

TGF- β Prodomain Alignments Reveal Unexpected Cysteine Conservation Consistent with Phylogenetic Predictions of Cross-Subfamily Heterodimerization

Robert G. Wisotzkey* and Stuart J. Newfeld^{†,1}

*Sema4 Genomics, Stamford, Connecticut 06902 and [†]School of Life Sciences, Arizona State University, Tempe, Arizona 85287-4501

ORCID ID: 0000-0003-1400-7978 (S.J.N.)

ABSTRACT Evolutionary relationships between prodomains in the TGF- β family have gone unanalyzed due to a perceived lack of conservation. We developed a novel approach, identified these relationships, and suggest hypotheses for new regulatory mechanisms in TGF- β signaling. First, a quantitative analysis placed each family member from flies, mice, and nematodes into the Activin, BMP, or TGF- β subfamily. Second, we defined the prodomain and ligand via the consensus cleavage site. Third, we generated alignments and trees from the prodomain, ligand, and full-length sequences independently for each subfamily. Prodomain alignments revealed that six structural features of 17 are well conserved: three in the straitjacket and three in the arm. Alignments also revealed unexpected cysteine conservation in the “LTBP-Association region” upstream of the straitjacket and in β 8 of the bowtie in 14 proteins from all three subfamilies. In prodomain trees, eight clusters across all three subfamilies were present that were not seen in the ligand or full-length trees, suggesting prodomain-mediated cross-subfamily heterodimerization. Consistency between cysteine conservation and prodomain clustering provides support for heterodimerization predictions. Overall, our analysis suggests that cross-subfamily interactions are more common than currently appreciated and our predictions generate numerous testable hypotheses about TGF- β function and evolution.

KEYWORDS Activin; alignments/trees; arm/bowtie/straitjacket; BMP; cleavage site; heterodimer

SECRETED TGF- β family members perform a myriad of tasks during development and homeostasis, while mutations disrupting TGF- β pathways can lead to disease. The mouse genome encodes 33 TGF- β family members, the fly encodes seven, and the nematode encodes five (Kahlem and Newfeld 2009). Structurally, TGF- β family members share an amino-terminal signal sequence, a long prodomain involved in regulation that is cleaved before secretion but remains associated, and a short biologically active ligand that binds

to cell surface receptors. The ligand of TGF- β proteins contains a stereotypical pattern of six cysteines, with a subset containing seven or nine cysteines that form a disulfide bond-based cystine knot structure [reviewed in Hinck *et al.* (2016)].

One means of generating hypotheses for multigene families is to ascertain evolutionary relationships via phylogenetics. This approach has successfully predicted new mechanisms of TGF- β regulation twice. First, Smad linker phosphorylation was predicted (Newfeld and Wisotzkey 2006) then validated by experiment in mice and flies (Fuentelba *et al.* 2007; Quijano *et al.* 2011). Second, monoubiquitylation of Smad4 was predicted (Konikoff *et al.* 2008) then validated by experiment in frogs, mice, and flies (Dupont *et al.* 2009; Morsut *et al.* 2010; Stinchfield *et al.* 2012).

Twenty years ago the first phylogenetic study of TGF- β ligands employed fly, mouse, and nematode proteins, before all three genomes were available (Newfeld *et al.* 1999). To date, all ligand phylogenetic studies have been done with an artificially shortened ligand that begins at the first conserved

Copyright © 2020 Wisotzkey and Newfeld

doi: <https://doi.org/10.1534/genetics.119.302255>

Manuscript received April 26, 2019; accepted for publication December 5, 2019; published Early Online December 16, 2019.

Available freely online through the author-supported open access option.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supplemental material available at figshare: <https://doi.org/10.25386/genetics.11350061>.

¹Corresponding author: Arizona State University, Mail Code 4501, Tempe, Arizona 85287-4501. E-mail: newfeld@asu.edu

cysteine (the cystine knot). Historically this was necessary because the cleavage sites that separate the prodomain from the ligand were not defined. This biased the analysis toward the most highly conserved region. Nevertheless, the resultant clustering of the TGF- β superfamily into two large subfamilies (BMP and Activin + TGF- β) that functionally appeared to rely on distinct sets of receptors and receptor-associated Smads was intellectually satisfying.

The first full-length tree of TGF- β family members utilizing the same species was published 10 years ago, after all three genomes were available (Kahlem and Newfeld 2009). Discrepancies with the previous cystine knot tree were noted. The prodomain sequences responsible for full-length vs. cystine knot tree discrepancies have not been identified.

While the prodomain has long been known to be required for proper folding and dimerization of the ligand (Gentry and Nash 1990; Gray and Mason 1990), formal definition of cleavage sites came later (Degnin *et al.* 2004; Kunnappu *et al.* 2009). Proprotein/latent complex crystal structures of TGF- β 1 (Shi *et al.* 2011), BMP9 (Mi *et al.* 2015), and Inhibin- β a (also called Activin-A; Wang *et al.* 2016) and the solution structure of Myostatin (also called GDF8; Walker *et al.* 2018) have identified functional features such as the Latency Lasso and the bowtie.

We hypothesized that the discrepancy between the full-length and the cystine knot trees was due to conserved prodomain sequences involved in dimerization. However, the perception was that prodomains were to degenerate to be confidently aligned. To test our hypothesis, we developed a new approach that began with a quantitative analysis of each family member from fly, mouse, and nematode that sorted them into one of three subfamilies: Activin, BMP, and TGF- β . Then we employed the biochemically defined consensus cleavage site to separate each full-length protein into ligand and prodomain. We generated annotated alignments to examine structural conservation. Lastly, trees of the prodomain, biochemically defined ligand and full-length sequences were created from each individual subfamily, an Activin + TGF- β subfamily and from all family member alignments.

The implementation of the consensus cleavage site led to the movement of a highly degenerate region between the cleavage site and the first cysteine out of the prodomain and into the ligand. This resulted in a reduction in the resolution of our ligand trees (vs. cystine knot trees), but an increase in resolution of prodomain trees. In our view, cystine knot clustering suggests common receptor binding and common function, while prodomain clustering suggests heterodimerization and common regulation.

In the interest of brevity, we focus our analysis on fly proteins plus interesting observations for nematode family members and mouse Nodal. The prodomain alignments revealed that six structural features are well conserved: three in the straitjacket and three in the arm. Alignments also revealed unexpected cysteine conservation in the “Latent TGF- β Binding (LTBP) LTBP association region” upstream of the straitjacket and in β 8 of the arm in 14 proteins

belonging to all three subfamilies. In the prodomain trees, eight clusters across all three subfamilies were present that were not seen in the ligand or full-length trees, suggesting prodomain-mediated cross-subfamily heterodimerization. Consistency between cysteine conservation and prodomain clustering provides support for our heterodimerization predictions.

Materials and Methods

Sequences and subfamilies

For consistency with our previous papers we focus on the same three species (Newfeld *et al.* 1999; Kahlem and Newfeld 2009). The justification for this approach is that examining genetic model organisms with completely sequenced genomes and an established evolutionary divergence of over a billion years will provide metazoan scale explanatory power and a convenient platform for testing new hypotheses. The newest version of the longest isoform of each TGF- β protein from *Caenorhabditis elegans* (Ce, 5), *Drosophila melanogaster* (Dm, 7), and *Mus musculus* (Mm, 33) was identified. Two species are coelomates with three germ layers and a digestive tract with two openings: *M. musculus* is a deuterostome (blastopore becomes the anus) and *D. melanogaster* is a protostome (blastopore becomes the mouth). *C. elegans* is a pseudocoelomate with three germ layers and a digestive tract with one opening. The split between deuterostomes and protostomes was roughly 964 MYA and between coelomates and pseudocoelomates 1.298 billion years ago (Hedges *et al.* 2004). For consistency with previous papers, mouse GDNF was employed as an outgroup to root all trees. This is appropriate because GDNF shares pattern of cysteines with TGF- β family ligands yet signals strictly via a distinctive ternary complex with Ret tyrosine kinase receptors (*e.g.*, Jing *et al.* 1996). In contrast, Maverick primarily signals through TGF- β receptors but can also bind Ret (Myers *et al.* 2018). This clear distinction in affinity supports our interpretation of data for Maverick. Details on the 46 sequences are in Supplemental Material, Table S1.

Initial separation of fly and mouse TGF- β family members into the two well-known subfamilies Activin/TGF- β and BMP followed Newfeld *et al.* (1999). We then conducted an Informative Sites analyses in MegaX (Kumar *et al.* 2018) to rigorously separate sequences into distinct Activin and TGF- β subfamilies. This had not been done before and led to several changes from previous analyses (Kahlem and Newfeld 2009; Özüak *et al.* 2014). Alignments were generated for the Activin, BMP, and TGF- β subfamilies independently, an Activin + TGF- β combined subfamily, and all family members (five family/subfamilies total).

Separation of full-length sequences into two structural families (prodomain and ligand) was based on identifying the site analogous to the consensus Furin cleavage site in Dpp (Kunnappu *et al.* 2009; RNKR). For proteins where the sequence was not an exact match and/or there was more than choice, we picked the site closest to the first cysteine in the

ligand (Table 2). This approach is more rigorous than past analyses when ligands were defined for convenience at the first conserved cysteine (Newfeld *et al.* 1999). The spacer between the most proximal Furin site and first cysteine in Dpp is 14 residues (Table S3). To validate our cleavage site we checked for conservation in three pairs of congeneric species: *D. melanogaster* and *D. simulans*, *C. elegans* and *C. briggsae*, and *M. musculus* and *M. caroli*. We identified and aligned the region surrounding our chosen cleavage site via BLASTp. The analysis showed that that all fly and mouse cleavage sites are identical in both species, while nematodes showed minor differences in the site in three proteins.

Prodomain, ligand, and full-length trees were analyzed according to subfamily in the main paper. Trees were grouped according to structure (prodomain, ligand, and full-length) in Figures S1–S3. A cystine knot tree for all family members, where the ligand begins at the first cysteine, is included for comparison to the cleavage site defined ligand tree in Figure S2.

Alignments

Sequences from NCBI were aligned with default settings in Clustal Omega at EMBL-EBI (<https://www.ebi.ac.uk/Tools/msa/clustalo/>). Alignments depicting sequence conservation were generated in BoxShade3.21 (ch.embnet.org/software/BOX_form.html) as described (Newfeld and Wisotzkey 2006). The cutoff for shading was an identical or similar amino acid in half of the sequences. Similar amino acids are: D/E, K/R/H, N/Q, S/T, I/L/V, F/W/Y, and A/G (Smith and Smith 1990). A set of complete BoxShade alignments for the prodomain, with ungainly leaders and tails trimmed, are found in Figures S4–S8. Fully unedited prodomain as well as ligand and full-length BoxShade alignments are available upon request.

Activin subfamily: We analyzed 11 sequences (1 Ce, 2 Dm, and 8 Mm) plus mouse GDNF. The prodomain alignment was 983 amino acids including gaps, and there were 185 informative sites without gaps. The ligand alignment was 204 amino acids including gaps, and there were 76 informative sites without gaps. The full-length alignment was 1147 amino acids including gaps, and there were 262 informative sites without gaps.

TGF- β subfamily: We analyzed 12 sequences (2 Ce, 2 Dm, and 8 Mm) plus mouse GDNF. The prodomain alignment was 838 amino acids including gaps, and there were 230 informative sites without gaps. The ligand alignment was 167 amino acids including gaps, and there were 101 informative sites without gaps. The full-length alignment was 929 amino acids including gaps, and there were 358 informative sites without gaps.

Activin + TGF- β subfamily: We analyzed 23 sequences (3 Ce, 4 Dm, and 16 Mm) plus mouse GDNF. The prodomain alignment was 1116 amino acids including gaps, and there

were 24 informative sites without gaps. The ligand alignment was 214 amino acids including gaps, and there were 76 informative sites without gaps. The full-length alignment was 1302 amino acids including gaps, and there were 101 informative sites without gaps.

BMP subfamily: We analyzed 22 sequences (2 Ce, 3 Dm, and 17 Mm) plus mouse GDNF. The prodomain alignment was 787 amino acids including gaps, and there were 335 informative sites without gaps. The ligand alignment was 166 amino acids including gaps, and there were 108 informative sites without gaps. The full-length alignment was 870 amino acids including gaps, and there were 415 informative sites without gaps.

All family members: We analyzed 45 sequences (5 Ce, 7 Dm, and 33 Mm) plus mouse GDNF. The prodomain alignment was 1265 amino acids including gaps, and there were 554 informative sites without gaps. The ligand alignment was 229 amino acids including gaps, and there were 142 informative sites without gaps. The full-length alignment was 1414 amino acids including gaps, and there were 641 informative sites without gaps. Cystine knot alignment was 168 amino acids including gaps, and there were 114 informative sites without gaps.

Phylogenetics

Trees were created in MrBayes3.2 (Ronquist *et al.* 2012; mrbayes.sourceforge.net/). The “prior amino acid model” was set to BloSum (a matrix of empirically derived amino acid substitution frequencies; Henikoff and Henikoff 1992) and the “rate of variation across sites” was set to a gamma distribution (this distribution has an L-shape with a few sites evolving rapidly, while most sites are conserved; Yang 1993). Generation times were 200,000 for all trees except that Activin full-length was 100,000. The sample frequency was 100 with burn-in of 0.25.

For alignments with >150 informative positions (prodomain and full-length for all subfamilies except Activin + TGF- β) a posterior probability of 0.95 is statistically significant. For alignments with fewer informative positions, simulation studies (Alfaro *et al.* 2003) showed that the true tree contained branches with posterior probabilities of 0.50 for 25–50 informative positions (Activin + TGF- β prodomain tree), 0.65 for 50–100 informative positions (Activin ligand, TGF- β ligand, and Activin + TGF- β ligand trees), and 0.85 for 100–150 informative positions (BMP ligand, All ligand, cystine knot, and Activin + TGF- β full-length trees).

Data availability statement

Unedited BoxShade alignments are available upon request. All data necessary for confirming the conclusions are present within the figures, tables, and supplemental information. Supplemental material available at figshare: <https://doi.org/10.25386/genetics.11350061>.

Results

Informative sites analysis and phylogenetics

Given the discordance between previous full-length and cysteine knot trees, we began by placing family members rigorously into subfamilies (Table S1). We started with alignments of three sets of recent mammalian duplications that always cluster together and are always distinct from others representing the TGF- β , Activin, and BMP subfamilies (TGF- β 1–3; Inhibin- β a, β b, β c, and β e; and BMP2 and 4). Note that the phrase “recent mammalian duplicates” indicates only that these duplications are not present in flies and nematodes.

Then, we added sequences one at a time to each subfamily alignment using the most current version of Clustal Omega (McWilliam *et al.* 2013). Each of these “core plus one” alignments was then run through MegaX (Kumar *et al.* 2018) for a quantitative analysis of total alignment length, gap number, and number of informative sites. A sequence that reduced the number of informative sites by the smallest amount was added to that subfamily and the process repeated until every sequence was added. We did not find any sequences with similar effects on multiple subfamilies as would be expected if there were additional subfamilies.

To our knowledge, this is the first rigorous distinction of sequences within the TGF- β family based on alignment and not a tree-building algorithm. This removes a set of phylogenetic assumptions from the process. Overall, the alignments showed that as a group the TGF- β and Activin subfamilies are just as distinct from each other as they are from BMP, further indicating that there are three separate subfamilies. In the big picture, subfamily separation predates the divergence of flies, mammals, and nematodes, as each subfamily has at least one protein from each species.

The clear distinction between the TGF- β and Activin subfamilies in our sequence analysis is wholly consistent with structural differences between the subfamilies. For example, the TGF- β 1 prodomain crystal structure contains a “bowtie” formed by β 8 and β 9 as part of the closed-ring conformation of the dimer. The bowtie contains cysteines that facilitate dimerization by linking two arm domains together (Shi *et al.* 2011). The bowtie is missing from Inhibin- β a (Wang *et al.* 2016) in the Activin subfamily, whose prodomain structure displays a cross-armed conformation, and from BMP9 (Mi *et al.* 2015), whose prodomain structure exhibits a widely open conformation. Note that BMP9 in this report is present via its synonym GDF2.

Our informative sites analysis led to firm subfamily placement for all proteins in each species. For nematodes, TIG-3 was confirmed in the Activin subfamily; UNC-129 and DAF-7 in the TGF- β subfamily; and TIG-2 and DBL-1 in the BMP subfamily. For flies, the four non-BMP proteins were confidently placed in the Activin (Activin and Myoglianin) and TGF- β subfamilies (Maverick and Dawdle). For mice, Nodal is firmly placed in the BMP subfamily yet our trees will suggest a hypothesis to explain its ability to signal through the

Activin pathway receptors ActRIB and ActRIIA/B and signal transducer Smad2 [reviewed Schier (2009)].

Employing a Bayesian approach (Ronquist *et al.* 2012), these rigorous subfamily alignments were built into trees. Confirming our initial hypothesis, these trees were able to resolve conflicts between full-length and cysteine knot trees from prior publications. For example, here Gbb and Screw cluster in all trees indicating a recent duplication rather than the complex relationships that were shown previously. In addition, the current approach is better able to discern subtle distinctions between family members. For example, initial placement into subfamilies via informative sites led to 22 BMP proteins that is extended in the current full-length tree of all family members to 26 proteins. This 26-member BMP cluster encompasses two TGF- β and two Activin subfamily members, most likely as a result of previously unsuspected prodomain similarity.

Cleavage site fidelity and spacer variability

Parsing the full-length sequences into prodomain and ligand, before tree building, based on the consensus Furin cleavage site was not hard (Table S2). Only two of the 46 sequences (45 TGF- β family members plus the mouse GDNF outgroup) did not contain a region with strong similarity to the consensus RX[R/K]R (Degnin *et al.* 2004) upstream of the first cysteine of the ligand. TIG-3 (Activin subfamily) and Maverick (TGF- β subfamily) have only a single R in the right place. In cases where multiple cleavage sites were identified (*e.g.*, Dpp; Kunnapuu *et al.* 2009), we chose the closest R to the first cysteine for the separation. We conducted a similar analysis of known Tolloid cleavage sites that did not reveal any conservation.

To validate our choice of Furin cleavage sites we checked them for conservation in three pairs of congeneric species: *D. melanogaster* and *D. simulans*, *C. elegans* and *C. briggsae*, and *M. musculus* and *M. caroli*. The analysis showed that that all fly and mouse cleavage sites are identical in both species (43 of 46 proteins). Nematodes showed minor differences in three cleavage sites (DAF-7, TIG-2, and TIG-3). An examination of the consensus divergence times between each pair revealed that the fly estimate is 4.7 MYA, the mouse estimate is 4.8 MYA, and the nematode estimate is 60.2 MYA (Time-tree.org; Hedges *et al.* 2015). The finding of minor differences in a subset of nematode sequences (three out of five) is unsurprising given the much larger distance between the two species. The high frequency (94%) of identity across species in the cleavage site employed for our analysis provides increased confidence in its validity.

We found that the spacer region between the most proximal cleavage site and the first cysteine was hypervariable in length and content (Table S3). Length variation spanned the range from 2 residues (BMP15; TGF- β subfamily) to 80 residues (BMP3; Activin subfamily). However, in hypervariable regions any conservation likely is functional or evidence of recent duplication. For example, 8 of 11 Activin subfamily members have an acidic residue (D/E) immediately upstream of the first cysteine in the ligand (72% with 10% expected by chance). Only 1 of the other 35 sequences has a glutamic acid at this position.

The BMP subfamily has no obvious amino acid conservation but length identity is visible in the recently duplicated Gbb and Screw as well as the mammalian duplicates BMP6, 7, 8a,b. The TGF- β family is home to the two newest duplications as revealed by the presence of both sequence and length identity for Lefty1 and 2 and TGF- β 1–3. Overall spacer hypervariability is consistent with structural data showing it sits outside the prodomain-ligand complex (Mi *et al.* 2015).

The transition from a cystine knot defined ligand to a biochemically defined ligand had a dramatic effect on trees for each region. There was a loss of resolution in the ligand tree, as many proteins became unaffiliated. There was a concomitant increase in resolution of the prodomain tree, as a greater number of meaningful clusters are present when compared to the prior full-length tree (Kahlem and Newfeld 2009). Loss of resolution in the ligand tree is of little consequence as a cystine knot alignment of all family members yielded a familiar tree. The gain in resolution for the prodomain revealed numerous unexpected cross-subfamily clusters.

Trees and alignments of subfamily prodomains, ligands, and full-length proteins

Here data are discussed according to subfamily. For a distinct perspective, the supplemental figures display trees organized by structure (prodomain Figure S1, ligand including cystine knot Figure S2, and full-length Figure S3) and expanded alignments for each subfamily (Figures S4–S8).

Activin subfamily trees

This subfamily (Figure 1) is built upon the four Inhibin- β proteins that cluster together in all trees based on their recent origin and common ability to form heterodimers with Inhibin- α in the TGF- β subfamily (Walton *et al.* 2009). The significant cluster of Activin and Myoglianin seen only in the prodomain suggests that they have common regulation. The significant cluster of Activin with the four Inhibin- β proteins in the ligand suggests a common function. The significant cluster of Myoglianin and Myostatin/GDF11 in the full-length tree also suggests common function. Overall for the Activin subfamily, the similarity between the ligand tree and the full-length tree indicates that functional relationships of the ligand are driving its evolution.

Activin subfamily prodomain structural conservation

Known features such as α -helices and β -sheets were located on the alignment revealing pockets of structural conservation in the annotated Activin alignments (Figure 2 and Figure S4). The locations for α 1, the Latency Lasso, and α 2 are based on Inhibin- β a (Wang *et al.* 2016). The locations of the remaining features derive from our alignment of the Activin + TGF- β subfamily following TGF- β 1 (Shi *et al.* 2011).

The four features of the straitjacket domain (α 1, the Latency Lasso, α 2, and β 1) show the most conservation. There is a set of nine I/L/V residues, of which four are universal. There is also a universal proline in the Latency Lasso. β 1

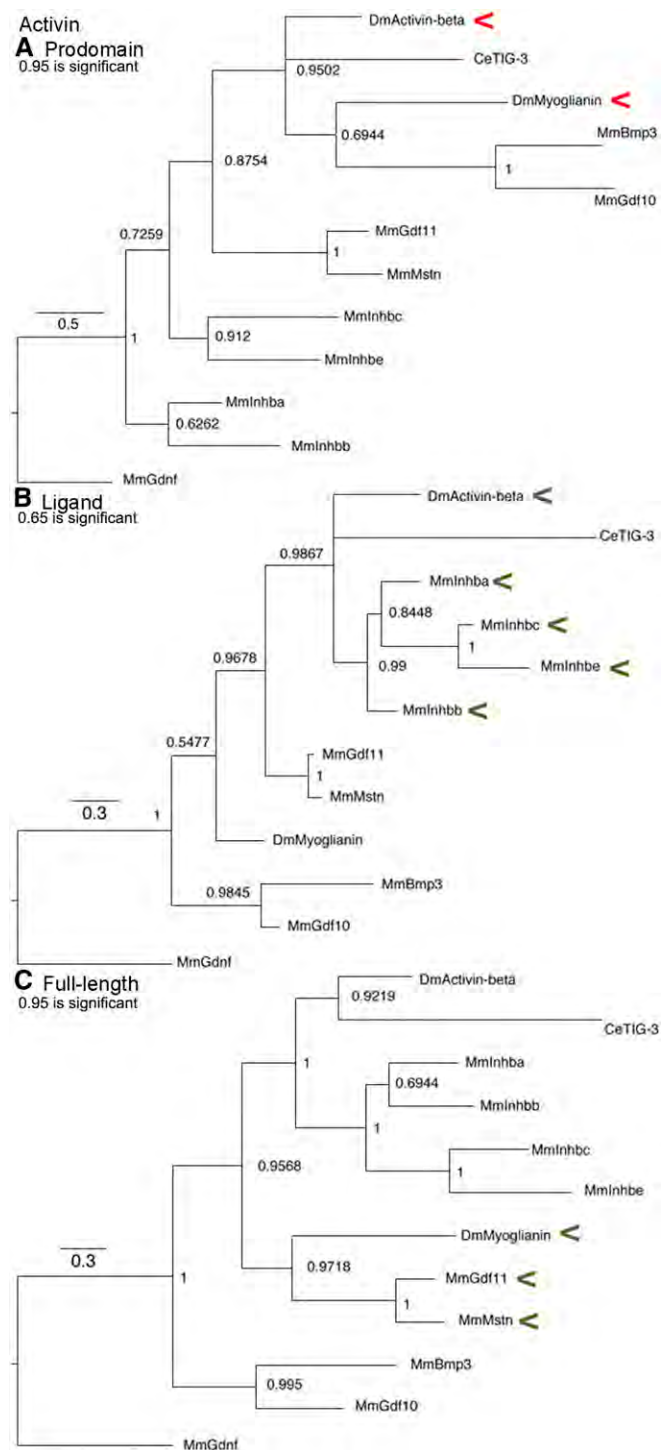


Figure 1 Activin subfamily trees. Bayesian trees of 11 sequences plus the outgroup are displayed. Accession numbers are in Table S1. Branch lengths are drawn to scale and a scale bar indicates the number of amino acid substitutions per site per unit length. Nodes with posterior probabilities ~ 0.50 are indicated. Red arrowheads indicate a cluster that may reflect common regulation and green arrowheads a cluster that may reflect common function. (A) Prodomain nodes ~ 0.95 are significant. The significant cluster of Activin and Myoglianin is unexpected. (B) Ligand nodes ~ 0.65 are significant. The significant cluster of Activin and the four Inhibin- β proteins was expected. (C) Full-length nodes ~ 0.95 are significant. The significant cluster of Myoglianin and Myostatin/GDF11 was expected.

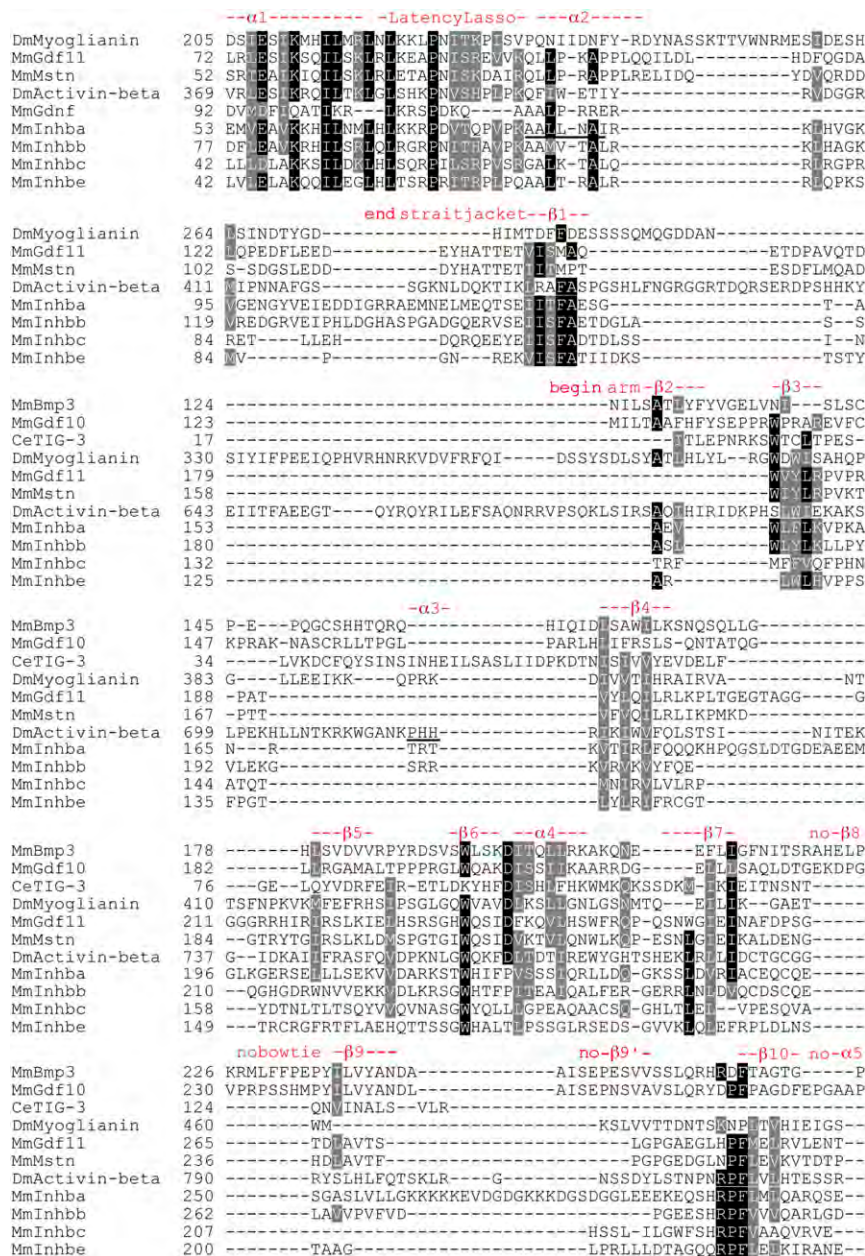


Figure 2 Focused Activin subfamily prodomain alignment indicating structural conservation. Sequences present only, as gaps for any row were omitted. Numbering is accurate. Black shading indicates an identical and gray shading biochemically similar amino acids at that position. As there is no available structurally annotated full-length sequence of Inhibin- β a, the locations and naming of structural features are derived from our alignment of the Activin + TGF- β subfamily that follows TGF- β 1 of Shi *et al.* (2011). Underlined α 2 indicates a location distinct in this subfamily from α 2 in TGF- β 1 and the underlined α 3 suggests that this feature of TGF- β 1 may be absent in the Activin subfamily. β 9' is not obvious and may be unique to the BMP subfamily.

contains two conserved I/L/V residues and a phenylalanine. In the arm domain, the helices β 2–10 and α 4 show less conservation. Most notable are three I/L/V residues in β 4, a near universal tryptophan in β 6 and near universal I/L/V residues in α 4 and β 7. This correlation of amino acid conservation with structural features had not been demonstrated rigorously in the Activin subfamily.

TGF- β subfamily trees

This subfamily (Figure 3) is built upon the three TGF- β proteins that cluster together in all trees, based on their recent origin and common regulation by LTBP (Rifkin *et al.* 2018). Neither prototypical TGF- β nor LTBP are present in flies and neither Maverick nor Dawdle has any relationship with them. The significant cluster of Dawdle and Inhibin- α seen only in

the prodomain suggests common regulation. Given the ability of Inhibin- α to form heterodimers and the previously noted cluster of Activin and the Inhibin- β group, Dawdle is a candidate as a heterodimerization partner with Activin. The ligand tree shows established clusters such as TGF- β 1–3 and Lefty1, 2 and new clusters such as Maverick with GDF15. Overall for the TGF- β subfamily, the full-length tree is distinct from the ligand and prodomain trees, indicating that functional and regulatory relationships are equally driving its evolution.

TGF- β subfamily prodomain structural conservation

Areas of structural conservation are evident in the annotated TGF- β alignments (Figure 4 and Figure S5). The locations and names of features derive from TGF- β 1 (Shi *et al.* 2011). The four features that compose the straitjacket domain (α 1, the

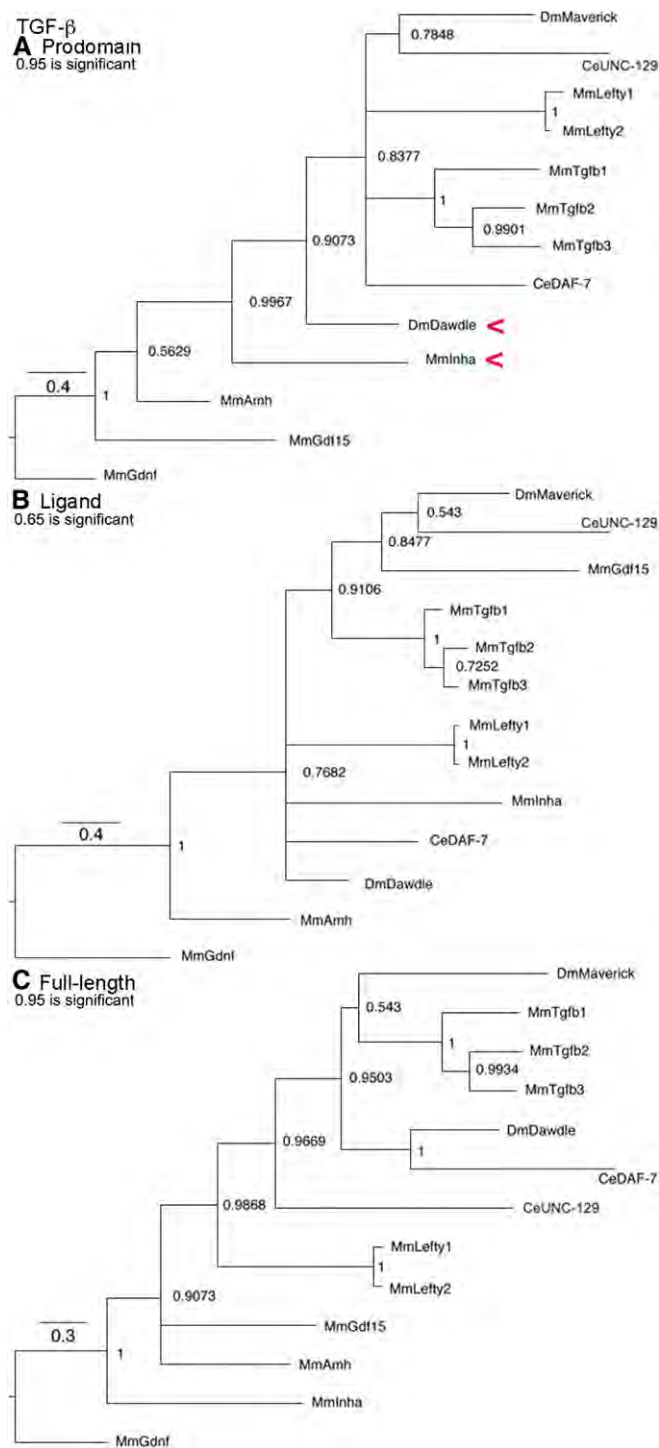


Figure 3 TGF- β subfamily trees. Bayesian trees of 12 sequences plus the outgroup are displayed as in Figure 1. Red arrowheads indicate a cluster that may reflect common regulation. (A) Prodomain nodes ~ 0.95 are significant. The significant cluster of Dawdle and Inhibin- α was unexpected. (B) Ligand nodes ~ 0.65 are significant. The significant clusters of TGF- β 1–3 and Lefty1 and 2 were expected. (C) Full-length nodes ~ 0.95 are significant. The significant clusters of TGF- β 1–3 and Lefty1 and 2 were expected.

Latency Lasso, α 2, and β 1) show less conservation than in the Activin subfamily. The first three features contain a set of nine

prominent I/L/V residues with only one near universal. The proline in the Latency Lasso is only modestly conserved and β 1 contains only a modestly conserved F/Y.

At the 5' end of the arm domain, there are places within helices β 2–6 and α 4 that show more conservation than the Activin subfamily. β 2 contains a near universal F/W, while β 3 contains a near universal Alanine and I/L/V. β 6 and α 4 each have a near universal tryptophan and a near universal I/L/V. At the 3' end of the arm, unexpectedly in β 8 (part of the distinctive bowtie) Maverick and Dawdle have a pair of cysteines that align with those in TGF- β 1–3, although the spacing is not the same (CxxC vs. CxC). β 10 is modestly conserved. The distinct patterns of conservation in the Activin and TGF- β subfamilies support the idea that they are separate.

Activin + TGF- β subfamily trees

The combined Activin + TGF- β subfamily (Figure 5) contains previously unsuspected relationships. The uniquely low threshold for node significance in the prodomain tree again demonstrates the distinct nature of these two subfamilies. The value needed for a node to attain statistical significance depends upon the number of informative sites in the underlying alignment. An informative site is one where an amino acid is present in virtually every family member with a different residue in at least two proteins. Thus, the large number of gaps needed to achieve prodomain alignment in the combined subfamily led to the smallest number of informative sites (i.e., no other tree has a significance threshold as low as 0.50).

In the prodomain, one cross-subfamily cluster contains three of the four fly family members (Activin, Maverick, and Myoglianin). Further, as a group they are tightly tied in a second cross-subfamily cluster to TGF- β 1–3 and Myostatin/GDF11. The group of four Inhibin- β proteins is the next closest cluster. Dawdle ends up as a solo next to Inhibin- α . The clustering of the three fly proteins with Dawdle as an outlier is reminiscent of the Inhibin- β group's relationship with Inhibin- α , proteins that are known to heterodimerize. The analogy is that Dawdle can bind to Activin, Maverick, and Myoglianin and that these heterodimers have a distinct function (possibly inhibition) from the four homodimers (possibly activation).

On the other hand, in the ligand and full-length trees there are no cross-subfamily clusters, the Activin subfamily and TGF- β subfamily relationships are simply recreated. For example, Activin is with the four Inhibin- β proteins and Myoglianin is with Myostatin/GDF11. Overall for the Activin + TGF- β subfamily, similarity between the ligand tree and the full-length tree indicates that functional relationships of the ligand are driving its evolution.

Activin + TGF- β subfamily prodomain structural conservation

The combined Activin + TGF- β subfamily alignment contains four features where the subfamilies differ, further

		----- <u>α1</u> ----- <u>-LatencyLasso-</u> ----- <u>α2</u> -----
CeUNC-129	15	----IANCAKVDV-----LINETRDLHFKSSD-----PNVTSFHRSSHT-----
CeDAF-7	19	GLTFNCTNSGVCIEKMKQHRTEYIRNEIDQLNMKEAPKGLKPMDEPKSVYLEMY---R
DmMaverick	245	-----SKRVDETRKHLVLKGLGKKLPDMRKVNISQAEYS-----S
MmTgfb1	27	AAG-LSTCKTHDMELVKRRKRIEARGQILSKLRITASPPSQGEVPPGPIPEAVLALYNSTR
MmTgfb2	18	ALS-LSTCSTHDMQFMRKRIEARGQILSKLRITSPPEDEY-PEPDEVPPPEVISIYNSTR
MmTgfb3	21	SLS-LSTCTTHDFGHIKKKRVEARGQILSKLRITSPPEPS-V-MTHVPYQVLALYNSTR
MmLefty1	20	EAL-T-----GEQ-----VLGSLQQLQDQFPVLDAKADVEGVIPSHVRTQYV-
MmLefty2	20	AAM-T-----BEQ-----VLSSLQQLQDSQAFTLDSADVEEMAIPTHVRSQYV-
MmInha	204	CSC-SGRPET-----
DmDawdle	233	CES-NRQVEHTEEQOLTHLRIEFYKQQLLEKLRKESPKVSA---VELPKPIFDGM----
MmGdf15	118	AYR-VHRAALLTPTARPWDITRPL-KRALSLRGPRAPALRLRL---TPPPDLA-----
MmAmh	351	AAT-EREPMPIHGPAAPWAAGLQF-RVAVELQAAASELRDLPLPTAPPLARL----
		endstraitjacket--β1--
CeUNC-129	53	-----TEHMKNLVENFIDEDSNEDGNLVRAIEPAVGKFEGQEVLVFDVE
CeDAF-7	76	DLLEKDEQDM-----GVEMSFYTAKD-----
DmMaverick	282	KYIEYLSRLR-----SNQEKGSYNNFMGASFT-----
MmTgfb1	86	DRVAGESAD-----PEPEADYYAKEV-----
MmTgfb2	76	DLLEKASRRAAACERERSDEEYAKEV-----
MmTgfb3	78	ELLEEMHGEREEGCTQETSESEYAKEI-----
MmLefty1	63	ALLOHSHASR-----SRGKRFSQNLREVA-----
MmLefty2	63	ALLOGSHADR-----SRGKRFSQNLREVA-----
DmDawdle	285	TLSHPDDST-----KNK-ELDDYARTS-----
MmGdf15	167	---MLPS-----GGTQLELR-----
MmAmh	405	LALCPNDSR-----SSGDLRALLL-----
		beginarm--β2-- <u>α3</u> -- <u>β3</u> --
CeUNC-129	143	TLKKIRVGGDENLE-----EYK---VINDATKS-VFDSYHLDAKQAVFRITREHSMRPPYA
CeDAF-7	102	-----NPSQLVAKEDTND-----LERSDLQATLTLSIEIPAK----
DmMaverick	356	---L---LRGEQDT-----MNILLHPTNAQDANFHHDKDEANVRMLLYSSSL----
MmTgfb1	126	-----DISH---SIY---MFENTSDIREAVPEPLSRALRLQRLKS-----
MmTgfb2	136	SQV---LCGYLDAIPPTFYRPYFRIVREDYST---MEKNASNLVKAERFRFRLQNPK----
MmTgfb3	126	-----GITSKVFRFNVSS---VEKNGTNTFRAEFRVLRVNPNP-----
MmLefty1	95	-----STHLLVFGIEQ---RLPPNSELVQAVLRVFOQEPVPR----
MmLefty2	95	-----STHLLVFGIEQ---RLPPNSELVQAVLRVFOQEPVPR----
DmDawdle	327	-----SNPSMCFTEKIDD---ADAEGFDVSTAVLWLFKNKQNR----
		-----β4-----β5-----β6-----α4--
CeUNC-129	249	-----LDNFDREPIKRNKGNKLSISEEISSDVMQGGCEPISREE-----
CeDAF-7	138	GMLQDVQVQVYEKNE---EG-S-MGEMVTSGIFA---TK-GSERISIQLPIDTKSWFTIS
DmMaverick	441	NSSQQLTKVYQQLLSANRRKITSRKTEFGNVGF--QETRTQWIEFD-VTKAVRSWLNKS
MmTgfb1	163	---SVEQHVELYQKYS---NNSWRYIGNALL---TPTDTPALSLFD-VGVVROWLNQG
MmTgfb2	187	ARVAEQRIELYQILKSKLTSPTQRYVDSVV---KTRAEGEGLSLFD-VTDVQENLHHK
MmTgfb3	161	SKRTEQRLELFQILRPDEH-IAQORYGGGNL---PTRGTAEGLSLFD-VTDVRENLRR
MmLefty1	140	SARARVTEWLRFRD---EGSNRTA-LIDSRLV---SIHESGAKAFD-VTEAVNFWQQLS
MmLefty2	140	SARARVTEWLRVRE---EGSNRTA-LIDSRLV---SIHESGAKAFD-VTEAVNFWQQLS
DmDawdle	370	-TSAQQTIVVSEVEV--DQKDSKYLSAATTHAIQSVNVQDENMKID-TEWPKKHVISHG
		-----β7-----β8-----bowtie-----β9-----
CeUNC-129	289	-----
CeDAF-7	190	PI-QGIFVKAMLDGRNV-----ALHPQ
DmMaverick	498	HENLGIE--IQCDKCKSIGA--RILSDFSPSTPPRSTASSDEHLNLMPLVNIIGHGTLNS
MmTgfb1	212	DGIQGFRFSAHCSCDSKDNK-----LHVEING---I
MmTgfb2	243	DRNLGFKISLHCCCTFVPSNNYIIPNKSEE-----LEARFAGIDGT
MmTgfb3	216	ESNLGLEISIHCPCHTFQPNGD-ILENVHEV-----MEIKFKGVNDNE
MmLefty1	192	RP--R-----
MmLefty2	192	RP--R-----
DmDawdle	426	EL--SHLIQITCGGCDVSDMEEIISVD-----
		no-β9'--β10--α5--
CeUNC-129	289	-----R---
CeDAF-7	211	QTTADVNDMRLQL-----S---TRPKG--S-RKRR---
DmMaverick	554	QQHGDADIHQIMLTNNRSDQYV-----HHRSNHDSTWR-----
MmTgfb1	240	SPKRRGDLGTIH-----DMNRFFLLMATPLERAQHLHSSRHRR-----
MmTgfb2	285	STYASGDQKTIKSTRKTSKGTTHLLMLLPSYRLSQSSRRKKR----
MmTgfb3	257	DDHGRGDLGRLK---KQKDHNHHLMMIPPHRLDSPGQGSQRKKR----
MmLefty1	195	-----QELLL-QVSVQ---REHLGPGTWSSHKLV---
MmLefty2	195	-----QELLL-QVSVQ---REHLGPGTWSSHKLV---
MmInha	213	-----TFELVAHTRAR--APSAGE--RARR----
DmDawdle	451	-----KDYRFFVIDMQNR--RR-----

Figure 4 Focused TGF-β subfamily prodomain alignment indicating structural conservation. Sequences presented as in Figure 2. Numbering is accurate. The locations and naming of structural features are derived from TGF-β1 of Shi *et al.* (2011). The underlined α2 indicates a location distinct from α2 in Inhibin-βa. Underlined α3 and β3 indicate these features are in the reverse order vs. the Activin subfamily. The underlined features β7 and β9 of TGF-β1 do not appear conserved outside its two siblings. Unexpectedly, in β8 Maverick and Dawdle have a pair of cysteines (red underline) that align with those in TGF-β1-3 (black underline), although the spacing is not the same (CxxC vs. CxC). β9' is not visible and may be unique to the BMP subfamily. The underlined α5 does not appear conserved.

supporting the conclusion that these are distinct (Figure 6 and Figure S6). The locations and names of features derive from TGF-β1 (Shi *et al.* 2011). The first two features of the straitjacket (α1 and Latency Lasso) are conserved. These contain seven I/L/V residues with one near universal. The proline in the Latency Lasso is also near universal. Each subfamily has a distinct location for α2 and for β1, although neither of the versions of α2 or β1 are conserved.

At the 5' end of the arm, β2 contains a near universal phenylalanine. β3 is the third feature distinct in the Activin and TGF-β subfamilies. β3 shows conservation in the TGF-β pattern that pulls in several Activin subfamily members anchored by an alanine and an I/L/V. β4 has three near universal I/L/V residues. β6 has a near universal tryptophan and α4 a near universal I/L/V. At the 3' end of the arm β7 is the fourth feature that is distinct, it shows conservation in the

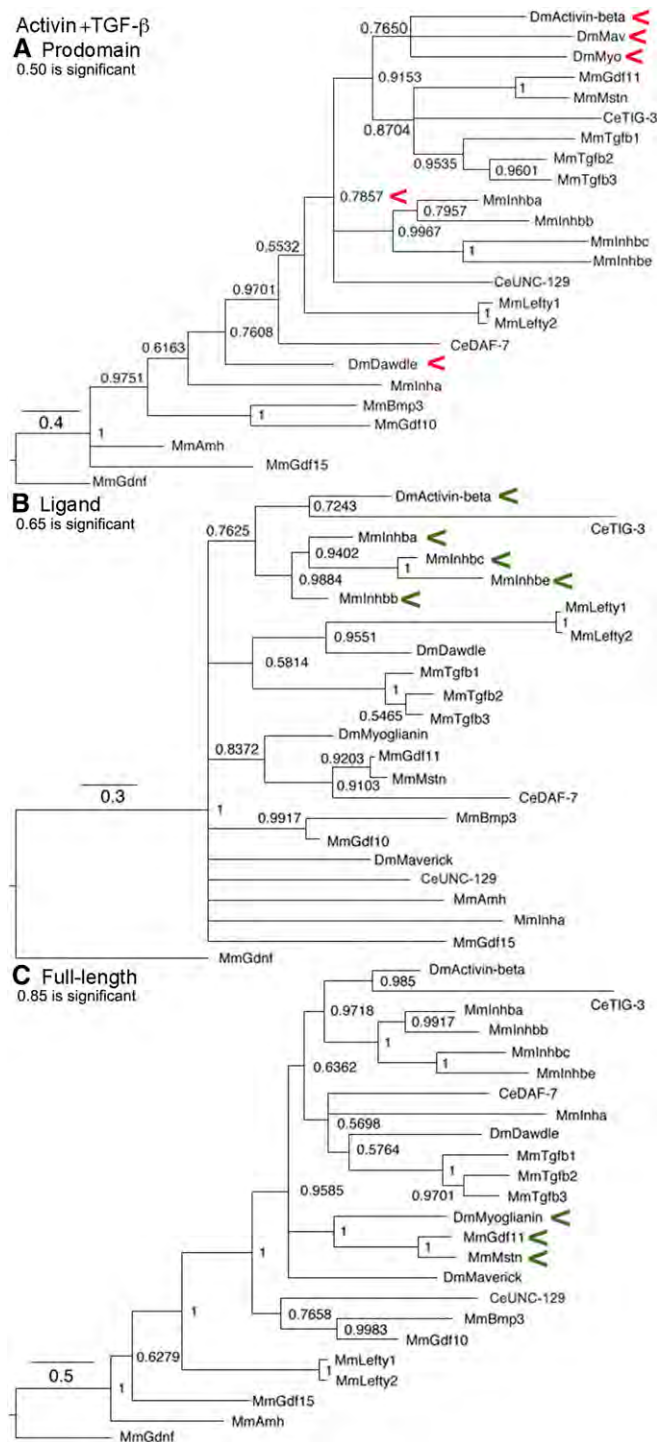


Figure 5 Activin + TGF- β subfamily trees. Bayesian trees of 23 sequences plus the outgroup are displayed as in Figure 1. Red arrowheads indicate a cluster that may reflect common regulation and green arrowheads a cluster that may reflect common function. (A) Prodomain nodes ~ 0.50 are significant. The significant cluster of Activin, Maverick, and Myoglianin that is clustered with the four Inhibin- β proteins and Dawdle's location near Inhibin- α were unexpected but consistent with cysteine conservation in the "LTBP-Association region" and $\beta 8$. (B) Ligand nodes ~ 0.65 are significant. The significant cluster of Activin and the four Inhibin- β proteins was expected. (C) Full-length nodes ~ 0.85 are significant. The significant cluster of Myoglianin and Myostatin/GDF11 was expected.

Activin pattern that pulls in several TGF- β subfamily members. $\beta 10$ has a near universal proline, phenylalanine, and three I/L/V residues.

Again unexpectedly in $\beta 8$, Maverick and Dawdle in the TGF- β subfamily are joined by Activin, Inhibin- βa , and Inhibin- βb in the Activin subfamily, with a pair of conserved cysteines having the same spacing (CxxC vs. CxC in TGF- $\beta 1-3$). In all eight proteins the first cysteine is aligned. The fact that Activin, Dawdle, Maverick, Inhibin- βa , and Inhibin- βb have a pair of similarly spaced cysteines in $\beta 8$ that mediates dimerization in TGF- $\beta 1$, is consistent with the prodomain clusters that suggested cross-subfamily heterodimerization of Activin with Dawdle and Maverick. Importantly, the cysteines suggest a biochemical mechanism by which heterodimerization can be achieved.

BMP subfamily trees

This subfamily (Figure 7) is the largest of the three and is built upon the BMP2/BMP4 proteins that cluster together in all trees based on recent origin. An important finding is that in all trees Gbb and Screw are in a cluster, with the same statistical significance as BMP2/4 and other recent duplications. It appears that Screw resulted from divergence after the duplication of Gbb, uniquely in the lineage leading to *Drosophila*. For example in *Aedes* mosquitos, also a Dipteran, there is no Screw but instead two copies of Gbb (Leiber and Luckhart 2004). Both Gbb and Screw form heterodimers with Dpp during development (e.g., Shimmi *et al.* 2005).

A corollary of Gbb/Screw clustering is that the within subfamily clustering of Gbb with mouse BMP5-8a,b proteins is now extended to Screw as is shown in all trees. A second corollary is that each of the BMP5-8a,b proteins may have the ability to heterodimerize with BMP2/4 yielding as many as 10 possible combinations. To date, only two of these heterodimer pairs have been reported: BMP2/BMP7 in zebrafish dorsal-ventral axis formation (Little and Mullins 2009), and BMP2/BMP6 in mammalian osteogenesis (Loozen *et al.* 2018). Outside this group, heterodimers of mammalian BMP10/GDF9 regulate vascular remodeling (Tillet *et al.* 2018).

Nodal has distinct but not significant clusters in the prodomain with GDF5-7 that link significantly to two mammalian pairs BMP15/GDF9 and GDF1/GDF3 and the triplet GDF5/GDF6/GDF7. Heterodimers of BMP15/GDF9 were seen *in vitro* in rat follicle cell assays that signaled through a cross-subfamily complex of the BMP Type II receptor BMPR2 and the Activin Type I receptor ACVR1B (McIntosh *et al.* 2008). It was recently reported that Nodal heterodimers with GDF1 are required for mesoderm induction in zebrafish (Montague and Schier 2017). Based on extensive coexpression of GDF1 and its duplicated partner GDF3, the authors propose Nodal/GDF3 heterodimers are functional in other developmental contexts.

Interestingly, the overall topology of the BMP prodomain tree is different from the others. In the ligand and full-length trees there are two asymmetric secondary clusters, everyone

		straitjacket--α1-----	--latencyLasso--	-----α2-----
DmMaverick	245	SKRVDETR-LKHLNPKGCGKKL-EDVRKVNISQ----	EYSSKYIEYLSRLRS-----	
DmMyoglianin	202	AKVDSHSS-LKMHILMLNKKL-ENNTKPI-----	SVQNIIDNFYRDYASS-KT•	
MmGdf11	69	SRELHHS-LKSLILSKRLKEA-ENHSREVVKQLPKAPPLQQLDLHDFQGDALQ-P•		
MmMstn	49	TRYSPHA-LKIQILSKRLPETA-ENHSKDAIRQLLPRAPPLRELIDQYDVQDDDS-S•		
CeTIG-3	1	MSTSR-----		
MmTgfb1	41	VKKRRTA-LKQILSKRLASP-ESQGEVPPG-----	PLPEAVLALYNSTRDRVA-G•	
MmTgfb2	32	FMRKRHA-LKQILSKRLTSP-PEDY-PEPD-----	EVPPEVISIYNSTRDLLQ-E•	
MmTgfb3	35	IKKKRHA-LKQILSKRLTSP-PEPS-V-MT-----	HVPYQVLAALYNSTRELLE-E•	
CeUNC-129	25	-----L-INETRDLLHFKSSDPM-----	TSFHS-----	
CeDAF-7	34	MKQHRTA-YLRNETIDDLNKEA-EKGLKPMDEPMKSVYLE-M-YRDLLEK-D-----	EQ•	
MmBmp3	56	VSEHMLWLYDSYSGSSRQATRT-EGSQLPGQPPL-----		
MmGdf10	71	VAIHMLRYEKNRRG-----	APP-----	
MmInhba	50	SQPEMVA-VKHHINMLHKKR-EDVTPVPKAAALLNAIRKL-----	H-----	VG•
MmInhbb	74	VDGDFHA-VKHHILSKRLRGR-ENTHAVPKAAMVTAIRKL-----	H-----	AG•
MmInhbc	39	QRELLHL-AKKSILDKLHESQR-FISRPVSRGALKTALQRL-----	R-----	GP•
MmInhbe	39	ERALVHL-AKQQLIEGHHSR-ERTRPLPQAALTRALRL-----	Q-----	PK•
DmActivin-beta	546	MNE-----	NALKKSTYFIDINHSDNKTHTG•	
MmLefty1	27	-----LGSLLQQLDQPP-EVLDKADVEGMVPSHVRTQVALLQHSAS-----	R•	
MmLefty2	27	-----LGSLLQQLDQSA-ETLDSADVEEMAIFTHVRSQYVALLQGSAD-----	R•	
DmDawdle	247	LTHLRHA-VKQILSKRLKES-PKYSA---VELPKPIFGM---	TLSHFPDS-----	T•
MmGdf15	132	ARPDWTR-PL-KRALSRGPRP-PAIRLRL-----	TPPDLA-----	MLPS-----
MmAmh	365	SAPWAAGL-QS-RVAVELQAAS-ELRDLPLPLPTAPPLARL-----	LALCFNDS-----	R•
		TGF-β Activin	α3	TGF-β
		β1	β2	β3
DmMaverick	377	SYFN-LHPL-DANRHHDKIDEANVRLMLTY-SSSLATN-QQILTKYQLLSANRR----		
DmMyoglianin	350	DYTG-IPPE-VFRFCHDS-----	SYSDS•	YATLHRY-KDMVTTHAIRKAVANTT•
MmGdf11	160	EYHA-----	CHRFSPKV-MFTKIL-KQQLWY•	ATVYLQILRLKPLTGEG•
MmMstn	139	DYHA-----	CFKFSKSI-QYNKV•	KQQLWY•
CeTIG-3	38	DLYG-----	CFQSSINSINHEILSASL-----	TNSIVVYE--VDEL-----
MmTgfb1	133	DYYA-----	MFNTSDIREAVPEPLIS•	RELRLQ-VEQHEVYQKYS-----
MmTgfb2	157	EYYA-GMLD-IREPQST-MEKNASNV-KDEFRL•	AEORHEVYILKSKDLT•	
MmTgfb3	131	EYYA-----	VFRNVSS-VEKNGTNLF•	RAEFRLL-TEQRHEVYILKSKDLT•
CeUNC-129	109	-----LVGVDP-EHBYRR-RDSFARR-RS-----	OTIK-----	KIRVGG-----
CeDAF-7	125	AKQDV-TTVSPEI-PAKDS-----	QDQIQYQEKNE-----	DG•
MmBmp3	129	HTGNL-TLVFNGE-LVNI-----	SSC-----	QRLQD-SANILKSN-----
MmGdf10	128	YFNL-AFHEVSE-PRWPRAREFC-----	NPTEELIKGAE-TWKK-----	LPARHIFRSL-Q•
MmInhba	135	ITSAE-THEPESK-EGSDL-SVVE•	RAEVMFF•	TKTHRFQOQKHQGSGLD•
MmInhbb	162	ISSAE-RYEFVSN-EGNQN-LFW•	CASLWY•	RKVRKQYQ•
MmInhbc	114	ISSAD-RDEPFSG-RMASG-MEVR•	QTRFMFF•	ENRHLVLRP-----
MmInhbe	109	ISSAT-MTFQSP-LWSHH•	Y•	H•
DmActivin-beta	659	EYFN-ITSAE-IEESSAQN-RRVPS-QKESI•	PHSLWIE-HRKHMVQFLSTSINITEKG•	
MmLefty1	98	EYAG-----	LVFGEQ-RLPPN-SENV•	CAVLRFF-ARVTEWLR-----
MmLefty2	98	EYAG-----	LVFGEQ-RLPPN-SENV•	CAVLRFF-ARVTEWLR-----
DmDawdle	332	-----CFTEKDD-ADAEG-FDS•	TVLWLF-AQCTVWSE-----	VEVDQ•
		β5	β6	β7
		β8	β9	β10
DmMaverick	467	IEFGNVGFQETRTQWEPF-VKAKRSWLNKS-HENNGIE•	CDCKKSIG-ARILS	
DmMyoglianin	421	FRHSIPSLGQWAVT-LKSLGNLGS-----	NTQELIKGAE-TWKK	
MmGdf11	222	LKIELHLSRSHHSD-FKQVHSHFRQP-QSNWIE	NAFDPSTDLA	
MmMstn	193	LKLDMSPTGIGQSID-VKTVLQNLKQP-ESNGIE	KALDENGDLA	
CeTIG-3	84	FE-----	IRETLDKXFD-SHLFHKWMKQK-SSDMK	BITNSNTQNVINALSVLR
MmTgfb1	187	LT-----	PTDTPMELSD-VGVQRWLNQD-DGTQGFESAHCSKSDKNK	
MmTgfb2	218	VK-----	TRAEGLMSD-VDAQEWLHKK-DRNGPKSLHCCCTFVPSNNYIIP	
MmTgfb3	191	LP-----	TRGTAEMLSD-VDTREWLLRR-ESNGIE	SICHCPTFPQNGD-ILE
CeUNC-129	156	EE-----	VKIVMDATKSVFDSY--HLDA-KQAVPRITREHSK--M-----	
CeDAF-7	170	SERISIQLPIDTVKSMETISPGIFVKAMLDGRNVAPHQQTAD-----		
MmBmp3	181	VDVVRPYRDSVMSLSE-HQLRKAKONEE-FLGNFTSRAHELPRKM-----		
MmGdf10	185	GAMALTPPPRGLWAKF-SSIKAAARRDGE-LLSAQDTEKDPGVPR-----		
MmInhba	204	LLLEKVDARKSTHIFP-SSSQRLLDQK-KSSDVRHA-CEQCQESGAS-----		
MmInhbb	216	WNVVEKKVDLKRSGHTFP-DEAQAALFERG-ERRNLNQ-CDSCQELAVV-----		
MmInhbc	163	TLTSQYVVQVNASGMYQLL-GPEAQACSQGHLT-EL--VPESQVA-----		
MmInhbe	154	FRTFLAEHQTSSGHALT-PSSSLRSDESG-VVKQLEFRPLDLNSTAAG-----		
DmActivin-beta	746	ASFPQVDPKNLGAQKFL-EDTREWYGHSTHERLID-CTGCGGRYSL-----		
MmLefty1	165	SRLV-----	SIHESGAKAF-DEANFWQQLSRPR-----	
MmLefty2	165	SRLV-----	SIHESGAKAF-DEANFWQQLSRPR-----	
DmDawdle	395	AKTAIQSVNVQDEWKKID-LEWPKHWISGHLSH-LIQHT-CGCGDVSDME-----		
		bowtie--β9--	no-β9'-	β10--
		α5--		
DmMaverick	520	DFSPSTPPR-----	STASSDEHLNLMFVNIIGHGTLN---	
DmMyoglianin	463	-----SLVTTD-----	NTSKNELTHIEIG-----	
MmGdf11	269	-----VTSLGP-----	GAEGLHFEVRLVLE-----	
MmMstn	240	-----VTFPGP-----	GEDGLNEFEKVTD-----	
CeTIG-3	135	-----		
MmTgfb1	232	-----LHVEING-----	ISPKRRG-DLGTIH-----	DMNREFILMATPLERAQ---
MmTgfb2	269	-----NKSEELARFAGID-----	GTSTYASG-DQRTIKSTRKTSKTEHLLMLLPSYRLE---	
MmTgfb3	241	-----NVHEVMEIKFGVD-----	NEDDHGRG-DLGRLK-----	KQKDHNNHLLMMIPPHRLD---
CeUNC-129	191	-----	RPYA-----	EMIRKSTPEVYYSKVNHTLDTVSV---
CeDAF-7	216	-----		VDNRLQLSTRPK---
MmBmp3	229	-----LFFPEPYI-----	LVYANDAAISEPESVSVSLQRHRDFTAGTG---	PR---
MmGdf10	233	-----PSSHMPYI-----	LVYANDLAISEPNSVAVSLQRYDFFPAGDFEPGAAPN---	
MmInhba	254	-----LVLLGKKKKKEVDGDKKKDGSDDGLEEKQSHRPFMLQARQS---		
MmInhbb	266	-----PVFVD-----		PGEESHREFFVQARLG---
MmInhbc	207	-----	HSSL-ILGWFSHREFFVAAQVRVE---	
MmInhbe	204	-----	LPRLLLDTAGQORFEHVKIRAN---	
DmActivin-beta	794	-----HLFQTSKLGRN-----	SSDYLSTNPNRPFVHTESS---	
MmLefty1	195	-----		QFLH-QVSVQ---
MmLefty2	195	-----		QFLH-QVSVQ---
MmInha	213	-----		TPFLAHTRAR---
DmDawdle	445	-----EITS-----		VKDQYREFVDMQNR---

Figure 6 Focused Activin + TGF-β subfamily prodomain alignment indicating structural conservation. Sequences presented as in Figure 2 except gaps are omitted for brevity and indicated by columns of black dots. Numbering is only approximate. Locations and naming of structural features are from TGF-β1 (Shi et al. 2011). Underlined α2 sequences of TGF-β1 and Inhibin-βa indicate distinct locations. Underlined β1 in TGF-β1 shows it is in a distinct location in Inhibin-βa. Underline of α3 in TGF-β1 indicates absence of conservation. The notation TGF-β with β3 indicates this feature approximates the TGF-β1 pattern (underlined). The notation Activin with β7 indicates this feature approximates the Activin pattern (underlined) not the underlined TGF-β1 pattern. In β8 of the bowtie of TGF-β1-3, the conserved pair of cysteines (CxC) is underlined in black. In β8 Maverick and Dawdle are joined by Activin, Inhibin-βa, and Inhibin-βb, with a conserved pair of cysteines of distinct spacing (red underline; CxxC). Underlined β9, β9', and α5 are not conserved.

vs. BMP15/GDF9, suggesting one predominant function. The prodomain tree has two symmetric secondary clusters. One cluster is BMP proteins related to Dpp, Gbb, and Screw, while the other is GDF proteins plus Nodal. Note that BMP10 and BMP15 (also known as GDF9b) have names that do not fit

their association with GDF proteins. The prodomain tree suggests that perhaps each major cluster has a distinct mode of regulation.

In the ligand tree Nodal is significantly paired with nematode DBL-1 and then loosely with the mammalian pair

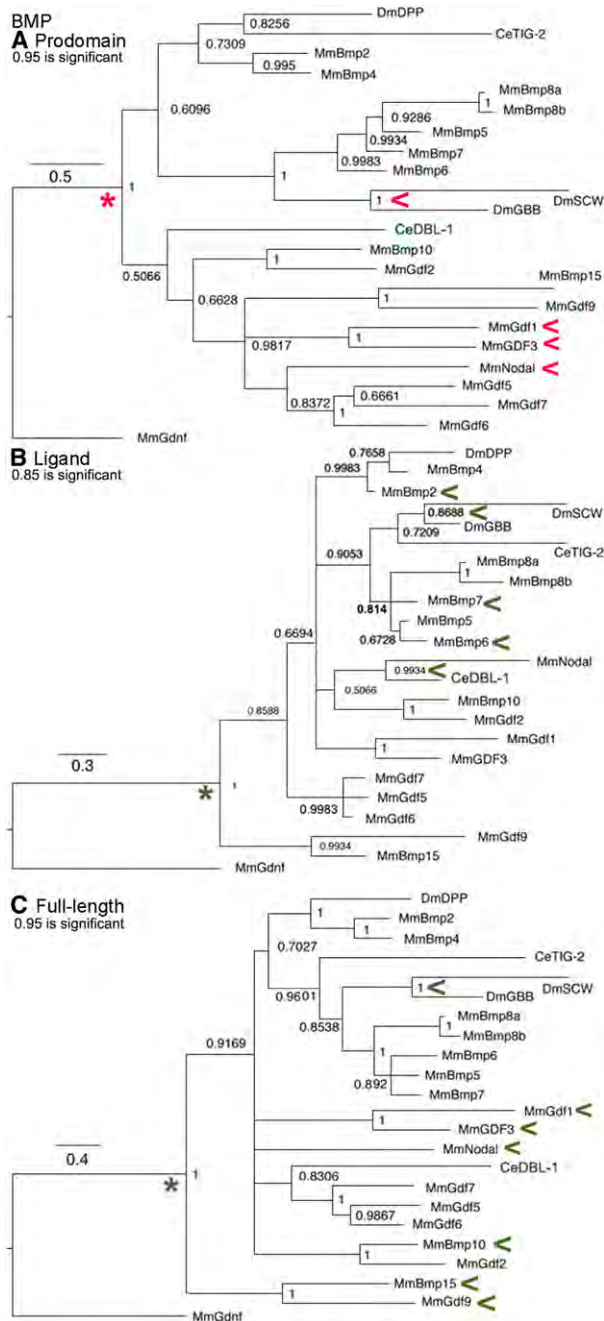


Figure 7 BMP subfamily trees. Bayesian trees of 22 sequences plus the outgroup are displayed as in Figure 1. Red arrowheads indicate a cluster that may reflect common regulation and green arrowheads a cluster that may reflect common function. (A) Prodomain nodes ~0.95 are significant. The significant cluster of Gbb/Screw was unexpected. The cluster of heterodimerizing Nodal and GDF1/GDF3 was expected. Red asterisk indicates node leading to two symmetric secondary clusters. (B) Ligand nodes ~0.85 are significant. The significant cluster of BMP2-8a, b with Gbb/Screw was expected and consistent with functional heterodimers of BMP2-BMP6 and BMP2-BMP7 that have been reported. The significant cluster of Nodal and DBL-1 was unexpected. Green asterisk indicates node leading to two asymmetric secondary clusters. (C) Full-length nodes ~0.95 are significant. Clustering of GBB/Screw, BMP10-GDF9, BMP15-GDF9, and Nodal-GDF1/GDF3 are consistent with heterodimerization that has been reported. Green asterisk indicates node leading to two asymmetric secondary clusters.

BMP10/GDF2. The tight association between mouse Nodal and DBL-1 in the absence of any fly protein is curious. Nodal's distinct ligand and prodomain associations lead it to be a loner in the full-length tree. Otherwise, the full-length BMP tree largely shows previously established clusters. Overall for the BMP subfamily, similarity between the ligand and full-length trees indicate that functional relationships of the ligand are driving its evolution.

BMP subfamily prodomain structural conservation

The BMP subfamily appears more homogeneous than the Activin or TGF- β subfamilies in the annotated alignments (Figure 8 and Figure S7). Homogeneity is evident in a larger number of conserved residues and a greater frequency of identical residues. The locations and names of features in this subfamily derive from BMP9 (Shi *et al.* 2011; included here as GDF2). All of the features of the straitjacket (α 1, Latency Lasso, α 2, and β 1) display strong homogeneity. α 1 and the Latency Lasso contain 10 conserved I/L/V residues with three that are near universal. There is a pair of near universal F/Y residues in α 2. An F/Y and the adjacent S/T in β 1 are moderately conserved.

At the 5' end of the arm strong conservation is visible. β 2 contains near universal F/Y and I/L/V residues. Conservation is present between β 2 and β 3 with a near universal I/L/V and a modestly conserved proline. β 3 has a stretch of seven consecutive conserved residues, a degree of continuous conservation not seen previously. This stretch includes two near universal I/L/V residues, near universal R/K, alanine and glutamic acid residues, as well as modestly conserved R/K and F/Y. β 4 also has a stretch of seven residues highly conserved in 66% of family members: four I/L/V residues, an S/T, and an F/Y. β 5 is only moderately conserved with one near universal I/L/V. The β 6 to α 4 region has a highly conserved stretch of seven consecutive residues including near universal tryptophan, phenylalanine, aspartic acid, and S/T. There is a near universal tryptophan at the distal end of α 4.

At the 3' end of the arm, β 7 has three moderately conserved I/L/V residues. β 8 and β 9 (the distinctive bowtie of TGF- β 1) are not conserved. β 9' is a unique BMP feature containing a moderately conserved arginine and an I/L/V. β 10 contains two near universal I/L/V residues and a modestly conserved phenylalanine. Overall, in the BMP subfamily the 5' ends are more highly conserved than the 3'.

All family members trees

Trees of the whole family (Figure 9) including a cystine knot tree to compare to the biochemically defined ligand tree are shown. A comparison of the latter two trees demonstrates the loss of resolution resulting from adding the degenerate spacer to the cystine knot alignment.

In the prodomain trees, there are six cross-subfamily clusters. One has Activin in a cluster with Gbb and Screw that heterodimerize with Dpp, although the node is just short of significant (0.78 vs. 0.95 for significance). What this modest shortfall means is that this is the best, but not the only

straitjacket		-α1-----		-LatencyLasso-----		-----α2-----		--β1---	
DmSCW	30	IYQKRLSEQMEVDID	DGDGRRRQAEP	--	NLHNSASKELLEVNEHSE	•	ASCNSLIFSS		
DmGBB	56	LSEDDKLDSYED	FEFLGHAERETHLSSHQL	--	SLRKSASKEFLDLYRYTA	•	DESDTHMIFLN		
MmBmp8a	35	LGARERRDMQRE	AVGGLPGRRPRAQPA	--	RQPASAPLMLLYRYATD	•	GRADIVMSFVN		
MmBmp8b	35	LGVRPRDMQRE	REVGLPGRRPRAQPA	--	QQPASAPLMLLYRYATD	•	DRADIVMSFVN		
MmBmp6	68	LKTHKREMQKE	SVGLGPHRRPLHGLQ	--	PGRLKSAPLMLLYRYASN	•	NDADIVMSFVN		
MmBmp5	43	LRNHERREIQRE	SVGLGPHRRPFPSPG	--	KQASSAPLMLLYRYASN	•	NDADIVMSFVN		
MmBmp7	46	LRQERREMQRE	SVGLGPHRRPFPLOG	--	KHNSAPLMLLYRYASN	•	TDADIVMSFVN		
CeDBL-1	42	ADQHASATRRGL	VRKGLGEHVF	--	V-QT--	•	GPSIDVQHMMDDND		---VWRHYP
MmBmp10	50	DFNTLLQSMKNEF	KTUNESDIE	--	V-QD--	•	TGRVDFEMLLYRYKAT	•	PSANTVRSFKN
MmGdf2	53	DLQMFLENMKVD	FRSNEGIF	--	S-QD--	•	KTRAEFPMMLLYRYNTT	•	PASNTVRSFV
CeTIG-2	38	QATDK	-----IGEQRELFNDINE	--	NGPAV--	•	KANNYVSTMKRLYKQEN	•	LSADRLVSHMA
MmGdnf1	61	PAED	-----HSLGHRVE	--	F--AL--	•	TSDSNMEDYPCDDVMD		
DmDPP	213	PPSTLVIEKST	SLFNKKRPE	--	K-ID--	•	RSKIIIEEKKLAEIMG	•	KSANTVRSFTH
MmBmp2	41	PSEDVLSEFELR	SMFGKQRE	--	T--PS--	•	KD-VVFPMLLYRYRHSG	•	SRANTVRSFHH
MmBmp4	43	QSHELLRDFEAT	DMFGRRR	--	Q--PS--	•	KS-AVIDMMLLYRYLQSG	•	SRANTVRSFHH
MmBmp15	35	PSILLDAKEA	-----PG-----	•	KEMKQ--	•	WPQGVFLMLLYRYRSAD	•	IGAKVRLVKP
MmGdf9	55	PPLFKVLSDR	-----RG-----	•	ETPKL--	•	QPDSTRALYMKKLYKYAT	•	HLNTVRLKSP
MmGdf1	25	PAAILLVQ	-----LGPEAFRSVPTHR	--	--PVPP--	•	VMWRPDRRD	•	VANTVVRHFD
MmGDF3	20	QSDLLQF	-----LGLEKASP-HRFQ	--	--PVPR--	•	VVRKTIAREAA	•	VRNDVQLLDP
MmNodal	13	ACWALLHPR	-----APTAAALPL--WTRGQ	--	--PSSPSPLAMMSLYRDLPL	•	---	•	DIRSLQA
MmGdf5	124	PSSFLQSKTR	-----EP-CTPRE	--	--KEPFR--	•	PPPIITPHEMLLYRYTSD	•	GLANTVTSFID
MmGdf6	81	-----	QH-QGQEP	--	C--R--	•	GLRVVPHEMLLYRYTYSI	•	KSANTVTSFVD
MmGdf7	65	-----	PR-ARRAA--GSGFR	--	--NGSVVPHHMLLYRYASAG	•	GRVDITGTETD		
arm -β2-----		-----β3-----		--α3---		--β4---			
DmSCW	131	MHTTNTND	--VPVDLSVCAMLRIYK	OPSLVD	--R--	RANFTVSVRK	•	DNRQ	
DmGBB	176	RRLWFQDN	--VPNDNYVMAELRIY	QANEGKWLTA	---	NREFTIVYATGTGLG			
MmBmp8a	123	KEFHEDLQ	--PAGEAVTAABERLY	EPSTHPL	---	NTTLHISMHEVQEHNS			
MmBmp8b	123	KEFHEDLQ	--PAGEAVTAABERLY	EPSTHPL	---	NTTLHISMHEVQEHNS			
MmBmp6	233	KEFKENLQ	--PAGEAVTAABERLY	DCVVGSE	--K--	NQTFILISIVYQEHQH			
MmBmp5	178	KEFRFDLQ	--PHGPAVTAABERLY	DKGNHRE	--E--	NETIKIEVQIKEYTH			
MmBmp7	153	KEFRFDLQ	--PAGEAVTAABERLY	NDYIRERF	--D--	NETFQIVQVQEHSG			
CeDBL-1	106	LLSNLNLAAARNAHNE	VRKALTLRLRRNNK	---	---	ARRSGNISYFFEDD			
MmBmp10	126	YPLLENNV	--EPHHEVVMALRIY	TLVQRDRMMY	---	DGVDKRTIIEVESADG			
MmGdf2	130	HILILENT	--TPRHECTRAELRIY	VSQNDVDST	---	HGLEGSVWVDVEDSET			
CeTIG-2	120	YSIRPAKEH	--VPAKEGOSIVRAQ	---	---	LRHIGLQGISPVFF			
DmDPP	303	FRILHEDVKS	--PADEKPKAELR	DTTALDALSQQVVASRSSANTRYQ	---	LYDITRVGV			
MmBmp2	129	RRFFNLSS	--VPSDEPTSAELR	QFEQIQAELGNSSSFQHR	---	INHEHKKPAA			
MmBmp4	139	FRFLENSS	--PENEVTSSELRL	PREQVDQGPDEQGFHR	---	INHEHKKPAA			
MmBmp15	113	QTLDEPLAS	--NOVAYEIRAPVYR	QLHL-VNYH	---	---	•	CHVETVV	
MmGdf9	138	VDLLENNR	--VTAMHHLKSVLYTL	NNSASSST	---	---	•	CCDLVV	
MmGdf1	117	WTVVDELN	--VETTERPARRLE	RLAESEDTGGW	---	EL-----	•	SVAL	
MmGDF3	108	KVLYENLSA	--KEKAKDMAQLT	DLGPRSYNLR	---	ELVVAS	•	VQDRGV	
MmNodal	74	WTFTEDEFS	--SQEEDVMAELRL	QLPGMDIPEG	---	PL--TD--	•	EFQAKG	
MmGdf5	214	QRVYFDTSA	--EKEGCLGAELRL	KKPLDVAKPA	---	VP--SSG--	•	RVAQLKL	
MmGdf6	149	QKYLEFVST	--SKDEBVGAEELRY	QAPPTFW	---	GL--PAR--	•	PHQL-L	
MmGdf7	132	QSFLEDMSS	--SEADPVMAELRL	RRRSEPD	---	RD--SA--	•	DEPRLLL	
-----β5-----		-----β6---		--α4---		--β7---			
DmSCW	179	SYRDLGS	-----VNT---	TSSQRWDEFLD	---	TLRWLHNKGLQ	•	RRNBLRIS	•
DmGBB	228	TMEPLSS	-----VNT---	TGDYVWHELNVT	---	GLHEMLVKSQN	•	HGYHGAHAVNR	•
MmBmp8a	172	DLFELDL	-----QTD---	RSGDEWVLDYTA	---	ASDWLLNHHD	•	GLRLYVET	•
MmBmp8b	172	DLFELDL	-----QTD---	RSGDEWVLDYTA	---	ASDWLLNHHD	•	GLRLYVET	•
MmBmp6	283	DLFELDT	-----RVW---	WASEEWEFDTA	---	TSNLMVVTQHN	•	MGLQLSVT	•
MmBmp5	228	DLFELDT	-----RKT---	QALDVWVFDITV	---	TSNLMVVTQHN	•	MGLQLCAET	•
MmBmp7	203	DLFELDE	-----RTU---	WASEEWEVFDITA	---	TSNLMVVTQHN	•	MGLQLSVT	•
CeDBL-1	153	FOLES	-----RS---	VDNLTEWDFDVTA	---	AFLRTRNRSFF	•	IDLPEDEIEETQS	•
MmBmp10	177	M-LVLVS	-----TEN---	YGTNSMETFDVTD	---	ATRWQKSGPST	•	HQLEHUESRQNA	•
MmGdf2	181	TKTFLVS	-----QDT---	RDEWTETLEVS	---	AVKEWVRADST	•	NKNKLEVTQ	•
CeTIG-2	156	-----	-----VLV---	SSDDPTVVDVDT	---	MVDWHLQLST	•	LFTVTARAST	•
DmDPP	359	SYLLEDT	-----KTV---	RLNSTDTSLDVQP	---	AVDWLASPORN	•	YGLVWRTVRSLK	•
MmBmp2	180	VTRILLDT	-----RLV---	NQNTSQMESFDVTP	---	AVMWTTQGTN	•	HGFVVEAHLEENP	•
MmBmp4	190	ITRILLDT	-----RLV---	HHNVTRWETFDVTP	---	AVLWTRKQPN	•	YGLVWTHLQTR	•
MmBmp15	155	PSSKSG	-----SKP---	SPMSKWTEID	---	HCIOQKL	•	WNRKGRS	•
MmGdf9	181	EAMSSGR	-----APPAPYSFT	LKKHRT	---	ETDVTSLQLPLV	•	TSSERSI	•
MmGdf1	158	GPELTL	-----RVPAPGVL	---	---	---	•	TAVANASVPT	•
MmGDF3	157	VGRILFL	-----RSVPGFQGL	---	CFNQG	---	---	ALKDWSSNRLKN	•
MmNodal	121	PADCLRIWMETTV	VP-SQVTFAS	STVLVTK	---	PLSKWLKD	•	PR--	•
MmGdf5	260	PAALILV	-----RSVP-GL-	DGSEWEVFD	---	WKLERNF	•	--KN-SA--	•
MmGdf6	192	PLALLA	-----RTLD	PQGTQAEVFDVWQGLRP	---	QP-WK--	•	---QCLEA	•
MmGdf7	176	TAHLHL	-----RAAE-PLG-	GARWEAFDVT	---	DAVQSHRWPRASR	•	--KFCVLV--	•
noβ8 nobowtie noβ9		-β9'-		--β10--		noα5			
DmSCW	217	DSQLSTFAAG	-----	LVTPQ	---	LVGYFNGPELLVKIQ	•	KLRFKR	•
DmGBB	266	DREVKLDDIG	-----	LHRRK	---	MIGFRRGP	•	ELIKATAHSSHRSKR	•
MmBmp8a	211	GHSMDPGLAG	-----	LGGRQ	---	MVTHFRASQSPVRAPR			
MmBmp8b	211	GHSIDPGLAG	-----	LGGRQ	---	MVGHFRANQSPVRAPRTAR			
MmBmp6	322	GLHVNPRAG	-----	LVGRD	---	MVAFKASEVHVVRTR			
MmBmp5	267	GRSINVKSAG	-----	LVGRH	---	MVAFKASEVLLRSVR			
MmBmp7	242	QGSINPKLAG	-----	LVGRH	---	MVAFKATEVHLRSIR			
CeDBL-1	191	S-----	-----	SL	---	PLVVFSDLSSEPSSVRK	•	R-----	
MmBmp10	218	ED-----	-----GR--GQLE	---	---	LVVVFSDDQSDNK	•	EQ-KEELNE	
MmGdf2	221	HR-----	-----SC--DTLD	---	---	FFVVFSDNDRSNGTK	•	ET-RLELKE	
CeTIG-2	195	-----	-----N--DEL	---	---	LVHALKDEADGPKKRSRRSASTT			
DmDPP	401	P-----	-----PH--HHVR	---	---	RRSGADE	•	LEFTYTDGGRHKA	•
MmBmp2	222	G-----	-----V--SK--	---	---	RRVRSRSLHQ	•	LVVTFGHDGKGHP	•
MmBmp4	233	T-----	-----H--QG--	---	---	QHVRRSRSLFQ	•	LVVTFGHDGGRHT	•
MmBmp15	200	-----	-----KGNET	---	---	REFRWH	•	LVVYENDTDDRVOGK	•
MmGdf9	225	-----	-----QVPED	---	---	GVFSMP	•	LVVLYNDTSTQAYHS	•
MmGdf1	198	-----	-----SLHFGATA	---	---	AC-----	•	GL-----	
MmGDF3	199	YSRVTVQPEN	-----	PC-----	---	DR-----	•	LVVVTLDPRCLP	•
MmNodal	169	WHQ-----	-----PYTPPVV	---	---	---	•	LVVYSNRPOEQRL	•
MmGdf5	302	E-----	-----RGRAVDLRGLGFERT	---	---	---	•	LVVVEGRTKKRDLFFNE	•
MmGdf6	231	WGLDAGDTGARARG	PQPPPLDLRSLGFG	---	---	---	•	LVVVFTRSQRKN	•
MmGdf7	222	T-----	-----ASESSPLALRLR	---	---	---	•	LVVVISSRTORKESLFE	•

Figure 8 Focused BMP subfamily prodomain alignment indicating structural conservation. Sequences presented as in Figure 6 with gaps indicated by columns of black dots and approximate numbering. The locations and naming of structural features are from BMP9 (Mi *et al.* 2015), shown here by its synonym GDF2. The absence of α3 is indicated by the underlining in GDF2. The features β7 and β9' are conserved in this subfamily, while β8, β9 (the distinctive bowtie), and α5 of TGF-β1 are not.

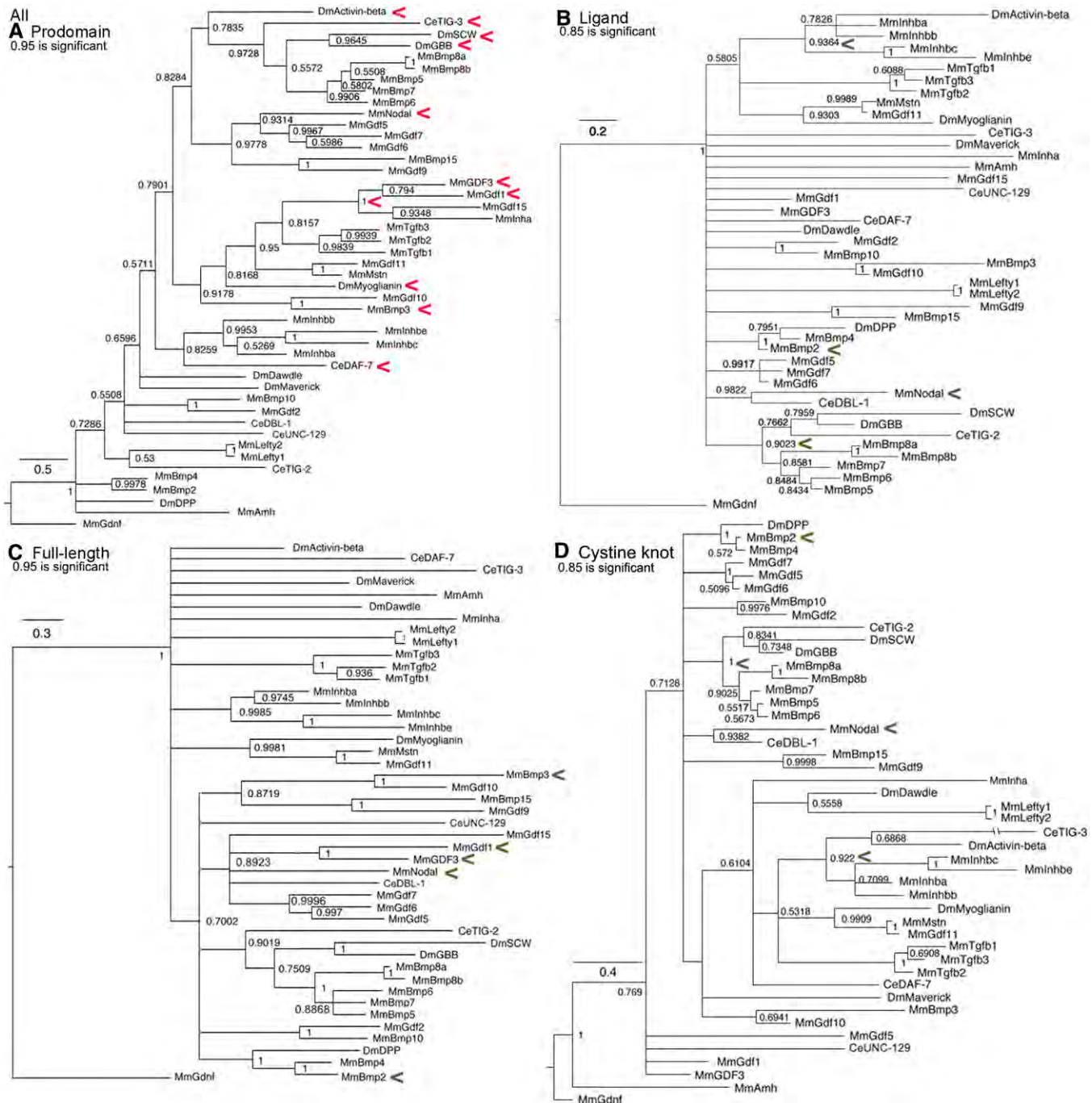


Figure 9 All family members trees. Bayesian trees of all 45 sequences plus the outgroup are displayed as in Figure 1. Red arrowheads indicate a cluster that may reflect common regulation and green arrowheads a cluster that may reflect common function. (A) Prodomain nodes ~0.95 are significant. The not quite significant cross-subfamily cluster of Activin, TIG-3, Gbb, and Screw with Nodal was unexpected but three are known to heterodimerize and two have conserved cysteines. The absolute cluster of GDF3/GDF1 with GDF15/Inhibin- α and this group's not quite significant cluster with Myoglianin was unexpected. The not quite significant cluster of DAF-7 with the four Inhibin- β proteins was unexpected but is consistent with "LTBP-Association region" cysteine conservation. The cluster of BMP3/GDF10 with Myoglianin was unexpected. (B) Ligand nodes ~0.85 are significant. Several significant clusters are expected such as the four Inhibin- β proteins and Dpp/BMP2/BMP4. The significant cluster of Nodal and DBL-1 was unexpected. (C) Full-length nodes ~0.95 are significant. The not quite significant clustering of BMP3 and GDF1 with all of the BMP subfamily proteins was unexpected. The not quite significant cluster between Nodal and GDF1/GDF3 was expected. (D) Cystine knot nodes ~0.85 are significant. Several significant clusters are expected, such as Activin with four Inhibin- β proteins, Dpp/BMP2/BMP4 and Gbb/Screw/BMP5-8a,b. A significant cluster of Nodal and DBL-1 was unexpected.

possible placement of Activin in the tree. This unexpected cluster forms a second (although still not significant, 0.83 vs. 0.95), larger cross-subfamily cluster with Nodal and its

closest GDF relatives in the BMP subfamily. While not conclusive, the first cluster suggests the hypothesis that the prodomains of Activin, Gbb, and Screw are similar as a

result of the shared ability to heterodimerize with Dpp. The inclusion of BMP5–8a,b and Nodal with its closest GDF relatives in the larger second cluster suggests a level of similarity not found in the wider family. The most parsimonious hypothesis for this similarity is intracluster heterodimerization, implying numerous functional heterodimers yet to be identified.

A third cross-subfamily cluster in the prodomain is strongly although not significantly connected to the one above (0.79 vs. 0.95 for significance). This group contains the BMP subfamily members GDF3/GDF1 joined with absolute confidence (node = 1.0) to the TGF- β subfamily members Inhibin- α /GDF15. Together these form a fourth cross-subfamily cluster with significant, or just below significant, nodes to the TGF- β proteins and five Activin subfamily members (Myoglianin, Myostatin/GDF11, and BMP3/GDF10). This cluster suggests that these prodomains are similar to Inhibin- α as result of the ability to heterodimerize. This prediction is supported by Nodal heterodimers with GDF1/GDF3 (Montague and Schier 2017). The prodomain clustering of Nodal partners GDF1/GDF3 with the TGF- β subfamily members Inhibin- α /GDF15 could explain the ability of BMP subfamily Nodal/GDF1 heterodimers to signal via the TGF- β subfamily receptor ActRIIA.

A fifth cross-subfamily cluster in the prodomain is DAF-7 (TGF- β subfamily) with heterodimerizing Inhibin- β proteins. A sixth cross-subfamily cluster is TIG-3 (Activin subfamily) with heterodimerizing Gbb/Screw. These clusters suggest possible heterodimerization for these nematode proteins, the only prediction of this type for this species.

For the biochemically defined ligand tree, aside from recent duplications such as Lefty1/2, only three small secondary and three small tertiary clusters are seen. These are all composed of the most conserved proteins such as Activin/Inhibin- β group, Dpp/BMP2/4, and Gbb/Screw/BMP5–8a,b. Ten proteins are solos and Nodal is again paired with DBL-1.

By comparison, the cystine knot tree shows better resolution with only five proteins as solos. Unsurprisingly, secondary clusters of the same highly conserved proteins are visible. Surprisingly, Nodal is again paired with DBL-1. This is because the spacer of Nodal and DBL-1 shows unexpected conservation. The 11 amino acids closest to the first cysteine in DBL-1 contain seven of the 10 amino acids in the Nodal spacer (Table S3). This likely explains their consistent pairing in the biochemically defined ligand and cystine knot trees.

In the full-length tree, the Activin subfamily members BMP3/GDF10 are not quite significantly associated with the expanded group of BMP subfamily members. However, they show the same level of association with the TGF- β subfamily in the prodomain and are solos in the cystine knot and ligand trees. This combination of placements suggests that for BMP3/GDF10 their regulation and function share features of distinct subfamilies that will need to be identified by experiment. The pair of BMP subfamily members GDF1/GDF3 that heterodimerize with Nodal also have

features of multiple subfamilies. They are significantly clustered to the TGF- β subfamily members Inhibin- α /GDF15 in the prodomain tree, solos in the cystine knot and ligand trees, and not quite significantly associated with the expanded group of BMP subfamily members in the full-length tree. As noted above, the association of GDF1/GDF3 prodomains with Inhibin- α /GDF15 may explain why Nodal heterodimers with GDF1/GDF3 can signal through ActRIIA.

The placement of Nodal and DBL-1 as solos in the full-length tree is distinct from the ligand and cystine knot trees where they are a significant pair and the prodomain tree where they are essentially unlinked. These two proteins likely have homologous receptors but distinct regulation and it is a specific combination of regulation and function not found in any fly protein.

In the full-length tree, four fly proteins Activin and Myoglianin in the Activin subfamily, plus Dawdle and Maverick in the TGF- β subfamily, are solos (with Myoglianin sticking to its mammalian partners). This contrasts with the prodomain tree where Activin is in a weak BMP cluster and Myoglianin in a weak TGF- β cluster. Alternatively, these two are weakly linked to an Activin + TGF- β cluster in the cystine knot and ligand trees. Their distinct placement in these trees suggests the four fly proteins have mechanisms of regulation yet to be identified. Overall, the many solos in the full-length trees result from dissimilarity between the cystine knot, ligand, and prodomain trees. This indicates that functional and regulatory relationships are equally driving the evolution of the TGF- β family.

All family members prodomain structural conservation

The conservation pattern in the subfamilies is reiterated in the annotated All family members alignment (Figure 10 and Figure S8). In the straitjacket α 1, the Latency Lasso and α 2 display strong conservation. These contain nine conserved I/L/V residues with the second, third and fourth nearly universal and nearly always leucine. There is a near universal proline in the Latency Lasso and a tyrosine in α 2.

The α 2 region of TGF- β 1 contains three conserved I/L/V residues (one universal) and a tyrosine. These are not conserved in the TGF- β subfamily or the Activin + TGF- β subfamily. Alternatively, two to four of these amino acids (VL_{LY}) are nearly universal in the BMP subfamily. BMP conservation has driven the alignment of All family members to identify these amino acids in α 2 of the other subfamilies. In other words, the All family members alignment erases the distinction in the location of α 2 in the Activin and TGF- β subfamilies. In β 1, one I/L/V and a phenylalanine present in the Activin + BMP subfamilies are absent in the TGF- β subfamily, maintaining the prior distinction in β 1 location and conservation.

The 5' end of the arm is not well conserved. β 2 contains only a modestly conserved phenylalanine. β 3 is conserved in an Activin + BMP pattern that draws in several TGF- β subfamily members, although not TGF- β 1. The previously identified stretch of seven conserved residues in β 3 is reduced to

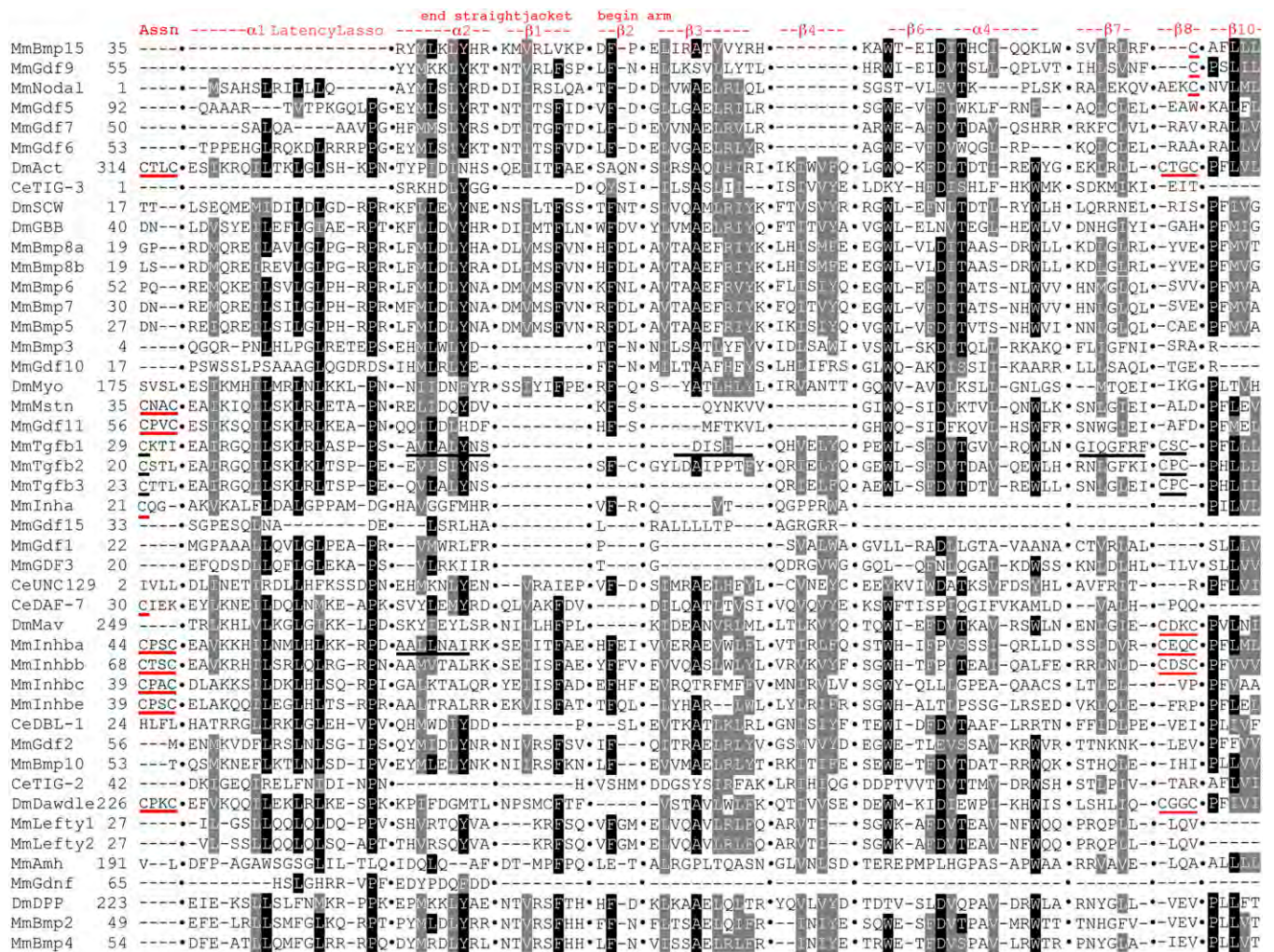


Figure 10 Focused All family members prodomain alignment indicating structural conservation. Sequences presented as in Figure 6 with gaps indicated by columns of black dots and approximate numbering. Only highly conserved features are shown for brevity. The locations and naming of structural features are derived from TGF-β1 and BMP9 as noted above with the latter shown here by its synonym GDF2. In the “LTBP-Association region”, the cysteines in TGF-β1–3 that form a disulfide bond with LTBP are underlined in black. The unexpectedly conserved pairs of cysteines in Activin, Dawdle, the four Inhibin-βs, and Myostatin/GDF11 are underlined in red. Note that the first cysteine of each pair aligns with the cysteine of TGF-β1–3. The conserved cysteines of Inhibin-α and DAF-7 are also underlined in red. The two underlines of α2 show that TGF-β1 and Inhibin-βa have been pulled into alignment by the homogeneity of the Activin + BMP subfamilies. The Activin + BMP version of β1 is conserved but the TGF-β1 version is not. The underlining of TGF-β1 in β3 and β7 shows that the Activin + BMP version is conserved, while the TGF-1β version is not, although TGF-β subfamily members have been pulled into alignment by the homogeneity of the Activin + BMP subfamilies. In β8, the pair of cysteines in TGF-β1 and its siblings (CxC) are underlined in black. The unexpectedly conserved pairs of cysteines in Activin, Maverick, Dawdle, Inhibin-βa, and Inhibin-βb (CxC) are underlined in red. Note that the first cysteine of the CxC aligns with the first cysteine of the CxC. The unexpectedly conserved single cysteines of Nodal, BMP15/GDF9, are also underlined in red. Note that these single cysteines align with the second cysteine of the CxC.

six having lost the glutamic acid. β4 conservation is also reduced, with its stretch of seven residues now only three: two I/L/V residues and an F/Y. β4 is absent in Nodal and seven other BMP subfamily members. The middle region of the arm, β6 to α4 is the best conserved part, yet it too shows a reduction. The highly conserved stretch of seven consecutive residues is at most four and often just two or three. The near universal tryptophan that was previously the first of the seven conserved residues is now separated by two or more non-conserved amino acids from the core of two to four residues (phenylalanine, aspartic acid, I/L/V, and threonine). The

aspartic acid is well conserved. The near universal tryptophan at the distal end of α4 is still present.

At the 3' end of the arm conservation is also limited. β7, like β3, is conserved in an Activin + BMP pattern that draws in several TGF-β subfamily members, although not TGF-β1. β7 retains two moderately conserved I/L/V residues of three previously. β10 is the best conserved feature in the region, perhaps anchoring the 3' end of the protein. There are two near universal and a modestly conserved I/L/V plus a modestly conserved proline. Overall, the 5' end of the straitjacket (α1, Latency Lasso, and α2) plus the central β6 to α4 and 3'

end of the arm ($\beta 10$) are the six best conserved features of the 17 prodomain features in the TGF- β family.

Within $\beta 8$ of the bowtie of TGF- $\beta 1$ is the previously noted conservation of a pair of cysteines in eight Activin + TGF- β subfamily members (either CxxC or CxC with the position of the first cysteine aligned). Unexpectedly, three proteins in the BMP subfamily (BMP15/GDF9 and Nodal) have a single cysteine in $\beta 8$ that aligns with the second cysteine of the CxxC (underlined in Figure 10 and Figure S8, page 4). All members of the BMP trio with a conserved cysteine are known to heterodimerize: BMP15/GDF9 with each other (McIntosh *et al.* 2008) and Nodal with GDF1 (Montague and Schier 2017). Also, these three form a significant secondary cluster in the BMP prodomain tree. The confluence for these three proteins of conserved cysteines in the alignments, a significant prodomain cluster and experimentally demonstrated heterodimerization serves as a proof of principle for our approach and heterodimerization predictions for other family members.

The discovery of cysteine conservation in $\beta 8$ in all three subfamilies reminded us that TGF- $\beta 1$ –3 prodomains are often covalently linked via a cysteine bridge to LTBP (Rifkin *et al.* 2018). We easily identified the conserved solo cysteines in the “LTBP-Association region” near the amino terminus of TGF- $\beta 1$ –3 (underlined in Figure 10 and Figure S8, page 1). In our alignment, the “LTBP-Association region” contains a conserved pair of cysteines in Activin, the four Inhibin- β s, Dawdle, and the duplicated pair Myostatin/GDF11. The first cysteine of the pair in each of these eight proteins from the Activin and TGF- β subfamilies is aligned with the cysteine of TGF- $\beta 1$ –3. Further, a single cysteine is present in Inhibin- α and DAF-7 that also aligns with the cysteine of TGF- $\beta 1$ –3. A total of 10 proteins in the Activin + TGF- β subfamilies appear capable of covalent linkages via the “LTBP-Association region”.

Taken together, a total of 14 proteins from all three subfamilies display cysteine conservation in regions associated with dimerization ($\beta 8$) or protein–protein interactions (“LTBP-Association region”). Many of these cysteine containing proteins are predicted by prodomain clustering to heterodimerize such as Activin and Dawdle. Interestingly, both of these proteins have conserved cysteines in $\beta 8$ and the “LTBP-Association region”, suggesting the possibility of multiple heterodimerization partners.

Discussion

Prodomain structure conservation

Across the prodomain alignments, distinctions in the conservation of structural features between the subfamilies are seen in both the straitjacket and arm domains. In the straitjacket there are discrepancies between the Activin and TGF- β subfamilies in the locations of $\alpha 2$ and $\beta 1$. At the boundary of the straitjacket and arm, a third distinction between the Activin and TGF- β subfamilies is the order of $\alpha 3$ and $\beta 3$ (Activin has $\beta 3$ first and TGF- β has $\alpha 3$ first). In the arm there are three additional differences. $\beta 3$ and $\beta 7$ show dissimilarities

between the Activin + BMP subfamilies and the TGF- β subfamily. $\beta 9'$ is distinct between the BMP and Activin + TGF- β subfamilies. If any functional differences are engendered by these structural distinctions, then they are unknown at this time.

The discovery of a conserved pair of cysteines in $\beta 8$ in five Activin + TGF- β subfamily members and a single conserved cysteine in three BMP subfamily members is exciting. From an evolutionary perspective two points can be made. First, the presence of conserved cysteines in $\beta 8$ in all subfamilies suggests that prodomain participation in protein–protein interactions is an ancient mechanism. The closed-ring conformation of TGF- $\beta 1$ employing a bowtie in $\beta 8/\beta 9$ to mediate dimerization is a recent innovation built upon this foundation. Second, the nonuniversality of $\beta 8$ cysteine conservation suggests significant within-subfamily structural variation between cysteine-bearing and noncysteine-bearing proteins. Structures are known only in the Activin subfamily for Inhibin- $\beta\alpha$ that has these cysteines and in the BMP subfamily only for BMP9 that does not have them. Analysis of additional family members may reveal additional conformations.

Similar excitement is generated by the discovery of a conserved pair of cysteines in the “LTBP-Association region” in eight Activin + TGF- β subfamily members and a single conserved cysteine in two additional members of the TGF- β subfamily. Interestingly, no BMP subfamily members have conserved cysteines here. One caveat is that prodomain length upstream of the straitjacket in the BMP subfamily is highly variable (from one residue in Nodal to 223 in Dpp). There might be functional cysteines that are not close enough to the Activin + TGF- β subfamily cysteines to be captured in the alignment.

The presence of conserved cysteines in all Inhibin- β s and Inhibin- α suggests the obvious hypothesis that they participate in cross-subfamily heterodimerization. A structural analysis of these heterodimers should be fruitful. A logical extension of this hypothesis is the heterodimerization of Activin and Dawdle. For the latter, beyond the “LTBP-Association region” this hypothesis is supported by three pieces of evidence: LTBP does not exist in flies and cannot utilize this cysteine, Activin–Dawdle heterodimerization was predicted in numerous prodomain trees, and these proteins also share $\beta 8$ cysteine conservation. When four lines of computational evidence converge, confidence in the hypothesis is very high.

Predicted heterodimerization

Although previously we noted *a priori* that we considered prodomain clustering as evidence of heterodimerization and thus common regulation, our review of existing literature in light of the identified clusters suggests that heterodimerization can also influence function. There are a number of examples where heterodimers function distinctly from constituent homodimers, with the Inhibins being the most prominent. Here we consider prodomain clustering, whether within or between subfamilies, to suggest new hypotheses for common regulation and/or distinct function via heterodimerization.

One hypothesis for a distinct function of heterodimers suggests a mechanism for TGF- β ligands' ability to stimulate their receptors to activate non-Smad pathways such as the MAP-kinase, Rho-like GTPase, and PI3-kinase/AKT pathways. Currently, cell type-specific accessory proteins such as Par6 are considered responsible for a receptor's choice of signal transduction pathway [reviewed in Zhang (2009)]. Prodomain clustering suggests that the choice may also be influenced by ligand heterodimers, a possibility that has not been previously considered.

In addition to non-Smad pathway activation, functional discrepancies for subfamily dedicated receptors and receptor-associated Smads have been noted. In flies, the Activin/TGF- β dedicated Type I receptor Baboon can signal through the BMP Smad protein Mad (Gesualdi and Haerry 2007; Peterson *et al.* 2012; Peterson and O'Connor 2013). Also in flies, BMP ligands can bind to the Activin/TGF- β Type II receptor Wit (Lee-Hoeflich *et al.* 2005). In mammals, Inhibin- β homodimers can bind to the Type II receptor BMPRII (Rejon *et al.* 2013). The mechanisms underlying these cross-subfamily interactions remain largely unknown. One could speculate that they are influenced by heterodimers resulting from cross-subfamily prodomain similarity. Nodal heterodimerization may serve as an example as its partners GDF1/GDF3 have prodomains that cluster with Inhibin- α /GDF15 suggesting a mechanism for Nodal signaling through ActRIIA.

Overall, we identified six cross-subfamily and two within-subfamily clusters that suggest previously unsuspected heterodimers. In every cross-subfamily cluster at least one protein with an unexpectedly conserved cysteine is involved. For example Activin, that has both association region and β 8 cysteines participates in multiple cross-subfamily clusters.

Predicted fly heterodimers for activin

To date, there is no evidence in the literature that consideration has been given to the possibility that Activin functions as a heterodimer. This is surprising since Activin owes its name to its closest relatives, the heterodimerizing Inhibin- β proteins (Inhibin- β a synonym is Activin-A). The prodomain trees contain clusters that suggest multiple heterodimer partners for Activin.

First is the prodomain cross-subfamily cluster of Activin, Myoglianin, and Maverick in the Activin + TGF- β tree. This cluster has strong statistical support. In the same tree Dawdle is adjacent to Inhibin- α , the heterodimerization partner of Activin's Inhibin- β relatives. The Dawdle and Inhibin- α relationship is statistically significant. This pair of clusters suggests that Activin as well as Myoglianin and Maverick can form heterodimers with Dawdle. The heterodimerization predicted by this cross-subfamily cluster is strongly supported by structural conservation: conserved cysteines in β 8 of Activin, Maverick, and Dawdle; and conserved cysteines in the "LTBP-Association region" of Activin and Dawdle.

Second is the prodomain cross-subfamily cluster in the All family members tree of Activin with Gbb and Screw, two proteins known to heterodimerize with Dpp. While not quite

at statistical significance, the explanatory power of this cluster is welcome. Recently Dpp from imaginal tissues was shown to circulate in the hemolymph to reach the prothoracic gland where it influenced steroid hormone biosynthesis via its typical pathway (Setiawan *et al.* 2018). Circulation is an unprecedented role for Dpp, but a well-established one for Activin (e.g., Gibbens *et al.* 2011). This cross-subfamily cluster suggests that circulating Dpp is actually a heterodimer with Activin and that Dpp targets the prothoracic gland via an Activin-based mechanism. This adds to the suggestion that Activin does many of its jobs as a heterodimer.

Predicted heterodimers for nodal and nematode proteins with potential convergence

For Nodal, our data may explain one of its puzzles but also reveals a new one. A cross-subfamily cluster in the All family members prodomain tree may explain Nodal's ability to signal through the TGF- β receptor ActRIIA. This is a cross-subfamily cluster that links the BMP subfamily members GDF3/GDF1 with absolute confidence to the TGF- β subfamily members Inhibin- α /GDF15. These four are in a second larger cross-subfamily cluster with the prototype TGF- β proteins and five Activin subfamily members. All the proteins in the larger cluster, except GDF3/GDF1, signal through TGF- β receptors such as ActRIA and ActRIIA (ten Dijke *et al.* 1994). GDF1/GDF3 may be included in this cluster because, like the others, their prodomain provides the ability to signal through TGF- β receptors. Nodal heterodimers with GDF1/GDF3 that signal via TGF- β receptors could explain Nodal signaling through ActRIIA.

The new puzzle is embodied in the statistically supported cluster of the BMP subfamily members Nodal and DBL-1 in the All family members ligand and cystine knot tree but not in any prodomain or full-length tree. Contributing to this cluster in our ligand and cystine knot trees is the unexpected conservation of their spacers. Another level of incongruity for this cluster is the absence of a fly protein. Two logical explanations for this cluster are that Nodal/DBL-1 are identical by descent and the fly counterpart has been lost, or convergent evolution of Nodal/DBL-1 based on shared function.

Evidence supporting convergence is the fact that conservation of the Nodal/DBL-1 spacer region (70% similarity, 40% identity) exceeds that of documented homologs Dpp/BMP2 (50% similarity, 15% identity) and Dpp/BMP4 (21% similarity; 0% identity), notwithstanding the 30% longer divergence time between nematodes and mammals than between flies and mammals (Hedges *et al.* 2004). We hesitate to speculate on the basis for convergence as it could be due to receptors or coreceptors either known or yet to be identified, or a completely unanticipated feature of their signaling pathways.

Additional clarity regarding DBL-1 comes from the BMP subfamily trees. In every tree, the unstudied TIG-2 is substantially closer to Dpp/BMP2/BMP4 than DBL-1. Thus DBL-1 is not the BMP2/BMP4 homolog, even though this outdated view is enshrined in GenBank (#AAC27729).

A new hypothesis for DAF-7 is provided by two sets of cross-subfamily clusters in the All family members prodomain tree:

DAF-7 with the Activin subfamily heterodimerizing Inhibin- β group and TIG-3 with heterodimerizing Gbb/Screw. Together these two clusters suggest the possibility of cross-subfamily heterodimerization between DAF-7 and the unstudied TIG-3.

In summary, the prodomain alignments revealed that six structural features are well conserved: three in the straitjacket and three in the arm. Alignments also revealed unexpected cysteine conservation in the “LTBP-Association region” upstream of the straitjacket and in β 8 of the bowtie in 14 proteins from all three subfamilies. In prodomain trees, eight clusters across all three subfamilies were present that were not seen in the ligand or full-length trees, suggesting prodomain-mediated cross-subfamily heterodimerization. Consistency between cysteine conservation and prodomain clustering provides support for heterodimerization predictions. Overall, our analysis suggests that cross-subfamily interactions are more common than currently appreciated, and our predictions generate numerous testable hypotheses about TGF- β function and evolution.

Acknowledgments

The Newfeld laboratory is supported by the National Institutes of Health (grant OD024794).

Literature Cited

- Alfaro, M., S. Zoller, and F. Lutzoni, 2003 Bayes or bootstrap? a simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Mol. Biol. Evol.* 20: 255–266. <https://doi.org/10.1093/molbev/msg028>
- Degnin, C., F. Jean, G. Thomas, and J. Christian, 2004 Cleavages within the prodomain direct intracellular trafficking and degradation of mature bone morphogenetic protein-4. *Mol. Biol. Cell* 15: 5012–5020. <https://doi.org/10.1091/mbc.e04-08-0673>
- Dupont, S., A. Mamidi, M. Cordenonsi, M. Montagner, L. Zaccagna *et al.*, 2009 FAM/USP9x a deubiquitinating enzyme essential for TGF- β signaling controls Smad4 monoubiquitination. *Cell* 136: 123–135. <https://doi.org/10.1016/j.cell.2008.10.051>
- Fuentealba, L., E. Eivers, A. Ikeda, C. Hurtado, H. Kuroda *et al.*, 2007 Integrating patterning signals: wnt/GSK3 regulates the duration of the BMP/Smad1 signal. *Cell* 131: 980–993. <https://doi.org/10.1016/j.cell.2007.09.027>
- Gentry, L., and B. Nash, 1990 The pro domain of pre-pro-transforming growth factor-beta 1 when independently expressed is a functional binding protein for the mature growth factor. *Biochemistry* 29: 6851–6857. <https://doi.org/10.1021/bi00481a014>
- Gesualdi, S., and T. Haerry, 2007 Distinct signaling of Drosophila Activin/TGF-beta family members. *Fly (Austin)* 1: 212–221. <https://doi.org/10.4161/fly.5116>
- Gibbens, Y., J. Warren, L. Gilbert, and M. B. O'Connor, 2011 Neuroendocrine regulation of Drosophila metamorphosis requires TGF-beta/Activin signaling. *Development* 138: 2693–2703. <https://doi.org/10.1242/dev.063412>
- Gray, A., and A. Mason, 1990 Requirement for Activin A and transforming growth factor-beta 1 pro-regions in homodimer assembly. *Science* 247: 1328–1330. <https://doi.org/10.1126/science.2315700>
- Hedges, S., J. Blair, M. Venturi, and J. Shoe, 2004 A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evol. Biol.* 4: 2. <https://doi.org/10.1186/1471-2148-4-2>
- Hedges, S., J. Marin, M. Suleski, M. Paymer, and S. Kumar, 2015 Tree of life reveals clock-like speciation and diversification. *Mol. Biol. Evol.* 32: 835–845. <https://doi.org/10.1093/molbev/msv037>
- Henikoff, S., and J. G. Henikoff, 1992 Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. USA* 89: 10915–10919. <https://doi.org/10.1073/pnas.89.22.10915>
- Hinck, A., T. Mueller, and T. Springer, 2016 Structural biology and evolution of the TGF- β family. *Cold Spring Harb. Perspect. Biol.* 8: a022103. <https://doi.org/10.1101/cshperspect.a022103>
- Jing, S., D. Wen, Y. Yu, P. Holst, Y. Luo *et al.*, 1996 GDNF-induced activation of the Ret protein tyrosine kinase is mediated by GDNFR-alpha a novel receptor for GDNF. *Cell* 85: 1113–1124. [https://doi.org/10.1016/S0092-8674\(00\)81311-2](https://doi.org/10.1016/S0092-8674(00)81311-2)
- Kahlem, P., and S. J. Newfeld, 2009 Informatics approaches to understanding TGFbeta pathway regulation. *Development* 136: 3729–3740. <https://doi.org/10.1242/dev.030320>
- Konikoff, C., R. Wisotzkey, and S. J. Newfeld, 2008 Lysine conservation and context in TGFbeta and Wnt signaling suggests new targets and general themes for posttranslational modification. *J. Mol. Evol.* 67: 323–333. <https://doi.org/10.1007/s00239-008-9159-4>
- Kumar, S., G. Stecher, C. Knyaz, and K. Tamura, 2018 MegaX: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35: 1547–1549. <https://doi.org/10.1093/molbev/msy096>
- Kunnapuu, J., I. Bjorkgren, and O. Shimmi, 2009 The Drosophila Dpp signal is produced by the cleavage of its proprotein at evolutionarily diversified furin recognition sites. *Proc. Natl. Acad. Sci. USA* 106: 8501–8506. <https://doi.org/10.1073/pnas.0809885106>
- Lee-Hoeflich, S., X. Zhao, A. Mehra, and L. Attisano, 2005 The Drosophila type II receptor Wishful thinking, binds BMP and Myoglianin to activate TGFbeta family signaling pathways. *FEBS Lett.* 579: 4615–4621. <https://doi.org/10.1016/j.febslet.2005.06.088>
- Leiber, M., and S. Luckhart, 2004 Transforming growth factor-betas and related genes in mosquito vectors of human malaria parasites: signaling architecture for immunological crosstalk. *Mol. Imm.* 41: 965–977. <https://doi.org/10.1016/j.molimm.2004.06.001>
- Little, S., and M. Mullins, 2009 BMP heterodimers assemble hetero-type I receptor complexes that pattern the DV axis. *Nat. Cell Biol.* 11: 637–643. <https://doi.org/10.1038/ncb1870>
- Loozen, A., A. Vandersteen, F. Kragten, W. Öner, M. Dhert *et al.*, 2018 Bone formation by heterodimers through non-viral gene delivery of BMP2/6 and BMP2/7. *Eur. Cell. Mater.* 35: 195–208. <https://doi.org/10.22203/eCM.v035a14>
- McIntosh, C., S. Lun, S. Lawrence, A. Western, K. McNatty *et al.*, 2008 The proregion regulates the cooperative interactions of BMP15 and GDF9. *Biol. Reprod.* 79: 889–896. <https://doi.org/10.1095/biolreprod.108.068163>
- McWilliam, H., W. Li, M. Uludag, S. Squizzato, Y. Park *et al.*, 2013 Analysis tool web services from the EMBL-EBI. *Nucleic Acids Res.* 41: W597–W600. <https://doi.org/10.1093/nar/gkt376>
- Mi, L., C. Brown, Y. Gao, Y. Tian, V. Le *et al.*, 2015 Structure of BMP9 pro-complex. *Proc. Natl. Acad. Sci. USA* 112: 3710–3715. <https://doi.org/10.1073/pnas.1501303112>
- Montague, T., and A. Schier, 2017 Vg1-Nodal heterodimers are the endogenous inducers of mesendoderm. *eLife* 6: e28183. <https://doi.org/10.7554/eLife.28183>
- Morsut, L., K. Yan, E. Enzo, M. Aragona, A. Soligo *et al.*, 2010 Negative control of Smad activity by ectoderm/Tif1-gamma patterns the mammalian embryo. *Development* 137: 2571–2578. <https://doi.org/10.1242/dev.053801>
- Myers, L., H. Perera, M. Alvarado, and T. Kidd, 2018 The Drosophila Ret gene functions in the stomatogastric nervous system with Maverick TGF β ligand and the Gfr1 co-receptor. *Development* 145: dev157446. <https://doi.org/10.1242/dev.157446>
- Newfeld, S. J., and R. Wisotzkey, 2006 Molecular evolution of Smad proteins, pp. 15–35 in *Smad Signal Transduction*, edited by

- by Heldin, C., and P. tenDijke. Springer, Dordrecht, Netherlands. https://doi.org/10.1007/1-4020-4709-6_1
- Newfeld, S. J., R. Wisotzkey, and S. Kumar, 1999 Molecular evolution of a developmental pathway: phylogenetic analyses of transforming growth factor-beta family ligands, receptors and Smad signal transducers. *Genetics* 152: 783–795.
- Öziak, O., T. Buchta, S. Roth, and J. Lynch, 2014 Ancient and diverged TGF- β signaling components in *Nasonia vitripennis*. *Dev. Genes Evol.* 224: 223–233. <https://doi.org/10.1007/s00427-014-0481-0>
- Peterson, A., and M. B. O'Connor, 2013 Activin receptor inhibition by Smad2 regulates *Drosophila* wing disc patterning through BMP-response elements. *Development* 140: 649–659. <https://doi.org/10.1242/dev.085605>
- Peterson, A., P. Jensen, M. Shimell, R. Stefancsik, R. Wijayatunge *et al.*, 2012 R-Smad competition controls Activin receptor output in *Drosophila*. *PLoS One* 7: e36548. <https://doi.org/10.1371/journal.pone.0036548>
- Quijano, J., M. Stinchfield, and S. J. Newfeld, 2011 Wg signaling via Zw3 and Mad restricts self-renewal of sensory organ precursor cells in *Drosophila*. *Genetics* 189: 809–824. <https://doi.org/10.1534/genetics.111.133801>
- Rifkin, D., W. Rifkin, and L. Zilberberg, 2018 LTBP in biology and medicine: LTBP diseases. *Matrix Biol.* 71–72: 90–99. <https://doi.org/10.1016/j.matbio.2017.11.014>
- Rejon, C. A., M. A. Hancock, Y. N. Li, T. B. Thompson, T. E. Hébert *et al.*, 2013 Activins bind and signal via bone morphogenetic protein receptor type II (BMPRII) in immortalized gonadotrope-like cells. *Cell. Signal.* 25: 2717–2726. <https://doi.org/10.1016/j.cellsig.2013.09.002>
- Ronquist, F., M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling *et al.*, 2012 MrBayes3.2: efficient phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61: 539–542. <https://doi.org/10.1093/sysbio/sys029>
- Schier, A., 2009 Nodal morphogens. *Cold Spring Harb. Perspect. Biol.* 1: a003459. <https://doi.org/10.1101/cshperspect.a003459>
- Setiawan, L., X. Pan, A. Woods, M. B. O'Connor, and I. Hariharan, 2018 The BMP2/4 ortholog Dpp can function as an inter-organ signal that regulates developmental timing. *Life Sci. Alliance* 1: e201800216. <https://doi.org/10.26508/lsa.201800216>
- Shi, M., J. Zhu, R. Wang, X. Chen, L. Mi *et al.*, 2011 Latent TGF- β structure and activation. *Nature* 474: 343–349. <https://doi.org/10.1038/nature10152>
- Shimmi, O., and D. Umulis, H. Othmer, M. B. O'Connor. 2005 Facilitated transport of a Dpp/Scw heterodimer by Sog/Tsg leads to robust patterning of the *Drosophila* blastoderm embryo. *Cell* 120: 873–886 (erratum: cell 121: 493). <https://doi.org/10.1016/j.cell.2005.02.009>
- Smith, R., and T. Smith, 1990 Automatic generation of primary sequence patterns from sets of related protein sequences. *Proc. Natl. Acad. Sci. USA* 87: 118–122. <https://doi.org/10.1073/pnas.87.1.118>
- Stinchfield, M., N. Takaesu, J. Quijano, A. Castillo, N. Tiusanen *et al.*, 2012 Fat facets deubiquitylation of Medea/Smad4 modulates interpretation of a Dpp morphogen gradient. *Development* 139: 2721–2729. <https://doi.org/10.1242/dev.077206>
- ten Dijke, P., H. Yamashita, H. Ichijo, P. Franzen, M. Laiho *et al.*, 1994 Characterization of type I receptors for transforming growth factor-beta and Activin. *Science* 264: 101–104. <https://doi.org/10.1126/science.8140412>
- Tillet, E., M. Ouarné, A. Desroches-Castan, C. Mallet, M. Subileau *et al.*, 2018 A heterodimer formed by bone morphogenetic protein 9 (BMP9) and BMP10 provides most BMP biological activity in plasma. *J. Biol. Chem.* 293: 10963–10974. <https://doi.org/10.1074/jbc.RA118.002968>
- Walker, R., J. McCoy, M. Czepnik, M. Mills, A. Hagg *et al.*, 2018 Molecular characterization of latent GDF8 reveals mechanisms of activation. *Proc. Natl. Acad. Sci. USA* 115: E866–E875. <https://doi.org/10.1073/pnas.1714622115>
- Walton, K., Y. Makanji, M. Wilce, K. Chan, D. Robertson *et al.*, 2009 A common biosynthetic pathway governs the dimerization and secretion of inhibin and related transforming growth factor beta (TGFbeta) ligands. *J. Biol. Chem.* 284: 9311–9320. <https://doi.org/10.1074/jbc.M808763200>
- Wang, X., G. Fischer, and M. Hyvonen, 2016 Structure and activation of pro-ActivinA. *Nat. Commun.* 7: 12052. <https://doi.org/10.1038/ncomms12052>
- Yang, Z., 1993 Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Mol. Biol. Evol.* 10: 1396–1401. <https://doi.org/10.1093/oxfordjournals.molbev.a040082>
- Zhang, Y., 2009 Non-Smad pathways in TGF-beta signaling. *Cell Res.* 19: 128–139. <https://doi.org/10.1038/cr.2008.328>

Communicating editor: A. Clark

Supplemental Information (Three Tables and Eight Figures)

Table S1. Sequences, subfamilies and accession numbers.

Subfamily	Name (synonym) - Accession Number - Description - Species
Activin	CeTIG-3 NP_497318.2 transforming growth factor beta family [C elegans]
Activin	DmActivin-beta NP_651942.2 activin-beta [D melanogaster]
Activin	DmMyoglianin NP_726604.1 myoglianin, isoform A [D melanogaster]
Activin	MmBmp3 NP_001297606.1 bone morphogenetic protein 3 isoform 1 [M musculus]
Activin	MmGdf10 NP_665684.2 growth/differentiation factor 10 preproprotein [M musculus]
Activin	MmGdf11 NP_034402.1 growth/differentiation factor 11 preproprotein [M musculus]
	Synonym BMP11
Activin	MmInhba NP_032406.1 inhibin beta A chain preproprotein [M musculus]
Activin	MmInhbb NP_032407.1 inhibin beta B chain preproprotein [M musculus]
Activin	MmInhbc NP_034695.1 inhibin beta C chain preproprotein [M musculus]
Activin	MmInhbe NP_032408.2 inhibin beta E chain preproprotein [M musculus]
Activin	MmMstn NP_034964.1 growth/differentiation factor 8 preproprotein [M musculus]
11 Activin	Synonym GDF8
BMP	CeDBL-1 NP_504709.1 Dpp/BMP-Like [C elegans]
BMP	CeTIG-2 NP_504271.1 transforming growth factor beta family [C elegans]
BMP	DmDPP NP_477311.1 decapentaplegic, isoform A [D melanogaster]
BMP	DmGBB NP_477340.1 glass bottom boat, isoform A [D melanogaster]
BMP	DmSCW NP_524863.3 screw, isoform A [D melanogaster]
BMP	MmBmp2 NP_031579.2 bone morphogenetic protein 2 preproprotein [M musculus]
BMP	MmBmp4 NP_001303289.1 bone morphogenetic protein 4 preproprotein [M musculus]
BMP	MmBmp5 NP_031581.2 bone morphogenetic protein 5 preproprotein [M musculus]
BMP	MmBmp6 NP_031582.1 bone morphogenetic protein 6 preproprotein [M musculus]
BMP	MmBmp7 NP_031583.2 bone morphogenetic protein 7 preproprotein [M musculus]
BMP	MmBmp8a NP_001242948.1 bone morphogenetic protein 8A isoform 1 [M musculus]
BMP	MmBmp8b NP_031585.2 bone morphogenetic protein 8B preproprotein [M musculus]
BMP	MmBmp10 NP_033886.2 bone morphogenetic protein 10 preproprotein [M musculus]
BMP	MmBmp15 NP_033887.1 bone morphogenetic protein 15 preproprotein [M musculus]
BMP	MmGdf1 NP_032133.2 embryonic growth/differentiation factor 1 [M musculus]
BMP	MmGdf2 NP_062379.3 growth/differentiation factor 2 preproprotein [M musculus]
	Synonym BMP9
BMP	MmGdf3 NP_032134.2 growth/differentiation factor 3 preproprotein [M musculus]
BMP	MmGdf5 NP_032135.2 growth/differentiation factor 5 preproprotein [M musculus]
BMP	MmGdf6 NP_038554.1 growth/differentiation factor 6 preproprotein [M musculus]
BMP	MmGdf7 NP_001299805.1 growth/differentiation factor 7 isoform 1 [M musculus]
BMP	MmGdf9 NP_032136.2 growth/differentiation factor 9 precursor [M musculus]
BMP	MmNodal NP_038639.2 nodal preproprotein [M musculus]
22 BMP	
TGF-β	CeDAF-7 NP_497265.1 dauer larva development growth factor daf-7 [C elegans]
TGF-β	CeUNC-129 NP_501566.1 uncharacterized protein CELE_C53D6.2 [C elegans]
TGF-β	DmDawdle NP_523461.1 dawdle, isoform A [D melanogaster]
TGF-β	DmMaverick NP_524626.1 maverick [D melanogaster]
TGF-β	MmAmh NP_031471.2 muellerian-inhibiting factor precursor [M musculus]
	Synonym MIS
TGF-β	MmGdf15 NP_035949.2 growth/differentiation factor 15 preproprotein [M musculus]
TGF-β	MmInha NP_034694.3 inhibin alpha chain isoform 1 preproprotein [M musculus]
TGF-β	MmLefty1 NP_034224.1 left-right determination factor 1 preproprotein [M musculus]
TGF-β	MmLefty2 NP_796073.1 left-right determination factor 2 preproprotein [M musculus]
TGF-β	MmTgfb1 NP_035707.1 transforming growth factor beta-1 preproprotein [M musculus]
TGF-β	MmTgfb2 NP_001316036.1 transforming growth factor beta-2 isoform 2 [M musculus]
TGF-β	MmTgfb3 NP_033394.2 transforming growth factor beta-3 precursor [M musculus]
12 TGF-β	
45 total TGF-β family sequences	
outgroup	MmGdnf NP_034405.1 glial cell line-derived neurotrophic factor 1 [M musculus]

Table S2. Cleavage site identified computationally separating prodomain and ligand.

Subfamily	Name	Residue	Prodomain	Cleavage site	Ligand
Activin	CeTIG-3	125	NVINALSVLR	//	NAPNFDVMVF
Activin	DmActivin-beta	822	TESSRTRRVR	//	RRAVDCGGAL
Activin	DmMyoglianin	483	GSQKKHRRKR	//	SVYMDCTEND
Activin	MmBmp3	276	ALSVERRKKR	//	STGILLPLQN
Activin	MmGdf10	328	LKPRTGRKDR	//	KKKDQDTFTA
Activin	MmGdf11	287	VLENTKRSRR	//	NLGLDCDEHS
Activin	MmInhba	298	SEDHPHRRRR	//	RGLECDGKVN
Activin	MmInhbb	285	RLGDSRHRIR	//	KRGLECDGRT
Activin	MmInhbc	226	RVEGKHVRVR	//	RGIDCQGASR
Activin	MmInhbe	227	EPGAGRARRR	//	TPTCEPETPL
Activin	MmMstn	258	VTDTPKRSRR	//	DFGLDCDEHS
BMP	CeDBL-1	227	SEPSSVRRKR	//	SAQTGNSERK
BMP	CeTIG-2	258	KPNENERCQR	//	KGLYVDFDIL
BMP	DmDPP	463	GGGKGGRNKR	//	QPRRPTRRKN
BMP	DmGBB	316	AHSSHRSKR	//	SASHPRKRKK
BMP	DmSCW	262	VKIQKLRFKR	//	DLEKRRAGGG
BMP	MmBmp2	271	GHPLHKREKR	//	QAKHKQKRKL
BMP	MmBmp4	283	HTLTRRRAKR	//	SPKHHPQRSR
BMP	MmBmp5	307	ASEVLLRSVR	//	AASKRKNQNR
BMP	MmBmp6	362	VSEVHVRTTR	//	SASSRRRQQS
BMP	MmBmp7	282	ATEVHLRSIR	//	STGGKQRSQN
BMP	MmBmp8a	251	ASQSPVRAPR	//	AARPLKRRQP
BMP	MmBmp8b	254	SPVRAPRTAR	//	PLKKKQLNQI
BMP	MmBmp10	304	IDDSSARIRR	//	NAKGNKYCKKT
BMP	MmBmp15	258	ESSFLMRSVR	//	QACSIEDAS
BMP	MmGdf1	246	GPVGTCRTRR	//	LHVSFREVGV
BMP	MmGdf2	309	VGPLLARRKR	//	STGASSHCQK
BMP	MmGdf3	243	HCHPSSRKRR	//	AAISVPKGFC
BMP	MmGdf5	365	EYLFSQRRKR	//	RAPLANRQGG
BMP	MmGdf6	346	GKKSRLRCSR	//	KPLHVNFKEL
BMP	MmGdf7	348	AGRGHGRRGR	//	SRCRSKSLHV
BMP	MmGdf9	297	EVERSPRRRR	//	GQKAIRSEAK
BMP	MmNodal	234	ERGGWGRRQR	//	RHHLPDRSQL
TGF- β	CeDAF-7	225	TRPKGSRKRR	//	SHAKPVCNAE
TGF- β	CeUNC-129	280	FGEETSREER	//	ERIANEELAN
TGF- β	DmDawdle	457	IVIDMQNRRR	//	KSRQKRSINC
TGF- β	DmMaverick	577	HRSNHDSTWR	//	KDKWTNNCYK
TGF- β	MmAmh	437	GREGRGRTGR	//	SAGTGTGPGC
TGF- β	MmGdf15	179	LRVAAGRGRR	//	SAHAHPRDSC
TGF- β	MmInha	224	APSAGERARR	//	STPSVPWPWS
TGF- β	MmLefty1	211	GTWSSHKLVR	//	FAAQGTPDGK
TGF- β	MmLefty2	211	GTWSAHKLVR	//	FAAQGTPDGK
TGF- β	MmTgfb1	269	QHLHSSRHRR	//	ALDTNYCFSS
TGF- β	MmTgfb2	321	SQQSSRRKKR	//	ALDAAYCFRN
TGF- β	MmTgfb3	291	PGQGSQRKKR	//	ALDTNYCFRN
outgroup	MmGdnf	111	QAAALPRRER	//	NRQAAAASPE

Table S3. Spacer between cleavage site and first cysteine of the ligand is hypervariable.

Subfamily	Name	Residues between cleavage site and first cysteine	Length
Activin	CeTIG-3	NAPNFDVMVFQPNTVTAGTSD	21
Activin	DmActivin-beta	RRAVD	5
Activin	DmMyoglianin	SVYMD	5
Activin	MmBmp3	STGILLPLQNNELPGAIEYQYKEEGAWEEKPYKSLQTQPPEKSRN	80
Activin	MmGdf10	KKKORKGSHOKGOTLOFDEOTLKKARRKOWVEPRN	36
Activin	MmGdf11	KKKDQDTFTAASSQVLDLDFDEKTMQKARRRQWDEPRV	36
Activin	MmInhba	NLGLD	5
Activin	MmInhbb	RGLE	4
Activin	MmInhbc	KRGLE	5
Activin	MmInhbe	RGID	4
Activin	MmInhbe	TPT	3
Activin	MmMstn	DFGLD	5
BMP	CeDBL-1	SAQTGNSERKNRKKGR KHHNTEAESNL	27
BMP	CeTIG-2	KGLYVDFDILGWKQWVIAPEGFSAFY	26
BMP	DmDPP	QPRRPTRRKNHDDT	14
BMP	DmGBB	SASHPRKRKKSVPNNVPLLEPMESTRS	28
BMP	DmSCW	DLEKRRAGGGSPPPPPPPVDLYRPPQS	28
BMP	MmBmp2	QAKHKQRKRLKSS	13
BMP	MmBmp4	SPKHPQRSRKKNKN	15
BMP	MmBmp5	AASKRKNQNRNKSNSHQDPSRMPSAGDYNTSEQKQA	36
BMP	MmBmp6	SASSRRRQQSRNRSTQSQDVSRGSGSSDYNGSELKTA	37
BMP	MmBmp7	STGGKQRSQNRSKTPKNQEALRMASVAENSSSDQRQA	37
BMP	MmBmp8a	AARPLKRRQPKKTNELPHPNKLPGLFDDGHGSGRGREV	37
BMP	MmBmp8b	PLKKKQLNQLNQLPHSNKHLGILDDGHGSHGREV	34
BMP	MmBmp10	NAKGNV	6
BMP	MmBmp15	QA	2
BMP	MmGdf1	LHVSFREVGVHRWVIAPRGFLANF	24
BMP	MmGdf2	STGASSH	7
BMP	MmGdf3	AAISVPKGF	9
BMP	MmGdf5	RAPLANRQGKRPSKNLKR	19
BMP	MmGdf6	KPLHVNFKELGWDDWIIAPLEYEAYH	26
BMP	MmGdf7	SR	2
BMP	MmGdf9	GQKAIRSEAKGPLLTASFNLSEYFKQFLFPQNE	33
BMP	MmNodal	RHHLPPDRSQL	10
Unexpected conservation between MmNodal and CeDBL-1 in red			
TGF- β	CeDAF-7	SHAKPV	6
TGF- β	CeUNC-129	ERIANEELANDVRVVLQNKNR	22
TGF- β	DmDawdle	KSRQKRSIN	9
TGF- β	DmMaverick	KDKWTNN	7
TGF- β	MmAmh	SAGTGTGDP	9
TGF- β	MmGdf15	SAHAHPRDS	9
TGF- β	MmInha	STPSVPWPWSPAALRLLQRPPEEPAAHAF	29
TGF- β	MmLefty1	FAAQGTPDGKGQGEPQLELHTLDLKDYGAAQGN	32
TGF- β	MmLefty2	FAAQGTPDGKGQGEPQLELHTLDLKDYGAAQGN	32
TGF- β	MmTgfb1	ALDTNY	6
TGF- β	MmTgfb2	ALDAAY	6
TGF- β	MmTgfb3	ALDTNY	6
outgroup	MmGdnf	NRQAAAASPENSRRGKRRRGQRGKNRG	26

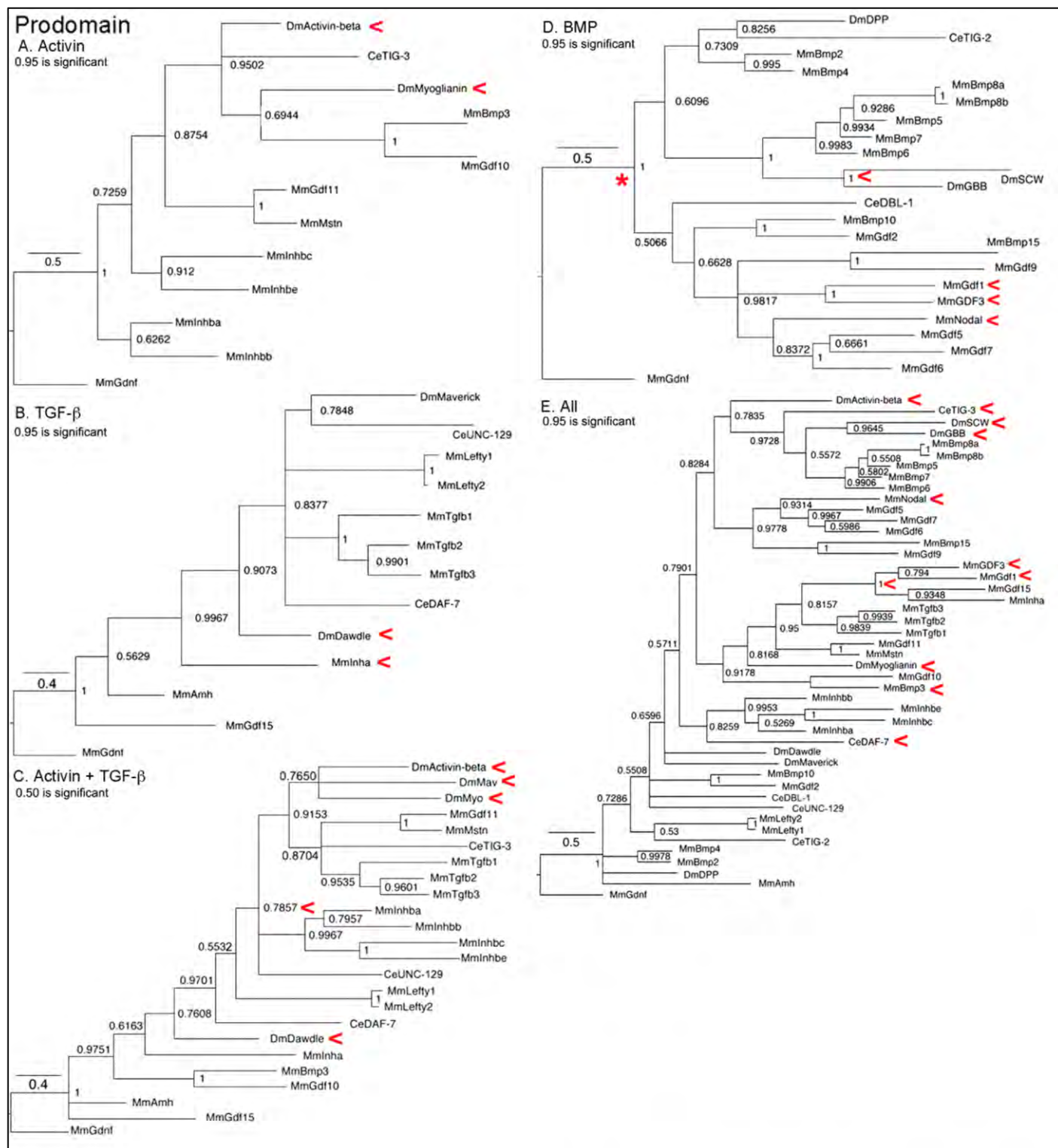


Fig. S1. Prodomain trees. Trees for this domain are displayed as in Fig. 1. Red arrowheads indicate a cluster that may reflect common regulation. For every tree except Activin+TGF- β nodes above 0.95 are significant. For Activin+TGF- β nodes above 0.50 are significant. A) Activin. The significant cluster of Activin and Myoglianin

is unexpected. B) TGF- β . The significant cluster of Dawdle and Inhibin- α was unexpected. C) Activin+TGF- β . The significant cluster of Activin, Maverick and Myoglianin that is clustered with the four Inhibin- β proteins and Dawdle's location near Inhibin- α were unexpected but consistent with cysteine conservation in the "Association region" and β 8. D) BMP. The significant cluster of Gbb/Screw in all trees was unexpected. The cluster of heterodimerizing Nodal and GDF1/GDF3 was expected. Red asterisk indicates node leading to two symmetric secondary clusters. E) All family members. The not quite significant cross-subfamily cluster of Activin, TIG-3, Gbb and Screw with Nodal was unexpected but three are known to heterodimerize and two have conserved cysteines. The absolute cluster of GDF3/GDF1 with GDF15/Inhibin- α and this group's not quite significant cluster with Myoglianin was unexpected. The not quite significant cluster of DAF-7 with the four Inhibin- β proteins was unexpected but is consistent with "Association region" cysteine conservation. The cluster of BMP3/GDF10 with Myoglianin was unexpected.

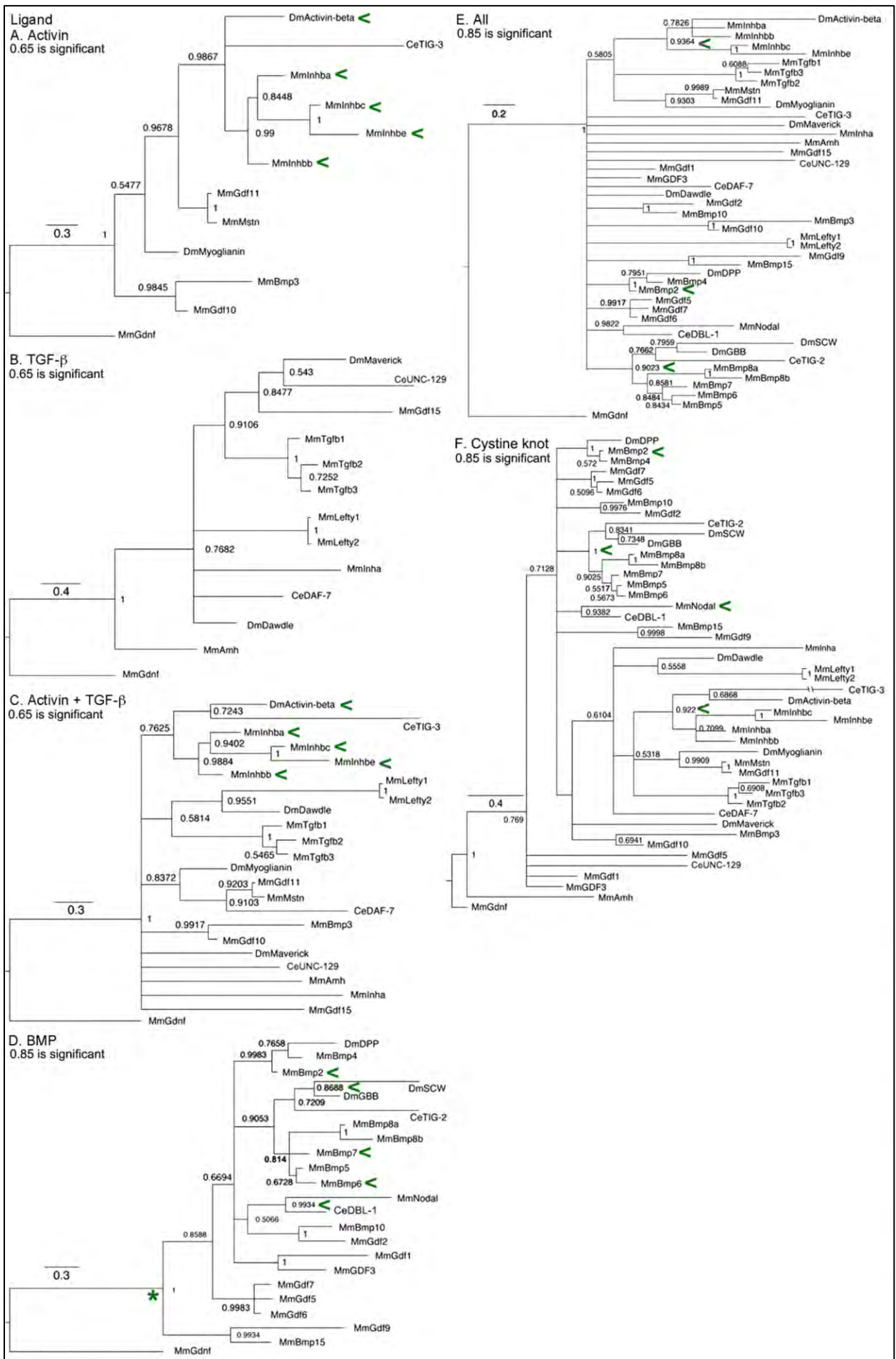


Fig. S2. Ligand trees. Trees for this domain are displayed as in Fig. 1. Green arrowheads indicate a cluster that may reflect common function. For Activin, TGF- β and Activin+TGF- β nodes above 0.65 are significant. For BMP subfamily and All family members trees nodes above 0.85 are significant. A) Activin. The significant cluster of Activin and the four Inhibin- β proteins was expected. B) TGF- β . The significant clusters of TGF- β 1-3 and Lefty1,2 were expected. C) Activin+TGF- β . The significant cluster of Activin and the four Inhibin- β proteins was expected. D) BMP. The significant cluster of BMP2-8a,b with Gbb/Screw was expected and consistent with functional heterodimers of BMP2-BMP6 and BMP2-BMP7 that have been reported. The significant cluster of Nodal and DBL-1 was unexpected. Green asterisk indicates node leading to two asymmetric secondary clusters. E) All family members. Several significant clusters are expected such as the four Inhibin- β proteins and Dpp/BMP2/BMP4. A significant cluster of Nodal and DBL-1 was unexpected. F) Cystine knot nodes above 0.85 are significant. Several significant clusters are expected such as Activin with four Inhibin- β proteins, Dpp/BMP2/BMP4 and Gbb/Screw/BMP5-8a,b. A significant cluster of Nodal and DBL-1 was unexpected.

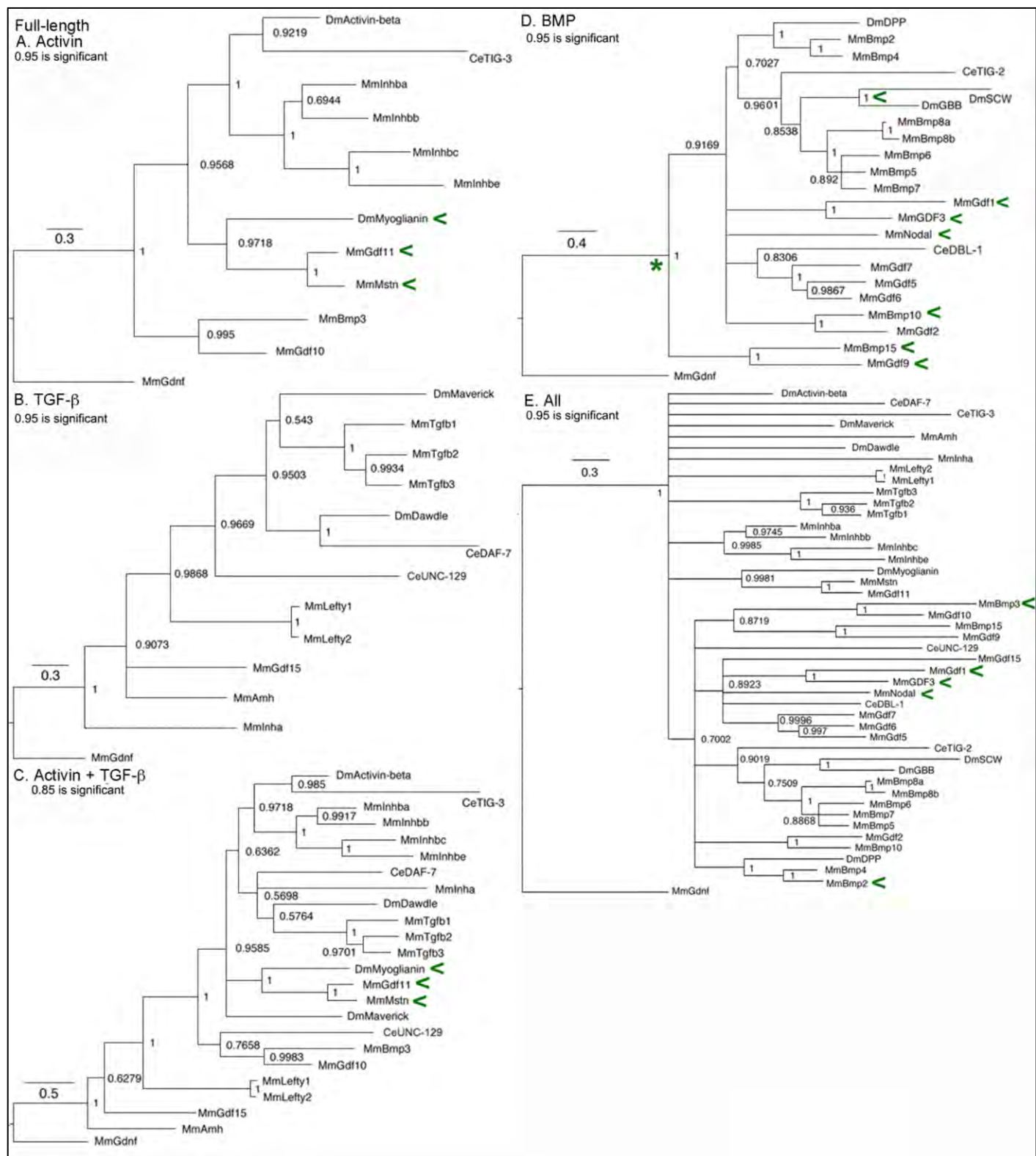


Fig. S3. Full-length trees. Trees for this domain are displayed as in Fig.1. Green arrowheads indicate a cluster that may reflect common function. For every tree except Activin+TGF- β nodes above 0.95 are significant. For Activin+TGF- β nodes above 0.85 are significant. A) Activin. The significant cluster of Myoglianin and

Myostatin/GDF11 was expected. B) TGF- β . The significant clusters of TGF- β 1-3 and Lefty1,2 were expected. C) Activin+TGF- β . The significant cluster of Myoglianin and Myostatin/GDF11 was expected. D) BMP. Clustering of GBB/Screw, BMP10-GDF9, BMP15-GDF9 and Nodal-GDF1/GDF3 are consistent with heterodimerization that has been reported. Green asterisk indicates node leading to two asymmetric secondary clusters. E) All family members. The not quite significant clustering of BMP3 and GDF1 with all of the BMP subfamily proteins was unexpected. The not quite significant cluster between Nodal and GDF1/GDF3 was expected.

			--α1----- -latencylasso- ----α2-----
MmBmp3	124		-----
MmGdf10	123		-----
CeTIG-3	1		-----
DmMyoglianin	205	DS	ESTKMHILMRLNEKKLENTTKPTSVQPNIIDNFY-RDYNASSKTTVWNRMESIDESH
MmGdf11	72	LR	ESTKSQLSLRLKEAPNISREIVFQLLP-KAPPLQQLDL-----HDFQGD
MmMstn	52	SR	EATKIQLSKLRLETAPNISKDAIRQLLP-RAPPLRELIDQ-----YDWQDD
DmActivin-beta	369	VR	ESTKRCQILTKLGLSHKPNVSHPIPKQFHW-ETIY-----RWDGGR
MmGdnf	92	DVM	FFQATIKR---LKRSPDKQ---AAALP-RRER-----
MmInhba	53	EM	EAVKKHILNMLHLKKRPDVTQPIPRALL-NAIR-----KTHVGK
MmInhbb	77	DE	EAUKRHILSLQLRGRNITHAMPRAAV-TALR-----KTHAGK
MmInhbc	42	LL	DLAKSILDKLHLSQRPIISRPVSRGALK-TALQ-----RIRGPR
MmInhbe	42	LV	IELAQQILEGLHLSRPREITRPLPQAALT-RALR-----RQPKS
			end straitjacket--β1--
MmBmp3	124		-----
MmGdf10	123		-----
CeTIG-3	1		-----
DmMyoglianin	264	S	INDTYGD-----HIMTDEDESSSSQMQGDDAN-----
MmGdf11	122	Q	PEDFLEED-----EYHATTETIEMAQ-----ETDPAVQTD
MmMstn	102	S	-SDGSLEDD-----DYHATTETIIMPT-----ESDFLMQAD
DmActivin-beta	411	M	IPNNAFGS-----SGKNLDQKTIKRAFASPGSHLFNGRGGRTDQSRERDPSSHKY
MmGdnf			-----
MmInhba	95	V	GENGYVEIEDDIGRAEMNELMEQTSEIIFAESG-----T-A
MmInhbb	119	V	REDGRVEIPHLDGHASPGADQQERVSEIISFAETDGLA-----S-S
MmInhbc	84	R	ET---LLEH-----DQRQEYELISFADTDLSS-----I-N
MmInhbe	84	V	---P-----GN---REKVISFATIIDKS-----TSTY
			begin arm-β2--- -β3---
MmBmp3	124		-----NILSATYFYVVELVN1---SLSC
MmGdf10	123		-----MILTAAFHFYSEPPRPAPRAEVFC
CeTIG-3	17		-----ITLEPNRKSWTCTPES
DmMyoglianin	330	S	IYIFPEEIQPHVRHNRKVDVFRFQI-----DSSYSDLSYATLHLYL-RGWDWSAHQ
MmGdf11	179		-----WVYLFPVPR
MmMstn	158		-----WTYLPVK
DmActivin-beta	643	E	ITFAEEGT---QYRQYRILEFSAQNRRVPSQKLSIRSQTIRIDKPHSTWEKAKS
MmGdnf			-----
MmInhba	153		-----AEV-----WFLMVPKA
MmInhbb	180		-----ASL-----WTYLLLPY
MmInhbc	132		-----TRF-----MFEWQFPHN
MmInhbe	125		-----AR-----WFLVPPS
			-α3- -β4--
MmBmp3	145	P	-E---PQGCSSHHTQEQ-----HIQIDSAWILKSNQSQLLG-----
MmGdf10	147	K	PRAK-NASCRLLTPLG-----PARLHIFRSL-SQNTATQG-----
CeTIG-3	34		-----LVKDCFYQSINSINHEILSASLIIDPKDNTNISTVYEVDELFI-----
DmMyoglianin	383	G	---LLEEIKK---QPRK-----DGVVTHRAIRVA-----NT
MmGdf11	188	-	PAT-----VYLOLRLKPLTGEGTAGG-----G
MmMstn	167	-	PTT-----VFQQLRLIKPMKD-----
DmActivin-beta	699	L	PEKHLNLTNRKKGANKPHH-----RKLWVQLSTSI-----NITEK
MmGdnf			-----
MmInhba	165	N	---R-----TRT-----KVTIRFQQQKHPQGS�DTGDEAEEM
MmInhbb	192	V	LEKG-----SRR-----KVRKYFQE
MmInhbc	144	A	TQT-----QNRVLVLRP-----
MmInhbe	135	F	PGT-----LYRFRCGT-----
			---β5- -β6-- -α4--- -β7--- no-β8
MmBmp3	178		-----HLSVDVVRPYRDSVSWLSKDTQLRKAKQNE-----EFLIGFNITSRAHELP
MmGdf10	182		-----LARGAMALTPPPRGLWQARDISSIKAARRDG-----EFLISAQLDTGEKDPG
CeTIG-3	76		-----GELQYV---DRFER-ETLDKYHFDISHFHKWMKCKSSDKY-KKIEITNSNT-----
DmMyoglianin	410	T	SFNPVKVMEFRHS-PSGLGQWVAVDLKSLLGNLGSMTQ---EELIK- GAET-----
MmGdf11	211	G	GGRRHIRRSLKIEHSRSGHWQSIDFKQVHSHWFRCP-QSNWGEINAFDPGSG-----
MmMstn	184	-	GTRYTGRLSKLDSPGTGIMQSIDVKTILQNWLKCP-ESNLGEIKALDENG-----
DmActivin-beta	737	G	--IDKAIIFRASFOQDPKNLGWQKEDLDTIREWYGHTSHEKLRLLDCTGCGG-----
MmGdnf			-----
MmInhba	196	G	LKGERSEILLSEKVLDARKSTWHIFPVSSSQRLLDG-GKSSLDRIACEQCQE-----
MmInhbb	210	--	QHGDRNVVVEKKVLDKRSQWHTFPTTEALQALFER-GERRNLDVQDCSCQE-----
MmInhbc	158	--	YDTNLTLSQYVQVNASGWYQLLGPEAQACSC-GHLTLET---VPESQVA-----
MmInhbe	149	--	TRCRGFRFLAEHQTTSSGHALTPSSGLRSEDS-GVVKLQEFRLDLNS-----
			no bowtie -β9--- no-β9'- --β10- no-α5
MmBmp3	226	K	RMLFFPEPYLVYANDA-----AISEPESVSSLRHREDFTAGT---P
MmGdf10	230	V	PRPSSHMPYLVYANDL-----AISEPNSVAVSLQRYDFFPAGDFEPGAAP
CeTIG-3	124		-----QNTINALS---VLR-----
DmMyoglianin	460		-----WM-----KSLVTTDNTSBNPTVHIEIGS---
MmGdf11	265	T	DIAVTS-----LPGAEGLHFPFEDRVLENT---
MmMstn	236	H	DAVTF-----PGGEDGLNPFLEVKVTDTP---
DmActivin-beta	790	-	RYSLHLFQTSKLR---G-----NSSDYLSTNPNRPFVHTESSR---
MmGdnf			-----
MmInhba	250		-----SGASLVLLGKKKKKEVDGDKKKDGSDDGLEEEKEQSHRPFVQARQSE---
MmInhbb	262		-----LAIVPVFVD-----PGEESHRRPFVQARLGD---
MmInhbc	207		-----HSSL-ILGWFSHRPFVAAQVRVE---
MmInhbe	200		-----TAAG-----LPRLLLDTAGQQRPFELKIRANE---

Fig. S4. Expanded Activin subfamily prodomain alignment indicating structural conservation. Sequences from Fig. 2 with no omissions, similar underlining and accurate numbering. Ungainly leaders and trailers were trimmed for brevity.

			-----α1-----	-----LatencyLasso-----	-----α2-----	
CeUNC-129	15	----	IANCAKDV	-----LNETTRDLHFKSSD-----	PNVTSFHRSSHT-----	
CeDAF-7	19	GLTFNCTNSG	CIKMKQHRTEYKNEELDQLNKEA	PKGLKPMDEPKSVYLEMY---	R	
DmMaverick	245	-----	SKRVDETRRHLV	KKGLGKKLDMRKVNISQAEYS	-----S	
MmTgfb1	27	AAG-LSTCKT	DMVLVKKRIEA	RGQLSKRLASPE	SQGEVPPGPPEAVLALYNSTR	
MmTgfb2	18	ALS-LSTCST	DMQFMKRKIEA	RGQLSKRLKTSPE	EDY-PEDEVPEVISIYNSTR	
MmTgfb3	21	SLS-LSTCTT	DFGHIKKRVEA	RGQLSKRLTSP	PEPS-V-MTHVPYQVLALYNSTR	
MmLefty1	20	EAL-T-----	GEO-----	ILGSELQQLQDQPE	VLDKADVEGVIPSHVRTQYV-	
MmLefty2	20	AAM-T-----	EEO-----	ILSSELQQLQSQAT	LDSDADVEEAPITHVRSQYV-	
MmInha	204	CSC-SGRPET	-----	-----	-----	
DmDawdle	233	CES-NRQVEH	TEQLTHLRIEFV	QOITEKIRKES	PKVSA---VEPKPIFDGM---	
MmGdf15	118	AYR-VHRALL	LTPTARPWDITRPL	KRALSRGPRA	ALRLRL---TPPPDLA-----	
MmGdnf	-----	-----	-----	-----	-----	
MmAmh	351	AAT-EREEMP	THGPASAPWAAGLQ	RVAVE	TQAAASELRDLPGLEPTAPPLLARL---	
			end straitjacket--β1--			
CeUNC-129	53	-----	TEHMKNL	ENFIDE	SDSNEDGNLVRAIEPAVGKFEQEVLVFDVE	
CeDAF-7	76	DL	EKDEQDM-----	GVMSFYTAKD-----	-----	
DmMaverick	282	KY	EYLSRLR---SNQEKGSY	NNFMGASFT-----	-----	
MmTgfb1	86	DR	AGESAD-----PEPE	PEADYAKEV-----	-----	
MmTgfb2	76	DL	QEKASRRAAACERERSDEEY	AKEV-----	-----	
MmTgfb3	78	EL	EMHGEREEGCTQETSESEY	AKEI-----	-----	
MmLefty1	63	AL	QSHASR---SRGKRFSQN	REVA-----	-----	
MmLefty2	63	AL	QSHADR---SRGKRFSQN	REVA-----	-----	
MmInha	213	-----	-----	-----	-----	
DmDawdle	285	-T	SHPDDST---KNK-ELDDY	ARTS-----	-----	
MmGdf15	167	---	MLPS-----GGTQ	LELR-----	-----	
MmGdnf	-----	-----	-----	-----	-----	
MmAmh	405	-L	ALCPNDSR---SSGDPL	RALL-----	-----	
			begin arm --β2--	-----α3--	-----β3--	
CeUNC-129	143	TLKKIRVGGDENLE	-----EYK--VI	DATKS-VFDSYHLDAKQ	AVFRITREHSMKRPYA	
CeDAF-7	102	-----	NPSQLVAKEDVTND	-----LERSD	LQALTYISIEIPAK---	
DmMaverick	356	--L--LRGEQDT	-----MNILLH	PELTNAQDANFHHDK	IDEANVRMLLYSSSL---	
MmTgfb1	126	-----	DISH---SIY---MF	NTSDIREAVPEPPL	SRAEIRQRLKS-----	
MmTgfb2	136	SQV--LCGYLDAIPTFYRPFYRIVRE	DVST---	MEKNASNVKAEFR	VRFLQNPK---	
MmTgfb3	126	-----	GITSKVFR	NNVSS---VEKNGTN	FRAEFRVLRVPNS---	
MmLefty1	95	-----	STHLLVGE	MEQ---RLPPNSE	VQAVLRVQEPVPR---	
MmLefty2	95	-----	STHLLVGE	MEQ---RLPPNSE	VQAVLRVQEPVPR---	
MmInha	213	-----	-----	-----	-----	
DmDawdle	327	-----	SNPSMCFT	EKIDD---ADAEGFD	STAVLWLFKNKQNR---	
MmGdf15	-----	-----	-----	-----	-----	
MmGdnf	-----	-----	-----	-----	-----	
MmAmh	-----	-----	-----	-----	-----	
			-----β4--	-----β5--	-----β6--	
CeUNC-129	249	-----	LDN	DREPIKRNKGN	NSLSEEISSDVMQGE	PEPSREE---
CeDAF-7	138	GMLQDVQ	OVYEKNE--G-S-MGEM	YTSGIFA---TK-GSERIS	IOLPIDTKSWFTIS	
DmMaverick	441	NSSQQLT	KKVYQLLSANRRRKITSRKTE	FNGVGF--QETRTQ	WIEFDVTKAWRSWLNSKS	
MmTgfb1	163	--SVEQH	EYLYQKYS---NNSWRY	GNELL---TPTDTP	EWLSFDVTGVVRQWLNQG	
MmTgfb2	187	ARVAEQR	EYLYQILKSKLTSP	QRYDSKV---KTRAEGE	WLSFDVTDVQEWLHHK	
MmTgfb3	161	SKRTEQR	EYLYQILRPD	H-IAQORYGGN---	PTRGTAEWLSFDVTDVREWLLRR	
MmLefty1	140	SARARVT	EWLRFRD--G	SNRTA-LDSRLV---	SIHESGWKAFD-VTEA	
MmLefty2	140	SARARVT	EWLRFRE--G	SNRTA-LDSRLV---	SIHESGWKAFD-VTEA	
MmInha	213	-----	-----	-----	-----	
DmDawdle	370	-TSAQQT	EVSEVEV--GQKDSKYL	SAAATTAQSVNVQDE	WMKID-EWPE	
MmGdf15	-----	-----	-----	-----	-----	
MmGdnf	-----	-----	-----	-----	-----	
MmAmh	-----	-----	-----	-----	-----	
			-----β7--	-----β8--	-----bowtie-----	
CeUNC-129	289	-----	-----	-----	-----β9--	
CeDAF-7	190	PI-QGIFVKAMLDGRNV	-----	-----	ALHPQ	
DmMaverick	498	HENLGIE--IQ	CKCKSIGA--RILSDFSPST	PPRSTASSDEHLNLM	VPVNIIGHGTINS	
MmTgfb1	212	DGIQGFRESAH	CSCDSKDNK-----	-----	LHVEING---I	
MmTgfb2	243	DRNLGFKISLH	CPCTFVPSNNYIIPNKSEE	-----	LEARFAGIDGT	
MmTgfb3	216	ESNLGLEISIH	CPCHTFQPNGD-ILENVHEV	-----	MEIKFKGVDNE	
MmLefty1	192	RP--R-----	-----	-----	-----	
MmLefty2	192	RP--R-----	-----	-----	-----	
MmInha	213	-----	-----	-----	-----	
DmDawdle	426	EL--SHLIQIT	CGGCDVSDMEEIISVD	-----	-----	
MmGdf15	-----	-----	-----	-----	-----	
MmGdnf	-----	-----	-----	-----	-----	
MmAmh	-----	-----	-----	-----	-----	
			no-β9'	-----β10--	-----α5--	
CeUNC-129	289	-----	-----	-----	R---	
CeDAF-7	211	QTTADVDMRLQL	-----	S---	TRPKG--S-RKRR---	
DmMaverick	554	QQHGDADIHQIMLTNNRS	DQYV-----	-----	HHRSNHDSWTW---	
MmTgfb1	240	SPKRRGDLGTIH	-----DMNR	EFLLMATPLERAQHL	SSRRHR---	
MmTgfb2	285	STYASGDKTIKSTRKKTSGKT	PHLLMLLLPSYRLES	QQSSRRKKR---	-----	
MmTgfb3	257	DDHGRGDLGRLK	-----KQKDHNN	PHLLMMIPPHRLDS	PGQGSQRKKR---	
MmLefty1	195	-----	-----	CELL--QVSQV--REHLG	PGTWSSHKLV---	
MmLefty2	195	-----	-----	CELL--QVSQV--REHLG	PGTWSSHKLV---	
MmInha	213	-----	-----	TEFLVAHTRAR--APSAGE	--RARR---	
DmDawdle	451	-----	-----	KDYRE	FFLIDMQNR--RR---	
MmGdf15	-----	-----	-----	-----	-----	
MmGdnf	-----	-----	-----	-----	-----	
MmAmh	-----	-----	-----	-----	-----	

Fig. S5. Expanded TGF-β subfamily prodomain alignment indicating structural conservation. Sequences from Fig. 4 including underlining and presented as in Fig. S4.

		--α1-----	--LatencyLasso-	-----α2-----
DmMaverick	245	SKRVDETR-IRHLVLKGLGKKL	PDTRKVNISQA	---EYSSKYIEYLSRLRS-----
DmMyoglianin	202	AKVDSIES-LKMHILMRNLKKL	PNITKPI	-----SVPQNIIDNFYRDYNASS-KT
MmGdf11	69	SRELRLS-LKSQILSKRLKEA	PNISREVVQQLLPKAPPLQQILDLHDFQGDALQ-P-	
MmMstn	49	TRYSLREA-LKIQILSKRLETA	PNISKDAIRQLLPRAPPLRELIDQYDVQRDDS--S-	
CeTIG-3	1	MSTSR		
MmTgfb1	41	VKRKRLEA-IRGQILSKRLASP	PSQGEVPPG	-----PLPEAVLALYNSTRDRVA-G-
MmTgfb2	32	FMRKRLEA-IRGQILSKRLTSP	PEDY-PEPD	-----EVPPEVISIYNSTRDLLQ-E-
MmTgfb3	35	IKKKRLEA-IRGQILSKRLTSP	PEPS-V-MT	-----HVPYQVLALYNSTRELLE-E-
CeUNC-129	25	-----L-INTETRDLEHFKSSDPN		-----TSFHRS-----
CeDAF-7	34	MKQHRTE-YLKNELIDQNMKEA	PKGLKPMDEPMKSVYLE-M-YRDLLEK-D----	EQ
MmBmp3	56	VSEHMLWLYDRYSGSSRQATRT	PGSQLPGPQPL	
MmGdf10	71	VAIHMERLYEKYNRRG		-----APP-----
MmInhba	50	SQPEMVEA-VKKHILNMHLLKKR	PDNTQVPVKAALLNAIRKL	-----H-----VG
MmInhbb	74	VDGDFLEA-VKRHILSRQLRGR	PNITHAVPKAAMVTALRKL	-----H-----AG
MmInhbc	39	QRELLLL-ARKSLDKHLSQR	PIISRPVSRGALKTALQRL	-----R-----GP
MmInhbe	39	ERALVLEL-AKQQLLEGLHLSR	PRITRPLPQAALTRALRL	-----Q-----PK
DmActivin-beta	546	MNE		-----NALKKSTYPIDINHSDNKTHTG
MmLefty1	27	-----VLGSLLQQQLDQF	PVLDKADVEGMVIPSHVRTQYVALLQSHAS	-----R
MmLefty2	27	-----VLSSLLQQQLSQA	PTLDSADVEEMAIPTHVRSQYVALLQGSHAD	-----R
MmInha	213			
DmDawdle	247	LTHLRLEF-VKQQLLEKRLKES	PKNSA	---VELPKPIFDGM---TLSPDDSDS---T
MmGdf15	132	ARPWDLTR-PL-KRALSRGPR	PAIRLRL	---TPPPDLA---MLPS---
MmGdnf				
MmAmh	365	SAPWAAGL-QR-RVAVEQAQAS	ELRDLPLGLPPTAPPLARL	----LALCPNDS----R
		end straitjacket -β1-- TGF-β		
DmMaverick	293	-----NQEKGSYFNFMGASFTRDHLFLSITTNFNDISNKRRLHRRSLKKIN		
DmMyoglianin	251	TVVNRMESIDESHLISINDTYGDHIMTDFDESSSSQMQGDDANTVNEFLIDLNKNQAKKS		
MmGdf11	125	EDF	-----LEEDEYHAT	-----TETVISMAQETDPAV
MmMstn	104	DGS	-----LEDDDYHAT	-----TETIITMPTESDFLM
CeTIG-3	6	-----KHDLYGG		-----VLDKITLEPNRK---
MmTgfb1	91	ESAD	-----PEPEPEADYYAK	-----EVTRVLMVDRNNAIY
MmTgfb2	81	KASRRAAA-CERERSDEEYYAK		-----EVYKIDMPSHLPSET
MmTgfb3	83	MHGEREEG-CTQETSSEYYAK		-----EIHKFDMIQGLAEHN
CeUNC-129	50	--SHT		
CeDAF-7	84	DMGVE	-----MSFYTA	-----KDPSYGEN
MmBmp3	90	RGG		-----NTVRSFRAAAA
MmGdf10	90	GGG		-----NTVRSFRARLE
MmInhba	94	KVG		-----EN-----G-----Y
MmInhbb	118	KVR		-----ED-----G-----R
MmInhbc	83	RRE		-----T-----
MmInhbe	83	SMV		
DmActivin-beta	572	KNG-E	-----MSHNDYEFNDYSVQ	-----THDKNRYHEGRSSIGYQPAIHN
MmLefty1	73	SRGKR	-----FSQNLRVAGRFVLS	-----ET-----S---
MmLefty2	73	SRGKR	-----FSQNFREVAGRFVLS	-----ET-----S---
MmInha	213			
DmDawdle	294	KNK-E	-----LDDYYARTSKKFILL	-----NR--EEVECNARDGKSNP---
MmGdf15	171	-GGTQ	-----LELR	-----LRVAAGRGR
MmGdnf				
MmAmh	414	SSGDP	-----LRALLL	-----LKALQG
		end straitjacket -β1-- Activin		
DmMaverick	342	RLNQNPKKHQNYGD	-----LLRGEQD	-----TMNILLHPLTN-----AQ
DmMyoglianin	311	DIPINTNDE-EYES	-----ILSHISS	-----IY-----IPPEIQPHVRHNRKVD
MmGdf11	152	Q	-----TD	-----GS-----PLC
MmMstn	131	Q	-----AD	-----GK-----PKC
CeTIG-3	25	-SWTCLTPE		-----SLVKD
MmTgfb1	122	EK	-----TK	-----DI-SH-----SIY---
MmTgfb2	117	VCPVVTPSGSLGS	-----FCSRQSQ	-----VLC---GLDAI-PPTFYRPYFR
MmTgfb3	119	ELAVC	-----PK	-----GITSK
CeC53D6.2	53	-LTE	-----HMKNLYENFIDEDSNEDGNLVRAIEPAVGKFEGQEVLV	-----DVEG-FDSHESIMRA
CeDAF-7	103			-----PSQLVAKSDVTN-DLERSDILQA
MmBmp3	104	-GTP	-----QTKG	-----LHTENLTS-LTKSENILSA
MmGdf10	104	-MI	-----DQKP	-----VYFENLTS-MQDSEMILTA
MmInhba	101	VEIE	-----DDIGRRA	-----EMNELMEQTSEIITAESEG---T---ARK
MmInhbb	125	VEIP	-----HLDGHAS	-----PGADGQERVSEIISAEETD-GLAS---SRV
MmInhbc	87	-LLE	-----H	-----DQRQEEYEIISADTD-LSSI---NQT
MmInhbe	86	-P		-----GN---REKVISATII-DKST-STYRS
DmActivin-beta	613	IEYE	-----NQKGHHESFADDHEN	-----IDHEDFFGNTQEIIITAESEG-TQYR---QYR
MmLefty1	96			-----TH-----
MmLefty2	96			-----TH-----
DmDawdle	330			-----SM-----
MmGdf15				
MmGdnf				
MmAmh				

Fig. S6. Page1

		begin arm -β2-- α3 TGF-β -----β3--	
DmMaverick	377	DANFHHDKIDEANVRLMLY-----SSSLATNFRRPGSRKNKISQISGNDNI	
DmMyoglianin	350	VRFQIDS-----SYSDIS-----YATLHYLRGWDWIS-----	
MmGdf11	160	-CHFHSPKV---MPTKVL-----KAQLWYLRPVPRP-----	
MmMstn	139	-CFFKSSKI---QYNKV-----KAQLWYLRPVKTP-----	
CeTIG-3	38	CFQMSINSINHEILSASL-----IDPKD-----	
MmTgfb1	133	-FFNTSDIREAVPEPPLS-----RAELRORLKS-----	
MmTgfb2	157	IRFDVST---MEKNASNV-----KAEFRVRLQNPKA-----	
MmTgfb3	131	VRFNVSS---VEKNGTNLF-----RAEFRVLRVPNPSS-----	
CeUNC-129	109	ETHFYRR--RDSFARR-RS---RQIR-----A---K-----	
CeDAF-7	125	TTVTSHEI--PAKDS-----	
MmBmp3	129	TLTYFYGE--LVNI---SFS-----P---E-----	
MmGdf10	128	AFHFYSEP--PRWPRAREMEC-----K---PR-----	
MmInhba	135	THFEFSK--EGSDL-SVE-----RAEVWFLKVPKAN-----	
MmInhbb	162	RHYFFVSN--EGNQNLV-----QASLWYLLKLPYV-----	
MmInhbc	114	RHEFHFSG--RMASG-MEVR-----QTRFMFFVQFPHNA-----	
MmInhbe	109	MUTFOVSP--LWSSH---Y-----HAR---WLHVPPSF-----	
DmActivin-beta	659	IEEFSAQN--RRVPS-QKLSIRSAQIHIRIDKPHSLWLEKAKSLPE-----	
MmLefty1	98	LVVFGMEQ--RLPPN-SEIV-----QAVLRVFOEPVP-R-----	
MmLefty2	98	LVVFGMEQ--RLPPN-SEIV-----QAVLRVFOEPVP-R-----	
MmInha	213	-----	
DmDawdle	332	CTFEKIDD--ADAEG-FDVS-----TAVLWLFKNQN-R-----	
MmGdf15		-----	
MmGdnf		-----	
MmAmh		-----	
		Activin-α3 ---β4--	
DmMaverick	425	ERHCNFGDVLNQS---NKNSSQQTIKYQLLSANRR-----RKITSRK-	
DmMyoglianin	379	-AHQ---PGLLEEI---KKQPRKDVVTHRAIRVANTT-----SFNPK---VKMFE-	
MmGdf11	189	-----ATVYLQILRLKPLTGEGTAGGGGGRRHIRIRS-	
MmMstn	168	-----TTVFVQILRLIKPMKD-----GTRYTGIRS-	
CeTIG-3	62	-----TNISVVEY---VDEL-----FGELQYVDR---	
MmTgfb1	163	-----SVEQHVEYQKYS-----NNSWRYLGNRL-	
MmTgfb2	188	-----RVAEQRIEYQILKSKDLT-----SPTQRYIDSKV-	
MmTgfb3	162	-----KRTEQRIEYQILRPDEH-----IAKQRYIGGKN-	
CeUNC-129	132	-----SVCVNEYCRQ---QTLK-----KIRVGGDENL-	
CeDAF-7	138	-----GMLQDQVQVYKNE-----DGS---MGEMVTSGIFATKG	
MmBmp3	147	-----PQGCSH--HTQRQHVIDSAWILKSN-----QSLLGLHLS---	
MmGdf10	150	-----AKNASCRLLTPGLPARLHIFRSL-S-Q-----NTATQGLLR---	
MmInhba	166	-----RTRTKVTRTFQQQKHPQGS�DTGDAAEMGLKERSE---	
MmInhbb	193	-----LE-----KGSRRKVRKYVQE-----QGHGDR---	
MmInhbc	145	-----TQ-----T---NIRMLVLRP-----YDTNL---	
MmInhbe	136	-----PG-----T---HYDRFRCGT-----TRCRG---	
DmActivin-beta	702	-----KHLNTRKRWGANKPHHRKIVWFQLSTSNITEKGIDKA-----IIFR---	
MmLefty1	128	-----TALRRQ-KRLSPHSARARITHEWLR-----FRDDGSNRT-----A-LID-	
MmLefty2	128	-----TALRRF-ERLSPHSARARITHEWLR-----VREDGSNRT-----A-LID-	
MmInha	213	-----	
DmDawdle	362	-----TDTASV-NS---TSAQQTIVSE-----VEVDQQKDS-----KYLSA---	
MmGdf15		-----	
MmGdnf		-----	
MmAmh		-----	
		--β5-- --β6-- --α4-- Activin --β7-- --β8-- bowtie	
DmMaverick	467	IEFGNVGFQETRTQWIEFDVTKARSWLNKS-HENIGIEIQ--CDKCKSIG--ARILS	
DmMyoglianin	421	FRHSIPSGLGQWVAVDLKSLLGNLGS---NTQELIKGAE-TWMK---	
MmGdf11	222	LKIELHSRSGHWQSIDFKQVHSWFRQP-QSNWGIEINAFDPSTGDLA-----	
MmMstn	193	LKLDMSPGTGIWQSIDFKTVLQNLKQP-ESNIGIEIKALDENGHDLA-----	
CeTIG-3	84	FE---IRETLDKYHEDLSHLFHKWMKQK-SSDKMIKEITNSNTQNVINALSVLR	
MmTgfb1	187	LT---PTDTPBWLSPFVWGVVQWLNQGDGIGQGFRESAHCSGCDKDNK-----	
MmTgfb2	218	VK---TRAEGBWLSPFVTDVQEWLHKK-DRNIGFRKSLHCPCTFPVPSNNYIIP	
MmTgfb3	191	LP---TRGTABWLSPFVTDVREWLLRR-ESNIGLESIHCPCHTFQPNGD-ILE	
CeUNC-129	156	EE---YKVIWDATKSWFDSY---HLDA--QAVFRITREHSK---M-----	
CeDAF-7	170	SERISIQLPIDTVKSWFTISPIQGIQFVKAMLDGRNVAHPQQTAD-----	
MmBmp3	181	VDVVRPYRDSVSLSKELTQLRKAKQNEE-FLGFNTSRAHELPRM-----	
MmGdf10	185	GAMALTPPPRGLWQAKDTSIKAAARRDGE-LLSAQDTEKDPGVPR-----	
MmInhba	204	LLLSEKVVDAKSTWHIFPVSSQRLLDQG-KSSDVRDA--CEQCQESGAS-----	
MmInhbb	216	WNVVEKKVDLKRSGWHTFPLTEAQLALFERG-ERRNLDAQ--CDSCQELAVV-----	
MmInhbc	163	TLTSQYVVQVNASGWYQLLGPEAQAAACSQG-HLTTEL--V--PESQVA-----	
MmInhbe	154	FRTFLAEHQTTSSGWHALTLPSSGLRSEDG-VVKQLEFR--PLDLNSTAAG-----	
DmActivin-beta	746	ASFQVDPKNLQWQKEDLTDTREWYGHTSHEKRLRID--CTGCGGGRYSL-----	
MmLefty1	165	SRLV---SIHESGWKAFD-VTEANFWQQLSRPR-----	
MmLefty2	165	SRLV---SIHESGWKAFD-VTEANFWQQLSRPR-----	
MmInha	213	-----	
DmDawdle	395	AKTIAIQSVNVQDEWPKIDLEWPKHWISGHESHLH-LIQHT--CGGCDVSDME-----	
MmGdf15		-----	
MmGdnf		-----	
MmAmh		-----	

Fig. S6 Page 2

		bowtie-- β 9--	no- β 9'--	-- β 10--	-- α 5--
DmMaverick	520	DFSPSTPPR-----	STASSDEHLNMPVNIIGHGTLN---		
DmMyoglianin	463	-----SLVVTTD-----	NTSKNELTHIEIG-----		
MmGdf11	269	-----VTSLGP-----	GAEGLHPFCEIRVLE-----		
MmMstn	240	-----VTFPGP-----	GEDGLNPFLEMKVTD-----		
CeTIG-3	135	-----	-----		
MmTgfb1	232	-----LHVEING-----	ISPKRRG-DLGTIH-----	DMNRPFELMATPLERAQ---	
MmTgfb2	269	NKSEELEARFAGID--	GTSTYASG-DQKTIKSTRKKTSGKTEHLLMLLPSYRLE---		
MmTgfb3	241	NVHEVMEIKFKGVD--	NEDDHGRG-DLGRLK---KQKDHNNPHIIMMIIPPHRLD---		
CeUNC-129	191	-----	RPYA-----	EMIRKSTPFLVYSKVNHTLDTVSV	
CeDAF-7	216	-----	-----	VDNRLQLSTRPK---	
MmBmp3	229	-----LFFPEPYI---	LVYANDAAISEPESVSVSSLQRHDETAGTG-----	PR---	
MmGdf10	233	-----PSSHMPYI---	LVYANDLAISEPNSVAVSLQRYDEPPAGDFEPGAAPN---		
MmInhba	254	-----LVLLGKKKKKEVDGDGKKKDGSDGGLLEEKEQSHRPFVMIQARQS-----			
MmInhbb	266	-----PVFVD-----	-----PGEESHRRPFVVVQARLG-----		
MmInhbc	207	-----	HSSL-ILGWFSHRPFVAAQVRVE-----		
MmInhbe	204	-----	LPRLLLDTAGQQRPFLEMKIRAN-----		
DmActivin-beta	794	-----HLFQTSKLRGN-----	SSDYLTNPNRPFVLVHTESS-----		
MmLefty1	195	-----	-----	CEFLI-QVSVQ-----	
MmLefty2	195	-----	-----	CEFLI-QVSVQ-----	
MmInha	213	-----	-----	TPFLVAHTRAR-----	
DmDawdle	445	-----EIIS-----	-----	VDKDYRPFVDMQNR-----	
MmGdf15		-----	-----	-----	
MmGdnf		-----	-----	-----	
MmAmh		-----	-----	-----	

Fig. S6. Expanded Activin+TGF- β subfamily prodomain alignment indicating structural conservation. Sequences from Fig. 6 including underlining and presented as in Fig. S4.

		-----α1----- -LatencyLasso-----	
DmSCW	30	-----IYQKRPLSEQMEDIIDLGDRERRQAEP-----	
DmGBB	56	-----LSEDDKLDVSYETIEFLGIAERETHLSHQL-----	
MmBmp8a	35	-----LGARERRDMQREITAVLGLPGRRPRAPQAAA-----	
MmBmp8b	35	-----LGVREPRDMQREIREVLGLPGRRSRAPVGAA-----	
MmBmp6	68	-----LKTHEKREMQKEITSVLGLPHRRRPLHGLQQPQPPVLPPQQQQQQQQQQ	
MmBmp5	43	-----LRNHERREIQREITISILGLPHRRRPFSPG-----	
MmBmp7	46	-----LRSQERREMQRREITISILGLPHRRRPHLQG-----	
CeDBL-1	42	-----ADQHASHATRRGLLRKLGLEHVE--V--QT-----	
MmBmp10	50	-----DFNTLLQSMKNEFTKTLNLSDF--V--QD-----	
MmGdf2	53	-----DLQMFLNKMKVDFIRSNLSGIE--S--QD-----	
CeTIG-2	38	-----G--QATDK--IGEQIRELFNIDINE--NGPAV-----	
MmGdnf	61	-----PAED-----HSLGHRRE--F--AL-----	
DmDPP	213	-----KD--KLK--PDPSTLVEIEKSLSLFNKKRPE--K--ID-----	
MmBmp2	41	-----LSR--PSEDVLSEFELRLSMFGLKQRE--T--PS-----	
MmBmp4	43	GGR--RSG--QSHELLRDFEATLQMFGIRRE--Q--PS-----	
MmBmp15	35	---LADDPTLPSILDLAKEA---PG-----KEMKQ-----	
MmGdf9	55	---GTDRLSGLLPFLKVLSDR---RG-----ETPKL-----	
MmGdf1	25	PAPASMG---PAAALLQV-----LGLPEAERSVPTHR-----	
MmGDF3	20	-GQTSEF---QSDLLQF-----LGLKAEPS-HRFQ-----	
MmNoda1	13	-----ACWALLHPR---APTAAALPL--WTRGQ-----	
MmGdf5	124	GKGASSKAGSAPSSFLKKTR---EP-GTPRE--KEPFR-----	
MmGdf6	81	G-----QH-QGQEP--G--R-----	
MmGdf7	65	G-----PR-AVRRRA--GSGFR-----	
		-----α2-----	
DmSCW	59	-----NLHNSASKELLEVYNEISEDQEPKEVLHQR-----	
DmGBB	87	-----SLRKSAPKSLDVMYHRTAEGLSDQDEDDDYERGHRS-----	
MmBmp8a	67	-----RQPASAPTEMLDLYHMTDDDDGGGP-----	
MmBmp8b	67	-----QQPASAPTEMLDLYRAMTDDSGGGTP-----	
MmBmp6	117	TAREEPPGRLKSAPTEMLDLYNASENDDEEDGASEGVGQEPGSHGGASSSQLRQSPGA	
MmBmp5	72	-----KQASSAPTEMLDLYNASEDNPEESEYLVRSVLAGEAKETRGYPASPNGY	
MmBmp7	75	-----KH-NSAPTEMLDLYNASEVESGPDG-----	
CeDBL-1	68	-----GPSIDVQHMWDIYDDNDVD-----	
MmBmp10	76	-----TGRVDPPEMLDLYNKFATDRT-----	
MmGdf2	79	-----KTRAEPQIMDLYNRYTDDKS-----	
CeTIG-2	64	-----KANNYVSTYMKRLYKOTENYEHGENHNEE-EVN-----	
MmGdnf	77	-----TSDSNMPEDYPRQEDDMDFIQATIKRLKRPDK-----Q-----	
DmDPP	244	-----RSKIIIEPPMKLYAETMGHELD-----SVNI-----	
MmBmp2	70	-----KD-VVVPPEMLDLYRHSQGPGA-----PAPD-----	
MmBmp4	75	-----KS-AVIPPDMRDLRLQSGESEE--EEQS-----Q-----	
MmBmp15	59	-----WPQGYPLRYMLDLYRSAD-PHGHPRE-----	
MmGdf9	80	-----QPDSRALYMKKLYRYAT-KEGVKPK-----	
MmGdf1	54	-----PVPP---VMWRLRRRDPOEARVGRPL-----	
MmGDF3	47	-----PVPR---VLRKIRAREA-AAASGASQ-----	
MmNoda1	36	-----PSSPSPLAYMLSLYDPLP-RA-----	
MmGdf5	158	-----PPPITPHEIMLSLYRTSD-ADKGGNS-----	
MmGdf6	92	-----GLRVVPHHEIMLSLYRTYSI-AEKLGINA-----	
MmGdf7	79	-----NGSVVPHHEIMLSLYRTAG-RAPVAAA-----	
		end straitjacket--β1--- -β1'--	
DmSCW	89	---H-----KRSLDDDLISNEDRQEIASCNSILFSSRLKP---QL---DNELD	
DmGBB	125	---RRSADLEEDEGEQQNFITDLDKRAIDESDTIMFLNKRHHNVDEL---RHEHG	
MmBmp8a	93	---QAHLGRADVMSEFVNIVERRTLGY---QEPHW	
MmBmp8b	93	---QPHLDRADLMSFVNIVERRTLGY---QEPHW	
MmBmp6	177	AHSLNRKSLAPGPGGGASPLTSAQDSAFLNADVMSEFVNIVEYKEFSP---HQRHH	
MmBmp5	124	AHRLHLP--PRTPLTTQSPPLASLHDTNFLNADVMSEFVNIVERCKDFSH---QRRHY	
MmBmp7	100	-QGFSYP--YKAVFSTQGPPLASLQDSHFLTDADVMSEFVNIVEHKEFFH---PRYHH	
CeDBL-1	89	---WVRHMYPKIIEONEG---F-----	
MmBmp10	98	---SMPSANTVRSFKNEDLFSQPVTF--F-NGLRK	
MmGdf2	101	---STPASNTVRSFSDAISTAAT---EDFFFQK	
CeTIG-2	96	---AWLSADRVSHMAQEVSH--H--R-LDDGS	
MmGdnf	112	AAALPRR-----ER-----	
DmDPP	271	---PKP-----GLLTKSANTVRSFTHKDSKID-DR---F-PHHHR	
MmBmp2	96	---HRL-----ERASRANTVRSFHHEEAVEELPE---M-SGKTA	
MmBmp4	102	GTGLEYP-----ERPASRANTVRSFHHEEHLNIPG---T-SESSA	
MmBmp15	85	---NRTIGAKVRLVKPSANTVRRPR-GSW---HV	
MmGdf9	106	---SRSHLYNTVRLFSPCAQQQAPS-NQVTGPLPM	
MmGdf1	78	--RPCHV-----EELGVAENVRHIPDSGLSSRPAQPARTSGLCPE	
MmGDF3	70	--DLCYV-----KELGVRNLQLLPDQGFFLNTQKPF-QDGSCLQ	
MmNoda1	57	---DIIRSLQAQDVD-----VTGQN	
MmGdf5	185	---SV-----KLEAGLANTITSFIDKQDPRG---PAVRK	
MmGdf6	119	---SF-----FQSSKSANTITSFVDRGLDLSH---TPLRR	
MmGdf7	105	---SGHGRVDTITGETDQATQETA---AAEPG	

Fig. S7 Page 1

		β 10-	no- α 5-	
DmSCW	251	<u>LV</u> GYFNGPELLVVKIQKLRFKR----		
DmGBB	302	<u>LV</u> GFERGPelikatahssHRSKR----		
MmBmp8a	245	<u>LV</u> TFERASQSPVRAPR----		
MmBmp8b	245	<u>LV</u> GFERANQSPVRAPRTAR----		
MmBmp6	356	<u>LV</u> AFKVSSEVHVRTTR----		
MmBmp5	301	<u>LV</u> AFKASEVLLRSVR----		
MmBmp7	276	<u>LV</u> AFKATEVHLRSIR----		
CeDBL-1	219	<u>LV</u> IVESDLSEPSSVRRK-R----		
MmBmp10	252	<u>LV</u> VESDDQSDNK--EQ-KEELNELITHEQDLDLSDA---FF-SGPD--EEALLQMR-		
MmGdf2	252	<u>LV</u> VESNDRSNGTK-ET-RLELKEMIGHEQETMLVKTAKNAYQVAGESQEEGLDGY---		
CeTIG-2	214	<u>LV</u> IALKDEDA GPPKKRSRRSASTTPI SAPP MRQVKRSE SAYFEK-PNENERCQ----		
MmGdnf				
DmDPP	440	<u>LV</u> FTYTDDGRHKA--RSIRDVSGGEGGGKGG-----RNKR-----		
MmBmp2	261	<u>LV</u> TFGHGDKGHP--LHKREKR-----		
MmBmp4	272	<u>LV</u> TFGHGDRGHT--LTRRRAKR-----		
MmBmp15	231	<u>LV</u> LYNDTDDRQVGK-----L-----LARGQEELTDR		
MmGdf9	253	<u>LV</u> LYLNDTSTQAYHS-----W-----QSLQSTW---RPLQHPGQAGVAAR		
MmGdf1	220	<u>LV</u> VTLDPRLCF-----		
MmGDF3	234	<u>LV</u> VTLNPKHCH-----		
MmNodal	198	<u>LV</u> IYSNRPQEQRL-----GGATLLWEAESSW---RAQE--GQL----		
MmGdf5	335	<u>LV</u> IVFGRTKKRDLFFNE-----IKARSGQDDKT VY EY-----		
MmGdf6	280	<u>LV</u> VETRSQRKN-LFTE-----MHEQLGSAEAGAEGSW---PAPS--GSP----		
MmGdf7	263	<u>LV</u> VISSRTQRKESLFRE-----IRAQARALRAA---AE---PPPD--PGP----		

Fig. S7. Expanded BMP subfamily prodomain alignment indicating structural conservation.

Sequences from Fig. 8 including underlining presented as in Fig. S4.

		Association	-----al-----	-LatencyLasso-
MmBmp15	35			LADDP
MmGdf9	55			GTRSG
MmNodal	1			SAHSRLILQ
MmGdf5	92			PPRSGGPETKPGPSSQTR--QAAAR--TTPFKGQLGGKASSKAG
MmGdf7	50			ARAPGP--SADQA--AAVPG
MmGdf6	53			QRTPOESAEGR--TPPEHGLRQKDRRRPFG
DmActivin-beta	314	SHKGCTLCHESGKPNYTKDNPHDYASDEVRLSE	KROLTKGGLSH-KEN--VSHPL	
CeTIG-3	1			MS
DmSCW	17	TTYVTT--N--NHIE--M--PIYQKRPLSEQME	TDIDPDGD-RFR--RQAEF	
DmGBB	40	GIYIDN--GKDQTIMH--R--VLSEDDKLDYSYEL	LEFGCHAE-RPT--HLSSH	
MmBmp8a	19	GGGHGP--REPHTCFQ--R--RLGARERRDQRE	LAVGCPG-RFR--PRAQP	
MmBmp8b	19	GGGHLs--HPPHVFPQ--R--RLGVREPRDQRE	REVGLPG-RFR--SRAPV	
MmBmp6	52	EQPPFQ--SSSSGFLY--R--RLKTHEKREKQRE	LSVGLPH-RFR--PLHGL	
MmBmp7	30	DFSLDN--EVHSSFIH--R--RLRSQERREKQRE	LSIGLPH-RFR--PHLQG	
MmBmp5	27	GGIGDN--HVHSSFIY--R--RLRNHERREKQRE	LSIGLPH-RFR--PFSFG	
MmBmp3	4			ARGLLCLWLGYFCLNLAQQR-PNHLPLGRETETS--DRATG
MmGdf10	17			MVPLLLLRLGAGCGHRRGSPSSSLPSAAGLQGRDS--QOSPG
DmMyoglianin	175	SRVQSVSLYRNTLINIESMLQRLREKAKVDSIES	KMHLMRNKKL-PN--ITKP	
MmMstn	35	KEGLCNACAWRQ-----NTRYRIEA	KIQLSKRLTA-PN--ISKD	
MmGdf11	56	EPDGCPCVCVWRQ-----HSRELKLES	KSQLSKRLKEA-PN--ISRE	
MmTgfb1	29	GLSTCRTID-----MELVKRRRIEA	RGOILSKRHASP-PS--QG-E	
MmTgfb2	20	SLSTCSTLD-----MDQFMRRKRIA	RGOILSKRLTSP-PE--DY-	
MmTgfb3	23	SLSTCTILD-----FGHIKKRVEA	RGOILSKRLTSP-PE--PS-	
MmInha	21	CQG-----PELV-R--ELVLAK	KALFLDALGPPAM-DG--EGGD-	
MmGdf15	33			SQG--DALA-MP--EQRPSPGPESQNA--DE--LR-
MmGdf1	22			TLA--PAP--ASMGPAAALQVTEPEA-R--SV-
MmGDF3	20			GQ-----TSEFQDSDLQFGLCEKA-PS--P-
CeUNC-129	2	RRLPVILLLSV-----F--SIANCAKVDVLD	INETHRLDHFSSDEN--VTSE-	
CeDAF-7	30	CIEKMKQHR-----TEYKNE	LDQCNKE-APK--GLKP-	
DmMaverick	249			DETRKHLKGGGKK-LPD--MRKV-
MmInhba	44	PDCPSCALATL-----PKDGPNSQFEMVEA	KKHHLNMHKKK-RPD--VTPQ-	
MmInhbb	68	DTCTSCGGGG-----PEELGRVDGDFLEA	KRRHLSRLARG-RFN--ITHA-	
MmInhbc	39	GPCPCWGAIF-----QRELLDLAKKS	SLDKHLQ-RFI--LSRP-	
MmInhbe	39	SACPCSCGGPTL-----ERALVLELAKQ	TEGRLTS-RFR--ITRP-	
CeDBL-1	24	SAILHLFLIS-----FTPMSAADQHASHATRRG	LKKGLDEH-VFV--Q-	
MmGdf2	56			M-----FLENKVDPLRSNLSG-IPS--Q-
MmBmp10	53			T-----LLQSKNEFKTNLSD-IPV--Q-
CeTIG-2	42			DKAGEORELEN--DI--NEN--GP-
DmDawdle	226	SAGGCPKCESNRQ-VEHITE--EQLTH-LRIEF	VQOHLKRLKE-SPK-----	
MmLefty1	27			IL--GSLOOQLDQ-PEV--LDKA-
MmLefty2	27			VL--SSLOOQLDQ-ATV--LDSA-
MmAmh	191	RYLV--L-----TVDFP--AGAWSGSG	LIL-LQ-----P-	
MmGdnf	65			HSLSHRR-VF-----A-
DmDPP	223			LVEIE-KSLSLFNKKR-PK-----I-
MmBmp2	49			LSEFE-LRELMSFGLKQ-RPT-----P-
MmBmp4	54			LRDFE-ATLLOMFGLR-REQ-----P-
MmBmp15	53			GKEMKQWPQGYPLRYLRLNHRADPHGHPRE-
MmGdf9	74			GETPKLPDSRALYYKKLTKYATKEGVKPK-
MmNodal	30			LWTRGQSPSPSPLAYLSRYRDLFRA--
MmGdf5	152			PKEPFRPPPIITPHEYLSRYRTLSDADRKGNS
MmGdf7	73			AGSGFRNGSVVPHFESRYSLAGRAPVAAA-
MmGdf6	89			PG--RGLRVVPEHYLSKTYISIAEKLGINA
DmActivin-beta	515	MPIELKSHHNSSPKELKSGAVRKVNGINGTQM	NENALKSTYFDINHSIDNKTHTGKG	
CeTIG-3	3			TSRKHDVGGVL-----
DmSCW	65			SKFLPEVNEISEDQEPKEV-----
DmGBB	93			PKFLDYVHRITAEGLSDQ-----
MmBmp8a	73			PLFLDYVHAMTDDDDGGPP-----
MmBmp8b	73			PLFLDYVAMTDDSGGGT-----
MmBmp6	131			PLFLDYVNLNSNDEEDGA-----
MmBmp7	80			PMFLDYVNAMAVEESGPDG-----
MmBmp5	78			PMFLDYVNAMASEDNPEES-----
MmBmp3	56			VSEHLDVND-RYSGSSRVQAT-----
MmGdf10	71			VAIHRLRRE-KYNNRG-----
DmMyoglianin	235			NLDNLYRDYNASSKTTVM-----
MmMstn	87			PLREDDCDVQDDDS-S-DGS-----
MmGdf11	107			PLQQLDHDQFGDALQP-EDF-----
MmTgfb1	76			AAALYNSTRDRVAG-ESA-----
MmTgfb2	66			EVSTYNSTRDLQEE-KAS-----
MmTgfb3	68			CAALYNSTRELLEE-MHG-----
MmInha	63			RHAUGGFMRHTSEPEEE-DVS-----
MmGdf15	65			LSRLHANQSRDS-NSE-----
MmGdf1	58			VWRFLRRRDQEA-RVG-----
MmGDF3	51			LPKIRAREA-AA-ASG-----
CeUNC-129	53			LTEHKNNENFIDESDN-EDG-----
CeDAF-7	66			MKSVYELVRLLEKDEQ-DMG-----
DmMaverick	279			YSSKYEYLSRLSNQEK-GNS-----
MmInhba	79			PKAALNATRKLHVGVG-ENG-----
MmInhbb	103			PKAALNATRKLHAGVVR-EDG-----
MmInhbc	68			SRGAKTALQRLRGPRRE-T-----
MmInhbe	68			PQAAKTRALRLQPKSMV-----
CeDBL-1	73			VPQRWDVDDNDVDW-----
MmGdf2	84			PQYVLDNNRYTTDK-----
MmBmp10	81			PPEYVLDNNKFATDR-----
CeTIG-2	65			ANN-----
DmDawdle	276			LKPLFDGMTLSHPDDST-KNK-----
MmLefty1	53			IPSHVORTQVALLQSHS-----
MmLefty2	53			IPTHRSQVALLQGSHA-----
MmAmh	222			LSIDQLQ--AFLFGSDSR-----
MmGdnf	82			MPEDYPDQDDVMDFIQA-----
DmDPP	249			IFEPKKKFAEIMGHELD-----
MmBmp2	74			VPPYVLDNRRHSQPGA-----
MmBmp4	79			IPDYVRDRLQSGEEEE-----

Fig. S8 Page 1

			end straitjacket TGFβ-β1---
MmBmp15	85	-----	-----
MmGdf9	106	-----	-----
MmNodal	57	-----	-----
MmGdf5	185	SVK-----	-----
MmGdf7	105	-----	-----
MmGdf6	119	SFF-----	-----
DmActivin-beta	575	EMSHN-D---YEFNDYSVQTHDKNR-YHEGRSSIGYQPAIHNI-EYENQ-KG----	-----
CeTIG-3	15	-----	-----
DmSCW	85	-----LHQR-----	-----H-----KRSI-----
DmGBB	113	-----DEDDDYERGHRS-----	-----RRSADLEEDEGE-----
MmBmp8a	93	-----	-----
MmBmp8b	93	-----	-----
MmBmp6	151	-----SEGVGQEPGSHGGASSSQLRQPSGAAHSLNRKSLLAGPGGG-----	-----
MmBmp7	100	-----	-----QGFSYP--YKAVFST-----
MmBmp5	98	-----EYLVRVSLAGEAKETRKGPASPNGYAHRLHLP--PATPLTT-----	-----
MmBmp3	77	R-----TPGSQL-----PGP--QPLRGG--NTVRSFRAAAAGTP--	-----
MmGdf10	87	-----	-----APPGGG--NTVRSFRARLEMI-----
DmMyoglianin	254	NRMESIDESHLSINDTYGDHIMTDFDESSSSQMGDDA--NTVNEFLIDLKNG--	-----
MmMstn	107	-----	-----LEDDDY--HATTETIITMPT-----
MmGdf11	128	-----	-----LEEDEV--HATTETIVISMAQ-----
MmTgfb1	94	D-----	-----PEPEPEADY--YAKEVTRVLMVDRN-----
MmTgfb2	84	R-----	-----RA-----AACERERSDEEY--YAKEVYKIDMPSHL-----
MmTgfb3	86	E-----	-----RE-----EGCTQETSESEY--YAKEIHKFDMIQGL-----
MmInha	83	Q-----	-----AILFPATGATCEDQP-AA-----RG--LAQE-----
MmGdf15	81	-----	-----PSPDPAVRILSPEVRLGS-----HGQLLRV-----
MmGdf1	75	-----	-----RPLR--PCHVEELGV--AGNIVRHI-----
MmGDF3	67	-----	-----ASQD--LCYVKELGV--RGNLLQLL-----
CeUNC-129	74	NL-----	-----
CeDAF-7	87	VEN-----	-----SFY-----TAKDP-----
DmMaverick	300	YFN-----	-----NFMGASFTRDLH---FLSITTNGF---NDISNKLRRHRSLSKKI
MmInhba	100	YV-----	-----EIEDDIGRA-----
MmInhbb	124	RV-----	-----EIPHLDGHAS-----
MmInhbc	87	-----	-----LLEH-----
MmInhbe	86	-----	-----P-----
CeDBL-1	90	-V-----	-----RHYY-----
MmGdf2	100	-----	-----
MmBmp10	97	-----	-----
CeTIG-2	68	YV-----	-----STYMKRLYQLENY-----
DmDawdle	297	EL-----	-----DDYARTSKKFILL-----
MmLefty1	71	-----	-----SRSG-----
MmLefty2	71	-----	-----DRSG-----
MmAmh	238	-----	-----CFTRMTPTLVVL-----
MmGdnf	100	-----	-----TIKRLKRSPPDKQ-----
DmDPP	267	-----	-----SVNI-----
MmBmp2	92	-----	-----PAPD-----
MmBmp4	97	-----	-----EEQSQ-----
			end straitjacket Activin+BMP-β1---
MmBmp15	85	-----	-----NRTIGAKMLVLKPSANTVRPPRG---S---W---H
MmGdf9	106	-----	-----SRSHLYNTYRLSPCAQQEQAPSN---Q---VTGPLE
MmNodal	57	-----	-----DIHRSLOAQDQD---Q---VTGQ
MmGdf5	188	-----	-----LEAGLANHTSLEIDKGQDDR---G-PAVR
MmGdf7	105	-----	-----SGHGRVDTTGTGTDQATQDE---TAAAEF
MmGdf6	122	-----	-----QSSKSANTTSVDRGLDDL---SHTPLR
DmActivin-beta	621	---HHESFADDHENI--DHEDFFGNTQELITAEEGTQ-----	---YRQYR-----
CeTIG-3	15	-----	-----D---KI--TLEPNRKSWTCLTPESLV
DmSCW	94	---DDILIS---N--EDRQEIASCNSLTSRLKPE---QLDNELD	-----
DmGBB	137	---QKNFIT---D--LDKRAIDESDITSLNRKHNV--DEL-RHEHG	-----
MmBmp8a	93	-----	---QHLGRADLMSVNMVERDR--TLGYQEPHW
MmBmp8b	93	-----	---QHLGRADLMSVNIIVERDR--TLGYQEPHW
MmBmp6	193	---GASPLTS---A--QDSAFLNADMMMSVNLVEYDK--EFSPHQRHH	-----
MmBmp7	113	---QGFLAS---L--QDSHFLTDADMMMSVNLVEHDK--EFFHPRYHH	-----
MmBmp5	138	---QSPFLAS---L--HDTNFLNADMMMSVNLVERDK--DFSHQRHY	-----
MmBmp3	107	Q-----	-----T-----
MmGdf10	106	D-----	-----Q-----
DmMyoglianin	307	AKKSDIPINTNDEEYESILS---HISSEYITPEEIOPH-V--RHNK	-----
MmMstn	125	---ESDFLMQADGKP-----	---K-----
MmGdf11	146	---ETDPAVQTDGSP-----	---L-----
MmTgfb1	118	NAIYEK---T-----	---K-----
MmTgfb2	113	PSETVCVVVTF-----	---S-----
MmTgfb3	115	AEHNELAVC--P-----	---K-----
MmInha	107	AEGLF-----	-----
MmGdf15	107	NRASLS-----	-----
MmGdf1	96	PDSGLS-----	-----
MmGDF3	88	PDQGF-----	-----
CeUNC-129	76	-----	---RALEPAV---GK--FE--G-----
CeDAF-7	98	-----SYGE-----	---NPSQLAKQDVTNDLER-----
DmMaverick	341	NRLNQNPKK--HONYGDLLRGEQDTMNI--LHPLTNAQD--A-----	---VTGPLE
MmInhba	112	-----	---E-MNELMEQTSEHITSAES---G---T--A-----
MmInhbb	136	-----	---P-GADGQERVSEHITSAET---DG-LAS--S-----
MmInhbc	91	-----	---DQRQEYEHITSAET---DL-SSI--N-----
MmInhbe	87	-----	---GN--REKHSATI---ID-KSTSY-----
CeDBL-1	95	-----	---PKEIIDNEGFLLS-----
MmGdf2	100	-----	---SSTPAS--NIHRSFVEDAISTAATE-DFPFQ-----
MmBmp10	97	-----	---TSMPSA--NIHRSFKNEDLFSQPVTF--NGLR-----
CeTIG-2	84	EHGE---N-----	---H--NEEVNANW---LS-----
DmDawdle	313	NREE---V---ECNR-ARDGKS--NPSMCTFKIDDADAEG--FD-----	-----
MmLefty1	76	-----	---KRSQNLREVA-GRFLVSETS-----
MmLefty2	76	-----	---KRSQNFREVA-GRFLMSETS-----
MmAmh	250	PP-A---E---PSPQ-PAHGQL--DT-MPQPQPLSLE-PE--ALPHS	-----
MmGdnf	112	AA-A---L---PRRE-R-----	-----
DmDPP	271	-----	---PKPG-LLTKSA--NTVRSFTHKDSKID-DR--FPHHH-----
MmBmp2	96	-----	---HRLA-RAASRA--NTVRSFHHEAVEELPE--MSGKT-----
MmBmp4	102	GT-G---L---EYPE-RPASRA--NTVRSFHHEEHLENIPG--TSESS	-----

Fig. S8 Page 2

	begin	arm-β2----	----β3----	TGF-β---α3--	TGF-β-β3----
MmBmp15	112	VQTLDE-PLASNQVAYE	IRATVYRQHLHL-VNYHL	---SCHVETWVPKCRTH	---
MmGdf9	137	MVDLLE-NLDRVTAMEH	ILKSVLYLTNNSSSTV	---TCMCDLVV	---
MmNodal	73	NWTFLE-DFSFLSQEED	VWAEELRLQPGMDIPT	EGPLTIDIFHQAKGD	---
MmGdf5	213	KQRYVE-DISAL-EKDG	LCAGELRLKRPPLD	VAKPAVPSGRVAQLKSSC	---
MmGdf7	131	GQFLLE-DVSSLSEADE	VNAELRLRRRSPEPD	---RDSATLLPRLLSTC	---
MmGdf6	148	RQKYLE-DVSTLSDKEE	VGAEELRLRQAPPTFW	---GLPARPLHLQ-LFPC	---
DmActivin-beta	659	ILEFSAQNRVPSQKLS	RSQAHLRLIDKPHSLWIEKAKSL	---PE----	KHLL--N
CeTIG-3	36	KDCFOHSINSINHE	---ILSASLH	---	DF----
DmSCW	131	-MHITENTNDVPVD	-LSVCAMLRLTKQPSLVD	---	---
DmGBB	176	-RRLMEVSNVPND	-NYVMAEELRLQANANEGKWL	---	---
MmBmp8a	123	-KEFHEDLTQIPAG	-EATAEELRLKEPSTHPL	---	---
MmBmp8b	123	-KEFHEDLTQIPAG	-EATAEELRLKEPSTHPL	---	---
MmBmp6	233	-KEFHEDLTQIPAG	-EATAEELRLKDCVVGVSF	---	---
MmBmp7	153	-REFREDLSKIPEG	-EATAEELRLKDYIRERF	---	---
MmBmp5	178	-KEFHEDLTQIPHG	-EATAEELRLKDKGNHRF	---	---
MmBmp3	109	KGLHTE-NLTSLTKSEN	ILSATVYFVIGELVNI	---SLSCP-E--PQCSHHTQR	---
MmGdf10	108	KPVYFE-NLTSMQDSEM	ILTAFFHSEPPRPW	---RAREVFCKPRAKNASCRLLTPG	---
DmMyoglianin	348	VDVFRF-QIDSSYSDL	-YATDHLRLGWDWI	---SAHQPGLLLEIKKQP	---
MmMstn	138	CCFFHE-SSKI	---QYNKVVKQALWI	---YLRFVKT	---PTTVFVQLLR
MmGdf11	159	CCHFHE-SPKV	---MFTKVLKALWV	---YLRFVPR	---PATVYLQILR
MmTgfb1	126	-----DISH	---SIY	---MFFNTSDIREAVPEPPLLSRAELR	---
MmTgfb2	126	GSLGSE-CSRQSQVLCGYL	PIPPITYRPFYRI	---VRFDVST	---MEKNASNLVKAEFR
MmTgfb3	126	-----GITSKV	---FRFNVSS	---VEKNGTNLFAEFR	---
MmInha	113	---TYVE-RPSQHIRSHQ	---IT	---SAQLWF	---HTG-LGRKS
MmGdf15	113	---QGL	---PEAYRVHRAALLLP	---TAREWD	---ITRPLKRALS
MmGdf1	102	---SRP	---AQPARTSG	---LCPWT	---VVFDSLNVF
MmGDF3	94	---LNT	---QKPF-QDG	---SCLQKV	---LYFNLSAIKE
CeUNC-129	89	QEVLMG-DVEGFDSHES	MRRELRLRRRDSFARRRSRQIRAKSV	---	---
CeDAF-7	119	-----SDILCAT	---TSIEIP	---A	---KDSG
DmMaverick	379	-----NFHMDKDEANVR	MLLYSSSLATNFRFGPSRKN	---	KISQISGNDN
MmInhba	133	RKTLHEEISKEGSDLSV	ERAEMLLKV	---KAN	---R
MmInhbb	160	RVRLYEFVSNEGNQNL	FVCAVWLKLL	---PYVLEKG	---S
MmInhbc	112	QTRLESHFSGRMASGME	ROTFRMFVQFP	---HNATQT	---
MmInhbe	107	RSMLEHLSPLWSHH	---YHAR	---LHLHP	---PSFPGT
CeDBL-1	109	-YNL	-SLAARNAHNEET	TRATKHLRRN	---NKARRS
MmGdf2	129	KHLIE	-NISIPRHEQ	TRAEELRLV	---S
MmBmp10	125	KYPLLE	-NVSIPHHEE	VMAEELRLT	---L
CeTIG-2	100	ADRIVSHMAQEVSHRL	DDGSYSERFAK	---E	---HVPAGEQSIV
DmDawdle	347	-----YSTAV	MLLKN	---KQNRDTASVN	---S
MmLefty1	96	THLLVGEQRLPPNSE	VCAVRLLEQ	---E	---FVPRTA
MmLefty2	96	THLLVGEQRLPPNSE	VCAVRLLEQ	---E	---FVPRTA
MmAmh	284	ADPFLE-TLTRLVR	-ARGPLTQASN	---T	---QLALDPGALAS
MmGdnf	302	RFRLHE	-DVKSI PADEK	KAEELRLTR	---D
DmDPP	128	ARRFFE	-NLSSVPSDEF	TSAEELRLTR	---E
MmBmp2	138	AFRFLS	-NLSSIPENEV	SSAEELRLTR	---E
MmBmp4	138	AFRFLS	-NLSSIPENEV	SSAEELRLTR	---E
Activin-BMP-β3 --β4----					
MmBmp15	162	-----	-----	LPS	-----SKSGSSKP
MmGdf9	181	-----	-----	KEA	-----MSSGRAPP
MmNodal	122	-----	-----	-----	PERDPADCLE
MmGdf5	263	-----	-----	-----	PSGRQPAALLD
MmGdf7	179	-----	-----	-----	PDEAGTAHLH
MmGdf6	195	-----	-----	-----	LS
DmActivin-beta	707	--TK	-----RKWGA-NKPHHRIK	WVQLSTSN	---I
CeTIG-3	60	-----	KDTNISV	VEVDE	---L
DmSCW	162	-----	RRANFTV	SRKLD	---N
DmGBB	210	-----	ANREFTV	VAIGT	---G
MmBmp8a	155	-----	NTTLHSS	EEVVQ	---E
MmBmp8b	155	-----	NTTLHSS	EEVVQ	---E
MmBmp6	265	-----	KNQTFLES	QVQLQ	---E
MmBmp7	185	-----	DNETFQNT	QVQLQ	---E
MmBmp5	210	-----	ENETIKS	QIIK	---E
MmBmp3	157	-----	QHIQID	SAILKS	---
MmGdf10	162	-----	LPARLH	IFRSL	---
DmMyoglianin	393	--RK	-----DIVVTI	HRAIRANTT	---
MmMstn	177	LIKPMKD	-----	-----	GTRYT
MmGdf11	198	LKPLTGEGTAGGG	-----	-----	GGGRRHI
MmTgfb1	157	LQRL	-----KS	-----SVEQHME	QKYS
MmTgfb2	179	VFRL	-----QNP	K-----ARVAEQRE	QILKSK
MmTgfb3	153	VLRV	-----PNPS	-----SKRTEQR	ELQILRPD
MmInha	149	SAPLLDLLVLSSGG	PMAPVPSLGQGP	PAVLH	---
MmGdf15	159	LTPPPDLAMLP	SGGTQLELRVAAGRGR	---	---
MmGdf1	141	LEAESE	---DTGGWEL	---SVAM	ADAE
MmGDF3	132	LGPRSY	---YNLRPELVVALSV	QDRG	GRSH
CeUNC-129	134	-----	-----	CNEV	CR
CeDAF-7	139	-----	MLQDVQ	QW	WEKNEDG
DmMaverick	424	IERHCNFGDVNLN	---QSNKNSSQ	LLTK	QALLSAN
MmInhba	169	-----	TKVTIR	QQQKH	PQGLDTGDEAEEMGLKGRSEL
MmInhbb	199	-----	RKVR	KV	FQE
MmInhbc	148	-----	MNR	RLVLRP	---
MmInhbe	139	-----	LYDR	RCGT	---
CeDBL-1	142	-----	GN	SS	FF
MmGdf2	161	-----	STHLEGS	VV	VDVLEDS
MmBmp10	157	-----	MYDGVDRK	TT	SEVLESA
CeTIG-2	140	-----	AQLR	HH	QGVSPV
DmDawdle	371	-----	SAQQT	V	SEV
MmLefty1	136	L	-----SPHSARAR	TT	---
MmLefty2	136	L	-----SPHSARAR	TT	---
MmAmh	323	-----	GL	AND	SDPALG
MmGdnf	343	-----	RTRYQ	L	W
DmDPP	169	-----	DN	BEI	IKPA
MmBmp4	179	-----	DN	BEV	MKPP

Fig. S8 Page 3

		-----β3-----	-----β6-----	-----α4-----	-----β7-----	TGFβ-β8
MmBmp15	173	-----SPMSKAMT-EI	-----THCI	-----CQKLNNRKG	-----SVPRFRM	-----COQ
MmGdf9	192	-----RAP	-----YSFTLKKHRAI	-----EID	-----TSLI	-----QPLVTSSERSIHSNFT
MmNodal	132	RIN	-----ETFTVI	-----PSQVTFASG	-----ST-VL	-----TK-----PLSKWLKDPRAEKQVSSRAEK
MmGdf5	274	V	-----RSV	-----PGL	-----DGSGAE	-----VDDKLF
MmGdf7	190	S	-----RAA	-----EPLG	-----GARME	-----AIOVTDAT
MmGdf6	202	A	-----RTL	-----DPQGP	-----TQAGAE	-----VDDWQGE
DmActivin-beta	744	-----FRASF	-----QVDPKNLGM	-----KID	-----TDTI	-----RENYGH
CeTIG-3	80	-----YDRF	-----EI	-----RETLDKY	-----HDDSHLF	-----HKMMKQ
DmSCW	184	-----IGSV	-----NTTSSQRCAL	-----EEN	-----TDTI	-----RYALHN
DmGBB	233	-----PSSV	-----NTTGDYVGM	-----ELN	-----TEGI	-----HEDLVK
MmBmp8a	178	-----FDLQ	-----TLRSGDEGAL	-----VLD	-----TAAS	-----DRALLN
MmBmp8b	178	-----FDLQ	-----TLRSGDEGAL	-----VLD	-----TAAS	-----DRALLN
MmBmp6	289	-----LSDR	-----VWASEEGAL	-----EED	-----TATS	-----NIAVVT
MmBmp7	209	-----LSDR	-----TIWASEEGAL	-----VDD	-----TATS	-----NIAVVT
MmBmp5	234	-----LSDR	-----KTQALDVGM	-----VDD	-----TATS	-----NIAVVT
MmBmp3	178	-----HDSVDV	-----VRPYRDSVSM	-----SKD	-----TQLI	-----RKAKQN
MmGdf10	182	-----LGRGAM	-----ALT	-----PPRGLM	-----AM	-----SII
DmMyoglianin	420	-----EFR	-----HSIPSGLGQV	-----AVQ	-----KSLI	-----GNLGSN
MmMtn	189	-----GRSLK	-----LDMSPGTGIM	-----SIO	-----KTVI	-----QNALKQ
MmGdf11	218	-----GRSLK	-----IELHSRSGHM	-----SIO	-----KTVI	-----HSMFRQ
MmTgfb1	183	-----G	-----NR	-----LLTPTDTP	-----PEAL	-----SEDTGVN
MmTgfb2	214	-----D	-----SK	-----VVKTRAE	-----GEM	-----SEDTDAI
MmTgfb3	187	-----G	-----GK	-----NLPTRGTAEM	-----SEDTDTI	-----RDELLR
MmInha	189	-----L	-----L	-----L	-----L	-----L
MmGdf15		-----	-----	-----	-----	-----
MmGdf1	170	-----	-----RVPAPPGVLL	-----RAE	-----LGTA	-----VAANAS
MmGDF3	169	-----F	-----LR	-----SVPGPQQL	-----Q	-----MLOGA
CeUNC-129	141	-----QOTLKIRVG	-----GDNLEEDKVI	-----E	-----PKS	-----FDS
CeDAF-7	173	-----ISILPI	-----	-----DTVKSMTIS	-----F	-----OGIEVKAMLDG
DmMaverick	461	-----K	-----TSRKIEFGNVGFQ	-----ETRTQMI	-----EED	-----TKAT
MmInhba	205	-----L	-----SEKVVVD	-----A	-----RKSTH	-----I
MmInhbb	217	-----N	-----VEKKVD	-----L	-----KRS	-----G
MmInhbc	164	-----L	-----TSQYVQ	-----V	-----NAS	-----G
MmInhbe	155	-----R	-----TFLAEHQ	-----T	-----TSS	-----G
CeDBL-1	159	-----C	-----ESRSV	-----	-----DNL	-----TEMI
MmGdf2	190	-----F	-----VS	-----QDI	-----RDEGAE	-----TLAV
MmBmp10	185	-----V	-----VS	-----TEIYGTN	-----SEME	-----T
CeTIG-2	163	-----P	-----GDTV	-----V	-----LVS	-----SDPTVVT
DmDawdle	392	-----L	-----SAAKTIA	-----	-----IQSVNVQ	-----DEM
MmLefty1	162	-----L	-----DSRLVS	-----I	-----H	-----ESG
MmLefty2	162	-----L	-----DSRLVS	-----I	-----H	-----ESG
MmAmh	346	-----L	-----SPAA	-----	-----ATEREPM	-----FLHGPAS
MmGdnf		-----	-----	-----	-----	-----
DmDPP	368	-----L	-----D	-----TKTVR	-----L	-----N
MmBmp2	189	-----L	-----D	-----TRLVN	-----Q	-----N
MmBmp4	200	-----L	-----D	-----TRLVH	-----H	-----N
only TGFβ-β8----bowtie----only TGFβ-β9----- BMP only -β9'-						
MmBmp15	210	-----Q	-----KG	-----	-----NET	-----REFRWH
MmGdf9	235	-----D	-----QV	-----	-----PED	-----GVFSMP
MmNodal	184	-----Q	-----	-----	-----PYTPPVV	-----
MmGdf5	312	-----	-----	-----	-----RGRAVD	-----LRGLGFERT
MmGdf7	232	-----	-----	-----	-----ASESSPLAL	-----RRLGFGWMP
MmGdf6	242	-----E	-----LDAGD	-----TGAR	-----	-----ARGPQQPPFLDL
DmActivin-beta	787	-----	-----	-----	-----RYSL	-----HLFQTS
CeTIG-3	122	-----N	-----	-----	-----TONV	-----IN
DmSCW	225	-----I	-----G	-----	-----DSQL	-----STFAAG
DmGBB	276	-----N	-----NRP	-----	-----DREV	-----KLDDIG
MmBmp8a	220	-----A	-----D	-----	-----GHS	-----M
MmBmp8b	220	-----E	-----D	-----	-----GHS	-----I
MmBmp6	331	-----R	-----D	-----	-----GLHV	-----N
MmBmp7	251	-----L	-----D	-----	-----GQS	-----I
MmBmp5	276	-----G	-----D	-----	-----GRS	-----I
MmBmp3	224	-----L	-----PKRMLFFPEPY	-----ILVYANDAA	-----ISEPESV	-----VSS
MmGdf10	228	-----P	-----GVP	-----PRPSSHMPY	-----ILVYANDL	-----AISEPNSVAVSLQRY
DmMyoglianin	459	-----T	-----W	-----	-----MKSLV	-----VTTDNT
MmMtn	235	-----G	-----H	-----	-----GHD	-----LAVTFPGP
MmGdf11	264	-----G	-----T	-----	-----GTD	-----LAVTSLGP
MmTgfb1	226	-----D	-----NK	-----	-----LHVEING	-----I
MmTgfb2	257	-----V	-----PSNHYIIPNKS	-----	-----EEL	-----EAFAGIDGT
MmTgfb3	230	-----Q	-----PNGD	-----ILENVH	-----	-----EVMEIKFKGVDNE
MmInha	191	-----	-----	-----	-----	-----DDHGRG
MmGdf15		-----	-----	-----	-----	-----DLGRK
MmGdf1	203	-----	-----	-----	-----SLHFG	-----ATAACG
MmGDF3	210	-----D	-----RY	-----	-----SRVTQ	-----PENPCD
CeUNC-129	186	-----E	-----H	-----	-----SKMRPY	-----
CeDAF-7	214	-----A	-----D	-----	-----ADV	-----DNRLQ
DmMaverick	510	-----I	-----GARIL	-----	-----IGARIL	-----SOFSPST
MmInhba	247	-----S	-----GASLV	-----	-----LGKK	-----K
MmInhbb	259	-----L	-----AVVPV	-----	-----FVD	-----
MmInhbc	204	-----Q	-----VA	-----	-----	-----HSSL
MmInhbe	197	-----I	-----NS	-----	-----TAAG	-----
CeDBL-1	202	-----T	-----QSSSL	-----	-----SSLPYA	-----
MmGdf2	230	-----Q	-----	-----	-----SHRES	-----
MmBmp10	227	-----R	-----QN	-----	-----QAEDTG	-----
CeTIG-2	203	-----S	-----	-----	-----	-----
DmDawdle	438	-----S	-----	-----	-----DMEEI	-----
MmLefty1	204	-----Q	-----RE	-----	-----HLGPGT	-----
MmLefty2	204	-----Q	-----RE	-----	-----HLGPGT	-----
MmAmh	385	-----S	-----EL	-----	-----RDL	-----PGL
MmGdnf		-----	-----	-----	-----	-----PFTA
DmDPP	408	-----R	-----TV	-----	-----RSLKPA	-----
MmBmp2	229	-----A	-----HL	-----	-----EENPGV	-----
MmBmp4	240	-----T	-----HL	-----	-----HQRTH	-----

Fig. S8 Page 4

		--β10--	---α5-TGFβ only	
MmBmp15	223	-MTSLDVAF	LYFNDDTD	-----QGK--L-----
MmGdf9	247	---LSVF	SELYLNDTSTQA	-----YHS--W-----QS-L-----Q--S
MmNodal	193	---ASTNV	MYSNRPQEQ	-----QL-----GGAT-LLWE-A-----E--S
MmGdf5	328	--RQVHEKA	SVFGRTKKRD	-----LFFNEIARSQGD-DKTV-Y-----E--Y
MmGdf7	254	GGTAAEER	ALNISRTRQKE	-----SLFREIQAARAL-RAA-----
MmGdf6	273	--RPPQER	ALVFTRSQRKN	-----LFTEMEQLGSA-EAAG-A-----E--G
DmActivin-beta	809	LSTNPNR	FLVHTES	-----SRTRVVR-----
CeTIG-3				
DmSCW	242	ASRTSL	EEVGYFNGPELLV	-----KIQKLRFKR-----
DmGBB	294	V-DDEF	QFFVGGFFRGPELIK	-----ATAHSSVHRSKR-----
MmBmp8a	237	A-PRSR	QFFVWTFFRASQSPV	-----RAPR-----
MmBmp8b	237	A-PRSR	QFFVWTFFRASQSPV	-----RAPRTAG-----
MmBmp6	348	G-PYDK	QFFVGAFFKVSEVHV	-----RTTR-----
MmBmp7	268	G-PQNK	QFFVGAFFKATEVHL	-----RSIR-----
MmBmp5	293	G-PQSK	QFFVGAFFKASEVLL	-----RSVR-----
MmBmp3	269	L--DSHV	R-----E-----	ALSVEREKKR
MmGdf10	278	SSADPR	VVR-----R-----	AAQVSKPLQDNELPGLDE-RFAPALHAQNFI
DmMyoglianin	472	---SKNEL	TWHIEIG	-----SQQKRRKKR-----
MmMstn	246	-GEDGL	NFFVEKVT	-----TPKRSR-----
MmGdf11	275	-GAEG	LHFFVEVRL	-----NTKRSR-----
MmTgfb1	252	--DMNR	FEELMATPLERAQ	-----HLHSSRR-----
MmTgfb2	300	KKTSGK	TEELMLLPSYRL	-----SQSSRRKKR-----
MmTgfb3	269	KQKDH	NFFVEIMIPPHRLD	-----SPGQSQRRKKR-----
MmInha	191	---THE	LVLLRCPLCS	-----CSGAPETTP-----
MmGdf15				
MmGdf1	215	L---AEAS	LVTLDPRLC	-----LPLRRHTE-----
MmGdf3	226	LLRSL	HASLVTLNPKHCH	-----PSSRRK-----
CeUNC-129	195	EMIRK	STFPLVYSKVNHTLDTVSVMKQTEQTKRK	-----RDLGNE-ELREYYNYSIPLDND
CeDAF-7	223	-----	LSTR-----	PKGSRKR
DmMaverick	534	DEHLN	LAFFVNIIGHGT	-----LNSQH-----GDA-D---IH--QIMLTNNR
MmInhba	282	EKEQ	SHRPEELQARQS	-----EDHPHRR-----
MmInhbb	271	PGEES	HRRFVYVQARLG	-----DSRHRIR-----
MmInhbc	212	LGWFS	HRRFVYVAAQVRVE	-----GKHRVIR-----
MmInhbe	210	DTAGQ	RRELEKIRAN	-----EPGAGARR-----
CeDBL-1	214	--RAQS	AFLEVFSDLSE	-----PSSVRRK-----
MmGdf2	244	PPGSK	NIFFFVFSNDR	-----SNGKETLELKE---MT-GHEQETMLVKTA
MmBmp10	244	SAQNKH	PELVFSDQ	-----SNDK-EQEEELNE---LI-THEQDLDLSDA
CeTIG-2	205	NDELK	IEAPLVALKDE	-----DAGPPKSSRRSAS-TTPI-SAPFMRQKVRS
DmDawdle	448	SVDKDY	RFVMDQNR	-----RR-----
MmLefty1				
MmLefty2				
MmAmh	414	SSGDPL	RALEELKALG	-----LRAEWH---GRE-G---R-----
MmGdnf				
DmDPP	432	ERWQH	KCELETTYTDDG	-----RHKARSIDVSGGE-G---G-----
MmBmp2	253	HSWSQ	IRFVTFGHDG	-----KGHPHKKR-----
MmBmp4	264	GDWAQ	LPELVTFGHDG	-----RGHTLRTRAKR-----

Fig. S8. Expanded All family members prodomain alignment indicating structural conservation. Sequences from Fig. 10 including underlining from Figs. 2, 4, 6, 8 and 10 are presented as in Fig. S4. Note that on the top half of Fig. S8 Page 1 there is an exception to the presentation of accurate numbering (shown by a double backslash). We removed a completely nonconserved stretch of 25 residues from Activin for efficiency of presentation. This allowed the unexpectedly conserved cysteines in the Association-region of ten proteins to be displayed near the structural features of the straightjacket. In the Association region unexpected cysteines conservation is underlined in red with known cysteines in TGF-β1-3 underlined in black. Accurate numbering for Activin resumes with the next alignment set below. All other numbering is accurate throughout. Also note that in the top half of Fig. S8 Page 4, the dimerization region β8 displays unexpectedly conserved cysteines in eight proteins underlined in red. Known β8 region cysteines in TGF-β1-3 are underlined in black. The features of TGF-β1 not conserved are α3, β5, β9 and α5. Features of BMP9 that are not conserved are β5 and β9’.