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 evolu-
tionary 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 inhibi-
tors 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 diver-
gence 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

---

---

*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

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,
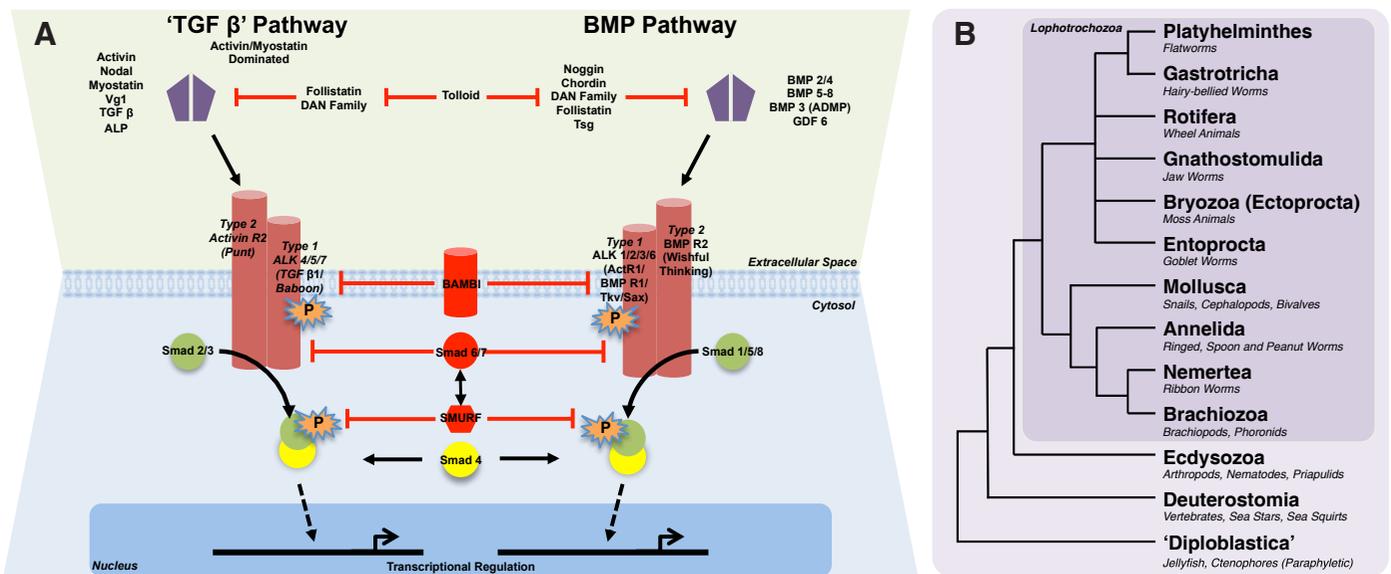

Fig. 1. Summary of TGF-β/BMP signalling cascades and Lophotrochozoan interrelationships. (A) *Canonical TGF-β/BMP signalling cascades: representation of the canonical signalling pathways for TGF-β-like and BMP-like cascades with inhibitors of signalling shown in red and operational signalling shown in 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.
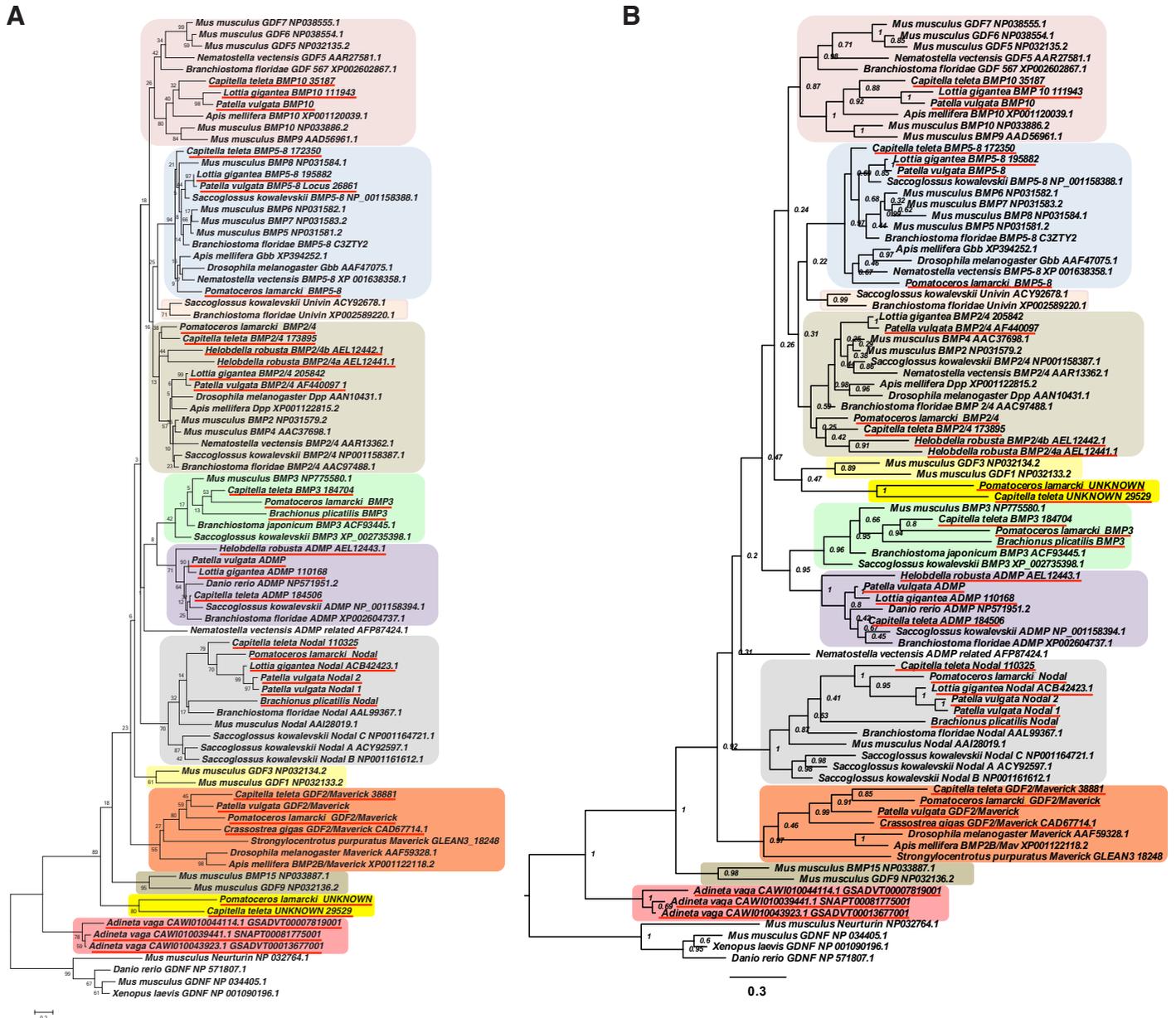

Fig. 2. Phylogeny of the BMP-like ligand class familial interrelationships across the Metazoa, as determined by (A) maximum likelihood (Tamura *et al.,* 2011) and (B) Bayesian (Huelsenbeck and Ronquist 2001) methods. *Alignment generated by MAFFT (Katoh and Standley 2013) using the L-INS-i strategy resulting 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.

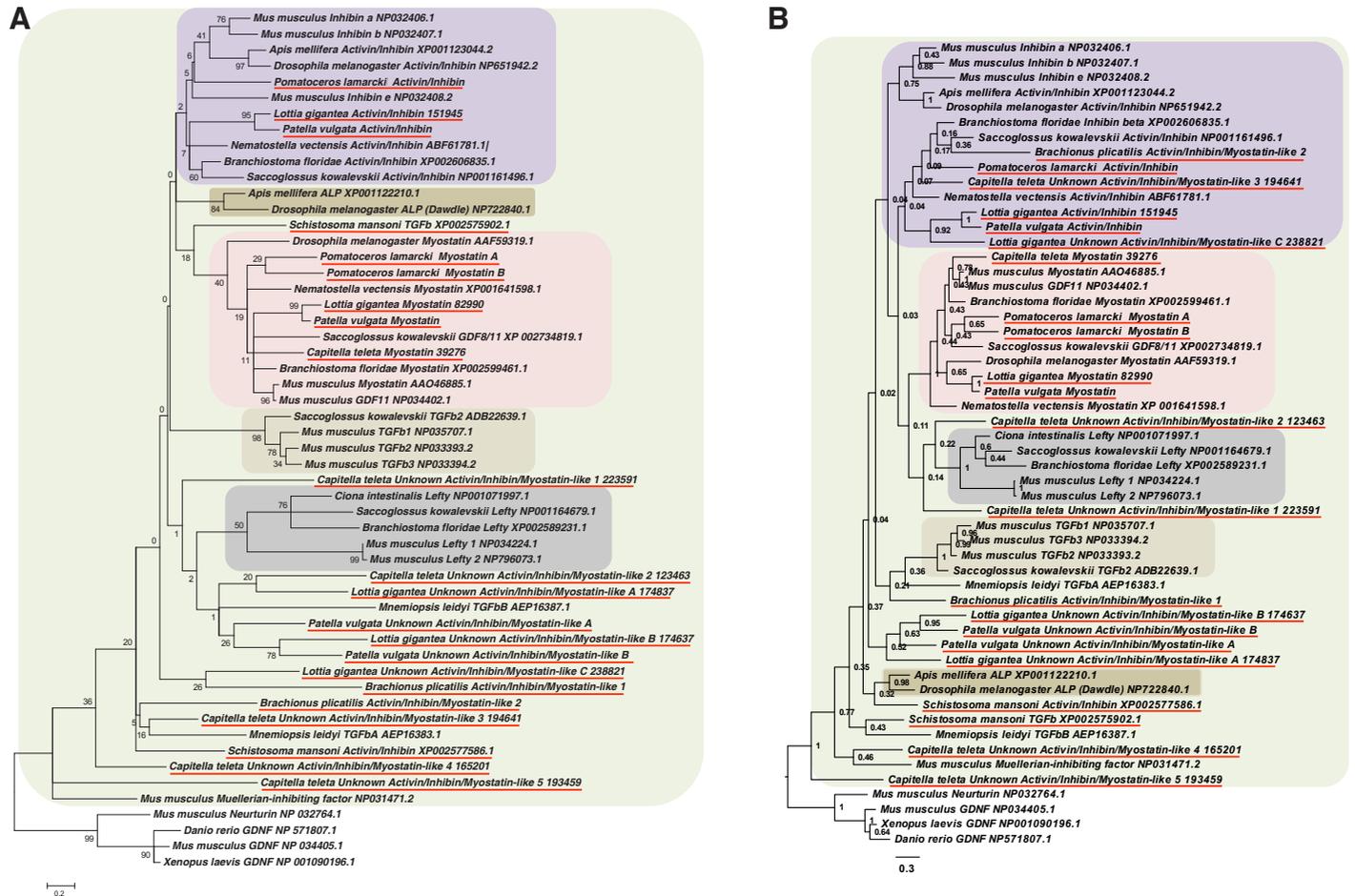The large number of ligands seen in *A. vaga* is the result of

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
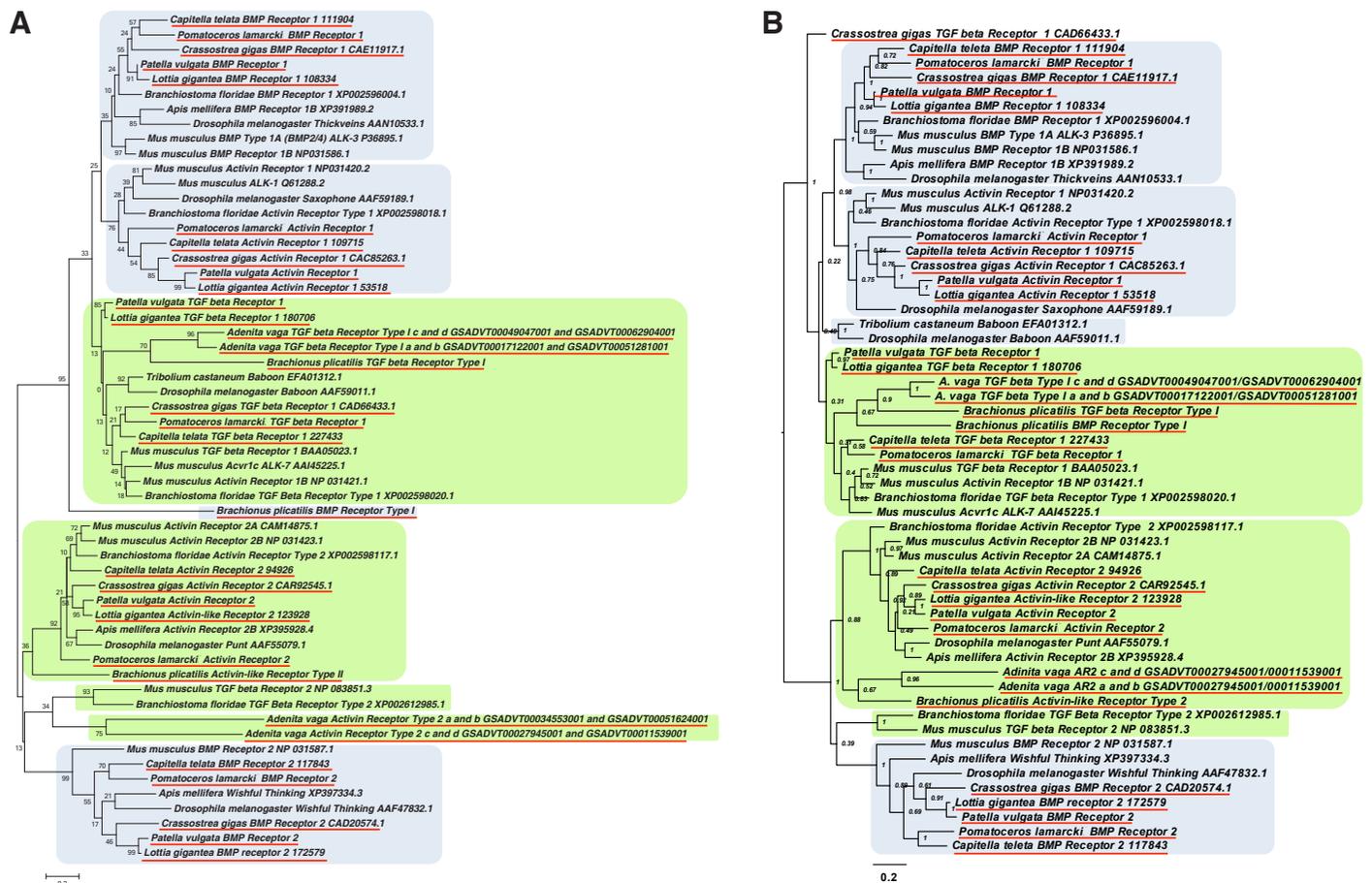

Fig. 5. Smad and Dad interrelationships across the Metazoa, as determined by (A) maximum likelihood (Tamura *et al.,* 2011) and (B) Bayesian (Huelsenbeck and Ronquist 2001) methods. *Alignment generated by MAFFT (Katoh and Standley 2013) using the G-iNS-i strategy, with the section used for analysis 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 neofunctionalisation 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
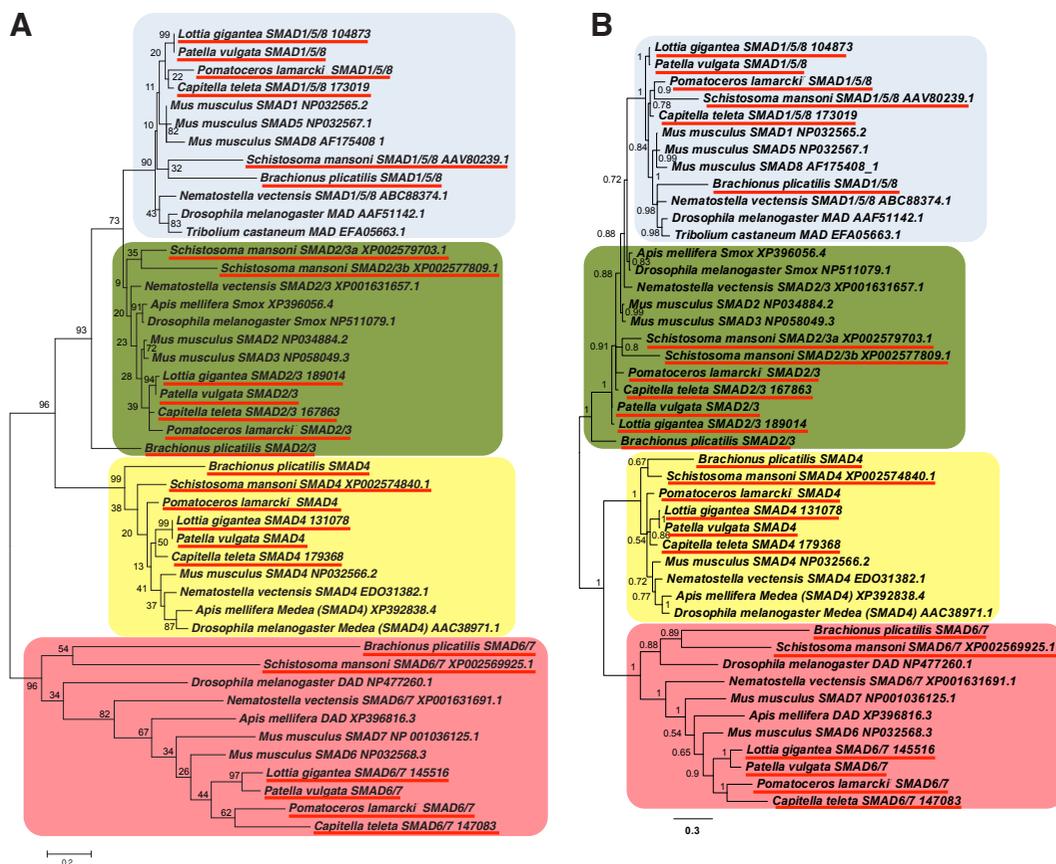

Fig. 6. DAN class interrelationships across the Metazoa, as determined by maximum likelihood (Tamura *et al.,* 2011) and Bayesian (Huelsenbeck and Ronquist 2001) methods. *Alignment generated by MAFFT (Katoh and Standley 2013) using the G-iNS-i strategy. Phylogenies calculated on the basis of an 87 informative 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.*
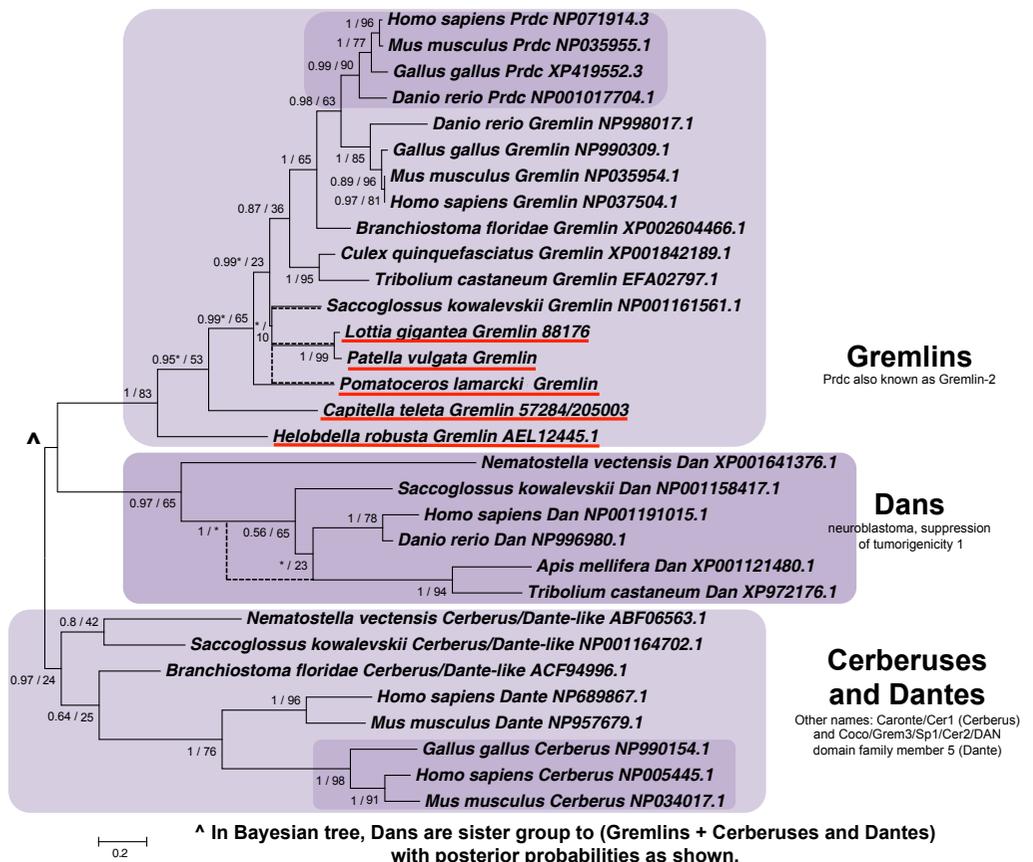
known to be active in *H. robusta*, targeting BMP 2/4 preferentially (Kuo and Weisblat 2011). Much investigation is required before inference can be made as to the wider roles of genes that remain in the Lophotrochozoa, given the wide range of actions that these genes are known to have in more established 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.*
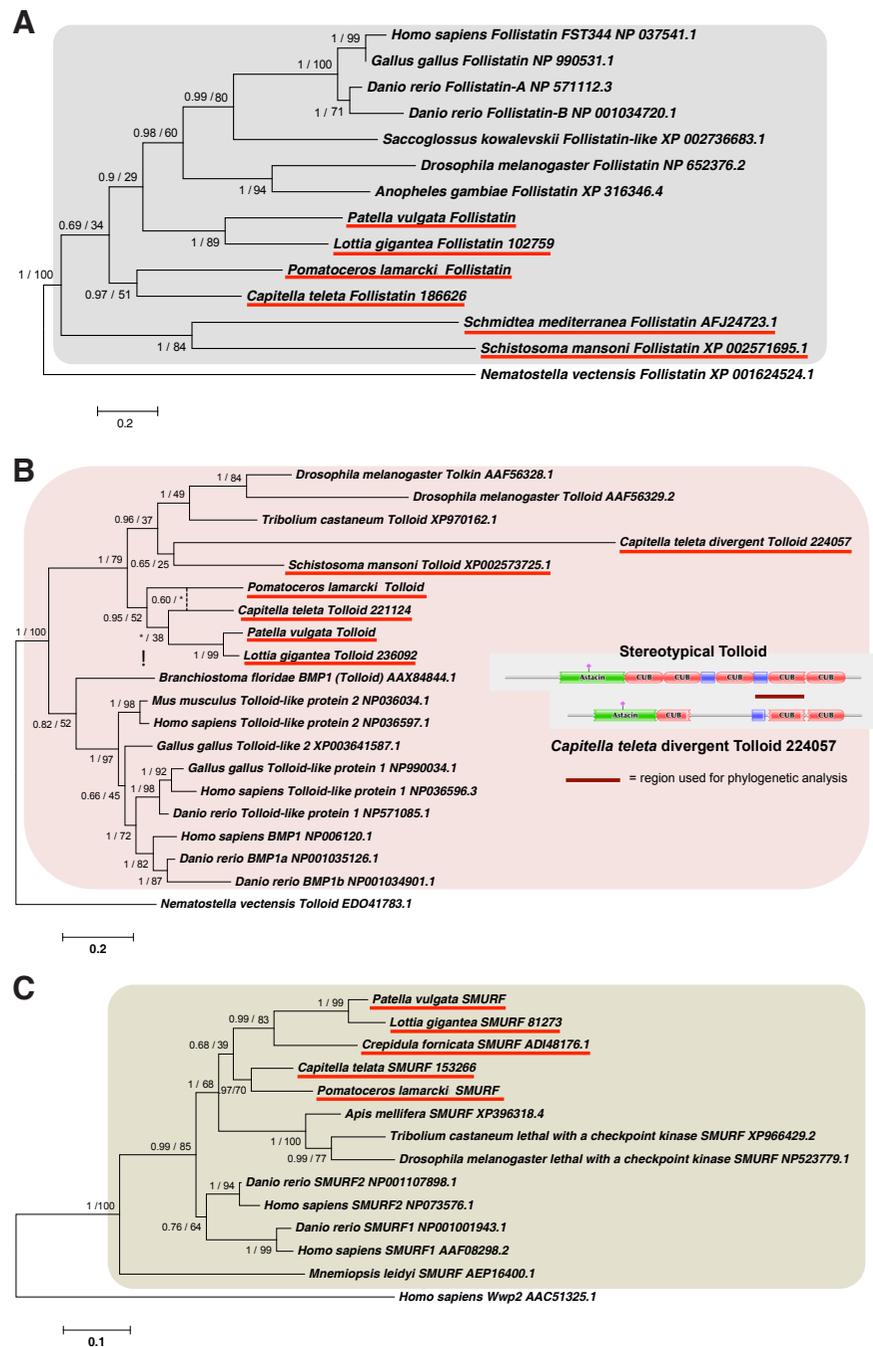
been found in sequenced invertebrate species. Our phylogenetic analysis (Fig. 8C) of these genes from species across the Metazoa suggests a paralogous relationship between the two homologues found in vertebrate model species.

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
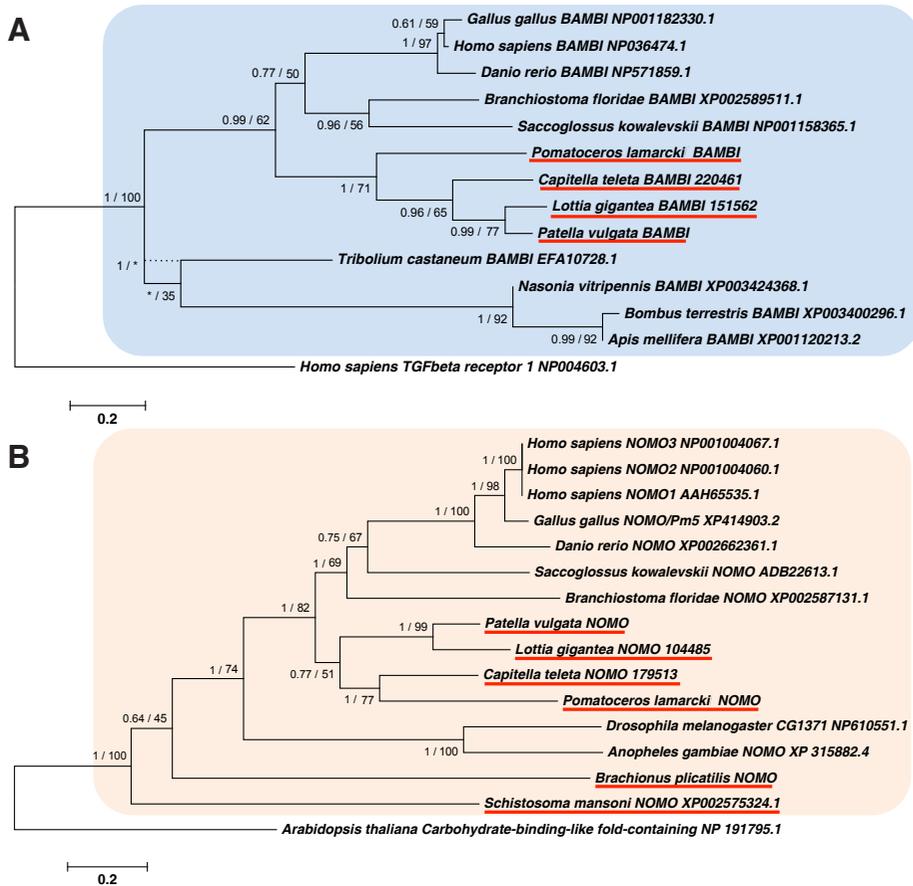

Fig. 9. Bambi (A) and NOMO (B) interrelationships across the Metazoa, as determined by maximum likelihood (Tamura *et al.,* 2011) and Bayesian (Huelsenbeck and Ronquist 2001) methods. *Phylogeny shown is the result of ML analysis, with differences in topology using Bayesian methods indicated with dotted 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 (Homolo-Gene: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
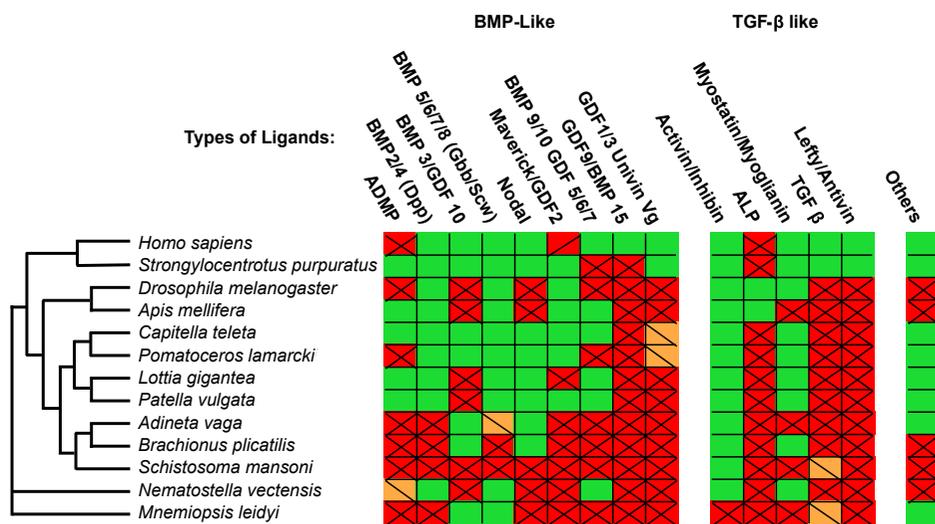

Fig. 10. TGF-β ligand family presence and absence in a range of metazoan species, as determined in the present manuscript or inferred from previously published work. *Only species with clearly informative TGF-β ligand identity known are listed above. Confirmed identity is shown in green (no lines), tentative 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 lat-ter 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.

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
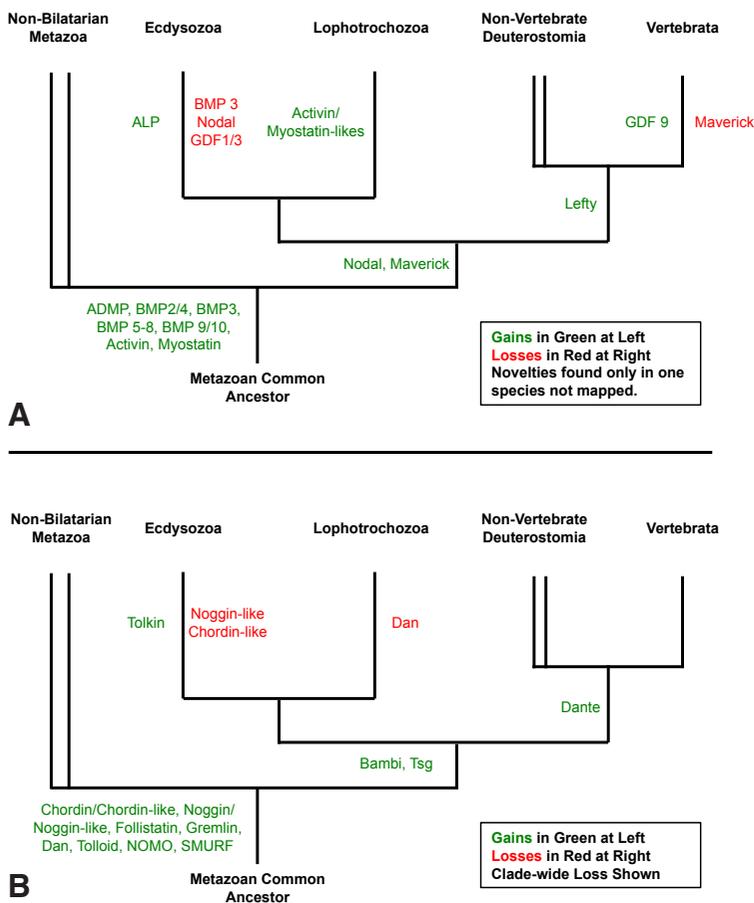

Fig. 11. Gain and loss of TGF-β ligands (A) and modulators of TGF-β signalling (B) across the Metazoa, mapped onto a schematic cladogram of the inter-relationships of these Superphyla. *Position of TGF-β itself is contingent on the true assignation of TGF-β status to* M. leidyi *protein sequence by Pang* et al.*, 2011, as this hypothesis is only weakly supported by our analysis.*

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.

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.

Further Related Reading, published previously in the *Int. J. Dev. Biol.*

Evolution, divergence and loss of the Nodal signalling pathway: new data and a synthesis across the Bilateria.
C Grande, J M Martin-Duran, N J Kenny, M Truchado-Garcia and A Hejnol.
Int. J. Dev. Biol. (2014) doi: 10.1387/ijdb.140133cg

On the origin of pattern and form in early Metazoans
Frederick W. Cummings
Int. J. Dev. Biol. (2006) 50: 193-208
http://dx.doi.org/10.1387/ijdb.052058fc

The first bilaterian organisms: simple or complex? New molecular evidence
J Baguna, I Ruiz-Trillo, J Paps, M Loukota, C Ribera, U Jondelius, M Riutort
Int. J. Dev. Biol. (2001) 45: S133-S134
http://dx.doi.org/10.1387/ijdb.01450133

Nodal signaling and the zebrafish organizer
A F Schier & W S Talbot
Int. J. Dev. Biol. (2001) 45: 289-298
http://www.ijdb.ehu.es/web/paper/11291859/

Functional analysis of the TGFbeta receptor/Smad pathway through gene ablation in mice
M J Goumans & C Mummery
Int. J. Dev. Biol. (2000) 44: 253-266.
http://www.ijdb.ehu.es/web/paper/10853822