Helobdella robusta* (Kuo and Weisblat 2011) and platyhelminthes (Gavino and Reddien 2011, Freitas *et al.,* 2007), while other aspects of the signalling pathway may be conserved (Molina *et al.,* 2011) although this is unclear (Kuo and Weisblat 2011). The identification of the full signalling complements of a number of lophotrochozoan species should provide a springboard for the disentanglement of this network.

Here we present a comprehensive catalogue of the components of TGF-β signalling in members of the lophotrochozoan Superphylum (Fig. 1B), allowing for the first time a Metazoa-wide understanding of the evolution and divergence of this crucial signalling pathway. We investigated the genomes of the mollusc *Lottia gigantea*, the annelid *Capitella teleta* (Simakov *et al.,* 2013), the bdelloid rotifer *Adenita vaga* (Flot *et al.,* 2013) and the planarian *Schistosoma mansoni* (Berriman *et al.,* 2009), along with transcriptomic and novel genomic resources for the monogont rotifer *Brachionus plicatilis*, limpet mollusc *Patella vulgata* and the serpulid annelid *Pomatoceros lamarcki.* We demonstrate the existence of the majority of the core TGF-β cassette in the Lophotrochozoa, confirming the conservation of this 'module' across evolutionary time, albeit with extensive diversification of the ligand complement in this lineage and loss of genes encoding extracellular inhibitors in some species.

Results

TGF-β ligands

TGF-β ligands participate in a diverse range of mechanisms in cellular specification and functionality. They are synthesized as relatively long precursor proteins, but an N-terminal propeptide

is cleaved during processing, leaving a short, 110-140 amino acid mature ligand. The mature ligand can be recognized by the characteristic conservation of at least six cysteine residues that when folded form a structure known as a cysteine knot, stabilized by three disulfide bonds (ten Dijke and Arthur 2007).

There are 33 distinct genes encoding TGF-β ligands in humans, seven in *D. melanogaster* and five in *Caenorhabditis elegans* (Humineicki *et al.,* 2009). These ligands can be split into two broad classes, the TGF-β/Activin/Myostatin class (generally referred to here as the TGF-β class) and the BMP (bone morphogenetic protein) class (Yamamoto and Oelgeschläger 2004). Generally these classes are mirrored by the signalling pathway through which they operate (see Fig. 1A for details) but this is not always the case - *Nodal*, for instance, is generally said to belong to the BMP class, but Nodal signals are transduced via the TGF-β pathway. We should note that nomenclature regarding whether these are 'families' or 'classes' varies in the literature. In this study we have followed the Linnaean convention, where classes are broader groupings – either TGF-β or BMP-related, while families refer to groupings of orthologous genes within the class set.

In many ways the complements of TGF-β ligands found in lophotrochozoan genomes are similar to those found in more classical model organisms. This similarity breaks down, however, in the TGF-β class, which appears to have undergone extensive divergence in this Superphylum. Initial attempts at making phylogenetic trees for TGF-β ligands were confounded by the diversity of ligand sequence in this class, which resulted in poor alignments and multiple gaps, and, ultimately, poorly-supported trees. Supplementary Fig.1 shows one such tree (Maximum Likelihood (ML), Whelan and Goldman (WAG) model, 1000 bootstrap replicates). To better ascertain phylogenetic relationships within and between TGF-β ligands, we have analysed the interrelationships of the TGF-β class in two steps - firstly, by assigning, on the basis of the preliminary tree shown in Supplementary Fig. 1, and by Blast identity (Altshul *et al.,* 1990), ligands to either the BMP class or the Activin/TGF-β class, and secondly by analysing these two classes separately.

The results of maximum likelihood and Bayesian inference of their phylogenetic relationships for the BMP class can be seen in Fig. 2. Both means of phylogenetic inference recover clear familial relationships between lophotrochozoan sequences and orthologues of many well-described genes. Bootstrap values are not always high, most likely due to the relatively short dataset (88 amino acid alignment) from which these samples were drawn. Posterior probabilities, however, clearly support many nodes on the Bayesian tree; for example, the ADMP clade has a posterior probability of 1 under Bayesian analysis, but bootstrap support of 71 under ML analysis.

The existence and expression of several members of the BMP class of the TGF-β ligand superclass in the Lophotrochozoa has already been established by prior studies, although the phylogenetic distribution of these studies has been scattered and the full lophotrochozoan complement was unclear (e.g. Nederbragt *et al.,* 2002, Freitas *et al.,* 2007, Grande and Patel 2009, Kuo and Weisblat 2011). It seems that annelids and molluscs have retained the majority of the diversity of the BMP class found in the Ecdysozoa and Deuterostomia, and in many cases better conservation is found than in ecdysozoan models. In contrast, *B. plicatilis* and,

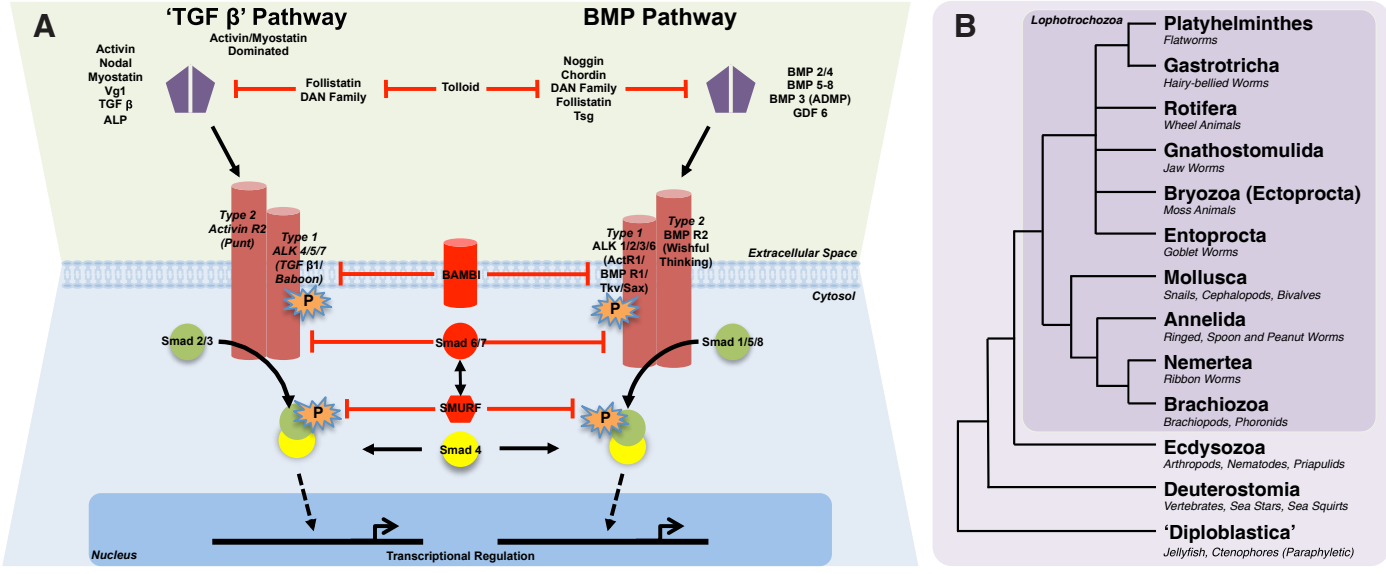


Fig. 1. Summary of TGF-β/BMP signalling cascades and Lophotrochozoan interrelationships. (A) *Canonical TGF-β/BMP signalling cascades: representation of the canonical signalling pathways for TGF-β-like and BMP-like cascades with inhibitors of signalling shown in red and operational signalling shown in black. Only ligands with well-known affinity to one or other signalling pathway listed, with each pathway operating through different combinations of Type I and Type II receptors, and hence signalling through either Smad 2/3 or Smad 1/5/8 proteins intracellularly. Ligands are regulated extracellularly by a diverse range of inhibitors, which can themselves be cleaved by Tolloid to release ligands and allow signalling to occur. Intracellular regulation of signalling can occur at the receptor level, with BMP and activin membrane-bound inhibitor (BAMBI) recruited in the place of functional Type I receptors, or intracellularly, through Smad 6/7 inhibition of signal transduction from receptors, SMURF-mediated degradation of Smad signalling proteins, or a range of further mechanisms not shown here.* **(B)** *Cladogram representing lophotrochozoan (boxed) and metazoan inter-relationships, based on Dunn* et al., *(2008). Non-bilaterally symmetrical metazoans are represented by the paraphyletic group "Diploblastica".*

in particular, *S. mansoni,* have apparently lost many of the more BMP-like members of the TGF-β ligand cassette. The criteria used as a basis for assignations of genes to particular families can be seen in the Materials and Methods section.

The apparent Ecdysozoan-specific loss of Nodal and BMP3/GDF10 is corroborated by our analysis, but we recover well-supported nodes containing lophotrochozoan orthologues for these gene families. For more detailed exploration of Nodal ligands, we refer the interested reader to Grande *et al.,* 2014 in this issue. More complex is the case of the Univin/VG-1/GDF1/3-like family. While these are not recovered as a monophyletic grouping in Fig. 2, they have appeared in previous studies as a well-supported clade of deuterostome-specific genes, for example in Lapraz *et al.,* (2006). The *H. robusta* BMP 2/4b sequence (AEL12442.1) has also been noted as showing some resemblance to the Univin/VG-1/GDF1/3-like family. An annelid gene, which we have named "UNKNOWN" in Fig. 2, groups with poor support with the GDF1/3 family in Bayesian trees. Better sampling is needed across the Lophotrochozoa in order to confirm whether these are truly orthologous to the Vg1/Univin/GDF1/3 genes in the Deuterostomia.

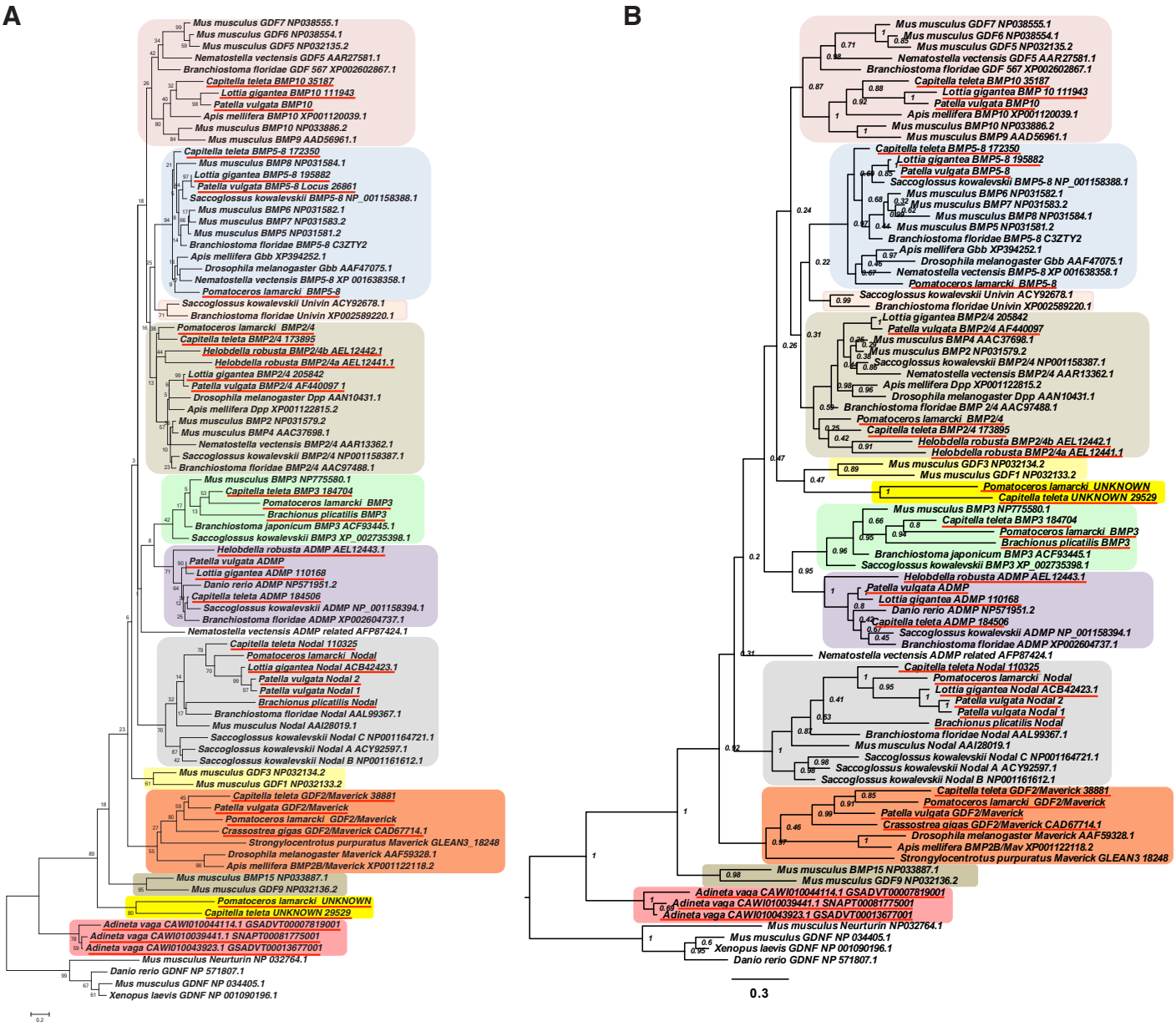


Fig. 2. Phylogeny of the BMP-like ligand class familial interrelationships across the Metazoa, as determined by (A) maximum likelihood (Tamura *et al.,* 2011) and (B) Bayesian (Huelsenbeck and Ronquist 2001) methods. *Alignment generated by MAFFT (Katoh and Standley 2013) using the L-INS-i strategy resulting in an 88 amino acid informative alignment of the mature ligand domains after the exclusion of gaps. Both phylogenies determined using the WAG model (ML: +4G) (Whelan and Goldman 2001). Bootstrap percentage (of 1,000 replicates) and posterior probabilities (after 3,000,000 generations) can be seen at the nodes of ML and Bayesian trees respectively. Lophotrochozoan sequences underlined in red. Phylogenies rooted with known Neuturin and GDNF outgroups. Sequences used in phylogenetic analysis, along with alignment, can be found in Supplementary File 1. Scale bars represent substitutions per site at given distances.*

Canonical homologues of BMP 2/4 (Dpp), BMP 5-8 (Gbb/Scw), BMP 9/10 (GDF5-7), ADMP and Maverick are also found in the Lophotrochozoa as reported by our trees (Fig. 2). These appear to be far better conserved in the Mollusca and Annelida than in other lophotrochozoan lineages examined. No homologues for GDF9/BMP15 can be found in our lophotrochozoan datasets, implying that this is a deuterostome or even chordate innovation (as suggested by the lack of such a ligand in Lapraz *et al.,* 2006).

Fig. 3 shows the inferred identity of the genes of the activin/myostatin/inhibin-like clade as determined by phylogenetic analysis. Clear and reproducible signals were found for a canonical Myostatin clade, especially in the case of Bayesian analysis, where Myostatins cluster together with a posterior probability of 1. *P. lamarcki* appears to have duplicated this gene, but this is not found in other annelids, and is likely lineage-specific. Support for an Activin/Inhibin clade is weak in both Bayesian and Maximum Likelihood phylogenies shown. Clades for TGF-β and Lefty homologues in deuterostomes are consistently recovered with good support. On the basis of the tree presented here, Lefty and TGF-β ligands *sensu stricto* seem to be deuterostome innovations, with the previous report of a tentative TGF-β homologue in *M. leidyi* (Pang *et al.,* 2011) standing as evidence against this. We have included this sequence in the tree shown in Fig. 3, and, as stated in Pang *et al.,* (2011), it is only weakly supported as a homologue of the TGF-β clade *sensu stricto*. Wider taxon sampling at the base of the Metazoa, and particularly in the Ctenophora beyond *M. leidyi*, would allow us to test whether TGF-β ligands (in the strictest sense) are indeed a deuterostome novelty.

Our investigations of the genome of *A. vaga* revealed a total of 10 ligands with gene models or transcript support for their existence. Five further genes were present in their entirety in the genome, with some gene models supporting their existence, but without evidence of transcription. As such, these were not included in the curated gene list appearing in that genome's publication (Flot *et al.,* 2013). However, as transcripts could be temporally restricted in appearance, or present in very low levels, we have included these sequences in Supplementary File 1 for consideration by interested parties. In particular, we note the existence of three complete *BMP 5-8* genes with no evidence of pseudogenisation.

The large number of ligands seen in *A. vaga* is the result of

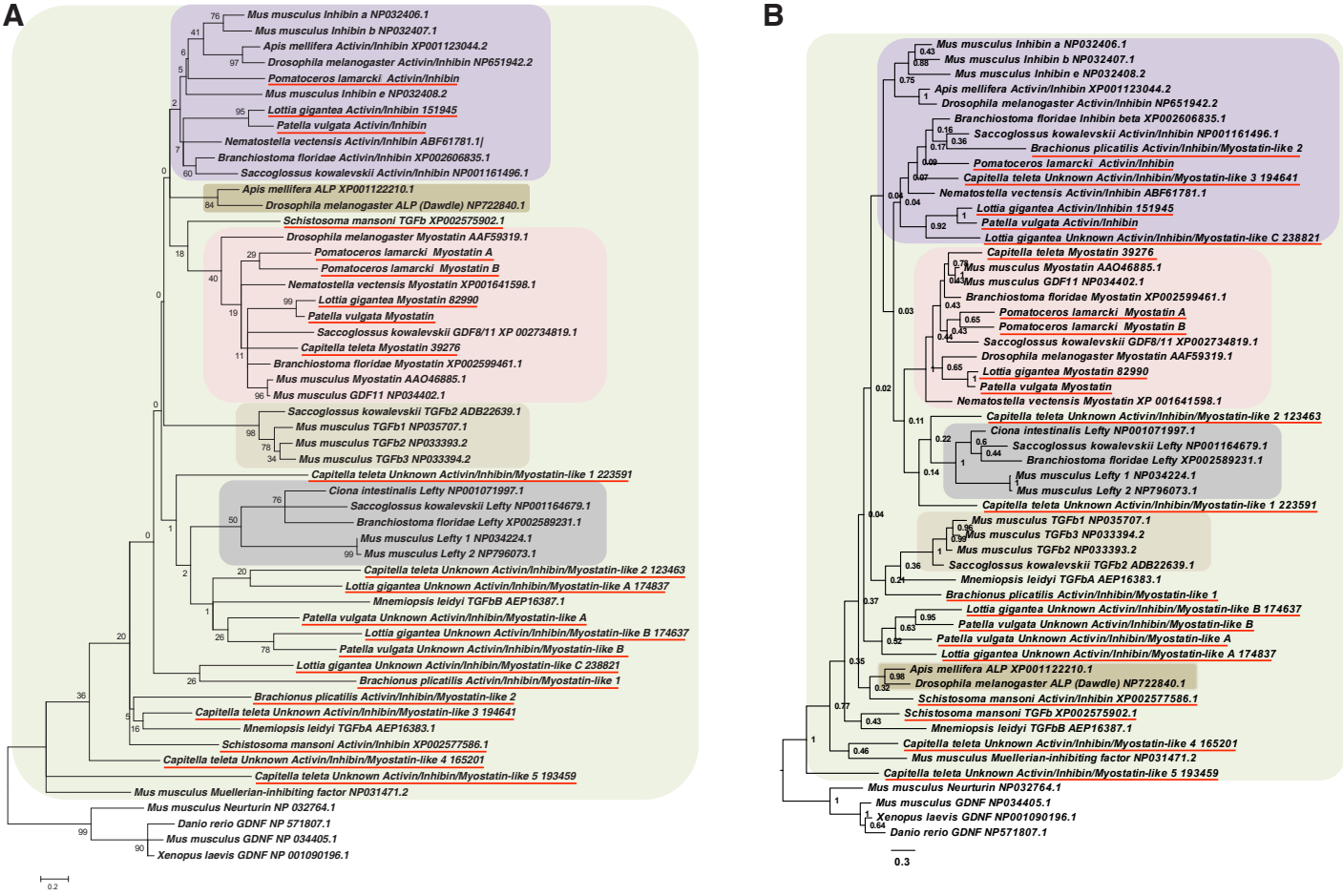


Fig. 3. TGF-β-like class familial phylogenetic interrelationships across the Metazoa, as determined by (A) maximum likelihood (Tamura *et al.,* 2011) and (B) Bayesian (Huelsenbeck and Ronquist 2001) methods. *Alignment generated by MAFFT (Katoh and Standley 2013) using the G-iNS-i strategy, resulting in a final 71 amino acid informative alignment. Both phylogenies determined using the WAG mode (Whelan and Goldman 2001). Bootstrap percentage (of 1,000 replicates) and posterior probabilities (after 10,000,000 generations before convergence) can be seen at the nodes of ML and Bayesian trees respectively. Sequences used in phylogenetic analysis, along with alignment, can be found in Supplementary File 1. Phylogenies rooted with known Neuturin and GDNF outgroups. Lophotrochozoan sequences underlined in red. Scale bars represent substitutions per site at given distances.*

two known rounds of lineage-specific whole genome duplication in that species (Flot *et al.,* 2013), in contrast to the unknown origin of gene duplications in other Lophotrochozoan clades. *A. vaga,* like *B. plicatilis,* possesses BMP 3, Nodal and Activin/Inhibin sequences. We have used *B. plicatilis* sequences to represent the Rotifera in our trees as its relatively slow rate of molecular evolution aided the resolution of other nodes in this figure, but have included these three closely-related *A. vaga* genes to show evidence that they cannot be categorised into known clades. We find the absence of BMP 2/4 (Dpp) homologues in the Rotifera as a whole particularly interesting, as these genes play a key role in establishing dorsoventral polarity in a phylogenetically broad range of species, in concert with Chordin.

It should also be noted that as well as these rotifer sequences, several other sequences are not shown in our analysis in Fig. 3, but are provided in Supplementary File 1. The first of these sequences is a *P. vulgata* Myostatin-like homologue, whose sequence was partially recovered from our genomic and transcriptomic data, but which covers only a portion of the mature ligand sequence and was therefore excluded from analysis. The second is a *C. teleta* gene, protein ID 198732, which appears to be a markedly divergent TGF-β class ligand. Its annotation in the *C. teleta* genome suggests that it has been noted as expressed in the EST studies used as the basis for gene prediction, but it appears to have lost a significant portion of the mature ligand region. This may be the result of pseudogenisation in progress, as without the portions of the ligand domains, it is unlikely the protein it encodes is fully functional in the same manner as canonical TGF-β class ligands. The oyster *C. gigas* also shows evidence of diversification in TGF-β class ligands (data not shown here, see Fleury *et al.,* 2008), although when these sequences are added to our analysis no further structure was added to our tree - these seem to be fast-evolving, highly derived sequences, with uncertain homology to the other lophotrochozoan data presented in this paper.

ALP seems to be an insect innovation, forming a clade with clear support and no orthologue seen in any lophotrochozoan or deuterostome species. Muellerian inhibiting factor has been suggested to be a deuterostome or even vertebrate-specific ligand, but weak support groups the *C. teleta* Unknown Activin/Inhibin/Myostatin-like 4 with this sequence in the Bayesian analysis. Whether this is a true relationship, or instead is a result of long-branch attraction is uncertain, but could be tested with increased genomic sequencing across the Lophotrochozoa.

As with the BMP-like ligands (shown in Fig. 2), TGF-β-like ligands seem to have been lost from both rotifer species considered and *S. mansoni,* with those that are found not falling into identifiable clades. This echoes the findings of previous studies in schistosomes (Freitas *et al.,* 2007). It has been suggested that *S. mansoni* could use host ligands as part of its signalling cassette (Osman *et al.,* 2006), which would explain low diversity of these ligands in this species. The lack of rotifer TGF-β ligands is more unusual, and future sequencing efforts in the Rotifera will reveal whether this loss is real, or an artefact of insufficient sequencing depth. It should be noted that the rotifer receptor complement is also slightly modified, and could reflect changes to ligand sequence and structure in concert with downstream aspects of this cascade.

Some support is found for a mollusc-specific clade of ligands, encompassing *L. gigantea* Unknown Activin/Inhibin/Myostatin-like A and B and similarly named sequences in *P. vulgata*. The roles of these genes *in vivo* are as yet uncatalogued, and no homologue is found in the genome of *C. gigas*. These sequences therefore could represent a gastropod or patellogastropod novelty. In some prior analyses (data not shown), platyhelminth sequences form a sister group to this clade with poor (< 20 bootstrap support), but further evidence from other lophotrochozoans is needed to determine whether this is a genuine relationship.

Outside of these clades, little signal can be recovered to support relationships between a diverse range of other ligands in the Lophotrochozoa. On the basis of the lack of clades forming, even between such relatively closely related species as *C. teleta* and *P. lamarcki*, and especially between *P. vulgata* and *L. gigantea*, it seems that these ligand sequences are highly variable between lophotrochozoan species. These sequences are named without reference to their evolutionary relationships, and the terms 'A, B, C' etc in multiple species are used to allow within-species numeration rather than any inference of orthology.

Some additional insight could potentially be drawn from the number of cysteine residues found in these sequences, as these residues have a characteristic distribution in some model organisms. In some vertebrates, TGF-β ligands *sensu stricto* and Inhibin β are said to have nine cysteines, while Inhibin α (along with BMPs and GDFs in the BMP-like clade) have seven. These pair to form four and three disulfide bonds respectively. The remaining cysteine residue forms such a bond only when the ligands dimerise to signal. Lefty proteins, as well as GDFs 3, 9 and BMP 15 are said to have only six cysteines - they do not bond covalently to form dimers (Derynck and Miyazono 2008, Moustakas and Heldin 2009). By our count, in the *M. musculus* and deuterostome sequences used in our analysis (Supplementary File 1), Lefty and Muellerian-inhibiting factor proteins possess seven cysteines (lacking the fifth and second respectively when counting from the N terminus of the mature ligand), while all other deuterostome ligands in our dataset possess at least eight cysteine residues. This may represent the ancestral condition, with the more derived form studied in detail by those papers referenced above. These cysteine positions can be clearly seen in the BMP alignment file in Supplementary File 1, in alignment positions 1, 2, 28. 32, 57, 58, 85, and 87 when present. For the TGF-β class alignment, these are positions 1, 2, 27, 31, 42, 43, 69, and 71.

The majority of our Lophotrochozoan ligand sequences possess eight cysteine residues. Of the uncategorised sequences seen in Fig. 3, the four *L. gigantea* and *P. vulgata* sequences mentioned as forming a weakly supported clade earlier have 7 cysteines (lacking the fifth as counted from the N terminus), as do *C. teleta* Unknown Activin/Inhibin/Myostatin-like 4 and *C. teleta* Unknown Activin/Inhibin/Myostatin-like 1, which lack the second from the N terminus. One sequence, *C. teleta* Unknown Activin/Inhibin/Myostatin-like 5, has only six cysteines, lacking both the second and the fifth. When cysteines are lost from ligands in vertebrates, they are also lost from the second and/or fifth position, which further reinforces that some cysteines are vital for maintaining the "cysteine knot" form of the active ligand, while others can be lost.

Unfortunately, given the uncertainty regarding the number of cysteines found in canonical groups, this character cannot be used to further classify our ligands. A better understanding of the interrelationships between these ligands is probably only to be drawn from denser taxon sampling across the Lophotrochozoa. At present, the short length of the mature ligand sequence and the

lack of conservation of sequence in the longer propeptide means that phylogenetic inference, by whatever means, is limited by a lack of information.

The lack of constraint on the non-cysteine portions of the sequences of these ligands is also interesting from a structural perspective. The sequence of BMP-like ligands seems highly constrained, probably because of their vital interactions with receptors and regulators of their activity. That TGF-β class ligands in the Lophotrochozoan clade can diversify so much, even between closely related species, when the remainder of the core signalling cascade remains relatively stable raises questions about how these ligands can successfully maintain their secondary and tertiary structures in the face of relatively large sequence changes. This is especially puzzling when so many other TGF-β class ligands have, presumably under purifying selection, maintained a relatively stable sequence over evolutionary time across the Metazoa.

To summarise, of the TGF-β superclass, TGF-β-like ligands appear to have diversified greatly in lophotrochozoans, at least in the lineage leading to molluscs and annelids (Fig. 1B), while the BMP class ligand complement is similar to that seen in more traditional model organisms. In contrast, loss in ligand complements can be seen in the Rotifera and Platyhelminthes. *TGF-β sensu stricto* and *Lefty* genes appear to be deuterostome novelties, as no evidence can be found for their presence in the genomes here examined. How these changes in ligand diversity have affected lophotrochozoan biology, and the degree to which known ligand families possess ancestrally shared roles mirroring those performed in other Superphyla, will be a topic of broad interest for developmental biologists in the future.

Serine/threonine kinase receptors

Each TGF-β ligand pair binds to Type I and II serine/threonine kinase receptors. When TGF-β ligands form a complex with representatives of both types of receptor, the intracellular kinase domains of the receptors are brought together and the Type I receptor is phosphorylated and activated (Massagué 1998). In humans, a total of seven Type I and five Type II receptors have been described, while *Tribolium castaneum, Apis mellifera* and *D. melanogaster* possess a total of five – three Type I, and 2 Type II (van der Zee *et al.,* 2008).

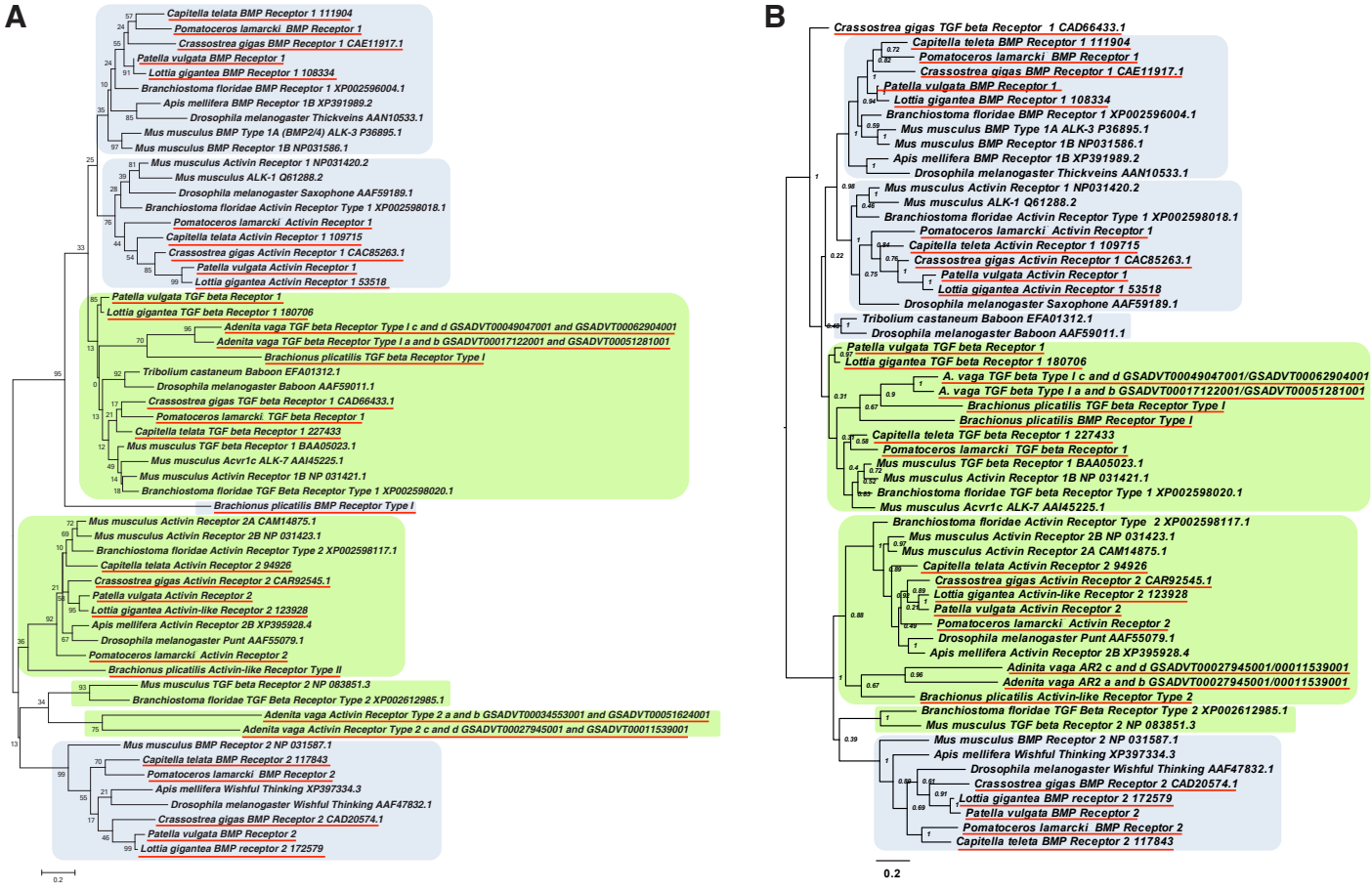


Fig. 4. TGF-β and BMP receptor molecule interrelationships across the Metazoa, as determined by (A) maximum likelihood (Tamura *et al.,* 2011) and (B) Bayesian (Huelsenbeck and Ronquist 2001) methods, and rooted at the midpoint. *Predominantly TGF-β-like and BMP-like cascade receptors shown in green and blue respectively. Alignment generated by MAFFT (Katoh and Standley 2013) using the G-iNS-i strategy resulting in a 136 informative amino acid alignment spanning the protein kinase domain (PFAM PF00069). Both phylogenies determined using the WAG model (ML +4G) (Whelan and Goldman 2001). Bootstrap percentage (of 1,000 replicates) and posterior probabilities can be seen at the nodes of ML and Bayesian trees respectively. Lophotrochozoan sequences underlined in red. Sequences used in phylogenetic analysis, along with alignment, can be found in Supplementary File 1. Scale bars represent substitutions per site at given distances.*

The serine/threonine kinase receptor complements of the Lophotrochozoa have previously been the subject of investigation, with those of *C. gigas* already described (Herpin 2005). Coupled with our extensive knowledge of the diversity of these molecules in the freshwater sponge *Ephydatia fluviatilis* (Suga *et al.,* 1999) and other non-bilaterian metazoans (Humineicki *et al.,* 2009, Pang *et al.,* 2011) these are perhaps the best-catalogued components of the TGF-β cascade. With some exceptions, the serine/threonine kinase receptor complement varies little across the Metazoa. Generally three Type I and two Type II receptors are found in any species, with TGF-β-like signalling occurring through a set pair of dimerised Type I and Type II receptors (TGF-β R1 and Act R2, also known by a diverse range of other names), while two Type I receptors (BMP R1 and the misleadingly named Act R1) can each be found in complex with BMP R2.

In vertebrates, considerably more diversity of receptor number exists, most likely due to the 2R whole genome duplications. Larger numbers of receptor also exist in invertebrate deuterostomes and the cnidarian *Nematostella vectensis*, most likely due to independent duplications in these lineages, some of which have been traced to specific nodes on the deuterostome tree of life (Humineicki *et al.,* 2009).

While the canonical five serine/threonine kinase receptors are found in all annelid and mollusc species examined (Fig. 4), these have diversified in the rotifers. Both *A. vaga* and *B. plicatilis* possess at least one Activin R2 and TGF-β R1 gene, although the four *A. vaga* Activin R2 genes are drawn toward the base of our ML tree by long branch attraction. One *B. plicatilis* sequence appears to be a divergent BMP R1 by Blast identity and in some of our ML trees (data not shown), although this gene is grouped with *A. vaga* and *B. plicatilis TGF-β R1* genes in the Bayesian tree shown in Fig. 4. This gene could therefore represent a diverged copy of a rotifer *TGF-β R1* gene after duplication in that lineage, or alternately rotifer receptors could be becoming more similar through gene conversion. The divergent nature of the TGF-β ligands found in these species may suggest that their receptor sequences are evolving to signal in a derived fashion, although the true causes of these changes requires further research.

Schistosome receptor complements have been studied elsewhere in depth (Davies *et al.,* 1998, Forrester *et al.,* 2004), and in many ways their complements represent a surprising finding, given the presence in *S. mansoni* of only two TGF-β-like ligands. These complements do not map exactly onto the canonical cassette, but, as hypothesised in Osman *et al.,* (2006) and earlier in this manuscript, their quantity, when compared to the few ligands encoded in its genome, may suggest that these molecules respond to host, rather than endogenous, signalling cues.

In short, the relative conservation of serine/threonine kinase receptor sequences within annelids and molluscs confirms the suggestion of Herpin *et al.,* (2005) with regard to the broad conservation of serine/threonine kinase receptor diversity across metazoan evolution.

Smad proteins

Smad proteins play a key role in transducing extracellular signals into an intracellular response. Of all the parts of the TGF-β signalling cascade investigated in the present study, it is these molecules that show the least amount of loss and disparity in number across

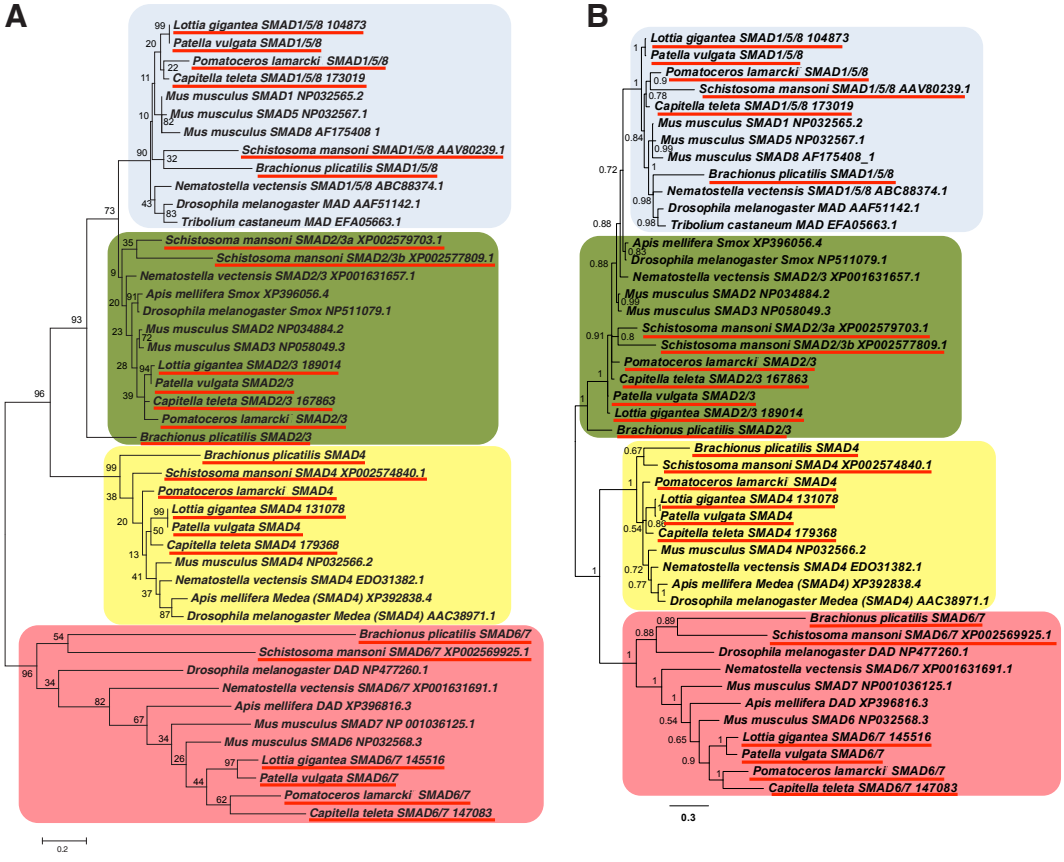


Fig. 5. Smad and Dad interrelationships across the Metazoa, as determined by (A) maximum likelihood (Tamura *et al.,* 2011) and (B) Bayesian (Huelsenbeck and Ronquist 2001) methods. *Alignment generated by MAFFT (Katoh and Standley 2013) using the G-iNS-i strategy, with the section used for analysis a 139 informative amino acid region spanning the MH2 domain (Pfam PF03166). Both phylogenies determined using the WAG mode (Whelan and Goldman 2001), rooted with the known Smad 6/7 clade (ML) and at midpoint (Bayesian). Bootstrap percentage (of 1,000 replicates) and posterior probabilities can be seen at the nodes of ML and Bayesian trees respectively. Lophotrochozoan sequences underlined in red. Sequences used in phylogenetic analysis, along with alignment, can be found in Supplementary File 1. Scale bars represent substitutions per site at given distances.*

the Metazoa, presumably due to the pleiotropic effects that would happen if loss were to occur. Smad molecules respond to a diverse range of incoming signals, as can be seen in Fig. 1A, and even if loss of one ligand occurs in a lineage, a Smad may still be responsible for passing on the message brought by another ligand. Such pleiotropy means that loss may be less tolerated by natural selection at this step of the signalling cascade than others.

All lophotrochozoan Phyla studied in the present investigation were found to contain at least one of each of the four major families of Smad molecule (Fig. 5). Interestingly, the rotifer *A. vaga* (but not *B. plicatilis*) has lost Smad 6/7 (Dad) and thus is unlikely to inhibit Smad signalling using this mechanism. The short branch lengths generally found outside the inhibitory Smad (Smad6/7 or Dad) clade also point to generally constrained selection on these molecules. Both *B. plicatilis* and *S. mansoni* show longer branch lengths for these sequences, however, which may be a result of co-evolution to interact with the divergent receptor cassettes also seen in these Phyla.

S. mansoni shows evidence of a lineage-specific Smad2/3 duplication, which is perhaps surprising in light of the reduced ligand complement of this species, and may represent a subfunctionalisation of roles previously performed by a single ancestral gene. We note that we do not recover a monophyletic clade of Smad 2/3 sequences in our Bayesian analysis, which is largely due to the small differences in sequence between the R-Smad clades.

Dan/Cerl/Coco/Prdc/Cerberus/Gremlin

Members of the wider DAN-like gene class sequester ligands, preventing them from binding to receptors and activating signalling cascades. This class has diverged into a large and confusingly named clade of genes, especially in vertebrates, where the 2R whole genome duplication event likely allowed sub- and neo-functionalisation to occur. The results of phylogenetic analyses of members of this gene class from species across the Metazoa (Fig. 6) reveal how this difficult-to-catalogue group has evolved.

It appears that the cassette of DAN-class members found in the common ancestor of deuterostomes and protostomes may have resembled that of *N. vectensis,* with two homologues giving rise to the diversity we see today. It is possible that the Cerberus/Dante family represents a deuterostome innovation - despite the placement of *N. vectensis* Cerberus/Dante-like (ABF06563.1) at the base of this clade. Given the widespread presence of Gremlins across the Metazoa, and previous study in this group in Cnidarians (Rentzsch *et al.,* 2006), it perhaps would be parsimonious to infer that it is in fact a Gremlin, rather than inferring loss of a cnidarian Gremlin and protostome Cerberus/Dante-like factors. We cannot distinguish between these alternatives definitively with the data available, but this hypothesis could be easily tested with the advent of broader sequence availability.

No wider DAN class genes can be found in schistosomes or in the rotifers *A. vaga* or *B. plicatilis*, and loss of portions of this class are prevalent in other species - *D. melanogaster,* for example, has no DAN class genes in its genome (van der Zee *et al.,* 2008), and DANs *sensu stricto* have been lost across the Lophotrochozoa (Table 1). DANs sequester a variety of ligands, with specificity varying depending on the DAN protein examined. Some are specific to the BMP-like signalling pathway, while others (such as Cerberus) can inhibit Nodal in the TGF-like pathway. Gremlin is

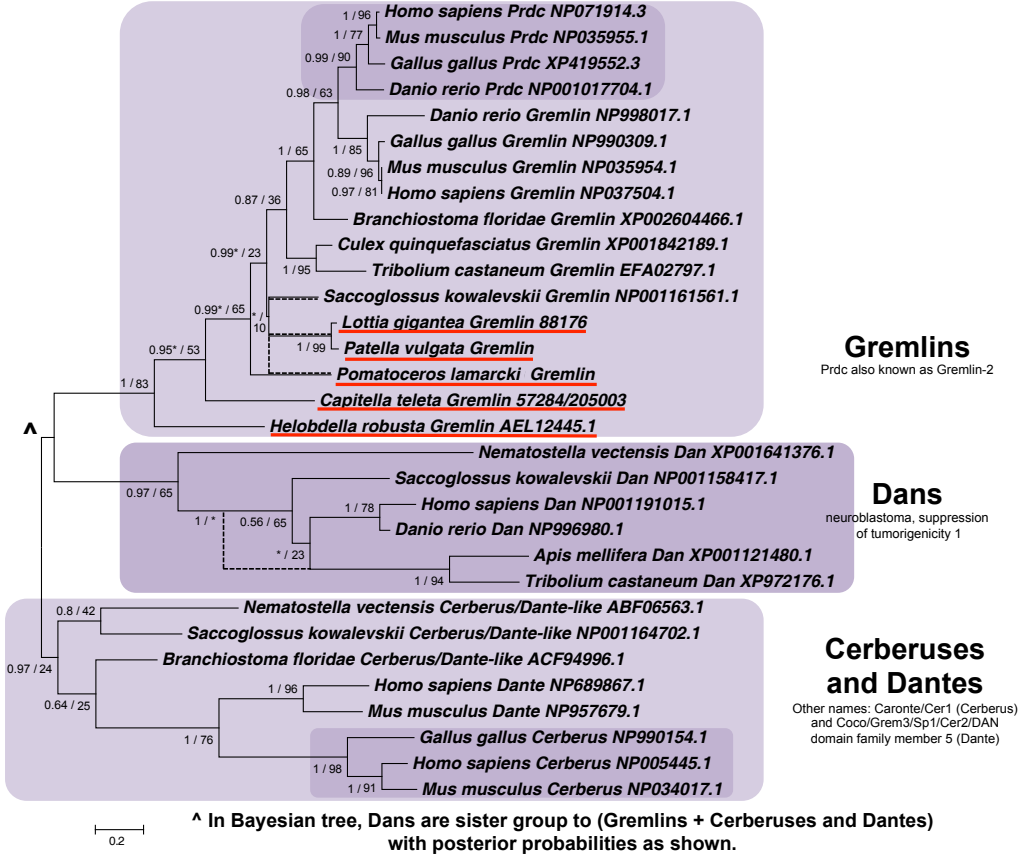


Fig. 6. DAN class interrelationships across the Metazoa, as determined by maximum likelihood (Tamura *et al.,* 2011) and Bayesian (Huelsenbeck and Ronquist 2001) methods. *Alignment generated by MAFFT (Katoh and Standley 2013) using the G-iNS-i strategy. Phylogenies calculated on the basis of an 87 informative amino acid alignment spanning the DAN domain (Pfam ID PF03045). Phylogeny shown is the result of ML analysis, with differences in topology using Bayesian methods indicated with a dotted line. Both phylogenies determined using the WAG mode (Whelan and Goldman 2001) and rooted at midpoint. Posterior probabilities/bootstrap percentage (of 1,000 replicates) and can be seen at the base of nodes. Lophotrochozoan sequences underlined in red. Sequences used in phylogenetic analysis, along with alignment, can be found in Supplementary File 1. Scale bars represent substitutions per site at given distances.*

known to be active in *H. robusta*, targeting BMP 2/4 preferentially (Kuo and Weisblat 2011). Much investigation is required before inference can be made as to the wider roles of genes that remain in the Lophotrochozoa, given the wide range of actions that these genes are known to have in more established model organisms.

Chordin

Chordin is a BMP regulatory molecule, which sequesters ligands by binding to them. It is best known for its role antagonising BMP 2/4 (also known as Dpp) in dorsoventral patterning. A similar molecule, Chordin-like, has also been identified, initially as a BMP 4 antagonist in the chick (Sakuta *et al.,* 2001).

Searches through the lophotrochozoan genomes examined in the present project suggest that both *Chordin* and *Chordin-like* genes are present within the Lophotrochozoa, as can be seen in Fig. 7A. Chordin itself seems to be lost across the Annelida. No Chordin or Chordin-like sequences were found in the rotifers *A. vaga* and *B. plicatilis* or in the Platyhelminthes, although other more derived families related to Chordin (eg CRIM) were not examined in the above analysis. The presence of clear *Chordin-like* genes in the Mollusca confirms the hypothesis that *Chordin-like* genes were present in the Bilaterian common ancestor, rather than being a deuterostome novelty, and is corroborated by the assignment of *T. adhaerens* and *Amphimedon queenslandica* Chordin-like homologues within this family by other authors, rather than as a canonical Chordin (*sensu* Richards and Degnan 2009).

Twisted gastrulation (Tsg)

Tsg is a modulator of BMP signalling, although its mode of action is yet to be fully understood. As well as binding BMPs, it has been suggested that Tsg might also act as a promoter of BMP signalling, by freeing ligands from Chordin after the Chordin-ligand complex has been cleaved by Tolloid (Oelgeschläger 2000). This could also occur in some of the lophotrochozoan clades presented here, although annelids lack canonical *Chordin* homologues while possessing *Tsg*. This mooted role could therefore be absent in this Phylum, although *Chordin-like* may be targeted by Tsg instead.

The phylogenetic relationships of a number of metazoan Tsg sequences can be seen in Fig. 7B. Mollusc and annelid species all possess a single Tsg homologue, while all schistosomes examined and the rotifers *B. plicatilis* and *A. vaga* appear to have lost theirs, as none can be found in their genomes or transcriptomes. This could be correlated to the markedly reduced BMP complements of these species – without the ligand diversity found in other species, it is perhaps unsurprising that regulatory mechanisms have also been lost.

Noggin

Noggins are proteins that act to interfere with canonical TGF-β signalling by sequestering ligands before they can bind to their receptors, as with Chordin above. Noggins primarily have been shown to interact with BMP-like ligands. Our data (Fig. 7C) confirms the study of Molina *et al.,* (2011), who posited the existence of two major clades of Noggin across the Metazoa, a canonical Noggin clade, and a less well-categorised Noggin-like clade. We find both kinds in all lophotrochozoans examined with the exception of *S. mansoni,* which has two Noggin-likes but no Noggin, and the rotifers *A. vaga* and *B. plicatilis*, which lack Noggins entirely. It is well documented that canonical Noggins and Noggin-like proteins are absent from the genomes of some ecdysozoan model organisms (van der Zee *et al.,* 2008) although these are found in some other arthropods (Duncan *et al.,* 2013): the presence of a Noggin in the

TABLE 1

		Ligands	Smads	Receptors	Chordin	Chordin-like	Noggin	Noggin-like	Follistatin	Gremlins	Dans	Dantes	Tsgs	Tolloids	BAMBI	NOMO	SMURFs	References
Deuterostomes	*Homo sapiens*	33	8	13	1	2	1	0	1	2	1	2	1	3	1	3	2	*Huminiecki et al 2009*
	Ciona intestinalis	10	5	7	1	1	1	0	0	1	0	0	1	1	0	1	1	*Hino et al 2003/Huminiecki et al 2009*
	Strongylocentrotus purpuratus	14	4	6	1	1	0	1	1	1	1	0	1	3	0	2	1	*Lapraz et al 2006*
Ecdysozoans	*Drosophila melanogaster*	7	4	5	1	0	0	0	1	0	0	0	3	2	0	1	1	*van der Zee et al 2008*
	Apis mellifera	7	4	5	1	0	0	0	0	0	1	0	1	1	1	1	1	*van der Zee et al 2008*
	Tribolium castaneum	8	4	5	1	0	0	0	1	1	1	0	1	1	1	1	1	*van der Zee et al 2008*
	Caenorhabditis elegans	5	7	3	0	0	0	0	0	1	0	0	0	1	0	1	0	*Savage-Dunn 2005, Huminiecki et al 2009*
Lophotrochozoans	*Capitella teleta*	16	4	5	0	1	1	1	1	1	0	0	1	2	1	1	1	
	Pomatoceros lamarckii	9	4	5	0	0	1	1	1	1	0	0	1	1	1	1	1	
	Lottia gigantea	10	4	5	1	1	1	1	1	1	0	0	1	1	1	1	1	
	Patella vulgata	12	4	5	1	0	1	1	1	1	0	0	1	1	1	1	1	
	Brachionus plicatilis	4	4	3	0	0	0	0	0	0	0	0	0	0	0	1	0	
	Adenita vaga	10*	10	8	0	0	0	0	2	0	0	0	0	0	0	2	0	*Flot et al 2013 *NB see Supplementary File 1*
Diploblasts	*Schistosoma mansoni*	2	5	5	0	0	1	2	1	0	0	0	0	1	0	1	0	
	Nematostella vectensis	6	4	6	1	0	1	1	1	1	1	0	0	1	0	1	0	*Huminiecki et al 2009, Saina and Technau 2009*
	Mnemiopsis leidyi	9	5	4	0	0	0	0	0	0	0	0	0	1	0	1	1	*Pang et al 2011*
	Trichoplax adherens	5	4	4	0	1	1	0	1	0	1	0	0	0	0	1	1	*Huminiecki et al 2009*

TGF-β signalling complements of a range of metazoan species, as determined in the present manuscript or from previously published work as cited at right, where well annotated examples of these complements have been determined from other studies. While no whole-genome analysis of this pathway has yet been performed in the Porifera, we refer the interested reader to Suga *et al.,* 1999, Adamska *et al.,* 2007, and Fig. 11 of Pang *et al.,* 2011, where a number of the key families listed above are described in this Phylum.

crustacean *Daphnia pulex* and Noggin and Noggin-like in the hemipterans *Acyrthosiphon pisum* and *Rhodnius prolixus* implies that this loss happened independently in some insects and nematodes.

The presence of Noggin-like sequences in schistosomes again raises questions as to their function – they could interact with the TGF-like ligands found in these species, unlike their role in BMP regulation in other Superphyla, or instead they may be involved in regulating exogenous signals or other processes entirely. We do not detect the diversity of Noggin sequences in other lophotrochozoans that are found in *S. mediterreanea,* where eight Noggin and Noggin-like sequences are found (Molina *et al.,* 2011). We therefore posit that this expansion is lineage specific, and may be related to species-specific roles for ligands, perhaps in regeneration.

Follistatin

Follistatin binds to and inhibits TGF-β ligands, particularly Activins, although it can bind to other ligands, even those in the BMP-like class. Canonical Follistatin sequences are found in the Lophotrochozoa, as can be seen in the phylogenetic tree presented in Fig. 8A. We note that Platyhelminthes possess Follistatin, despite not having a canonical Activin ligand for it to bind to. As noted earlier, it is suspected that parasitic platyhelminths can utilise host ligands, so it is possible that these Follistatins inhibit the action of these, but it is perhaps more likely that Follistatins bind the more derived Activin-like ligands found in these species. No Follistatin sequences could be identified within our *B. plicatilis* dataset, although two are present in *A. vaga* (CAWI010044267.1 and CAWI010043776.1).

Only Follistatin homologues with the classical Follistatin/Osteonectin-like EGF domain followed by three Kazal-type serine protease inhibitor domains were examined by the present study, with the exception of the *N. vectensis* homologue, which lacks a clearly identifiable EGF domain. Extensive diversification of this gene has taken place within the deuterostome lineage as early as the Ambulacraria/chordate split, with loss, rearrangement and gain of various domains resulting in a total of up to five follistatin-domain containing genes, descended from an ancestral Follistatin sequence. The nature, role and diversification of these are yet to be fully investigated.

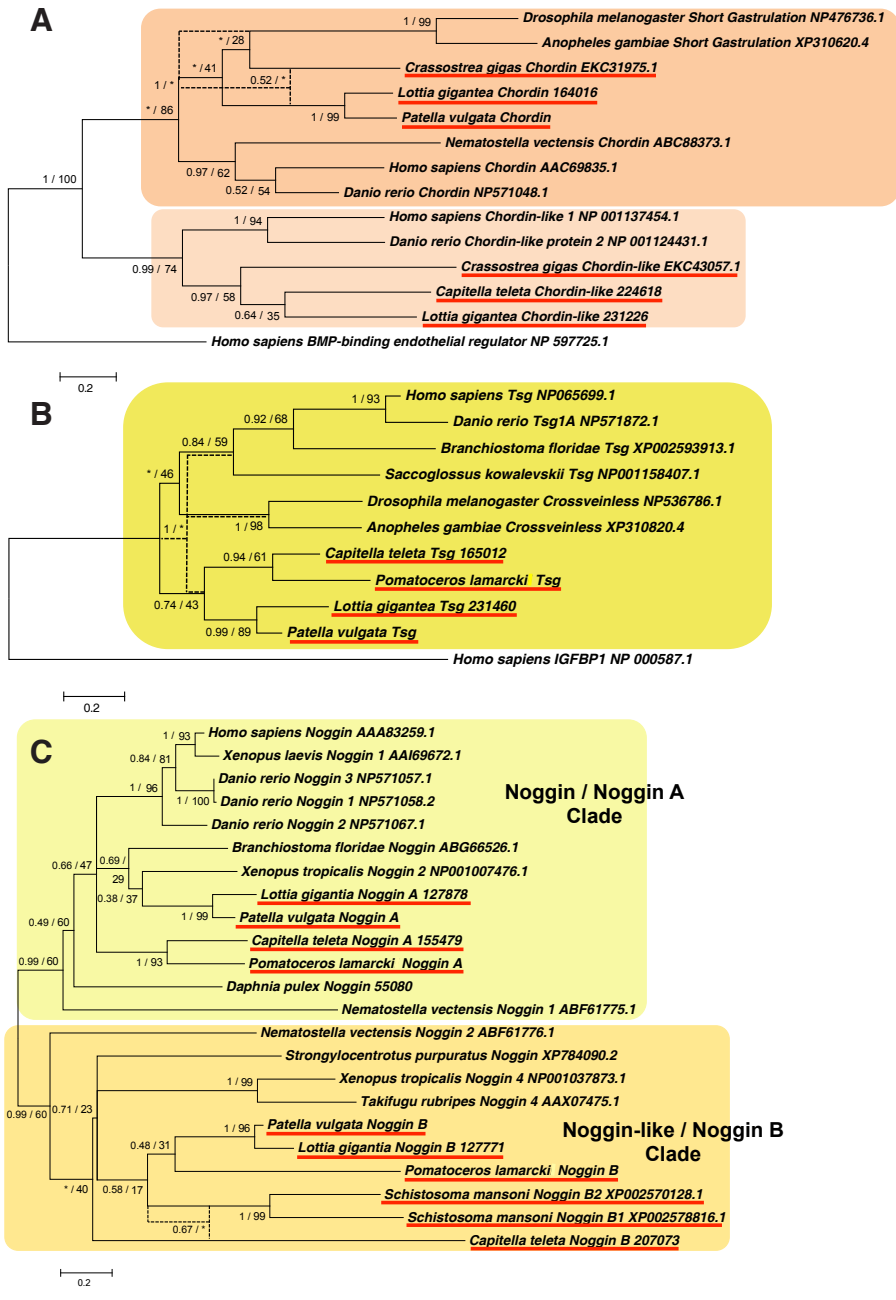


Fig. 7. Chordin (A), Twisted Gastrulation (B) and Noggin (C) interrelationships across the Metazoa, as determined by maximum likelihood (Tamura *et al.,* 2011) and Bayesian (Huelsenbeck and Ronquist 2001) methods. *Phylogeny shown is the result of ML analysis, with differences in topology using Bayesian methods indicated with a dotted line. Chordin phylogeny based on a 120 informative amino acid alignment of the von Willebrand factor type C/D domains and C8 domains generated by MAFFT (Katoh and Standley 2013) using the G-iNS-i strategy, analysed under the JTT model (Jones* et al., *1992, ML) and Dayhoff model (Dayhoff* et al., *1978, Bayesian) with tree shown rooted with* H. sapiens *BMP-binding endothelial regulator (NP 597725.1). Tsg phylogeny based on a 146 informative amino acid global alignment generated by MAFFT using the G-iNS-i strategy, rooted with* H. sapiens *IGFBP (NP 000587.1) after Vilmos* et al., *(2001), with both phylogenies determined using the WAG model (Whelan and Goldman 2001). Noggin phylogeny based on a 103 informative amino acid global alignment generated by MAFFT using the G-iNS-i strategy rooted at midpoint, followed by analysis using the WAG model in both analyses. Posterior probabilities/bootstrap percentage (of 1,000 replicates) and can be seen at the base of nodes. Lophotrochozoan sequences underlined in red. Sequences used in phylogenetic analysis, along with alignment, can be found in Supplementary File 1. Scale bars represent substitutions per site at given distances.*

Tolloids

Tolloids are found extracellularly, and cleave regulators of TGF-β signalling when they have formed complexes with free ligands, releasing the ligand they have bound and allowing signalling to occur. This action is vitally important in regulating a range of developmental processes, including the establishment of dorsoventral polarity, Fig. 8B shows the result of phylogenetic analysis of Tolloid sequences from a range of metazoan species, rooted with the homologue found in *N. vectensis.*

A single Tolloid homologue was found in both mollusc species examined, in *P. lamarcki*, and in all schistosome species (*S. mansoni* shown on tree). It appears to have been lost from rotifer genomes, although *A. vaga* possesses a number of similar metalloproteinases with high similarity to nematode NAS proteins and teleost hatching enzymes that may have been acquired by horizontal gene transfer from those species. *C. teleta* possesses two homologues, one with the canonical domain structure (inset, Fig. 8B) and one with a highly divergent structure, although this does not appear to be the result of mis-annotation of the *C. teleta* genome, and no equivalent is found in *H. robusta* or *P. lamarcki*. Two Tolloids (Tolloid and Tolkin) are found in *D. melanogaster*, and our analysis corroborates the suggestion of van der Zee *et al.,* (2008) that this Tolkin homologue is in fact a lineage specific paralogue, specific to the Diptera rather than found protostome-wide.

The roles played by Tolloid in lophotrochozoan development are yet to be fully explored, but Herpin *et al.,* 2007 have noted that *C. gigas* Tolloid is capable of ventralising zebrafish embryos. Furthermore, *C. gigas* Tolloid is maternally deposited into the oocyte, and may play an early role in patterning body axes in this species. A conserved role for Tolloid in dorsoventral polarity establishment in the Lophotrochozoa is therefore possible, although more investigation is required to confirm this, especially outside the Mollusca.

SMURFs

SMURF (SMAD specific E3 ubiquitin protein ligase) proteins regulate TGF-β signalling by a number of mechanisms, generally involving the targeting of R-Smads for degradation, but other roles have also been posited for these proteins. Two SMURF genes are characterised in vertebrates, but to date only single orthologues have

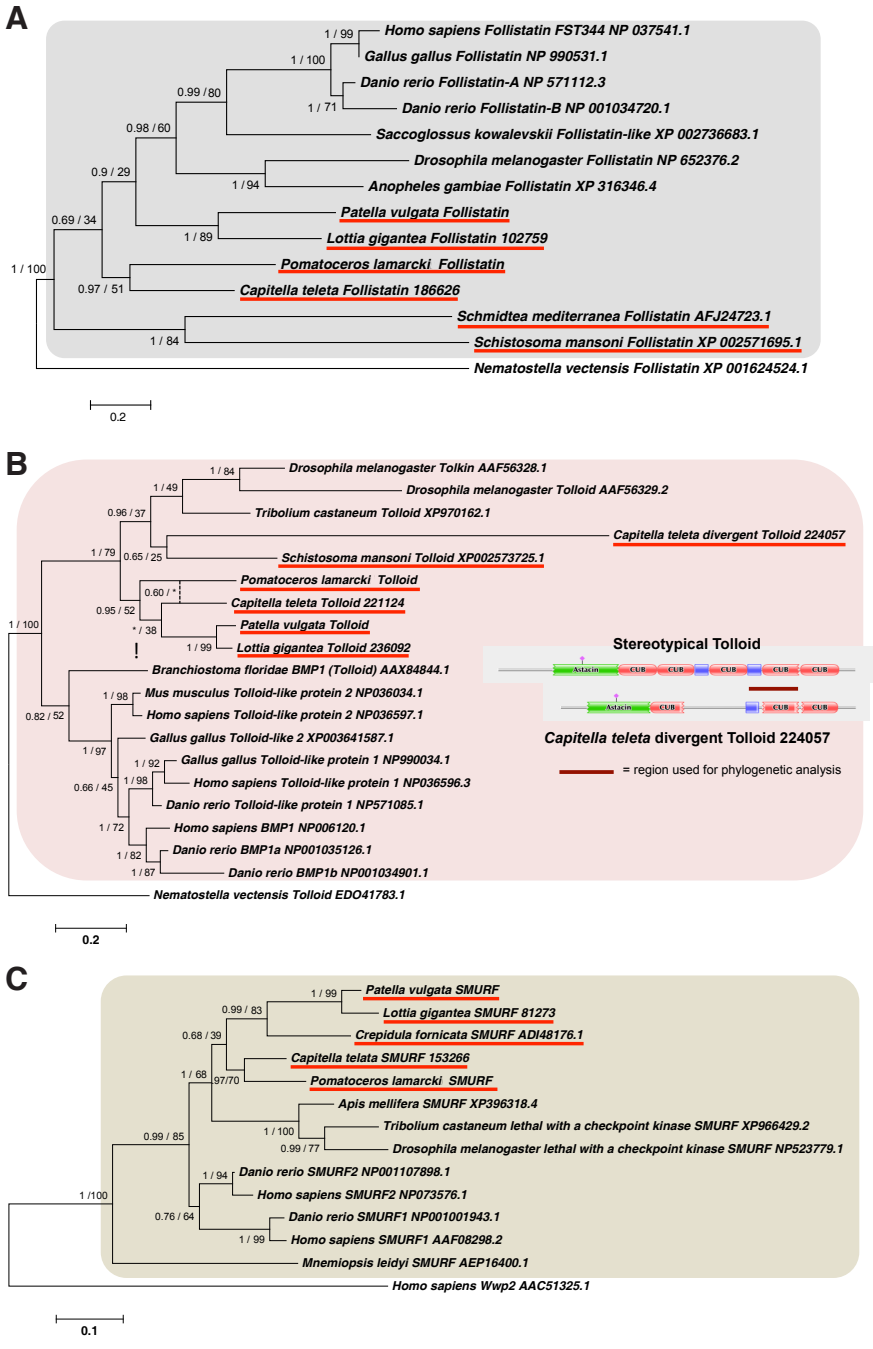


Fig. 8. Follistatin (A), Tolloid (B) and Smurf (C) interrelationships across the Metazoa, as determined by maximum likelihood (Tamura *et al.,* 2011) and Bayesian (Huelsenbeck and Ronquist 2001) methods. *Phylogeny shown is the result of ML analysis, with differences in topology using Bayesian methods indicated with a dotted line. Follistatin phylogenies inferred on the basis of a MAFFT alignment (Katoh and Standley 2013, E-INS-i strategy) with 227 informative sites, analysed under the WAG model (Whelan and Goldman 2001), rooted with* N. vectensis *Follistatin (XP 001624524.1). Tolloid phylogeny shown generated according to the JTT model (Jones* et al., *1992, ML)/Blosum model (Henikoff and Henikoff 1992, Bayesian) from a 150 informative amino acid alignment spanning the calcium-binding EGF domain and immediately proceeding the Cub domain as shown on inset, generated using MAFFT under the G-INS-i strategy and rooted with* N. vectensis *Tolloid (EDO41783.1). Inset shows the domain structure of canonical Tolloid proteins, along with that of the* C. teleta *divergent paralogue, presented from the Pfam database. The Smurf phylogeny was calculated under the JTT model, based on a 233 informative amino acid alignment generated using MAFFT using the G-INS-i strategy with outgroup specified as* H. sapiens *Wwp2 (AAC51325.1). Posterior probabilities/bootstrap percentage (of 1,000 replicates) and can be seen at the base of nodes. Sequences used in phylogenetic analysis, along with alignment, can be found in Supplementary File 1. Lophotrochozoan sequences underlined in red. Scale bars represent substitutions per site at given distances.*

been found in sequenced invertebrate species. Our phylogenetic analysis (Fig. 8C) of these genes from species across the Metazoa suggests a paralogous relationship between the two homologues found in vertebrate model species.

SMURF proteins appear to have originated within the early metazoan lineage, and clear homologues can be identified for these in *Hydra magnipapillata, M. leidyi* and *T. adhaerens,* although not to date in *N. vectensis.* While single canonical SMURF orthologues are readily identifiable in a variety of protostome species, no SMURF proteins can be found in the genomes of *C. elegans, A. vaga, B. plicatilis* or the schistosome species. These species do, however, possess other E3 ubiquitin-protein ligases that readily cluster with nedd-4-like E3 ubiquitin-protein ligases, and are probably homologues of this class of protein. We note that two *Crepidula fornicata* (Mollusca) SMURF homologues are present in the NCBI nr dataset, although one (ADI48175.1) is only a fragmentary sequence with 100% similarity to its homologue at the amino acid level, and as such was not used in our phylogenetic analysis.

BMP and activin membrane-bound inhibitor (BAMBI)

BMP and activin membrane-bound inhibitor is a pseudoreceptor that competes with true Type II receptors for ligand binding (Onichtchouk *et al.,* 1999). The presence of this gene has been noted in protostomes previously, particularly by van der Zee *et al.,* (2008) and Huminiecki *et al.,* (2009). Searches through the genomes of a variety of non-bilaterian metazoans revealed no trace of a BAMBI homologue, and it thus seems likely that BAMBI emerged on the lineage leading to the last common ancestor of protostomes and deuterostomes.

Fig. 9A shows the results of phylogenetic inference into the inter-relationships of BAMBI homologues from a variety of species in the Protostomia and Deuterostomia. No BAMBI orthologue could be identified in *C. elegans,* or in the Schistosome species or Rotifera sampled.

We note that a putative BAMBI sequence has been described in a previous publication in *S. mediterranea* (Gavino and Reddien 2011), but in our investigations this sequence (ADX42731.1) appears to more readily resemble canonical serine/threonine kinase receptors, clustering with these sequences in the course of phylogenetic analysis, although the partial sequence provided does not include the intracellular serine/threonine kinase domain required for signalling. We are unable to test whether this domain absence is the result of fragmentary sequence or is truly present in the complete protein, but if it is the latter this may represent the re-evolution of a trait present more generally across the Metazoa.

NOMO

NOMO (Nodal Modulator, previously known as pM5) is known for its role in directly antagonising Activin/Nodal signalling, in concert with Nicalin, with which it forms a transmembrane complex at the endoplasmic reticulum (Haffner 2004). These complexes are similar to complexes that regulate γ-secretase activity, but do not perform the same roles, as shown in Zebrafish rescue assays. Instead they have been shown to regulate the formation of mesendoderm by attenuating Nodal signalling. NOMO is, however, still under-researched, with little known of its molecular mode of

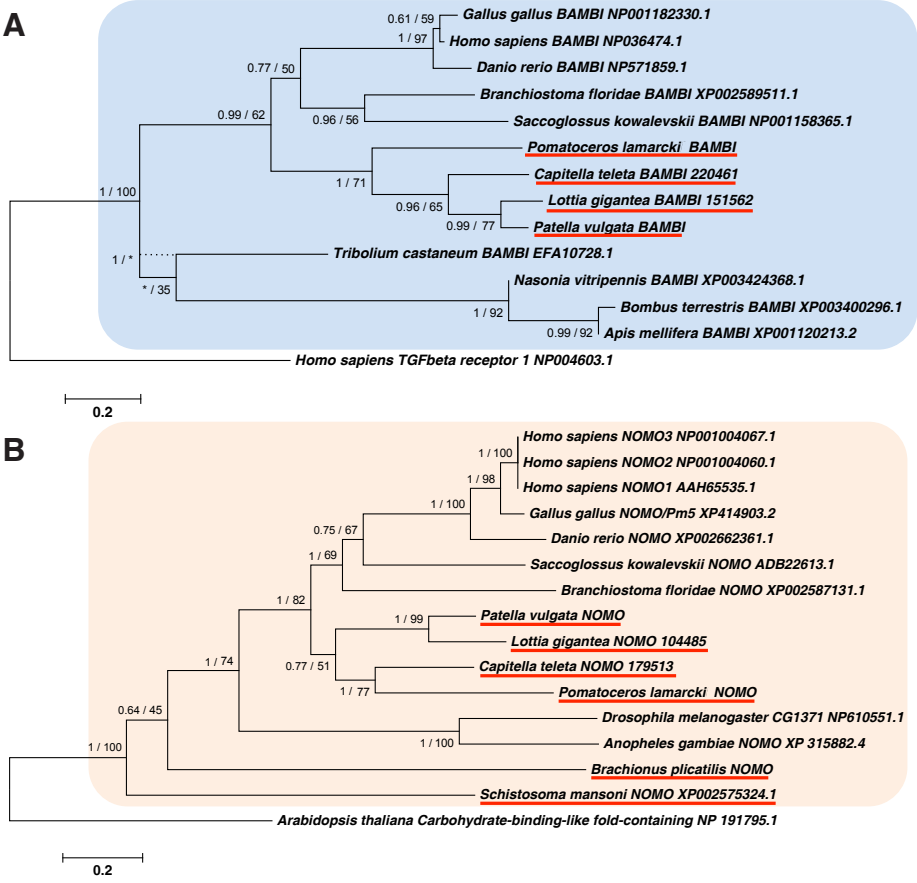


Fig. 9. Bambi (A) and NOMO (B) interrelationships across the Metazoa, as determined by maximum likelihood (Tamura *et al.,* 2011) and Bayesian (Huelsenbeck and Ronquist 2001) methods. *Phylogeny shown is the result of ML analysis, with differences in topology using Bayesian methods indicated with dotted lines. Bambi phylogenies determined using the JTT model (Jones* et al., *1992), NOMO phylogenies under the WAG model (Whelan and Goldman 2001). Bambi phylogeny based on 73 informative amino acid alignment created using MAFFT (Katoh and Standley 2013) under the L-INS-i model, spanning the transmembrane domain, incorporating some conserved regions both extra- and intracellularly. NOMO phylogeny based on G-INS-i alignment by MAFFT, with 439 informative positions. Bambi phylogeny rooted with* H. sapiens *TGF-β receptor 1 NP004603.1, NOMO tree with an apparent NOMO sequence found in* Arabidopsis thaliana *(NP_191795.1). Numbers at node reflect Bayesian posterior probabilities/bootstrap support (1,000 replicates, JTT model/WAG model, all default priors) respectively. Lophotrochozoan sequences underlined in red. Posterior probabilities/bootstrap percentage (of 1,000 replicates) and can be seen at the base of nodes. Sequences used in phylogenetic analysis, along with alignment, can be found in Supplementary File 1. Scale bars represent substitutions per site at given distances.*

action (Dettmer 2010).

Phylogenetic analysis of NOMO sequence from a range of metazoans can be seen in Fig. 9B, rooted with an apparent NOMO sequence found in *Arabidopsis thaliana.* NOMO is found in every genome examined in this work, implying a crucial role in cell signalling, even in species where *Nodal* is not found (for example, in the Ecdysozoa).

NOMO-like sequences are well described outside the Metazoa and are suggested to be found throughout the Eukaryota (HomoloGene:13810). It seems that NOMO is particularly poorly named given its conservation outside of clades where Nodal exists, and in many cases where TGF-β signalling is entirely absent. Some recent analyses have suggested that NOMO is involved in the regulation of nicotinic acetylcholine receptor functionality (Almedom *et al.,* 2009, for example), and this, along with the widespread conservation of NOMO throughout the Eukaryota, suggests that this complex has been recruited by metazoans for further regulation of TGF-β signalling – although much mechanistic work is required to untangle how this regulation is performed.

Discussion

General Lophotrochozoan TGF-β componentry and the evolution of TGF-β signalling

The results of our analysis of lophotrochozoan TGF-β cassettes can be seen in Table 1 and for TGF-β ligands in particular in Fig. 10. In general, the complements of the annelid and mollusc species considered in our analysis resemble that of invertebrate deuterostomes more than they resemble those of ecdysozoan models. Ligand diversity seems more pronounced than that seen in the Ecdysozoa, and many regulatory components, such as the *noggin* class, seem to be the result of Ecdysozoa-specific, rather than protostome-wide, loss.

The most striking differences between the cassettes of other metazoans and our sampled datasets are in the case of the rotifer *B. plicatilis* and the planarian *S. mansoni.* Extensive loss seems to have occurred at the ligand and regulatory levels in these Phyla, with Smad and receptor diversity relatively unchanged, particularly in *S. mansoni*.

These apparent lost genes could be missing from our datasets due to inadequate assembly of these genomes. This is more likely for *B. plicatilis* than for *S. mansoni*, as this genome results from a much deeper sequencing effort. Other planarian genomes also exist, allowing assessment by comparison to outgroups in cases of loss. Analyses of the *B. plicatilis* dataset, however, implies that the substantial majority of genes present in that species are recovered by the genome dataset used here, and while some missing genes are perhaps to be expected from our analysis, we would not expect these to be numerous.

It should be noted that the TGF-β componentry of *S. mansoni* is very similar to that of *S. japonicum* and in most cases to that of *S. mediterranea,* and it has been chosen as a representative and very well described member of its Phylum for the purposes of comparison. While some differences exist – *S. mediterranea*, for example, has eight Noggin sequences listed on Genbank – these appear to be the result of lineage specific loss or gain, and the *S. mediterranea* genome is still generally cited as a draft resource. It is possible that more basal planarian species will exhibit a more complete TGF-β complement than the species currently sampled.

Any loss of regulatory elements is of interest, as the fine-scale regulation of TGF-β ligands is known to play a variety of key roles in the establishment of bodily axes in other Phyla, for example, in the establishment of dorsoventral polarity and the germ layers in *Xenopus laevis* (Hill 2001). How lophotrochozoans accomplish these tasks, perhaps without the aid of these modulators of signalling, remains to be established.

While the internal phylogeny of the lophotrochozoan clade is yet to be fully resolved, most studies place molluscs and annelids as sister groups (in the Trochozoan clade, along with brachiopods, phoronids and nemerteans), with planarian and rotifer data implying that these Phyla are only distantly related to the Trochozoa *sensu stricto*. While our sampling may not allow us to trace the evolution of the lophotrochozoan TGF-β complement across the entirety of its constituent Phyla, we can be confident that the last common ancestor of the Trochozoa, Rotifer and Platyhelminthes will be relatively closely related in molecular complement to the *Urlophotrochozoan* (the common ancestor of all lophotrochozoans), and our sampling thus allows us to draw inference as to the cassette of that hypothetical organism, by mapping gain and loss onto a schematic phylogeny (Fig. 11).

Our work therefore suggests that we might expect the TGF-β signalling complement of the *Urlophotrochozoan* to be complete, with only the Dan clade completely missing from the ancestrally shared regulatory cassette across all lophotrochozoan Phyla, and no lophotrochozoan-wide loss seen in the ligand complement. Loss is, however, prevalent in gene families in the

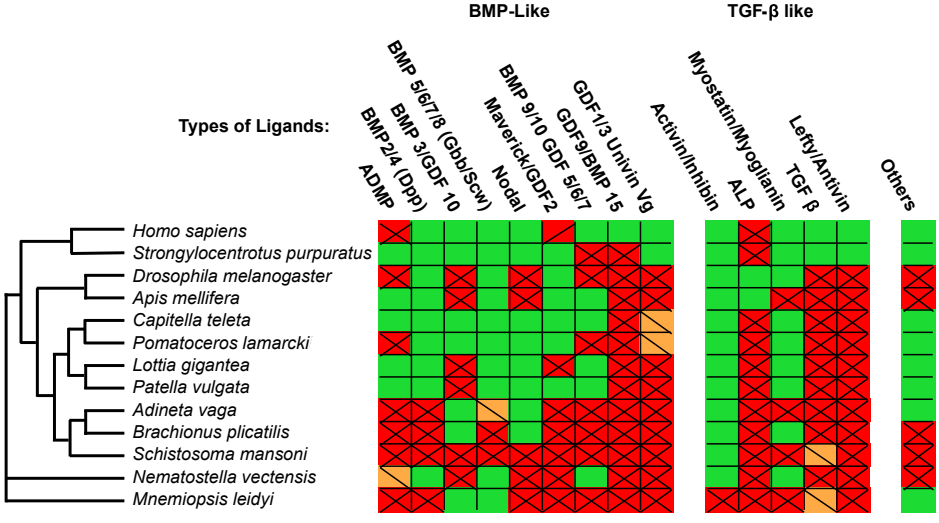


Fig. 10. TGF-β ligand family presence and absence in a range of metazoan species, as determined in the present manuscript or inferred from previously published work. *Only species with clearly informative TGF-β ligand identity known are listed above. Confirmed identity is shown in green (no lines), tentative identity in orange (single diagonal line) and absence in red (two diagonal lines, forming a cross).*

planarians and in the rotifer species sampled, implying that a diverse TGF-β complement may not be necessary for these clades, although the reasons why these genes have been lost while being so well conserved in other lineages remains unknown.

It should be noted that our findings suggest that the use of the phrase "TGF-β signalling" represents a historical accident, rather than a unified nomenclature reflecting a diversity of TGF-β ligands. The considerable momentum of the published literature means that this classification scheme is unlikely to be challenged, but we would suggest that "BMP signalling" is a more representative term for the ligand superclass and cascade as a whole, with TGF-β ligands being a probable deuterostome novelty. The TGF-β-like class would then perhaps be more properly termed the Activin/Myostatin-like class. However, we recognise the historical contingencies inherent in the current naming scheme, and have named our ligands accordingly in this manuscript.

In general, however, the TGF-β ligand signalling cassette is well conserved in the Lophotrochozoa, which will allow us to compare functional roles and expression of these elements between this Superphylum and the Deuterostomia and the Ecdysozoa – a practice that has not always been possible when only these latter Superphyla were described, as often one or other lacked a gene completely. This has and will continue to allow us to infer ancestral roles for a variety of genes, a vital step in understanding how animal life in general, and the TGF-β signalling cascade in particular, has evolved.

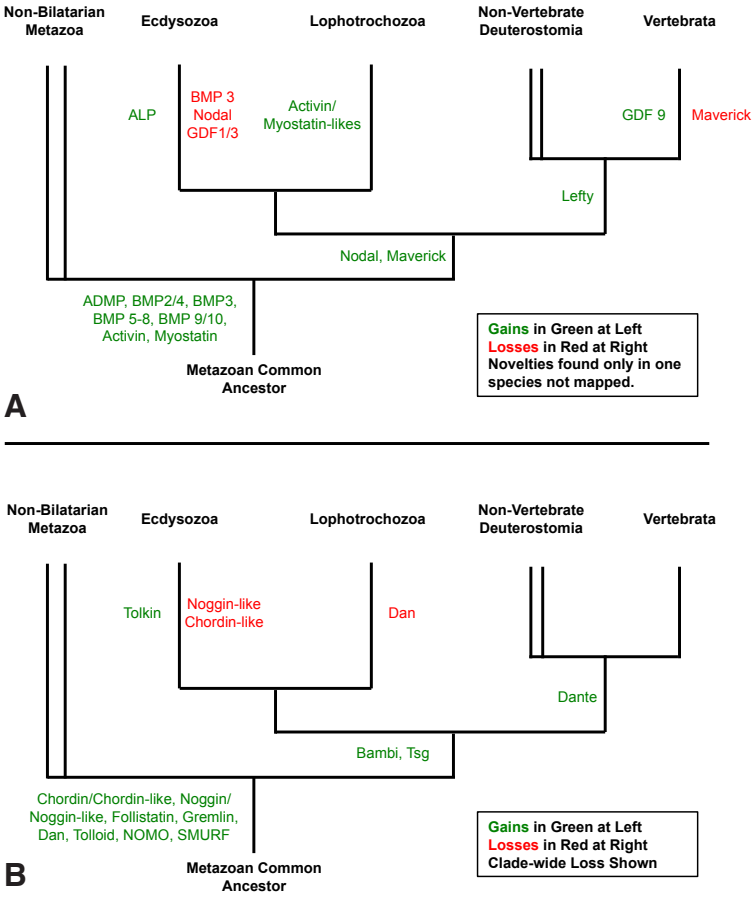


Fig. 11. Gain and loss of TGF-β ligands (A) and modulators of TGF-β signalling (B) across the Metazoa, mapped onto a schematic cladogram of the interrelationships of these Superphyla. *Position of TGF-β itself is contingent on the true assignation of TGF-β status to* M. leidyi *protein sequence by Pang* et al., *2011, as this hypothesis is only weakly supported by our analysis.*

Conclusions

Here we have presented the first systematic treatment of the TGF-β signalling cascade in the Lophotrochozoa. This work will provide the building blocks to allow us to understand a variety of developmental and homeostasis-related signalling cascades in these species, and also provide a valuable resource for tracing the ancestral functionality of the TGF-β pathway.

We have shown that while some aspects of the TGF-β signalling pathway are very highly conserved in the Lophotrochozoa when compared to other Superphyla and basal clades, some aspects, notably the TGF-β-like ligand complement, are highly derived, and differ greatly, even when compared between closely related species.

The loss of regulatory elements of the TGF-β signalling cascade is also particularly intriguing, as without the modulation of signalling provided by these molecules lophotrochozoans may have evolved alternative means of interpreting the levels of signal provided by TGF-β ligands, without relying on gradient control. Further research will be required to discern how this occurs.

This data will act as the starting point for a number of investigations into the function, evolution and diversification of these molecules in this under-represented Superphylum, and fills in the last major gap remaining in our understanding of the diversity of this cascade across metazoan life.

Materials and Methods

Gene identification

Gene sequences were derived from *P. lamarcki*, *P. vulgata*, *B. plicatilis* (Transcriptome sequences as described in Werner *et al.,* 2012, Kenny and Shimeld 2012, genomic sequence manuscripts in prep), *S. mansoni* (Berriman *et al.,* 2009), *L. gigantea* and *C. teleta* (Simakov *et al.,* 2013) genomic sequences, downloaded to a local server. *A. vaga* sequences were taken from http://www.genoscope.cns.fr. These were identified using tBlastn (Altschul *et al.,* 1990) of conserved regions of known gene sequence against each dataset. Genes thus putatively identified were then reciprocally blasted against the NCBI nr database using Blastx to further confirm their identity. Conversion into protein sequence was carried out using the EMBOSS Transeq (http://www.ebi.ac.uk/Tools/st/emboss_transeq/) tool, assuming standard codon usage.

Phylogenetic analysis

Gene sequences were aligned with those of known identity downloaded from the NCBI nr database, using MAFFT version 7 (Katoh and Standley 2013) under the G-INS-i strategy unless otherwise stated, and alignments saved in fasta format and imported into MEGA5 (Tamura *et al.,* 2011) for alignment curation. Conserved domains were identified and alignments trimmed to these areas for further analysis, with sections of alignment containing gaps excluded in all cases for the purpose of phylogenetic inference.

Maximum likelihood analysis was performed using MEGA5 (Tamura *et al.,* 2011) using the WAG model (Whelan and Goldman 2001) unless otherwise stated, 1,000 bootstrap replicates and all other default prior settings. Bayesian phylogenetic analysis was

performed with MrBayes v3.2.1-x64 software (Huelsenbeck and Ronquist 2001) using the WAG model (Whelan and Goldman 2001) of amino acid substitution unless otherwise stated, after initial identification using mixed models. The Monte Carlo Markov Chain search was run over 1,000,000 generations, unless otherwise stated in figure legends, and trees were sampled every 1,000 generations, with the first 25 % of trees thus gathered discarded as 'burn-in'. For the purposes of tree display, Contype was configured to Allcompat, but all other priors remained at default.

Gene sequences are firmly assigned homology to known genes where both maximum likelihood and Bayesian trees agree on the placement of these sequences within a clade of genes of known identity, and at least one phylogenetic reconstruction method shows node support with these genes of known homology with a bootstrap value greater than or equal to 80, or posterior probability support greater than or equal to 0.9. The suffix '-like' is affixed to gene names where these criteria are not met.

Acknowledgements

We give our thanks to all the members of our lab groups, whose time and discussions contributed much to this manuscript. Funds for sequencing P. lamarcki *and* P. vulgata *were provided by the Elizabeth Hannah Jenkinson Fund. Funds for sequencing* B. plicatilis *were provided by a Royal Society of New Zealand Marsden Grant (UOO0602) to PKD. NJK was supported by a Clarendon Fund studentship and much of this work was completed with the aid of a Santander Academic Travel Award in the lab of CG. CG is currently a "Ramon y Cajal" postdoctoral fellow supported by the Spanish MICINN and the UAM and funded by project CGL2011-29916 (MICINN). Our thanks also to Professor Henry, the editor of this special issue, and to two anonymous reviewers for their helpful comments on this work.*

References

ADAMSKA M, DEGNAN S M, GREEN K M, ADAMSKI M, CRAIGIE A, *et al.,* (2007). Wnt and TGF-β Expression in the Sponge *Amphimedon queenslandica* and the Origin of Metazoan Embryonic Patterning. *PLoS ONE* 2: e1031.

AL-SALIHI M A, HERHAUS L, SAPKOTA G P (2012). Regulation of the transforming growth factor β pathway by reversible ubiquitylation. *Open Biol* 2:120082. dx.doi.org/10.1098/rsob.120082.

ALMEDOM R B, LIEWALD J F, HERNANDO G, SCHULTHEIS C, RAYES D, PAN J, SCHEDLETZKY T, HUTTER H, BOUZAT C, GOTTSCHALK A (2009). An ER-resident membrane protein complex regulates nicotinic acetylcholine receptor subunit composition at the synapse. *EMBO J* 28(17): 2636-2649.

ALTSCHUL S, GISH W, MILLER W, MYERS E and LIPMAN D (1990). Basic local alignment search tool. *J Mol Biol* 215: 403-410.

BALEMANS W and VAN HUL W (2002). Extracellular regulation of BMP signaling in vertebrates: a cocktail of modulators. *Dev Biol* 250: 231-250.

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.

CHENG S K, OLALE F, BENNETT J T, BRIVANLOU A H and SCHIER A F (2003). EGF-CFC proteins are essential coreceptors for the TGF-β signals Vg1 and GDF1. *Genes Dev* 17: 31-36.

DAVIES S J, SHOEMAKER C B and PEARCE E J (1998). A Divergent Member of the Transforming Growth Factor β Receptor Family from *Schistosoma mansoni* Is Expressed on the Parasite Surface Membrane. *J. Biol Chem* 273: 11234-11240.

DAYHOFF M O, SCHWARTZ R M, ORCUTT B C (1978). A model of evolutionary change in proteins. In *Atlas of protein sequence and structure Volume 5* (Ed M O Dayhoff) National Biomedical Research Foundation, Washington DC, pp 345-351.

DERYNCK R and MIYAZONO K (2008). TGF-β and the TGF-β family. In *The TGF-b Family* (Eds. R DERYNCK and K MIYAZONO). Cold Spring Harbor Laboratory Press, New York, pp. 29-43.

DETTMER U, KUHN P-H, ABOU-AJRAM C, LICHTENTHALER S F, KRUGER M, KREMMER E, HAASS C, HAFFNER C (2010). Transmembrane Protein 147 (TMEM147) Is a Novel Component of the Nicalin-NOMO Protein Complex. *J Biol Chem* 285(34): 26174-26181.

DUNCAN E J, BENTON M A, DEARDEN P K (2013). Canonical terminal patterning is an evolutionary novelty. *Dev Biol* 377: 245-261.

DUNN C W, HEJNOL A, MATUS D Q, PANG K, BROWNE W E, SMITH S A, *et al.,* (2008) Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* 452: 745-749

FLEURY E, FABIOUX C, LELONG C, FAVREL P and HUVET A (2008). Characterization of a gonad-specific transforming growth factor-beta superfamily member differentially expressed during the reproductive cycle of the oyster *Crassostrea gigas. Gene* 410: 187-196.

FLOT J-F, HESPEELS B, LI X *et al.,* Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta vaga. Nature* 500: 453-457.

FORRESTER S G, WARFEL P W and PEARCE E J (2004). Tegumental expression of a novel type II receptor serine/threonine kinase (SmRK2) in *Schistosoma mansoni. Mol Biochem Parasitol* 136: 149-156.

FREITAS T C, JUNG E and PEARCE E J (2007). TGF-β signaling controls embryo development in the parasitic flatworm *Schistosoma mansoni. PLoS Path* 3: e52.

GAVINO M A and REDDIEN P W (2011). A Bmp/Admp Regulatory Circuit Controls Maintenance and Regeneration of Dorsal-Ventral Polarity in Planarians. *Curr Biol* 21: 294-299.

GRANDE C and PATEL N H (2009). Nodal signalling is involved in left-right asymmetry in snails. *Nature* 457: 1007-1011.

GRANDE C, MARTIN-DURAN J M, KENNY N J, TRUCHADO-GARCIA M and HEJNOL A (2014). Evolution, divergence and loss of the Nodal signalling pathway: new data and a synthesis across the Bilateria. *Int J Dev Biol* 58: 521-532.

HAFFNER C, FRAULI M, TOPP S, IRMLER M, HOFMANN K, REGULA JT, BALLY-CUIF L, HAASS C (2004). Nicalin and its binding partner Nomo are novel Nodal signaling antagonists. *EMBO J* 23: 3041-3050.

HELDIN C H and MOUSTAKAS A (2012). Role of Smads in TGF-β signaling. *Cell Tis Res* 347: 21-36.

HENIKOFF S and HENIKOFF J G (1992) Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. USA* 89: 10915-10919.

HERPIN A, LELONG C and FAVREL P (2004). Transforming growth factor-β-related proteins: an ancestral and widespread superfamily of cytokines in metazoans. *Dev Compar Immunol* 28: 461-485.

HERPIN A, LELONG C, BECKER T, ROSA F, FAVREL P and CUNNINGHAM C (2005). Structural and functional evidence for a singular repertoire of BMP receptor signal transducing proteins in the lophotrochozoan *Crassostrea gigas* suggests a shared ancestral BMP/activin pathway. *FEBS J* 272: 3424-3440.

HERPIN A, LELONG C, BECKER T, FAVREL P, and CUNNINGHAM C (2007). A tolloid homologue from the Pacific oyster *Crassostrea gigas. Gene Expr Patterns* 7 (6): 700-708.

HILL C S (2001). TGF-β signalling pathways in early *Xenopus* development. *Curr Opin Genet Dev* 11: 533-540.

HOLLEY S A, JACKSON P D, SASAI Y, LU B, DE ROBERTIS E M, HOFFMANN F M, and FERGUSON E L (1995). A conserved system for dorsal-ventral patterning in insects and vertebrates involving sog and chordin. *Nature* 376 (6537): 249-253.

HUELSENBECK J P and RONQUIST F (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754-755.

HUMINIECKI L, GOLDOVSKY L, FREILICH S, MOUSTAKAS A, OUZOUNIS C and HELDIN C H (2009). Emergence, development and diversification of the TGF-β signalling pathway within the animal kingdom. *BMC Evol Biol* 9: 28.

ITOH S and TEN DIJKE P (2007). Negative regulation of TGF-β receptor/Smad signal transduction. *Curr Opin Cell Biol* 19: 176-184.

JONES D T, TAYLOR W R, THORNTON J M (1992). The rapid generation of mutation data matrices from protein sequences. Computer applications in the biosciences: *CABIOS* 8 (3): 275-282.

KATOH K, STANDLEY D M (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* 30: 772-780.

KENNY N and SHIMELD S (2012). Additive multiple k-mer transcriptome of the keel-worm *Pomatoceros lamarcki;* (Annelida; Serpulidae) reveals annelid trochophore transcription factor cassette. *Dev Genes Evol* 222: 325-339.

KUO D-H and WEISBLAT D A (2011). A New Molecular Logic for BMP-Mediated Dorsoventral Patterning in the Leech *Helobdella. Curr Biol* 21: 1282-1288.

LAPRAZ F O, ROTTINGER E, DUBOC V R, RANGE R, DULOQUIN L, WALTON K, WU S-Y, BRADHAM C, LOZA M A, HIBINO T *et al.,* (2006). RTK and TGF-β signaling pathways genes in the sea urchin genome. *Dev Biol* 300: 132-152.

MASSAGUÉ J (1998). TGF-β Signal Transduction. *Ann Rev Bioc* 67: 753-791.

MASSAGUÉ J (2012). TGF-β signaling in context. *Nat Rev Mol Cell Bio* 13: 616-630.

MASSAGUÉ J, BLAIN S W and LO R S (2000). TGF-β signaling in growth control, cancer, and heritable disorders. *Cell* 103: 295-309.

MASSAGUÉ J, SEOANE J and WOTTON D (2005). Smad transcription factors. *Genes Dev* 19: 2783-2810.

MATUS D Q, PANG K, MARLOW H, DUNN C W, THOMSEN G H and MARTINDALE M Q (2006). Molecular evidence for deep evolutionary roots of bilaterality in animal development. *Proc. Natl. Acad. Sci. USA* 103: 11195-11200.

MOLINA M D, NETO A, MAESO I, GOMEZ-SKARMETA J L, SALO E and CEBRIA F (2011). Noggin and Noggin-Like Genes Control Dorsoventral Axis Regeneration in Planarians. *Curr Biol* 21: 300-305.

MOUSTAKAS A and HELDIN C H (2009). The regulation of TGF-β signal transduction. *Development* 136: 3699-3714.

NEDERBRAGT A J, VAN LOON A E and DICTUS W J (2002). Expression of *Patella vulgata* orthologs of engrailed and dpp-BMP2/4 in adjacent domains during molluscan shell development suggests a conserved compartment boundary mechanism. *Dev Biol* 246: 341-355.

OELGESCHLAGER M, LARRAIN J, GEISSERT D, & DE ROBERTIS E M (2000). The evolutionarily conserved BMP-binding protein Twisted gastrulation promotes BMP signalling. *Nature,* 405, 757-763.

ONICHTCHOUK D, CHEN Y G, DOSCH R, GAWANTKA V, DELIUS H, MASSAGUÉ J and NIEHRS C (1999). Silencing of TGF-β signalling by the pseudoreceptor BAMBI. *Nature* 401: 480-485.

OSMAN A, NILES E G, VERJOVSKI-ALMEIDA S, LOVERDE P T. *Schistosoma mansoni* TGF-β Receptor II: Role in host ligand-induced regulation of a schistosome target gene. *PLoS Pathog.* 2006. e54

PANG K, RYAN J F, BAXEVANIS A D and MARTINDALE M Q (2011). Evolution of the TGF-β signaling pathway and its potential role in the ctenophore, *Mnemiopsis leidyi. PLoS ONE* 6: e24152.

RENTZSCH F, ANTON R, SAINA M, HAMMERSCHMIDT M, HOLSTEIN T W and TECHNAU U (2006). Asymmetric expression of the BMP antagonists chordin and gremlin in the sea anemone *Nematostella vectensis:* Implications for the evolution of axial patterning. *Dev Biol* 296: 375-387.

RICHARDS G S and DEGNAN B M (2009). The dawn of developmental signaling in the Metazoa. *Cold Spring Harb Sym* 74, 81-90. doi: 10.1101/sqb.2009.74.028

ROSS S and HILL C S (2008). How the Smads regulate transcription. *Int J Biochem Cell Biol* 40: 383-408.

SAKUTA H, SUZUKI R, TAKAHASHI H, KATO A, SHINTANI T, IEMURA S-I, YAMAMOTO T S, UENO N, NODA M (2001). Ventroptin: A BMP-4 Antagonist Expressed in a Double-Gradient Pattern in the Retina. *Science* 293: 111-115.

SCOTT I C, BLITZ I L, PAPPANO W N, IMAMURA Y, CLARK T G, STEIGLITZ B M, THOMAS C L, MAAS S A, TAKAHARA K, CHO K W Y *et al.,* (1999). Mammalian BMP-1/Tolloid-Related Metalloproteinases, Including Novel Family Member Mammalian Tolloid-Like 2, Have Differential Enzymatic Activities and Distributions of Expression Relevant to Patterning and Skeletogenesis. *Dev Biol* 213: 283-300.

SHEN M M and SCHIER A F (2000). The EGF-CFC gene family in vertebrate development. *Trend Gene* 16: 303-309.

SHI Y and MASSAGUÉ J (2003). Mechanisms of TGF-β signaling from cell membrane to the nucleus. *Cell* 113: 685-700.

SIMAKOV O, MARLETAZ F, CHO S-J, EDSINGER-GONZALES E, HAVLAK P, HELLSTEN U, KUO D-H, *et al.,* (2013). Insights into bilaterian evolution from three spiralian genomes. *Nature* 493: 526-531.

SRIVASTAVA M, SIMAKOV O, CHAPMAN J, FAHEY B, GAUTHIER M E A, 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.

SUGA H, ONO K and MIYATA T (1999). Multiple TGF-β receptor related genes in sponge and ancient gene duplications before the parazoan/eumetazoan split. *FEBS Lett* 453: 346-350.

TAMURA K, PETERSON D, PETERSON N, STECHER G, NEI M and KUMAR S (2011). MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol.* 28: 2731-2739.

TEN DIJKE P and ARTHUR HM (2007). Extracellular control of TGF-β signalling in vascular development and disease. *Nat Rev Mol Cell Biol* 8: 857-869.

VAN DER ZEE M, DA FONSECA R N and ROTH S (2008). TGF-β signaling in Tribolium: vertebrate-like components in a beetle. *Dev Gene Evol* 218: 203-213.

VILMOS P, GAUDENZ K, HEGEDUS Z, and MARSH J L (2001). The Twisted gastrulation family of proteins, together with the IGFBP and CCN families, comprise the TIC superfamily of cysteine rich secreted factors. *Mol Pathol* 54: 317-323.

WAGNER G P (1996). Homologues, Natural Kinds and the Evolution of Modularity. *Am Zool*, 36: 36-43.

WERNER G D, GEMMELL P, GROSSER S, HAMER R, SHIMELD S M (2012). Analysis of a deep transcriptome from the mantle tissue of *Patella vulgata* Linnaeus (Mollusca: Gastropoda: Patellidae) reveals candidate biomineralising genes. *Mar Biotechnol* 15: 230-243.

WHELAN S, GOLDMAN N (2001). A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach. *Mol Biol Evol* 18: 691-699.

YAMAMOTO Y and OELGESCHLÄGER M (2004). Regulation of bone morphogenetic proteins in early embryonic development. *Naturwissenschaften* 91: 519-534.

Acknowledgements

Seb, a PhD is supposed to be a horrendous experience, but it was not. Instead, I was given some of the best and most challenging years of my life. Thank you for sharing your brilliance, providing a safe and comfortable environment to flourish in, and for guiding and supporting me as I stumbled down this winding road.

Cédric, Ferdi, and Jordi: who supported me daily and whose friendship I treasure.

Those who helped and supported me along the way (you know who you are).

My family, who are my center.

"Trust in the Lord with all your heart
and do not rely on your own insight;
in all your ways acknowledge Him,
and He will make straight your paths."
Proverbs 3:5-6