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

#### Robert G. Wisotzkey\* and Stuart J. Newfeld<sup>†,1</sup>

\*Sema4 Genomics, Stamford, Connecticut 06902 and <sup>†</sup>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- $\beta$  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- $\beta$  signaling. First, a quantitative analysis placed each family member from flies, mice, and nematodes into the Activin, BMP, or TGF- $\beta$  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  $\beta$ 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- $\beta$  function and evolution.

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

**S** ECRETED 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- $\beta$  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- $\beta$ regulation twice. First, Smad linker phosphorylation was predicted (Newfeld and Wisotzkey 2006) then validated by experiment in mice and flies (Fuentealba *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- $\beta$  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.

<sup>&</sup>lt;sup>1</sup>Corresponding 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- $\beta$  superfamily into two large subfamilies (BMP and Activin + TGF- $\beta$ ) that functionally appeared to rely on distinct sets of receptors and receptor-associated Smads was intellectually satisfying.

The first full-length tree of TGF- $\beta$  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; Kunnapuu *et al.* 2009). Proprotein/latent complex crystal structures of TGF- $\beta$ 1 (Shi *et al.* 2011), BMP9 (Mi *et al.* 2015), and Inhibin- $\beta$ 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 fulllength 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- $\beta$ . 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- $\beta$ 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- $\beta$  Binding (LTBP) LTBP association region" upstream of the straitjacket and in  $\beta$ 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-B 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-B 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- $\beta$  family members into the two well-known subfamilies Activin/TGF- $\beta$  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- $\beta$ 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- $\beta$  subfamilies independently, an Activin + TGF- $\beta$  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 (Kunnapuu *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-* $\beta$  *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- $\beta$  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- $\beta$ ) 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- $\beta$  prodomain tree), 0.65 for 50–100 informative positions (Activin ligand, TGF- $\beta$  ligand, and Activin + TGF- $\beta$  ligand trees), and 0.85 for 100–150 informative positions (BMP ligand, All ligand, cystine knot, and Activin + TGF- $\beta$  fulllength 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 cystine 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- $\beta$ , Activin, and BMP subfamilies (TGF- $\beta$ 1–3; Inhibin- $\beta$ a,  $\beta$ b,  $\beta$ c, and  $\beta$ 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- $\beta$  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- $\beta$  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- $\beta$  and Activin subfamilies in our sequence analysis is wholly consistent with structural differences between the subfamilies. For example, the TGF- $\beta$ 1 prodomain crystal structure contains a "bowtie" formed by  $\beta$ 8 and  $\beta$ 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- $\beta$ 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- $\beta$  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- $\beta$  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 cystine 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- $\beta$  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- $\beta$  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- $\beta$  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 (Timetree.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- $\beta$  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- $\beta$  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- $\beta$ 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- $\beta$  proteins that cluster together in all trees based on their recent origin and common ability to form heterodimers with Inhibin- $\alpha$  in the TGF- $\beta$  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- $\beta$  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  $\alpha$ -helices and  $\beta$ -sheets were located on the alignment revealing pockets of structural conservation in the annotated Activin alignments (Figure 2 and Figure S4). The locations for  $\alpha$ 1, the Latency Lasso, and  $\alpha$ 2 are based on Inhibin- $\beta$ a (Wang *et al.* 2016). The locations of the remaining features derive from our alignment of the Activin + TGF- $\beta$ subfamily following TGF- $\beta$ 1 (Shi *et al.* 2011).

The four features of the straitjacket domain ( $\alpha 1$ , the Latency Lasso,  $\alpha 2$ , and  $\beta 1$ ) 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.  $\beta 1$ 





**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- $\beta$  proteins was expected. (C) Full-length nodes ~0.95 are significant. The significant cluster of Myoglianin and Myostatin/GDF11 was expected.

		$-\alpha$ 1
DmMyoglianin MmGdf11 MmMstn DmActivin-beta MmGdnf	205 72 52 369 92	DSTESTKMHTLMRTNIKKLENTTKPISVPQNIIDNFY-RDYNASSKTTVWNRMESIDESH LRUESTKSQTLSKURIKEAPNISREVVKQLUP-KAPPLQQILDLHDFQGDA SRTEATKIQTLSKURJETAPNISKDATRQLP-RAPPLRELIDQYDQRDD VRUESIKRQTLTKIGISHKEN/SHPIPKQFIM-ETIYRVDGGR DVM FIQATIKRKKSPDKQAAAP-RER
MmInhba	53	EMVEAVKKHIINMIHIKKRPDVTQPVPKAAIL-NAIRKIHVGK
MmInhbb	77	DF BAVKRHTLSRLODRGRENTTHAVPKAAMV-TALRKUHAGK
MmInhbc	42	LLIDLAKKSTLDKLHUSORPILSRPWSRGALK-TALORURGPR
MmInhbe	42	LVLELAKOOT LEGTHUTSBERTTRET POAALT-RALE
		end straitjacketβ1
DmMyoglianin	264	SINDTYGDHIMTDF DESSSSQMQGDDAN
MmGdf11	122	QPEDFLEEDEYHATTETVISMAQETDPAVQTD
MmMstn	102	S-SDGSLEDDDYHATTETIITMPTESDFLMOAD
DmActivin-beta	411	MIPNNAFGSSGKNLDOKTIKI RAFASPGSHLFNGRGGRTDORSERDPSHHKY
MmTnhba	95	VGENGYVEIEDDIGBRAEMNELMEOTSEITTEAESG
MmTnhbb	119	VREDGRVEIPHLDGHASPGADGOERVSEIISEAETDGLASS
MmTnbbc	84	RETLIEHDOROFEYELLSEADTDLSS
MmTnbbo	91	
Platinde	04	WV
		begin arm-β2β3
MmBmp3	124	NILSTTYYVGELVNSLSC
MmGdf10	123	MILTAAFHFYSEPPROPROEVFC
CeTIG-3	17	TLEPNRKSWTC
DmMyoglianin	330	SIYIFPEEIOPHVRHNRKVDVFRFOIDSSYSDLSYAT HLYLRGODA SAHOP
MmGdf11	179	
MmMetn	158	
DmActivin-beta	643	FITTERECTOVDOVDTIFESAONDOVDSOKISTOSAO HTDTDKDKS
MmTnhha	153	
MmTnbbb	100	
Matabba	122	
Mathhbe	100	
Mminne	125	
		$-\alpha 3 \beta 4$
MmBmp3	145	P-EPQGCSHHTQRQHIQIDESAWELKSNQSQLLG
MmGdf10	147	KPRAK-NASCRLLTPGLPARLH IFRSLS-QNTATQG
CeTIG-3	34	LVKDCFOYSINSINHEILSASLIIDPKDTN
DmMyoglianin	383	GLLEEIKKOPBKDIVVTTHBAIBVANT
MmGdf11	188	
MmMstn	167	
DmActivin-beta	699	L DEKHI I NYKRKWCANKDHHDIKI WYFOL STSINIYEK
MmTnbba	165	
MmTnhhh	192	
MmTnbbc	111	
MmTnbbc	125	
Philitinne	100	I FOI
		β5β6α4β7- no-β8
MmBmp3	178	HISVDVVRPYRDSVSWLSKDITQLIRKAKQNEEFI <b>T</b> GFNITSRAHELP
MmGdf10	182	LIRGAMALTPPPRGLOOKDISSIIKAARRDGELLSAQLDTGEKDPG
CeTIG-3	76	GELOYVDRFER-ETLDKYHFDISHLFHKWMKOKSSDKM-IKIEITNSNT
DmMyoqlianin	410	TSFNPKVK FEFRHS PSGLGOWVAVDLKSLIGNLGSNMTOEILKGAET
MmGdf11	211	GGGRRHIR RSLKIE HSRSGHWOSIDFKOV HSWFROP-OSNWG E NAFDPSG
MmMstn	184	GTRYTG RSLKLD SPGTGIWOSIDVKTVLONWLKOP-ESNIGLE KALDENG
DmActivin-beta	737	GIDKAI FRASFOUDPKNLGWOKEDLTDT REWYGHTSHEKUR LUDCTGCGG
MmInhba	196	GLKGERSELLLSEKVVDARKSTWHIFPVSSSTORLLDO-GKSSUDVR ACEOCOE
MmInhbb	210	OGHGDRWNVVEKKVDLKRSGWHTFPTTEATOALFER-GERRUN, DVOCDSCOE
MmInhbc	158	
MmTnhbe	149	TRCRGFRTFLAEHOTTSSGQHALTIPSSGLRSEDS-GVVKIOLEFRPLDLNS
Think I Think 0	2.15	
	000	nobowtie $-\beta 9\beta 10 - no - \alpha 5$
MmBmp3	226	KRMLFFPEPYLVYANDAPAISEPESVVSSLQRHEDETAGTGP
MmGdf10	230	VPRPSSHMPY_LVYANDLAISEPNSVAVSLQRYDPAGDFEPGAAP
CeTIG-3	124	QNWINALSVLR
DmMyoglianin	460	WMWM-IEIGS
MmGdf11	265	LGPGAEGLH <mark>PF</mark> MELRVLENT
MmMstn	236	PGPGEDGLNPFLEVKVTDTP
DmActivin-beta	790	RYSLHLFQTSKLRGNSSDYLSTNPNRPFLVLHTESSR
MmInhba	250	SGASLVLLGKKKKKEVDGDGKKKDGSDGGLEEEKEQSH <mark>RPF</mark> IMIQARQSE
MmInhbb	262	PGEESHRPFVVVOARLGD
MmInhbc	207	HSSL-ILGWFSHRPFVAAQVRVE
	000	

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 in indicates an identical and gray shading biochemically similar amino acids at that position. As there is no available structurally annotated full-length seguence of Inhibin-Ba, the locations and naming of structural features are derived from our alignment of the Activin + TGF- $\beta$  subfamily that follows TGF- $\beta$ 1 of Shi et al. (2011). Underlined  $\alpha$ 2 indicates a location distinct in this subfamily from  $\alpha 2$  in TGF $-\beta$ 1 and the underlined  $\alpha$ 3 suggests that this feature of TGF-B1 may be absent in the Activin subfamily.  $\beta 9'$  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  $\beta 2$ –10 and  $\alpha 4$  show less conservation. Most notable are three I/L/V residues in  $\beta 4$ , a near universal tryptophan in  $\beta 6$  and near universal I/L/V residues in  $\alpha 4$  and  $\beta 7$ . 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- $\beta$  proteins that cluster together in all trees, based on their recent origin and common regulation by LTBP (Rifkin *et al.* 2018). Neither prototypical TGF- $\beta$  nor LTBP are present in flies and neither Maverick nor Dawdle has any relationship with them. The significant cluster of Dawdle and Inhibin- $\alpha$  seen only in the prodomain suggests common regulation. Given the ability of Inhibin- $\alpha$  to form heterodimers and the previously noted cluster of Activin and the Inhibin- $\beta$  group, Dawdle is a candidate as a heterodimerization partner with Activin. The ligand tree shows established clusters such as TGF- $\beta$ 1–3 and Lefty1, 2 and new clusters such as Maverick with GDF15. Overall for the TGF- $\beta$  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- $\beta$  alignments (Figure 4 and Figure S5). The locations and names of features derive from TGF- $\beta$ 1 (Shi *et al.* 2011). The four features that compose the straitjacket domain ( $\alpha$ 1, the



**Figure 3** TGF- $\beta$  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- $\alpha$  was unexpected. (B) Ligand nodes ~0.65 are significant. The significant clusters of TGF- $\beta$ 1–3 and Lefty1 and 2 were expected. (C) Full-length nodes ~0.95 are significant. The significant. The significant. The significant clusters of TGF- $\beta$ 1–3 and Lefty1 and 2 were expected. (C) Full-length nodes ~0.95 are significant. The significant clusters of TGF- $\beta$ 1–3 and Lefty1 and 2 were expected.

Latency Lasso,  $\alpha 2$ , and  $\beta 1$ ) 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  $\beta 1$  contains only a modestly conserved F/Y.

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

#### Activin + TGF-β subfamily trees

The combined Activin + TGF- $\beta$  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- $\beta$ 1–3 and Myostatin/ GDF11. The group of four Inhibin- $\beta$  proteins is the next closest cluster. Dawdle ends up as a solo next to Inhibin- $\alpha$ . The clustering of the three fly proteins with Dawdle as an outlier is reminiscent of the Inhibin- $\beta$  group's relationship with Inhibin- $\alpha$ , 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- $\beta$  subfamily relationships are simply recreated. For example, Activin is with the four Inhibin- $\beta$  proteins and Myoglianin is with Myostatin/GDF11. Overall for the Activin + TGF- $\beta$  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 +  $TGF-\beta$  subfamily alignment contains four features where the subfamilies differ, further

		$\alpha_1$
CeUNC-129	15	IANCAKWDVDLINETIRDLUHFKSSDPNVTSFHRSSHT
CeDAF-7	19	GLTFNCTNSGVCIEKMKOHRTEYLKNEILDOLNMKEARKGLKPMDPEMKSVYLEMYR
DmMaverick	245	SKRVDETRLKHLVLKGLGIKKLEDMRKVNISOAEYSS
MmTqfb1	27	AAG-LSTCKTIDMELVKRKRIEAIRGOILSKLRLASPESOGEVPPGPLPEAVLALYNSTR
MmTqfb2	18	ALS-LSTCSTLDMDOFMRKRIEAIRGOILSKLKLTSPEEDY-PEPDEVPPEVISIYNSTR
MmTafb3	21	SLS-LSTCTTI DEGHIKKKRVEAIRGOILSKURLTSPEEPS-V-MTHVPYOVLALYNSTR
MmLeftv1	20	EAL-TGROTLGST LOOLOUDOPEVI, DKADVEGVVTPSHVRTOVV-
MmLefty2	20	
MmTnha	204	CSC-SGRPET
DmDawdle	233	CES-NROVEHITEROLTHLETERVKOOTTEKTRIKESEKVSAVEI PKPTEDGM
MmGdf15	118	AYR-VHRALLILTPTARPWDTTRPL-KRALSURGPRAPALRLELTPPPDLA
MmAmh	351	AAT-EREPMPHGPASAPWAAGLOR-RWAVELOAAASELRDLPGLPPTAPPLLARL
11112 11111	551	
		end straitjacketβ1
CeUNC-129	53	-LTEHMKNLYENFIDEDSNEDGNLVRAIEPAVGKFEGQEVLVFDVE
CeDAF-7	76	DLLEKDEQDMGVEMSFYTAKD
DmMaverick	282	KYIEYLSRLRSNQEKGNSYFNNFMGASFT
MmTgfb1	86	DRVAGESADPEPEPEADYYAKEV
MmTgfb2	76	DLLQEKASRRAAACERERSDEEYYAKEV
MmTgfb3	78	ELLEEMHGEREEGCTQETSESEYYAKEI
MmLefty1	63	ALLQHSHASRSRGKRFSQNLREVA
MmLefty2	63	ALLOGSHADRSRGKRFSQNEREVA
DmDawdle	285	-TI SHPDDSTKNK-ELDDYYARTS
MmGdf15	167	MLPSGGTOLELR
MmAmh	405	- ALCENDSRSSGDPLBALLL
	100	
		beginarm $-\beta 2\alpha 3\beta 3$
CeUNC-129	143	TLKKIRVGGDENLEEYKVIMDATKS-VFDSYHLDAKQAVFRITREHSKMRPYA
CeDAF-7	102	LERSDILQATLTVSIEIPAK
DmMaverick	356	LLRGEQDTMNILLHEPLTNAQDANFHHDKIDEANVRLMLLYSSSL
MmTgfb1	126	DISHSIYMFCNTSDIREAVPEPPLLSRAELRLQRLKS
MmTgfb2	136	SQVLCGYLDAIPPTFYRPYFRIVRDVSTMEKNASNUVKAEFRVFRLQNPK
MmTgfb3	126	VEKNGTNLFRENVSSVEKNGTNLFREFRVLRVPNPS
MmLefty1	95	RLPPNSELVQAVLRLFQEPVPR
MmLefty2	95	RLPPNSELVQAVLRLFQEPVPR
-		
DmDawdle	327	ADAEGFDWSTAVLWWFKNKQNR
DmDawdle	327	ADAEGFDWSTEVLWWFKNKQNR
DmDawdle	327	β4 SNPSMCFT <b>FK</b> DDADAEGFDVST <b>F</b> VLW FKNKQNR β4β5β6α4
DmDawdle CeUNC-129	327 249	β4β5β6α4 β4β5β6α4 LDN DREPIKRKNGKKNSLSEEISSEDV@QGGGEERSREE
DmDawdle CeUNC-129 CeDAF-7	327 249 138	SNPSMCFTEK         DDADAEGFDWSTEVLW_FKNKQNR          β4        β6          β4        β6          CDNEDREPIKRKNGKKNSLSEEISSEDVØQCEGEEESSEE           GMLQDVQVVEKNE         G-S-MGEMVTSGIFATK-GSERISIQLPIDTØKSØFTIS
DmDawdle CeUNC-129 CeDAF-7 DmMaverick	327 249 138 441	
DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1	327 249 138 441 163	
DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb2	327 249 138 441 163 187	
DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb2 MmTgfb3	327 249 138 441 163 187 161	
DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb3 MmLefty1	327 249 138 441 163 187 161 140	β4        β6        β6        α4           GMLQDVQVQVYEKNE         G-S-MGEMVTSGIFATK-GSERISIQLPIDTVKS0FTIS           SNSQQITIKVYQLISANRRRKITSRK         FFGNVGFQETRTQWIED-VTKAWRS0LNKS          SVEQHVELYQKYSNNSWRYIGNRLITPTDTPEWLSFD-VTGVK00LNQG           ARVAEQRIELYQILKSK         LTSPQRYIDSKVVKTRAEGEWLSFD-VTGVVQ0LQULKSK           SKRTEQRIELFQILRDDH-IKQQYIGGKNIPTRTAEWLSFD-VTDVREDLLRR           SARARVTIEWLRFRD         GSNRTA-LIDSRLVSIHESGWKAFD-VTEAVNF0QLS
DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb2 MmTgfb3 MmLefty1 MmLefty2	327 249 138 441 163 187 161 140 140	$\label{eq:starter} =$
DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb2 MmTgfb3 MmLefty1 MmLefty2 DmDawdle	327 249 138 441 163 187 161 140 140 370	β4β5β6α4 β4β5β6α4 GMLQDVQVQVYEKNEG-S-MGEMVTSGIFATK-GSERISIQLPIDTVKSWFTIS NSSQUTIKVYQLISANRRKITSRKIEFGNVGFQETRTQWIEFD-VTGVVROLNQG ARVAEQRIELYQILSKILTSPTQRYIGRLLTPTDTPEWLSFD-VTGVVROLNQG ARVAEQRIELYQILKSKILTSPTQRYIGKNLTRTAEWLSFD-VTDVVREXLLRR SARARVTIEWLRFRDGSNRTA-LIDSRLVSIHESGWKAFD-VTEAVNFTQQLS SARARVTIEWLRFREGSNRTA-LIDSRLVSIHESGWKAFD-VTEAVNFTQQLS -TSAQQTIVVSEVEVQQKDSKYLSAAKTIAIQSVNVQDEWKKIDEWPIKH0ISGH
DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb2 MmTgfb3 MmLefty1 MmLefty2 DmDawdle	327 249 138 441 163 187 161 140 140 370	β4β5β6α4 β4β5β6α4 GMLQDVQVVQVYEKNE G-S-MGEMVTSGIFATK-GSERISIQLPIDTVKSØFTIS NSSQLITKVYQLLSANRRKKISSKIEFGNVGFQETRTQMIEFD-VKAVRSØLNKS -SVEQHVELYQKYSNNSWRYIGNKLLTPTDTPEVLSFD-VTQVVRQLNQG ARVAEQRIELYQILKSKILTSPTQRYIGSKVVKTRAEGEWLSFD-VTDVREØLLRR SARARVTIEWLAFRD GSNRTA-LIDSRLVSIHESGWKAFD-VTEAVNFØQLS SARARVTIEWLAFRD GSNRTA-LIDSRLVSIHESGWKAFD-VTEAVNFØQLS SARARVTIEWLAFRD GSNRTA-LIDSRLVSIHESGWKAFD-VTEAVNFØQLS -TSAQOTIVVSEVEV QQKDSKYLSAAKTTAIQSVNVQDEWMKID-IEWFKHISGH
DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb2 MmTgfb3 MmLefty1 MmLefty2 DmDawdle	327 249 138 441 163 187 161 140 140 370	β4      β5      β6      α4        β4      β5      β6      α4
DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb2 MmTgfb3 MmLefty1 MmLefty2 DmDawdle CeUNC-129	327 249 138 441 163 187 161 140 140 370 289	β4      β5      β6      α4        β5      β6      β6      β6
DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb2 MmLgfb3 MmLefty1 MmLefty2 DmDawdle CeUNC-129 CeDAF-7	327 249 138 441 163 187 161 140 140 370 289 190	β4      β5      β6      α4        β4      β5      β6      α4        β4      β5      β6      α4
DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb2 MmTgfb3 MmLefty1 MmLefty1 DmDawdle CeUNC-129 CeDAF-7 DmMaverick	327 249 138 441 163 187 161 140 140 370 289 190 498	β4      β5      β6      α4        β4      β5      β6      α4
DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb2 MmTgfb3 MmLefty1 MmLefty2 DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1	327 249 138 441 163 187 161 140 140 370 289 190 498 212	β4      β5      β6      α4        β4      β5      β6      α4
DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb2 MmLefty1 MmLefty2 DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb2	327 249 138 441 163 187 161 140 140 370 289 190 498 212 243	β4      β5       -β6       -α4        β4      β5       -β6       -α4
DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb2 MmLefty1 MmLefty2 DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb2 MmTgfb3	327 249 138 441 163 187 161 140 140 370 289 190 498 212 243 216	β4      β5       -β6       -α4        β4      β5       -β6       -α4
DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb3 MmLefty1 MmLefty2 DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb3 MmLefty1	327 249 138 441 163 187 161 140 370 289 190 498 212 243 216 192	β4      β5       -β6       -α4        β4      β5       -β6      α4
DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb2 MmTgfb3 MmLefty1 MmLefty2 DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb3 MmLefty1 MmLefty2	327 249 138 441 163 187 161 140 140 370 289 190 498 212 243 216 192 192	β4      β5      β6      α4        β4      β5      β6      α4
DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb2 MmLefty1 MmLefty2 DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb2 MmTgfb3 MmLefty1 MmLefty2 DmDawdle	327 249 138 441 163 187 161 140 140 370 289 190 498 212 243 216 2192 192 2426	β4      β5      β6      α4        β5      β6      β6      α4         SSQ0TT       LDN       DREPIKRKNGKMSLSEEISSEDVØQCG GEBSREE-         SMSQ0TT       KVYQLSANRRKISKISKEFGNVGF-QETRTØVIED-VIAVRSØLNKSØ        SVEQHVELYQKYSNNSWRYGRLITPTDTPEWLSFD-VIGVVROØLNQG         ARVAEQRIELYQILKSKELTSPTORYLDSKVVKTRAEGEWLSFD-VIDVVROØLNQG         SARARVT EWLRPD-H-IGSNRTA-LDSKVVSIHESGWAFD-VIDVVROØLS         SARARVT EWLRVRE GSNRTA-LDSKVVSIHESGWAFD-VIEAVNFØQQLS         -TSAQQT VVSEVEV QQKDSKYLSAAKTTALQSVNVQDEWMKID-TEWPIKHØISGH
DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb2 MmLefty1 MmLefty2 DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb2 MmTgfb3 MmLefty1 MmLefty2 DmDawdle	327 249 138 441 187 161 140 370 289 190 498 212 243 216 192 243 216 192 426	β4      β5       -β6      α4        β4      β5       -β6      α4
DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb2 MmLefty2 DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb2 MmTgfb3 MmLefty1 MmLefty2 DmDawdle CeUNC-129	327 249 138 441 187 161 140 370 289 190 498 212 243 216 192 243 216 289	β4      β5       -β6       -α4        β4      β5       -β6       -α4         GMLQDVQVQVEKNE       GG       -G       -β6       -α4         GMLQDVQVQVEKNE       GG       -G       -β6       -α4         SSQQLT       KVYQLSANRRKITSRK       FEGNGG       -QETRTQMIETDTWKSGITAK         SSQQLT       KVYQLSANRRKITSRK       FEGNGG       -VTGVVROLNQE         ARVAEQRIELYQILKSK       LTSPTQRYIDSKVV       FTRTAEGGWLSED-VTGVVROLNQE         ARVAEQRIELYQILKSK       LTSPTQRYIDSKVV       FTRTAEGGWLSED-VTDVREDLLRR         SARARVT       FWILKRE       GSNRTA-LIDSRLV       SIHESGWKAED-VTDAVQEDLLRR         SARARVT       FWILKPRD       GSNRTA-LIDSRLV       SIHESGWKAED-VTDAVFRQUS         SARARVT       FWILKPRE       GSNRTA-LIDSRLV       SIHESGWKAED-VTDAVFRQUS         SARARVT       FWILKPRE       GSNRTA-LIDSRLV       SIHESGWKAED-VTDAVFRQUS         SARARVT       FWILKPRE       GSNRTA-LIDSRLV       SIHESGWKAED-VTEAVFRQUS         SARARVT       FWILKPRE       GSNRTA-LIDSRLV       SIHESGWKAED-VTEAVFRQUS         STSAQQT       VVSEVEV       QQKDSKYLSAAKTIAIQSVNVQDEWKID-       FWILKPICOLS        G7-      G8
DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb2 MmTgfb3 MmLefty1 DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb3 MmLefty1 MmLefty2 DmDawdle CeUNC-129 CeDAF-7 CeUNC-129 CeDAF-7	327 249 138 441 163 187 161 140 370 289 190 498 212 243 216 192 192 426 289 211	β4      β5       -β6      α4        β4      β5       -β6      α4         GMLQDVQVQVYEKNE       GS-MGEMVTSGIFATK-GSERISIQLPIDTVKSØFTIS         NSSQQLTIKVYQLLSANRRKITSRKIEFGNVGFQETRTQMIEDD-VKAVRSØLNKS        SVECHVELVQKYSNNSWRYLGNELLTPTDTPEMLSD-VTGVVROLNQG         ARVAEQRIELYQILKSK LTSPTQRYLDSKVVKTRAEGEWLSD-VTDAVQEØLHHK         SKRTEQRIELYQILKSK LTSPTQRYLDSKVVSIHESGWKAPD-VTDAVQEØLHHK         SARARVT TEWLRFRD       GSNRTA-LLDSRLVSIHESGWKAPD-VTEAVNFØQUS         SARARVT TEWLRVRE       GSNRTA-LDSRLVSIHESGWKAPD-VTEAVNFØQUS         -TSAQQT TVVSEVEV       QQKDSKYLSAAKTIAIQSVNVQDEØMKID-EWPTKHDISG         PI-QGIFVKAMLDGRNV
DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb2 MmLefty1 MmLefty2 DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb3 MmLefty1 MmLefty2 DmDawdle CeUNC-129 CeDAF-7 DmMaverick	327 249 138 441 163 187 161 140 140 370 289 190 498 212 243 216 192 192 426 289 211 554	β4      β5      β6      β6        β4      β5      β6      β6        β4      β5      β6      β6        β4      β5      β6      β4
DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb2 MmLefty1 MmLefty2 DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmLefty1 MmLefty2 DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1	327 249 138 441 163 187 161 140 370 289 190 289 2243 216 192 243 2192 426 289 2151 525 4260	β4      β5       -β6      α4        β4      β5       -β6      α4        β5       -β6      α4
DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb2 MmLefty1 MmLefty2 DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb2 MmTgfb3 MmLefty1 MmLefty2 DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmLefty2 DmDawdle	327 249 138 441 163 187 161 140 370 289 190 498 212 243 216 192 243 216 192 242 426 289 211 554 426 289 211	β4      β5       -β6      α4        β4      β5       -β6      α4         GMLQDVQVQVYEKNE       GG       -G       -β6      α4         SSQQLT       KVYQLSANRRKKINSK       EGSFA       -FG
DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb2 MmTgfb3 MmLefty1 MmLefty2 DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb3 MmLefty1 MmLefty2 DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmLefty2 MmLefty2 MmLefty2 MmLefty3 MmLefty1 MmLefty2 MmLefty3 MmLefty1 MmLefty2 MmLefty3 MmLefty3 MmLefty3 MmLefty3 MmLefty3 MmLefty3 MmTgfb3 MmTgfb3 MmTgfb3 MmTgfb3	327 249 138 441 163 187 161 140 370 289 190 498 212 243 216 192 243 216 192 243 216 249 211 554 240 2257	β4      β5       -β6       -α4        β4      β5       -β6       -α4         GMLQDVQVQVYEKNE       GS-MGEM/TSGIFA       TK-GSERISIQLPIDTVKSØFTIS         SNSQQLTIKVYQLLSANRRKITSRKIEFGNVGFQETRTQMIEDD-VKAVRSØLNKS       -SVEQHVELVQKYSNNSWRYGRLL       TDTPDTPEMLSD-VKAVRSØLNKS        SVEQHVELVQILSK       LTSPTQRYLDSKVV       TTTDTPEMLSD-VKAVRSØLNKS         SARARVT EWLRFRD       GSNRTA-LIDSRLVSIHESGWKAPD-VTDAVQEØLHHK         SARARVT EWLRFRD       GSNRTA-LIDSRLVSIHESGWKAPD-VTEAVNFØQUS         SARARVT EWLRVRE       GSNRTA-LIDSRLVSIHESGWKAPD-VTEAVNFØQUS         SARARVT EWLKVRE       GSNRTA-LIDSRLVSIHESGWKAPD-VTEAVNFØQUS         SARARVT EWLKVRE       GSNRTA-LIDSRLVSIHESGWKAPD-VTEAVNFØQUS         SARARVT EWLKVRE       GSNRTA-LIDSRLVSIHESGWKAPD-VTEAVNFØQUS         SARARVT EWLKVRE       GQKDKSKYLSAAKTIAIQSVNVQDEWKID-         EVENKKISEC       LORTALOSVNVQDEWKID-         FSAQQT VVSEVEV       QQKDSKYLSAAKTIAIQSVNVQDEWKID-         EVENKKISEGSCDSKDNK-
DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb2 MmTgfb3 MmLefty1 DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmLefty1 MmLefty2 DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb2 MmTgfb1 MmTgfb3 MmTgfb3 MmLefty1	327 249 138 441 163 187 161 140 370 289 190 289 190 289 192 426 289 192 426 289 255 4 240 2857 195	β4      β5      β6      β6        β4      β5      β6      β6        β4      β5      β6      β6        β4      β5      β6      α4
DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb2 MmLefty1 MmLefty2 DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb2 MmTgfb3 MmLefty1 DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmLefty2 MmTgfb1 MmTgfb2 MmTgfb3 MmLefty1 MmTgfb2 MmTgfb3 MmLefty1 MmTgfb2 MmTgfb3 MmLefty1	327 249 138 441 163 187 161 140 370 289 190 289 2243 216 192 243 216 192 243 2151 5257 195	β4      β5       -β6      α4        β4      β5       -β6      α4         GMLQDVQVQVEKNE       GG-S-MGEM/TSGIFATK-GSERISIQLPIDTVKSGFTIS         SNSQQLTIKVYQLISANRRKITSRKIEFGN/GFQETRTQMIED-VTGVRQDLNQG         ARVAEQRIELYQILSANRRKITSRKIEFGN/GFQETRTQMIED-VTGVRQDLNQG         ARVAEQRIELYQILSANRRKITSRKIEFGN/GFQETRTQMIED-VTGVRQDLNQG         ARVAEQRIELYQILSANRRKTSRKITSRKIEFGN/GFQETRTQMIED-VTGVRQDLNQG         ARVAEQRIELYQILKSK       LTSPTQRYLDSKVVKTRAEGEWLSPD-VTDVRQLLRR         SARARVTEWLFRD       GSNRTA-LDSRLVSIHESGWKAPD-VTDAVQEDLHHK         SARARVTEWLEFRD       GSNRTA-LDSRLVSIHESGWKAPD-VTDAVNFDQQLS         SARARVTEWLREFR-       GSNRTA-LDSRLVSIHESGWKAPD-VTEAVNFDQQLS         -TSAQQTVVSEVEV       QQKDSKYLSAAKTTAIQSVNVQDEWMKID-TEWPIKH02GG        67-      68
DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmTgfb2 MmLefty1 MmLefty2 DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb2 MmTgfb3 MmLefty1 MmLefty2 DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmLefty2 DmDawdle CeUNC-129 CeDAF-7 DmMaverick MmTgfb1 MmLefty1 MmTgfb2 MmTgfb3 MmLefty1 MmLefty1 MmLefty1 MmLefty2 MmTgfb3 MmLefty1 MmLefty2 MmLe	327 249 138 441 163 187 161 140 370 289 190 2498 212 243 216 192 243 216 192 242 289 211 2554 426 285 257 195 257 195	β4      β5       -β6       -α4        β4      β5       -β6       -α4        β4      β5       -β6      α4

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-B1 of Shi et al. (2011). The underlined  $\alpha$ 2 indicates a location distinct from  $\alpha 2$  in Inhibin- $\beta a$ . Underlined  $\alpha 3$  and  $\beta 3$ indicate these features are in the reverse order vs. the Activin subfamily. The underlined features  $\beta 7$ and  $\beta 9$  of TGF- $\beta 1$  do not appear conserved outside its two siblings. Unexpectedly, in B8 Maverick and Dawdle have a pair of cysteines (red underline) that align with those in TGF- $\beta$ 1–3 (black underline), although the spacing is not the same (CxxC vs. CxC).  $\beta 9'$  is not visible and may be unique to the BMP subfamily. The underlined  $\alpha 5$  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- $\beta$ 1 (Shi *et al.* 2011). The first two features of the straitjacket ( $\alpha$ 1 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  $\alpha$ 2 and for  $\beta$ 1, although neither of the versions of  $\alpha$ 2 or  $\beta$ 1 are conserved.

At the 5' end of the arm,  $\beta 2$  contains a near universal phenylalanine.  $\beta 3$  is the third feature distinct in the Activin and TGF- $\beta$  subfamilies.  $\beta 3$  shows conservation in the TGF- $\beta$ pattern that pulls in several Activin subfamily members anchored by an alanine and an I/L/V.  $\beta 4$  has three near universal I/L/V residues.  $\beta 6$  has a near universal tryptophan and  $\alpha 4$ a near universal I/L/V. At the 3' end of the arm  $\beta 7$  is the fourth feature that is distinct, it shows conservation in the



**Figure 5** Activin + TGF- $\beta$  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- $\beta$  proteins and Dawdle's location near Inhibin- $\alpha$  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- $\beta$  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- $\beta$  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- $\beta$  subfamily are joined by Activin, Inhibin- $\beta a$ , and Inhibin- $\beta b$  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- $\beta a$ , and Inhibin- $\beta b$  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

straitja	acket	α1α2
DmMaverick	245	SKRVDETR-LKHLVEKGEGEKKL-PDWRKVNISQAEYSSKYIEYLSRLRS
DmMyoglianin	202	AKVDSIES-IKMHIIMRINIKKL-PNITKPISVPQNIIDNFYRDYNASS-KT.
MmGdf11	69	SRELRIES-IKSQILSKLRIKEA-PNISREVVKQLLPKAPPLQQILDLHDFQGDALQ-P-•
MmMstn	49	TRYSRIEA-IKIQILSKURETA-PNISKDAIRQLLPRAPPLRELIDQYDVQRDDSS-•
CeTIG-3	1	MSTSR
MmTgfb1	41	VKRKRIEA-IRGQILSKLRLASP-PSQGEVPPGPLPEAVLALYNSTRDRVA-G-•
MmTgfb2	32	FMRKRIEA-IRGQILSKLKUTSP-PEDY-PEPDEVPPEVISIYNSTRDLLQ-E-*
MmTgfb3	35	IKKKRVEA-IRGQILSKLRUTSP-PEPS-V-MTHVPYQVLALYNSTRELLE-E-•
CeUNC-129	25	TSFHRS*
CeDAF-7	34	MKQHRTD-YLKNEILDQUNMKEA-PKGLKPMDPEMKSVYLE-M-YRDLLEK-DEQ •
MmBmp3	56	VSEHMIWLYDRYSGSSRVQATRT-EGSQLPGPQPL
MmGdf10	71	VAIHMIRLYEKYNRRG·
MmInhba	50	SQPEMVEA-VKKHILNMLHUKKR-PDVTQPVPKAALLNAIRKLHVG·
MmInhbb	74	VDGDFIEA-WKRHILSRIQURGR-PNITHAVPKAAMVTALRKLHAG*
MmInhbc	39	QRELLIDL-AKKSILDKUHUSQR-PIISRPVSRGALKTALQRLRGP•
MmInhbe	39	ERALVINGL-AGQQILEGIHITSR-PRITRPLPQAALTRALRRLQPK•
DmActivin-beta	546	MNENALKKSTYPIDINHSIDNKTHTG•
MmLefty1	27	ILGS LQQLQLDQP-PVIDKADVEGMVIPSHVRTQYVALLQHSHASR•
MmLefty2	27	WLSSIMQQMQHSQA-WTHDSADVEEMAIPTHVRSQYVALLQGSHADR•
DmDawdle	247	LTHLRIDF-WRQQLDEKRKKKES-EKWSAVELPKPIFDGMTLSHPDDST.
MmGailo	132	ARPWDETR-PL-KRALSERGPRA-BAERLERLTPPPDLAMLPS
MINAMIN	305	GE-6 Activin arm TGE-6
	-	$\beta 1 \beta 1 - \beta 2 - \alpha 3 \beta 3$
DmMaverick	377	SYFN.LH.PL.DANHHDKIDEANVRLMLYSSSLATN.QQUTKVYQLLSANRR
DmMyoglianin	350	DTYG IFPE · VFR QUDSSYSDUS- · YATLHUY · KDUVUTUHRAIRVANTT ·
MmGdf11	160	EYHACHEHFSPKVMFTKWLKAQLWWYATVYLQILRLKPLTGEG.
MmMstn	139	DYHA••-CFEKFSSKIQYNKWV-•KAQLW Y•TTVFVQILRLIKPMKD-•
CeTIG-3	38	DLYG CFQYSINSINHEILSASLI · ·TNISIVVYEVDEL ·
MmTgfb1	133	DYYA WFONTSDIREAVPEPPLUS- • RAELRUQ • VEQHUEUYQKYS •
MmTgfb2	157	EYYA•-GYLD•IWREDVSTMEKNASNUV-•KAEFRWF•AEQRIEIYQILKSKDLT•
MmTgfb3	131	EYYA VFRENVSSVEKNGTNEF- RAEFRUL • TEQREE FQILRPDEH •
CeUNC-129	109	•LVFDV•EIHEYIRRRDSFARR-RS-•QTIKKIRVGG•
CeDAF-7	125	•AKEDV•TUTVSIEIPAKDS•••QDVQVQVYEKNEDG•
MmBmp3	129	•HTFNL•TUYEYVGELVNISUSC••QHIQIDUSAWILKSN•
MmGdf10	128	•YFENL•AFHEYSEPPRWPRAREWFC••LPARCHIIFRSLS-Q•
MmInhba	135	•ITPAE•TUHEEISKEGSDL-SVVE-•RAEVWUF•TKVTURUFQQQKHPQGSLDT•
MmInhbb	162	•ISMAE•RLYEFVSNEGNQN-LFVV-•QASLWLY•RKVRVKVYFQE•
MmInhbc	114	•ISEAD•RUEEHFSGRMASG-MEWR-•QTRFMFF•WNIRULVLRP•
MmInhbe	109	•ISBAT•MUTEQUSPLWSHHUY-•HARUW•UYURUFRCGT•
DmActivin-beta	659	EYFN•ITEAE•IMEESAQNRRVPS-QKMSI•PHSLWME•HRMKMWWFQLSTSINITEKG•
MmLeftyl	98	EVAG••LUVNGWEQRLPPN-SEUV-•QNVLRUF•ARVTHEWLRFRDDG•
MmLefty2	98	EVAG LUVEGVEQRLPPN-SEUV- QUVLREF ARWTEWLKVREDG •
Dundaware		
	332	Activin
	332	$\begin{array}{ccc}\beta &\alpha &\alpha &\beta &\beta &\beta &\alpha &\beta &\beta$
DmMaverick	467	$\begin{array}{c}\alpha\alpha + -\alpha + -\alpha + -\alpha + -\alpha + -\alpha + -\alpha$
DmMaverick DmMyoglianin	467 421	
DmMaverick DmMyoglianin MmGdf11	467 421 222	
DmMaverick DmMyoglianin MmGdf11 MmMstn	467 421 222 193	$\begin{array}{llllllllllllllllllllllllllllllllllll$
DmMaverick DmMyoglianin MmGdfl1 MmMstn CeTIG-3	467 421 222 193 84	
DmMaverick DmMyoglianin MmGdfl1 MmMstn CeTIG-3 MmTgfb1	467 421 222 193 84 187	
DmMaverick DmMyoglianin MmGdfl1 MmMstn CeTIG-3 MmTgfb1 MmTgfb2	467 421 222 193 84 187 218	$\begin{array}{llllllllllllllllllllllllllllllllllll$
DmMaverick DmMyoglianin MmGdf11 MmMstn CeTIG-3 MmTgfb1 MmTgfb2 MmTgfb2 MmTgfb3	467 421 222 193 84 187 218 191	
DmMaverick DmMyoglianin MmGdfll MmGdfll CeTIG-3 MmTgfb1 MmTgfb2 MmTgfb3 CeUNC-129	467 421 222 193 84 187 218 191 156	
DmMaverick DmMyoglianin MmGdfl1 MmMstn CeTIG-3 MmTgfb1 MmTgfb2 MmTgfb2 CeUNC-129 CeDAF-7	467 421 222 193 84 187 218 191 156 170	β6α4
DmMaverick DmMyoglianin MmGdfl1 MmMstn CeTIG-3 MmTgfb1 MmTgfb2 MmTgfb2 CeDAF-7 MmBmp3 WmC910	467 421 222 193 84 187 218 191 156 170 181	β6α4
DmMaverick DmMyoglianin MmGdfll MmGdfll MmTgfb1 MmTgfb3 CeUNC-129 CeUAF-7 MmBmp3 MmGdfl0	467 421 222 193 84 187 218 191 156 170 181 185	
DmMaverick DmMyoglianin MmGdfl1 MmMstn CeTIG-3 MmTgfb1 MmTgfb2 CeUNC-129 CeDAF-7 MmBmp3 MmGdf10 MmInbba	467 421 222 193 84 187 218 191 156 170 181 185 204 216	β6α4
DmMaverick DmMyoglianin MmGdf11 MmMstn CeTIG-3 MmTgfb1 MmTgfb2 MmTgfb2 CeUNC-129 CeUAF-7 MmBmp3 MmGdf10 MmInhba MmInhba MmInhbb	467 421 222 193 84 187 218 191 156 170 181 185 204 216 163	β6α4     Activit      Fβ5
DmMaverick DmMyoglianin MmGdfll MmGdfll MmTgfb1 MmTgfb3 CeUNC-129 CeUAF-7 MmBmp3 MmGdfl0 MmInhba MmInhbb MmInhbb MmInhbb	467 421 222 193 84 187 218 191 156 170 181 185 204 216 163 154	
DmMaverick DmMyoglianin MmGdfll MmTgfb1 MmTgfb1 MmTgfb3 CeUNC-129 CeDAF-7 MmBmp3 MmGdf10 MmInhba MmInhbb MmInhbb MmInhbc MmInhbc	467 421 222 193 84 187 218 191 156 170 181 185 204 216 163 154 746	β6α4
DmMaverick DmMyoglianin MmGdf11 MmMstn CeTIG-3 MmTgfb1 MmTgfb2 CeUNC-129 CeDAF-7 MmBmp3 MmGdf10 MmInhba MmInhbb MmInhbc MmInhbc MmInhbc MmInhbc MmInhbc	467 421 222 193 84 187 218 191 156 170 181 185 204 216 163 154 165	β6α4
DmMaverick DmMyoglianin MmGdfll MmGdfll MmTgfb1 MmTgfb3 CeUNC-129 CeDAF-7 MmBmp3 MmGdfl0 MmInhba MmInhbb MmInhbb MmInhbe DmActivin-beta MmLefty1 MmLefty1	467 421 222 193 84 187 218 191 156 170 185 204 216 163 154 746 165	
DmMaverick DmMyoglianin MmGdfll MmGdfll MmTgfb1 MmTgfb3 CeUNC-129 CeDAF-7 MmBmp3 MmGdf10 MmInhba MmInhbb MmInhbb MmInhbb MmInhbc MmLefty1 MmLefty2 DmDawdle	467 421 222 193 84 187 218 195 170 181 156 170 181 185 204 216 163 154 746 165 395	β6α4
DmMaverick DmMyoglianin MmGdf11 MmMstn CeTIG-3 MmTgfb1 MmTgfb2 CeUNC-129 CeDAF-7 MmBmg3 MmGdf10 MmInhba MmInhbb MmInhbc MmInhbc MmInhbc MmInhbc MmInhbc MmInhbc MmLefty1 MmLefty2 DmDawdle	467 421 222 193 84 187 218 191 156 156 163 154 746 165 395	β6α4
DmMaverick DmMyoglianin MmGdfll MmGdfll MmTgfb1 MmTgfb3 CeUNC-129 CeDAF-7 MmBmp3 MmGdfl0 MmInhba MmInhbb MmInhbb MmInhbe DmActivin-beta MmLefty1 MmLefty2 DmDawdle	467 421 222 193 84 187 218 191 156 170 181 155 204 216 163 154 165 165 395	β6α4
DmMaverick DmMyoglianin MmGdfll MmGdfll MmTgfb1 MmTgfb3 CeUNC-129 CeDAF-7 MmBmp3 MmGdfl0 MmInhbb MmInhbb MmInhbb MmInhbb MmLofty1 MmLefty2 DmDavdle DmMaverick	467 421 222 193 84 187 218 191 156 170 181 156 163 154 746 746 5395 5200	β6α4
DmMaverick DmMyoglianin MmGdfll MmGdfll MmTgfb1 MmTgfb3 CeUNC-129 CeDAF-7 MmBnp3 MmGdfl0 MmInhba MmInhba MmInhbb MmInhbb DmActivin-beta MmLefty1 MmLefty2 DmDawdle DmMaverick DmMaverick	467 421 222 193 847 156 170 181 185 204 216 163 154 746 165 395 520 463 395	β6α4
DmMaverick DmMyoglianin MmGdfll MmGdfll MmTgfb1 MmTgfb3 CeUNC-129 CeDAF-7 MmBmp3 MmGdfl0 MmInhbb MmInhbb MmInhbb MmInhbb MmLefty1 MmLefty2 DmDavdle DmMaverick DmMyoglianin MmGdfl1	467 421 222 193 193 191 150 181 150 181 154 204 216 163 154 165 165 395 520 463 269 520	
DmMaverick DmMyoglianin MmGdfll MmGdfll MmTgfb1 MmTgfb3 CeUNC-129 CeDAF-7 MmBmp3 MmGdfl0 MmInhbb MmInhbb MmInhbb MmInhbb MmLefty1 MmLefty2 DmDavdle DmMaverick DmMyoglianin MmGdfl1 MmStn cemp2 2	467 421 222 193 84 191 156 163 154 165 165 395 520 463 269 240	β6α4
DmMaverick DmMyoglianin MmGdfll MmGdfll MmTgfb1 MmTgfb3 CeUNC-129 CeDAF-7 MmBmp3 MmGdfl0 MmInhba MmInhbb MmInhbb MmLefty1 MmLefty1 DmDavdle DmMaverick DmMaverick DmMyoglianin MmGfl1 MmMstn CeTIG-3	467 421 222 193 84 187 218 191 156 170 181 156 165 395 520 463 395 520 463 240 135	β6a4
DmMaverick DmMyoglianin MmGdfll MmGdfll MmTgfb1 MmTgfb3 CeUNC-129 CeDAF-7 MmBmp3 MmGdfl0 MmInhbb MmInhbb MmInhbb MmInhbb MmLefty1 MmLefty2 DmDavdle DmMaverick DmMyoglianin MmGdfl1 MmGdfl1 MmTgfb1 MmTgfb1	467 421 222 193 84 187 218 84 191 156 170 0 181 185 204 216 395 520 463 269 240 240 240 240 240 240 240 240	β6α4
DmMaverick DmMyoglianin MmGdfll MmGdfll MmTgfb1 MmTgfb3 CeUNC-129 CeDAF-7 MmBmp3 MmGdfl0 MmInhbb MmInhbb MmInhbb MmInhbb MmLefty1 MmLefty2 DmDavdle DmMaverick DmMyoglianin MmGdfl1 MmTgfb1 MmTgfb1 MmTgfb2	467 421 222 193 84 187 218 191 156 170 181 156 163 154 746 165 395 520 463 269 240 135 269 241	β6α4
DmMaverick DmMyoglianin MmGdfll MmGdfll MmTgfb1 MmTgfb2 MmTgfb3 CeUNC-129 CeDAF-7 MmBmp3 MmGdfl0 MmInhbb MmInhbb MmInhbb DmActivin-beta MmLefty1 MmLefty1 DmDawdle DmMaverick DmMyoglianin MmGfl1 MmMstn CeTIG-3 MmTgfb2 MmTgfb2 MmTgfb2 MmTgfb2 MmTgfb2 MmTgfb2 MmTgfb2 MmTgfb2 MmTgfb2 MmTgfb2 MmTgfb3	467 421 222 193 84 187 218 84 191 156 160 163 154 165 165 165 165 202 463 240 135 222 240 135 222 240 135 240 240 125 240 240 125 240 240 165 165 240 240 165 240 240 165 240 240 165 240 240 165 240 240 165 240 240 165 240 240 240 240 240 240 240 240 240 240	βδα4
DmMaverick DmMyoglianin MmGdfll MmGdfll MmTgfb1 MmTgfb3 CeUNC-129 CeDAF-7 MmBmg3 MmGdfl0 MmInhbb MmInhbb MmInhbb MmInhbb MmLefty1 MmLefty2 DmDawdle DmMaverick DmMyoglianin MmGdfl1 MmGdfl1 MmTgfb1 MmTgfb3 CeUNC-129 CeDAF-7	467 421 222 193 84 191 156 63 395 204 463 269 240 135 232 269 241 191 191	β6α4
DmMaverick DmMyoglianin MmGdfll MmGdfll MmTgfb1 MmTgfb2 CeDXF-7 MmBmp3 MmGdfl0 MmInhbb MmInhbb MmInhbb MmInhbb MmLefty1 MmLefty2 DmDavdle DmMaverick DmMyoglianin MmGdfl1 MmTgfb1 MmTgfb1 MmTgfb3 CeUXC-129 CeDXF-7 MmTgfb3 CeUXC-129 CeDXF-7	467 421 222 183 84 187 218 191 156 170 181 185 204 216 165 1395 240 154 269 240 135 269 240 135 269 241 191 212 269 241 191 212 269 241 212 269 241 212 269 241 212 269 241 212 269 241 222 269 241 222 269 241 222 269 241 222 269 241 222 269 241 222 269 240 225 269 240 225 269 240 225 269 240 269 269 269 269 269 269 269 269 269 269	β6a4
DmMaverick DmMyoglianin MmGdfll MmGdfll MmTgfb1 MmTgfb2 MmTgfb3 CeUNC-129 CeDAF-7 MmBmp3 MmGdfl0 MmInhbb MmInhbb MmInhbb MmInhbb DmActivin-beta MmLefty1 MmLefty2 DmDawdle DmMaverick DmMyoglianin MmGfl1 MmTgfb2 MmTgfb2 MmTgfb2 MmTgfb3 CeUNC-129 CeDAF-7 MmBmp3 MmGdf10	467 421 193 84 191 156 150 181 150 181 152 165 165 165 165 165 232 240 135 240 135 240 135 240 240 135 240 240 135 240 240 135 240 240 125 240 240 125 240 240 125 240 240 125 240 240 240 240 240 240 240 240 240 240	β6a4
DmMaverick DmMyoglianin MmGdfll MmGdfll MmTgfb1 MmTgfb3 CeUNC-129 CeDAF-7 MmBmg3 MmGdfl0 MmInhbb MmInhbb MmInhbb MmLefty1 MmLefty2 DmDawdle DmMaverick DmMyoglianin MmGdfl1 MmTgfb1 MmTgfb3 CeUNC-129 CeDAF-7 MmBmg3 MmGdf10 MmTgfb3 MmTgfb3 CeUNC-129 CeDAF-7 MmBmg3 MmGdf10 MmTghba	467 421 2193 84 191 185 204 216 170 181 185 204 216 3154 165 165 165 165 165 204 240 240 232 269 240 191 191 223 224 191 191 223 232 241	β5β6α4
DmMaverick DmMyoglianin MmGdfll MmGdfll MmTgfb1 MmTgfb2 MmTgfb3 CeUNC-129 CeDAF-7 MmBmp3 MmGdfl0 MmInhba MmInhba MmLefty1 MmLefty2 DmDawdle DmMaverick DmMyoglianin MmGfl1 MmTgfb2 MmTgfb3 CeUNC-129 CeDAF-7 MmBmp3 MmTgfb2 MmTgfb3 Mm	467 421 193 84 191 185 204 218 191 185 165 165 165 165 165 204 240 135 520 463 240 135 202 240 135 202 240 240 233 254 26 26 26	β5β6α4
DmMaverick DmMyoglianin MmGdfll MmGdfll MmTgfb1 MmTgfb2 MmTgfb3 CeUNC-129 CeDAF-7 MmBmp3 MmGdfl0 MmInhbb MmInhbb MmInhbb MmInhbb DmActivin-beta MmLefty1 MmLefty2 DmDawdle DmMaverick DmMaverick DmMyoglianin MmGfl1 MmTgfb1 MmTgfb2 MmTgfb2 MmTgfb2 MmTgfb2 MmTgfb3 CeUNC-129 CeDAF-7 MmBmp3 MmGdf10 MmInhba MmInhba MmInhba	467 4211 193 84 191 156 170 185 204 218 156 163 154 165 165 165 165 165 165 204 204 229 240 135 232 241 191 121 222 241 191 223 2254 2254 2254 2254 2254 2254 2254	β5β6α4
DmMaverick DmMyoglianin MmGdfll MmGdfll MmTgfb1 MmTgfb3 CeUNC-129 CeDAF-7 MmBmg3 MmGdfl0 MmInhbb MmInhbb MmInhbb DmActivin-beta MmLefty1 MmLefty2 DmDawdle DmMaverick DmMyoglianin MmGdfl1 MmTgfb1 MmTgfb3 CeUNC-129 CeDAF-7 MmTgfb3 CeUNC-129 CeDAF-7 MmBmg3 MmGdfl0 MmInhbb MmInhbb MmInhbb MmInhbb	467 421 2193 84 181 191 185 204 2163 154 165 165 165 165 165 204 240 240 240 232 269 241 191 222 269 241 191 212 233 254 204 222 266 207 204	β5β6α4
DmMaverick DmMyoglianin MmGdfll MmGdfll MmTgfb1 MmTgfb2 MmTgfb3 CeUNC-129 CeDAF-7 MmBmp3 MmGdfl0 MmInhba MmInhbb MmInhbb DmActivin-beta MmLefty1 DmAaverick DmMyoglianin MmGfl1 MmTgfb1 MmTgfb2 MmTgfb1 MmTgfb2 MmTgfb3 CeUNC-129 CeDAF-7 MmBmg3 MmGdfl0 MmInhba MmInhbb MmInhbb MmInhbb MmInhbb MmInhbb	467 421 193 84 191 181 191 181 191 181 191 181 191 181 191 181 193 193 193 193 193 193 193 193 193 19	β5       -β6       -adtivin        β5       -β6       -adtivin        BFRSTPR      BFRSTPRSTPRSTPRSTPRSTPRSTPRSTPRSTPRSTPRSTP
DmMaverick DmMyoglianin MmGdfll MmGdfll MmTgfb1 MmTgfb3 CeUNC-129 CeDAF-7 MmBmp3 MmGdfl0 MmInhbb MmInhbb MmInhbb MmInhbb DmActivin-beta MmLefty1 MmLefty2 DmDawdle DmMaverick DmMaverick DmMyoglianin MmGdfl1 MmTgfb1 MmTgfb2 MmTgfb2 MmTgfb2 MmTgfb2 MmTgfb2 MmTgfb3 CeUNC-129 CeDAF-7 MmBmg3 MmGdf10 MmInhbb MmInhbb MmInhbb MmInhbb MmInhbc MmInhba	467 4212 193 84 191 185 204 218 191 185 204 165 165 165 165 165 165 165 204 204 204 204 204 204 204 204 204 204	β5β6α4
DmMaverick DmMyoglianin MmGdfll MmGdfll MmTgfb1 MmTgfb3 CeUNC-129 CeDAF-7 MmBmg3 MmGdfl0 MmInhbb MmInhbb MmInhbb DmActivin-beta MmLefty1 MmLefty2 DmDawdle DmMaverick DmMyoglianin MmGdfl1 MmTgfb1 MmTgfb3 CeUNC-129 CeDAF-7 MmTgfb3 CeUNC-129 CeDAF-7 MmTgfb3 CeUNC-129 CeDAF-7 MmBmg3 MmGdfl0 MmInhbb MmInhbb MmInhbb MmInhbb MmInhbb MmInhbc DmActivin-beta MmLefty1 MmLefty1 MmLefty2	467 421 193 84 191 185 204 218 191 185 204 216 3154 165 165 165 165 204 240 240 240 232 269 241 191 232 264 204 233 252 264 274 195 195 233 256 207 47 92 35 256 207 207 207 207 207 207 207 207 207 207	β5β6α4
DmMaverick DmMyoglianin MmGdfll MmGdfll MmTgfb1 MmTgfb3 CeUNC-129 CeDAF-7 MmBmp3 MmGdfl0 MmInhba MmInhbb MmInhbb DmActivin-beta MmLefty2 DmDawdle DmMaverick DmMyoglianin MmGfl1 MmTgfb1 MmTgfb3 MmTgfb1 MmTgfb3 MmTgfb1 MmTgfb3 MmTgfb1 MmTgfb3 MmTgfb1 MmTgfb3 MmTgf	467 421 193 84 191 181 191 181 191 181 191 181 191 181 191 181 193 194 216 163 194 216 204 216 216 216 216 216 216 216 216 216 216	β5       -β6       -activit        β5       -β6       -activit        BK      BK       -BK        BK      BK      BK        BK      BK      BK

Figure 6 Focused Activin + TGF- $\beta$  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-B1 (Shi et al. 2011). Underlined  $\alpha 2$  sequences of TGF- $\beta 1$  and Inhibinβa indicate distinct locations. Underlined β1 in TGF-β1 shows it is in a distinct location in Inhibin- $\beta a$ . Underline of  $\alpha 3$  in TGF- $\beta 1$ indicates absence of conservation. The notation TGF- $\beta$  with  $\beta$ 3 indicates this feature approximates the TGF-B1 pattern (underlined). The notation Activin with  $\beta$ 7 indicates this feature approximates the Activin pattern (underlined) not the underlined TGF- $\beta$ 1 pattern. In  $\beta$ 8 of the bowtie of TGF- $\beta$ 1–3, the conserved pair of cysteines (CxC) is underlined in black. In  $\beta 8$  Maverick and Dawdle are joined by Activin, Inhibin-Ba, and Inhibin-Bb, with a conserved pair of cysteines of distinct spacing (red underline; CxxC). Underlined  $\beta$ 9,  $\beta$ 9', and  $\alpha$ 5 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  $\sim$ 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 fulllength 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- $\beta$  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 ( $\alpha$ 1, Latency Lasso,  $\alpha$ 2, and  $\beta$ 1) display strong homogeneity.  $\alpha$ 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  $\alpha$ 2. An F/Y and the adjacent S/T in  $\beta$ 1 are moderately conserved.

At the 5' end of the arm strong conservation is visible.  $\beta 2$  contains near universal F/Y and I/L/V residues. Conservation is present between  $\beta 2$  and  $\beta 3$  with a near universal I/L/V and a modestly conserved proline.  $\beta 3$  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.  $\beta 4$  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.  $\beta 5$  is only moderately conserved with one near universal I/L/V. The  $\beta 6$  to  $\alpha 4$  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  $\alpha 4$ .

At the 3' end of the arm,  $\beta$ 7 has three moderately conserved I/L/V residues.  $\beta$ 8 and  $\beta$ 9 (the distinctive bowtie of TGF- $\beta$ 1) are not conserved.  $\beta$ 9' is a unique BMP feature containing a moderately conserved arginine and an I/L/V.  $\beta$ 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

strait	jacket	α1	Lat	encyLasso		α2	β1
DmSCW	30	IYQKRPLSEQMEM	IDILDLGDR	PRRQAEP	-NLHNSASI	KFLLEVYNEIS	E • ASCNSILIESS
DmGBB	56	LSEDDKLDVSYEI	LEFLGIAER	PTHLSSHQL	-SLRKSAP	KFLLDVYHRIT.	A • DESDIIMTELN
MmBmp8a	35	LGARERRDMORE	I AVLGI PGR	RPRAOPAA	-ROPASAPI	LEMIDLYHAMT	D • GRADI VMSFVN
MmBmp8b	35	LOVREPROMOREI	REVICIPCE	PRSRAPUCA	-OOPASAPI	FMT DT YRAMT	D. DRADI IMSEVN
MmBmp6	60	I VTUEVDEMOVE	SVI CI DUD	PDT UCT OO	DCDIVCAD	I PMI DI VNAL SI	N • NDADMUMSEUN
MmDmmE	4.2	IDNUEDDETODE	CTLCLDUD	DDECDC	KORGAN		
MINBINDS	43	LENHERREIQREI	I SILGLPHR	RPFSPG==	-KQASSAPI	LEMEDLYNAGA	S• NDADOVMSTVN
MmBmp/	46	LRSQERREMQREI	I SILGLPHR	RPHLQG	-KH-NSAPI	MEMLOLYNAMA	V • TDADWVMSEVN
CeDBL-1	42	ADQHASHATRRG	LRKLGLEHV	PVQT-	-GPSIDVP	QHMWDIYDDDN	DWVRHYYP
MmBmp10	50	DFNTLLQSMKNEF	KTLNLSDI	PVQD-	-TGRVDPP	EYMLELYNKFA	T•PSANIIRSEKN
MmGdf2	53	DLQMFLENMKVDF	RSENESGI	PSQD-	-KTRAEPP	QYMIDLYNRYT	T • PASNIVRSFSV
CeTIG-2	38	QATDKIGEQ	RELFNIDIN	PNGPAV-	-KANNYVS	TYMKRLYKOLE	N • LSADRIVSHMA
MmGdnf	61	PAED	-HSLGHRRV	FAL-	-TSDSNMP	EDYPROFIDDWM	D
DmDPP	213	PDPSTLVETEKS	SLENKRP	BKTD-	-RSKTTTP	EDWKKTYAFUM	C . KSANTWRSPTH
MmBmp2	41	DCEDVI CEEEI DI	CMECTKOP		-KD-WWWD	DYMINI VPDUC	C. CDANTUDGEUU
MINDINDZ	41	PSEDVLSEFELKL	SMEGLAQR		-KD-VVP	PIMEDLIRRHS	G• SRANTVRSTHH
MmBmp4	43	QSHELLRDFEATL	M M F G L K K K	QPS-	-KS-AVIPI	DYMRDLYRLQS	G• SRANTVRSFHH
MmBmp15	35	PSILDLAKEA	-PG	KEMKQ-	•-WPQGYPL	RYMERLYBRSA	D.IGAKOVRLVKP
MmGdf9	55	PPLFKVLSDR	-RG	ETPKL-	-QPDSRAL	YYMKKLYKTYA	T • HLYNTVRLESP
MmGdf1	25	PAAALLQV	LGLPEA	PRSVPTHR-		-VMWRLFRRRD	P• VAGNIVRHIPD
MmGDF3	20	ODSDLLOF	LGLEKA	SP-HRFO-	-PVPR	-VIRKIIRARE	A. VRGNILLOLLPD
MmNodal	13	-ACWALLHPR	-APTAAALP	LWTRGO-	-PSSPSPL	AVMEST YRDPL	PDITRELOA
MmCdf5	124	DCCELIKKTD	-FD-CTDDF	KEDED-	-DDDTTDU	TYMT OF YRTE	D. CLANTUTERID
Magalec	124	ESSEDBRAIK	OUL OCODD	REFER	- FFFTTFH	BUNK OF WERE	T KONIMUTOPUD
MIGGLO	01		-QH-QGQEP	GK-	-GLRVVPH	CHOLSHIN 115	
MmGdi /	65		-PR-AVRRA	AGSGFR-	-NGSVVPH	HEMBSLYRSLA	G•GRVDTHTGETD
	arm	-B2		B3	0	.3	64
DmSCW	131.	MUTTENTND	PUDT SHUO		CT VDP-	DANET	SWARD KEDNRO-
DIIISCW	170-	DDIWEDRON	VE VDLSEVQ	ANDATINOF	SEVDK-	NDDDDD	SVIRRDNRQ-
DWGBB	1/0.	RRLWEDVSN	VENDNYEVM	AELRIYQNA	NEGKWLTA-	NREFT	TVYALGTGTLG
MmBmp8a	123.	KEFHEDDUQ	I PAGBAVIA	AEFRIYKEP	STHPL	NTTLH	SMFEVVQEHSN
MmBmp8b	123.	KEFHEDLEQ	IPAGEAVTA	AEFRIYKEP	STHPL	NTTLH	SMFEVVQEHSN
MmBmp6	233.	KEFKFNLSQ	IPEGEAVTA	AEFRVYKDC	VVGSFK-	NQTFL	SIYOVLQEHQH
MmBmp5	178.	KEFREDI 0	TPHGEAVTA	AEFRIYKDK	GNHRFE-	NETIK	STYOTTKEYTN
MmBmp7	153.	PEEPEDI SK	TPECEAWITA	AFFRIVEDY	TREPE-D-	NETEOT	TWYOVIOFUSC
Pinibinp /	100-	KETKPULSK	LGLAVIA		IKEKED-	NEIFQI	IVIQVIQEN3G
CeDBL-1	106.	-LLSMNLSLAARI	NAHNDEVIK	ATLKLRLRR	NNK	ARRSGN	SIYFFEDD
MmBmp10	126.	YPLL NV-S	IPHHEEVVM	AELRLYTLV	QRDRMMY	DGVDRK	TIFEVLESADG
MmGdf2	130.	HILINN-S	IPRHEQITR	AELRLYVSC	QNDVDST	HGLEGS	VVYDVLEDSET
CeTIG-2	120.	YSIR AKEH	VPAKEGOSI	VRAQ		LR	HIQGIVSPVFF
DmDPP	303.	FRI.H.DWKS	PADEKIKA	AFTOTTEDA	LSOOVVASE	SSANRTRYO	LVYDTTRVGV-
MmDmm 2	120.	DDEEDNI GC	DODEFIC	AFTOTERFO	TOFATCNER	ECUD	NITVETTKDAA-
Philip 2	129.	RRFFFNL55	VPSDEFLIS	AELQIEREQ	TQEALGNSS	SrQnk	NTIETIKPAA-
MmBmp4	139.	FRETENTSS	IPENEVISS	AELRLFREQ	VDQGPDWEG	2GFHR	NIYEVMKPPA-
MmBmp15	113.	QTLDFPLASI	NQVAYELIR	ATVVYREQL	HL-VNYH	I	SCHVETWV
MmGdf9	138.	VDLLENLDR	TAMEHILK	SVLLYTLNN	SASSSST	V	TCMCDLVV
MmGdf1	117.	WTVVEDISN	EPTERPER	ARLEIRLEA	ESEDTGGW-	EL	SVAL
MmGDF3	108.	KVLYENISA	KEKAKUTM	AOTTIDLCP	RSYYNLRP-	ELVVA	SWWODRGV
MmNodel	74-	MIDEMEDICE	COPROLUM	ARTRIOTOC	DMDTDMEC	DI m	D TEUONKC
MININOGAL	14.	WIFITDESE	SQLADIVW	ALLR QLPG	PMDIPIEG-	PLI	D FHQAKG
MmGdf5	214.	QRYVEDIISAI	-EKDGLLG	AELRILRKK	PLDVAKPA-	VPSS	GRWAQLKL
MmGdf6	149.	QKYLEDVST	SDKEELVG	AELRLYRQA	PPTPW	GLPA	RPHHLQ-L
MmGdf7	122.						III
THUOUT /	TOT	QSFLEDVSS	LSEADEVVN	AELRVLRRR	SPEPD	RDSA	
Thuộc 1	152	QSFLEDVSS	SEADEVVN	AELRVLRRR	SPEPD	RDSA	
THROUL /	152	QSFLEDVSS	ISEADEVVN	AELRVLRRR β6	SPEPD	SA	β7
DmSCW	132 	QSFLEDVSS β5 YRIEGSV	SEADEVVN	AELRVLRRR β6 SQRGWLEFN	SPEPD α4- I <b>T</b> DTLF	RDSA	β7 RRNETRISI-G
DmSCW DmGBB	179•'S 228•T	·QSFLEDVSS β5 :YRIIGS MEPLSS	VNTTS: /NTTS	AELRVLRRR β6 SQRGWLEEN DYVGWLELN	SPEPD α4- LTDTLF VTEGLH	RDSA RYWLHNKGLQ• HEWLVKSKDN•	β7 RRNERRSI-G
DmSCW DmGBB MmBmp8a	179•'S 228•T 172•D	·QSFLEDVSS β5 YRLEGS MEPESS LEFEDI(	SEADEVVN	AELRVLRRR β6 SQRGWLEEN DYVGWLELN GDEGWLVLD	SPEPD α4 ITDTLF VTEGLH TTAASE	RDSA	β7 RRNETRISI-G HGIYIGAHAVNR-P LG RIYWETA-D
DmSCW DmGBB MmBmp8a	179•'S 228•T 172•E	QSFLEDVSS β5	SEADEVVN	AELRVLRRR β6 SQRGWLEIN DYVGWLELN GDEGWLVLD GDEGWLVLD	SPEPD Δ4- Δ Δ Δ Δ Δ Δ Δ Δ Δ Δ Δ Δ Δ Δ Δ Δ	RDSA RYWLHNKGLQ• HEWLVKSKDN• RWLLNHHKD•	β7 RRNELRISI-G HGIYIGAHAVNR-P LGLRIYVETA-D
DmSCW DmGBB MmBmp8a MmBmp8b	179•'S 228•T 172•D 172•D	QSFLEDVSS YRTLGS MEPLSS LFFLDL( LFFLDL(	SEADEVVN /NTTS: /NTTG 2TLRS( 2TLRS(	AELRVLRRR β6 SQRGWLEFN DYVGWLELN GDEGWLVLD GDEGWLVLD	SPEPD α4- ITDTLF VTEGLH ITAASE ITAASE ITAASE	RYWLHNKGLQ• HEWLVKSKDN• RWLLNHHKD• RWLLNHHKD•	β7 RRNETRISI-G HGYIGAHAVNR-P LGRTYVETA-D LGRTYVETE-D
DmSCW DmGBB MmBmp8a MmBmp8b MmBmp6	179 · S 228 · T 172 · D 172 · D 283 · D	QSFLEDVSS YRTIGS MEPLSS LFFLDL LFFLDL LFFLDL	VNTTS: VNTTGI 2TTRS( 2TTRS( RVVWA:	$\begin{array}{c} \textbf{AELRVLRRR} \\ \hline \textbf{\beta6} \\ \textbf{SQRGWLEEN} \\ \textbf{SQRGWLEEN} \\ \textbf{SDEGWLVLD} \\ \textbf{SDEGWLVLD} \\ \textbf{SEEGWLEFD} \end{array}$	SPEPD α4- TDTLF VTEGLH TTAASE TTAASE TTATSN	RYWLHNKGLQ • HEWLVKSKDN • RWLLNHHKD • RWLLNHHKD • HWLLNHHKD • HIWVVTPQHN •	RNETRISI-G RRNETRISI-G HGYIGAHAVNR-P LGRTYVETA-D LGRTYVETE-D MGLOISVVTR-D
DmSCW DmGBB MmBmp8a MmBmp8b MmBmp6 MmBmp5	179 • S 228 • T 172 • D 172 • D 283 • D 228 • D	QSFLEDVSS YRTIGS MEPLSS ULFFLDL ULFFLDL ULFLLDT FLLDT	SEADEWVN VNTTS: VNTTGI 2TTRS( 2TLRS( RVVWA: RKTQA:	AELRVLRRR β6 SQRCWLEIN SQRCWLEIN GDECWLVLD GDECWLVLD SEEGWLEFD LDVCWLVFD	SPEPD Δ TDTLF VTEGLH TTAASI TTAASI TTATSN TTVTSN	RDSA RYMLHNKGLQ• HEMLVKSKDN• RWLLNHHKD• RWLLNHHKD• JLWVVTPQHN• HWVINPQNN•	
DmSCW DmGBB MmBmp8a MmBmp8b MmBmp6 MmBmp5 MmBmp7	179 • S 228 • T 172 • D 172 • D 283 • D 228 • D 203 • D	QSFLEDVSS YRTLGS MEPLSS ULFFLDL ULFFLDL LFFLDT LFLLDT FLLDS	SEADE VN /NTTG /NTTG 2TIRS 2TIRS RVVWA RKTQA RTTWA	AELRVLRRR β6 SQRGWLEFN DYVGWLEIN GDEGWLVLD GDEGWLVLD SEEGWLVFD SEEGWLVFD SEEGWLVFD	SPEPD α4- TDTLF VTEGLH TAASE TAASE TATSN TVTSN TATSN TATSN	RDSA RD-SA RWLLNKSLQ• RWLLNHHKD• RWLLNHHKD• LLNVVTPQHN• RWVINPQNN•	$\beta7$ RRNE R SI-G HG Y GAHAVNR-P LG RY VETA-D LG RY VETE-D MG COS VTR-D LG COS VTR-D LG COS VETL-D
DmSCW DmGBB MmBmp8a MmBmp8b MmBmp6 MmBmp7 CeDBL-1	179 · S 228 · T 172 · D 283 · D 228 · D 228 · D 203 · D 153 · -	QSFLEDVSS YRIDGSV MEPLSSV LFFLDLC LFFLDTF LFFLDTF LFFLDSF CESSF	SEADE VN NTTS: NTTG 2TIRS( 2TIRS( RTWA RTWA RSVD	AELR VLRRR β6 SQRGWLEIN GDEGWLVID GDEGWLVID SEEGWLEFD LDVGWLVFD SEEGWLVFD NLTEWIDFD	SPEPD TDTLF VTEGLH TTAASI TTAASI TTATSN TVTSN TATSN TATSN VTAAFI	RDSA WILHNKGLQ • EMLVKSKDN • WILNHHKD • EWLLNHHKD • IWVTPQHN • HWVINPQNN • HWVINPQNN • HWVINPRHN • ERTNRISFF •	$\begin{array}{c}\beta^{7}\\ RNEPRISI-G\\ HGYIGAHAVNR-P\\ LGIRPYVETA-D\\ LGIRPYVETE-D\\ MGLOSSVTR-D\\ LGICPCAETG-D\\ LGICPCAETL-D\\ IDEPEDVEIEETOS$
DmSCW DmGBB MmBmp8a MmBmp8b MmBmp6 MmBmp7 CeDBL-1 MmBmp10	179 · S 228 · T 172 · D 283 · D 228 · D 228 · D 203 · D 153 · - 177 · M	QSFLEDV&S YRTEGS YRTEGS VMEPUSS UFFLDL UFFLDDTF UFFLDDSF UFFLDDSF FODSSF	SEA EVVN /NTTS: /NTRS( 2TIRS( 2TIRS( 2TIQA) RTQA RTQA RTVA RSVD	AELR ULRRR β6 SQRCWLEIN GDE CWLVLD GDE CWLVLD SEECWLVED SEECWLVFD SEECWLVFD SEECWLVFD NLTEWLDFD TNSEWETED	SPEPDα4-           ITDTLF           VTEGLF           ITAASI           ITAASI           ITATSN           ITVTSN           ITATSN           VTAAFI           VTAAFI           VTDATFI	RDSA RYMLHNKGLQ HEWLVKSKDN FWLLNHHKD SWLLNHHKD HWVVTPQHN HWVVNPQHN RTNRISFF KWOKSGPST	$\beta^{7}$ RRNEPRIS HGYYGAHAVNR-P LGIRYYGTA-D LGIRYYGTA-D LGIQUSVTR-D LGIQUSVTR-D LGIQUSVETL-D LGIQUSVETL-D LGIQUSVETL-D
DmSCW DmGBB MmBmp8a MmBmp8b MmBmp6 MmBmp7 CeDBL-1 MmBmp10 MmCdf2	179 · S 228 · T 172 · D 172 · D 283 · D 228 · D 203 · D 153 · - 177 · M	QSFLEDV&S 	SEALEVYN           /NTTS:           /NTRS:           2TTRS:           2TT	AELRULERR β6 SQRGWLEEN DYVGWLELN GDEGWLVLD GDEGWLVLD SEEGWLVFD SEEGWLVFD NLTEGIDFD INSEGGWTVFD INSEGGWTVFD	SPEPDα4- ITDTLF VTEGLF TTAASI ITATSN ITATSN ITATSN VTAAFI VTDAFI VTDAFI	RDSA RYWLHNKGLQ EWLLNHKD WLLNHHKD WLLNHHKD WVTPQHN HWVNPRHN RTNRISFF SWQKSGPST	$\begin{array}{c}\beta^{7}\\ RNEPRISI-G\\ HGYIGAHAVNR-P\\ LGRIYVETA-D\\ LGRIYVETR-D\\ MGIODSVTR-D\\ LGIQUCAETG-D\\ LGIQUCAETG-D\\ IDPPEOVEIEETQS\\ HQUEHHESRQNQA\\ NENERAL SALANDA$
DmSCW DmGBB MmBmp8a MmBmp8b MmBmp6 MmBmp7 CeDBL-1 MmBmp10 MmGdf2	132 179 ·S 228 ·T 172 ·D 172 ·D 283 ·D 203 ·D 153 ·- 177 ·M 181 ·T	QSFLEDV&S YRINGS YRINGS VEFDDLC LFFDDLC LFFDDLF LFFDDTF LFFDDSF LFFDSSF -FQDSS -FQUSS WTFUVS WTFUVS	SEA EUVN INT TGI INT RSI IT RSI IT RSI IT RAI IT VAI IT VAI IT VAI IT VGI IT VGI IT VGI IT VGI IT VGI IT VGI IT VGI IT	AELRVIERR β6 SQRCWIEIN DYVGWIEIN GDECWIVID GDECWIVID SEECWIFD LDVCWIVFD SEECWIFD NITEWIFT INTEWIFT RDECWETID	SPEPD	RYWLHNKGLQ • RYWLHNKGLQ • FWLLNHHKD FWLLNHHKD HWUVYPQHN HWVVNPRHN RRTNRISFF RRTNRISFF RRVQKSGPST GWVKADSTT	
DmSCW DmGBB MmBmp8a MmBmp8b MmBmp6 MmBmp7 CeDBL-1 MmBmp10 MmGdf2 CeTIG-2	132 179 ·S 228 ·T 172 · D 228 · D 228 · D 228 · D 238 · D 153 · - 177 · M 181 · T 156 · -	QSFLEDV@S           β5           YRTIGS           MEPUSS           LIFFUDL           LIFFUDL           LIFFUDL           LIFFUDL           LIFFUDL           LIFFUDL           FODS           -FODS           -FODS           -FOUS           -FOUS           -FOUS           -FOUS	SEALEVYN           /NTTS:           /NTTS:           /NTTS:           /NTRS:           2TIRS:           2TIRS:           2TIRS:           RTQA:           RTQA:           RTVA:           RTVA:           RTVA:           RTVA:           RTVA:           RSVD:           PEE          YG:           2DIS:	AELRVIERR β6 SQRCWIEIN GDEGWIVID SEECWIVID SEECWIVID SEECWIVFD NITEWIDFD NITEWIFT ROECWIVFD RDECWIVFD	SPEPD	RYMLHNKGLQ- HEWLVKSKDN RWLLNHHKD BWLLNHHKD BWLLNHHKD HWVINPQHN HWVVNPRHN RTNRISFF RWVKADSTT RWVKADSTT LWSHLQLST	
DmSCW DmGBB MmBmp8a MmBmp6 MmBmp7 CeDBL-1 MmBmp10 MmGdf2 CeTIG-2 DmDPP	132 179 ·S 228 ·T 172 · D 228 · D 228 · D 228 · D 238 · D 153 · - 177 · M 181 · T 156 · - 359 ·S	QSFLEDV&S YRILGS YRILGS VAMEDSS ULFIDDI ULFIDDT FUDS FUDS FUDS KTTEVS YLFUDT	SEA         EVVN           /NT TGI         ////////////////////////////////////	$\begin{array}{c} \textbf{AELR} & \textbf{ILRR} \\ & -\beta 6 - \\ \text{SQRCM} & \textbf{IEN} \\ \text{GDE W I VID} \\ \text{GDE W I VID} \\ \text{GDE W I VID} \\ \text{SEE W I VD} \\ \text{LDV W I VFD} \\ \text{SEE W I FD} \\ \text{NLTEW I DFD} \\ \text{NLTEW I DFD} \\ \text{NLTEW I VFD} \\ \text{NSDPT VTD} \\ \text{VSID} \\ $	SPEPD	HYMLHNKGLQ HEMLVKSKDN FWLLNHHKD FWLLNHHKD HWUVTPQHN HWVVNPRHN HWVVNPRHN HWVVNPRHN RTNRISFF RTNRISFF RTNRISFF RTNRISFF RWQKSGPST GWVKADSTT FWLASPQN	
DmSCW DmGBB MmBmp8a MmBmp6 MmBmp6 MmBmp7 CeDBL-1 MmBmp10 MmGdf2 CeTIG-2 DmDPP MmBmp2	132 179 · S 228 · T 172 · D 172 · D 228 · D 228 · D 228 · D 153 · - 153 · - 156 · - 359 · S 180 · V	QSFLEDV&S           β5           YRTIGS           MEPUSS           LFFUDI           YS           LFFUDI           LFFUDI           TRUD	SEARE VN           VNTTS:           VNTRS:           2TTRS:           2TTRS:           XKTQA:           XTTWA:           XSVD:           VEYG:           2DTYG:           2DTYG:           2DTYG:           2DTRL:           XTVRL:           XLVNQ:	$\begin{array}{c}\beta 6 - \\\beta 6 - \\ SQRCW LEN \\ SQRCW LEN \\ SQRCW LEN \\ SGRCW LEN \\ SEC W LV D \\ SEC W LV \\ SEC W \\$	SPEPD	YWLHNKGLQ EEGLVKSKDN EWLLNHHKD EWLLNHHKD EWLLNHHKD ILWVVTPQHN EWVINPQNN EWVNPQNN ERTNRISFF EWQKSGPST EWQKAGST EWLSEPQRN EWLSSPQRN	
DmSCW DmGBB MmBmp8a MmBmp8b MmBmp5 MmBmp7 CeDBL-1 MmBmp10 MmGdf2 CeTIG-2 DmDPP MmBmp2 MmBmp4	132 179 · S 228 · T 172 · D 283 · D 283 · D 283 · D 283 · D 283 · D 153 · - 177 · M 181 · T 156 · - 359 · S 180 · V 190 · I	QSFL         DVS	SEAREWYN           VNTTS:           VITRS:           2TIRS:           2TIRS:           XVUWA:           RSVDI           TEIYG:           VDI           VDI           VIT           XTS:           XTVRLI           XVNQI           XIVNQI	$\begin{array}{c}\beta 6 - \\\beta 6 - \\ SQRCWLEIN\\ SQRCWLEIN\\ GDECWLVID\\ GDECWLVID\\ GDECWLVID\\ SEECWLVID\\ SEECWLVD\\ SEECWLVFD\\ NLTEWIDFD\\ NLTEWIDFD\\ SEECWLVFD\\ NLTEWIJFD\\ SDDPTVFD\\ SDDPTFFD\\ SDDPTFFFD\\ SDDFFFFD\\ SDDFFFFFD\\ SDDFFFFFFFFD\\ SDDFFFFFFFFFF$	SPEPD	YWLHNKGLQ HEWLVKSKDN FWLLNHHKD HEWLVHKRD HEWLVHPQIN HEWVVPQIN HEWVVPRHN HEWVVPRHN HEWVVPRHN HEWVRGSGPST GWVRADSTT GWVRADST FWLASPQRN HEWTTQGHTN HEWTTQGHTN	β7 RRNER SI-G HGY GAHAVNR-P LGRRYVETA-D LGROVETD MGIODSVTR-D LGIOCAETG-D IDPPEDVEIEETQS HQUEHHESRQNQA NKNKETVQS LFVTRAST YGLNEVRTVRSLK HGFVTEVAHLEENP YGGAEVTHLHQTR
DmSCW DmGBB MmBmp8a MmBmp6 MmBmp6 MmBmp7 CeDBL-1 MmBmp10 MmGdf2 CeTIG-2 DmDPP MmBmp2 MmBmp4 MmBmp15	132 179 · S 228 · T 172 · D 283 · D 283 · D 228 · D 233 · D 153 · - 177 · M 181 · T 156 · - 359 · S 180 · V 190 · F 155 · F	QSFLEDV&S 	SEAREWYN           NTTS:           NTRS:           TTRS:           TTRS:           TTRS:           TTRS:           TTRS:           TTRS:           TTRS:           TTRS:           REVG           DDSP           TTSI           TTSI           TTSI           TTSI           RLVNQI           RLVSI           RLVSI           SSI	AELR (LERR           -β6           SOM LEIN           DYV WIELN           GDE WIVID           SEE WIVED           SEE WIVED           LDV WIVFD           SEE WIVFD           NTEWITS           DPT VTD           NTSONESFD           NTSONESFD           WYTRWETFD           WSK WTEID	SPEPD	YYWLHNKGLQ HEWLVKSKDN EWLLNHKD EWLLNHKD HWUNPQNN HWVNPQNN HWVNPQNN RTWNISFF GWVRADST	
DmSCW DmGBB MmBmp8a MmBmp8b MmBmp5 MmBmp7 CeDBL-1 MmBmp10 CeTIG-2 DmDPP MmBmp12 MmBmp14 MmBmp14 MmBmp15	132 179 ·S 228 ·T 172 · D 228 · D 228 · D 228 · D 228 · D 228 · D 238 · D 238 · D 153 · - 177 · M 181 · T 156 · - 359 ·S 180 ·V 190 · I 155 · F 181 · F	QSFLEDV&S           β5           WEPUSS           MEPUSS           ULFFUDL           DLFFUDL           DLFFUDL           POLES           VEFUDT           VEUT           VET           VET           VET           VET           VET           VET           VET           VET	SEAREWVN /NT TS; /NT RS; 2TI RS; 2TI RS; 2TI WA; 3TT WA; 3TT WA; 3TT WA; 3TT WA; 3TT S; 3TT S; 3TT S; 3TT S; 3TT S; 3TT S; 3TT S; 3TT S; 3TT S; 3TT	AELS (LERR - 66 SC (M EEN VV (W EEN SDE (W VID SEE (W VID SEE (W VID SEE (W VID SEE (W VID NITE (D FD NITE (SEF D NITE (SEF D NVTRWET FD MSKWTEID	SPEPD	YWLINKGLQ HENLVKSKDN EWLLNHRD EWLLNHRD HEWVYPQHN HEWVYPQHN HEWVYPQHN HEWYNPGNN BRTNRISFF BROKSGPST CWVRADST CWVRADST CWVRADST CWVRADST HEWSHLQLST FRUSHLQLST FRUSHLQLST FRUSHLQLST FRUSHLQLST FRUSHLQLST FRUSHCASPQRN HEWTREKQPN FUTSSERSI	
DmSCW DmGBB MmBmp8a MmBmp8b MmBmp5 MmBmp5 MmBmp7 MmBmp10 MmGdf2 CeTIG-2 DmDPP MmBmp2 MmBmp4 MmBmp15 MmGdf9	132 179 - S 228 - T 172 - D 283 - D 203 - D 153 177 - M 181 - T 156 - 359 - S 180 - V 190 - T 155 - P 181 - E	QSFL DV&S 	SEAREWYN           NTTS:           NTRS:           2TIRS:           2TIRS:           2TIRS:           2TIRS:           2TIRS:           2TIRS:           2TIRS:           2TI		SPEPD         α4           DTLF         VPEGLM           VPEGLM         A           TAASI         TA           TATSN         TA           TATSN         TA           VPAASI         TA           VPAASI         TA           VTATSN         TA           VTAASI         TA           VTAASI         TA           VTAASI         TA           VTAASI         TA           VTAASI         TA           VSAVI         VPAVI           VPAVI         VPAVI           VPAVI         THCIQQKI           VSSLOPIN         THCIQQKI           VS-LOPIN         CTAN	HYMLHNKGLQ HEWLVKSKDN FWLLNHHKD FWLLNHHKD HWVTNPQHN HWVTNPQHN HWVTNPQHN HWVTNPQHN HWVTNPRHN HWTNPQHN HWTCGHTN FWLASPQRN HWTREKQPN -WNKRGS- - TSSERSI - MNKGRS-	
DmSCW DmGBB MmBmp8b MmBmp6 MmBmp6 MmBmp7 CeDBL-1 MmBmp10 MmBmp10 MmBmp2 MmBmp2 MmBmp4 MmBmp15 MmBdf1 MmGdf2	179 • S 228 • T 172 • D 228 • D 228 • D 228 • D 203 • D 153 • - 155 • F 181 • E 155 • F 181 • E 158 • G	QSFLEDV&S           β5           YRTTGS           MEPDSS           MEPDSS           ULFTUDI           ULFTUDI           FOIDS           FOIDS	SEARE VVN MTTS: /NTRS: /TTRS: XKTWA: XKTWA: XKTWA: XKTWA: XKTWA: XKTWA: XKTWA: XKTWA: XKTWA: XKT	ADLE LARRA 	SPEPD	YYWLHNKGLQ HEOLVKSKDN HEOLVKSKDN HEWLLNHHKD SWLLNHHKD HWVVTPQHN HWVVTPQHN HWVVPRHN HWVVPRHN HWVVPRHN HWVVPRHN HWVTQGHST HWVTQGHST HWTQGHST HWTQG	
DmSCW DmGBB MmBmp8a MmBmp6 MmBmp7 CeDBL-1 MmBmp10 MmGdf2 MmBmp4 MmBmp4 MmBmp4 MmBmp4 MmBmp4 MmBmp4 MmBmp15 MmGdf9 MmGdf9	132 179 · S 228 · T 172 · D 228 · D 228 · D 228 · D 228 · D 233 · D 233 · D 233 · D 233 · D 233 · D 233 · D 238 · D 258 · S 180 · V 190 · I 155 · F 181 · E 158 · G 157 · V	QSFL         DVS	SEAREWVN           /NTTG           /NTRS           2TIRS           2TIRS           2TIRS           2TIRS           2TIRS           2TIRS           2TIRS           2TIRS           2TIRS           2TI		SPEPD	YWLINKGLQ HENLVKSKDN EWLLNHKD ILWVVTPQHN ILWVVTPQHN ILWVVNPRHN HENVNPRHN HENVNPRHN HENVSGPST COVRADST COVRADST EWTLGHTN EWTLQGHTN EWTLQGHTN -TSSERSI- -TSSERSI- CAANASVPCT CONSNRLKN	
DmSCW DmGBB MmBmp8b MmBmp6 MmBmp6 MmBmp7 CeDBL-1 MmBmp10 MmBmp10 MmBmp12 MmBmp2 MmBmp4 MmBmp4 MmBmp4 MmBmp4 MmBmp4 MmBmf3 MmBmf3 MmBdf1 MmBmf3 MmBm64	132 179 - S 228 - T 172 - D 283 - D 228 - D 203 - D 153 177 - M 181 - T 156 359 - S 180 - V 190 - I 155 - F 181 - E 158 - G 157 - V 121 - F	QSFL         DVS	SEAME VVN NTTS: /NTRS: /NTRS: /VIWA XKTWA XKTWA XKTWA XKTWA XKTWA XKT		SPEPD	YYWLHNKGLQ HEWLVKSKDN FWLLNHHKD FWLLNHHKD HWVINPQNN HWVINPQNN HWVINPQNN HWVINPQNN HWVINPKIN FRTNRISFF RWQKSGPST HASSPQRN HWTTQGHTN HWTQGHTN HWTQ	
DmSCW DmGBB MmBmp8a MmBmp6 MmBmp6 MmBmp7 CeDBL-1 MmBmp7 CeDBL-1 MmBmp2 CeTIG-2 DmDPP MmGdf2 MmBmp4 MmBmp4 MmBmp4 MmBmp15 MmGdf9 MmGdf1 MmGdf1 MmGDF3	132 179 - S 228 - T 172 - D 228 - D 233 - D 235 157 - N 155 - F 181 - E 158 - G 157 - V 260 - F 260 - F	QSFL         DVS	SEARE VVN           NNTTS;           NNTTG;           TIRS;           VNWA;           XTWA;           XTVD;           PEYG;           QDIS;           XTVRL;           RLVNQ;           PRS;           XTVS;           XTVS;           XTV	ADLR IL RR 	SPEPD	YWLINKGLQ HENLVKSKDN EWLLNHHKD SWLLNHHKD SWLLNHHKD HUVVTPQHN HENVINPQNN BRTNRISFF RWQKSGPST GWUKSGPST GWUKSGPST GWUKSGPST GWUKSGPST FWLSSPQRN HUNTTQGHTN FWTREKQPN WNRKGRS TSSRSI KUD-RR- -KN-SA-	
DmSCW DmGBB MmBmp8b MmBmp6 MmBmp6 MmBmp7 CeDBL-1 MmBmp10 MmBmp10 MmBmp10 MmBmp15 MmBmp2 MmBmp4 MmBmp4 MmBmf3 MmGdf3 MmGdf3 MmGdf1 MmGdf5	132 179 · S 228 · T 172 · C 238 · C	QSFL         DV&S <	SEARE VVN NTTS: /NTRS: /NTRS: /VTWA: /TTWA: /TTWA: /TTWA: /TTWA: /TTWA: /TTWA: /TTWA: /TT		SPEPD (1) (1) DTLK (1) DTLK (1) DTLK (1) DTLK (1) DALK (1) D	YYMLHNKGLQ HEMLVKSKDN EWLLNHHKD EWLLNHHKD EWLLNHHKD HWVINPQNN HWVINPQNN HWVINPQNN HWVINPRIN RTNRISFF RQXSGPST HASPQRN HWTTQGHTN HWTQGHTN	
DmSCW DmGBB MmBmp8a MmBmp6 MmBmp6 MmBmp7 CeDBL-1 MmBmp7 CeDBL-1 MmBmp2 MmBmp2 MmBmp4 MmBmp4 MmBmp4 MmBdf1 MmGdf1 MmGdf5 MmGdf5 MmGdf6 MmGdf6	179 - S 228 - T 172 - D 228 - D 238 - D 238 - D 238 - D 238 - D 238 - D 258 - D 155 - F 181 - E 155 - F 181 - E 155 - F 181 - E 155 - F 260 - F 192 176 - T	QSFLEDV&S           BDV&S           YRTIGS           MEPUSS           LFFUDI	SEAREWYN           MTTS;           MTRS;           TTRS;           XKTWA;           XKT		SPEPD	YYMLINKGLQ HEALVKSKDN EWLLNHHKD SWLLNHHKD SWLLNHHKD HLOVVTPQHN HLOVVTPQHN HWVVNPRIN ARTNRISFF RWQKSGPST CWVRADST CWVRADST CWVRADST CWVRADST CWLSPQRN HWTTQGHT AWTTQGHT AWTTQGHT HWTTQGHT ANASVPCT CDWSSNRLKN SWLKD-PR SA CWSA CONSTRUN CONSTRUCT CONST	
DmSCW DmGBB MmBmp8b MmBmp6 MmBmp5 MmBmp7 CeDBL-1 MmBmp10 MmBmp10 MmBmp10 MmBmp2 MmBmp2 MmBmp4 MmBmp4 MmBmp3 MmNoda1 MmGdf5 MmGdf5 MmGdf7	132 179 · S 228 · T 172 · D 228 · D 228 · D 228 · D 228 · D 233 · D 153 · - 153 · - 153 · - 155 · F 181 · T 156 · - 359 · S 180 · V 190 · I 155 · F 181 · E 158 · G 157 · V 121 · F 260 · F 192 · - 176 · T	QSFL         DVS	SEAME WVN           VNTTS;           VNTRS;           2TIRS;           2TIRS;           2TIRS;           2TIRS;           2TIRS;           2TIRS;           2TIRS;           2TI		SPEPD	YWLINKGLQ HEMLVKSKDN FWLLNHHKD ILWVTPQIN HEWLVNPQIN HEWVNPRHN HEWVNPRHN HEWVNPRHN HEWVNPRHN HEWSGSGPST GWVRADSTT GWVRADSTT FWLASPQRN HEWTTQGHTN - ONRKGRS- - TSERSI- STAANASVPCT CDSSNRLKN KWLD-PR- KN-SA- QP-WK-	
DmSCW DmGBB MmBmp8b MmBmp6 MmBmp6 MmBmp7 CeDBL-1 MmBmp10 MmBdf2 CeTIG-2 DmDPP MmBdf2 MmBmp4 MmBmp4 MmBmp4 MmBdf5 MmGdf7 MmGdf7	179 - S 228 - T 172 - D 228 - D 23 - D 153 177 - M 181 - T 155 - F 181 - E 158 - G 157 - V 121 - F 260 - F 192 176 - T	QSFLEDV&S           β5           MEPDS           MEPDS           JLFFUDI           JLFFUDI           JLFFUDI	SEAME WVN           MT TS;           MT TG;           TT RS;           XKT QA;           XKT RL;           XLV RL;           XLV RQ;           RLV RQ;           SV PGPQ: QQ;           SV PGPQ: QQ;           SV PGPQ: QQ;           SV PGQ: QQ;           SV PGQ: QQ;           RLV RQ;           RLV RQ;           RV PGQ: QQ;           RU QQ;           RU QQ;           RU	ADLR LARR 	SPEPD	YWULINKGLQ HEOLVKSKDN FWULINHHKD WULINHHKD WULINPQNN HWUVNPQNN HRTNRISFF HWUVNPKIN HWUVNPKIN HWUVNPKIN HWUVNPKIN HWUTYGHT HUSSHQLST FWULSPQRN -WITKGRS- -WITKGRS- -WITKGRS- -WITKGRS- -WITKGRS- 	
DmSCW DmGBB MmBmp8b MmBmp5 MmBmp5 MmBmp7 CeDBL-1 MmBmp10 MmGdf2 CeTIG-2 DmDPP MmBmp4 MmBmp15 MmGdf9 MmGdf9 MmGdf9 MmGdf5 MmGdf5 MmGdf7 DmSCW	179 - S 228 - T 172 - D 283 - D 172 - D 153 177 - M 181 - E 155 - F 185 - F 185 - S 180 - V 121 - F 260 - F 192 176 - T 217	QSFL         DVS	SEAREWVN           NTTS;           NTRS;           TIRS;           TIRS;           TIRS;           TIRS;           TIRS;           TIRS;           TIRS;           TIRS;           TI		SPEPD	YWLINKGLQ HEOLVKSKDN EWLLNHHKD SWLLNHHKD SWLLNHHKD HEOVVTPQHN HEOVVTPQHN HEOVVNPRHN HEOVVNPRHN HEOVVNPRHN HEOVTPQHN HEOVTQGHTN WSHLQLST WSHLQLST WSHLQLST WSHLQLST WSHLQLST WSHLQLST WSHLQLST HOTSSERSI -	
DmSCW DmGBB MmBmp8b MmBmp6 MmBmp6 MmBmp7 CeDBL-1 MmBmp10 MmBdf2 CeTIG-2 DmDPP MmBdf2 MmBmp4 MmBmp15 MmBdf4 MmBmp4 MmBdf4 MmBdf4 MmBdf5 MmGdf5 MmGdf5 MmGdf5 MmGdf5 DmSCW	179 S 228 T 172 C 283 C 228 C 172 C 283 C 228 C 203 C 233 C 235 C 177 M 181 T 155 F 181 C 155 F 181 C 155 V 121 F 260 F 260 F 192 C 176 C	QSFL         DVS	SEAME WVN           NTTS;           NTTG;           TTRS;           XKTQA;           XKTQA;           XKTQA;           XKTQA;           XKTQA;           XKTQA;           XKTQA;           XKTQA;           XKTQA;           XKT	ADLR LARR 	SPEPD	YYWLINKGLQ EEGLVKSKDN FWLLNHHKD SWLLNHHKD SWLLNHHKD IGWVYPQHN HWVYPQHN HWVYPQHN GWVRADST SWSHQLST SWSHQLST SWSHQLST SWLSSPQRN -ONRKGRS -ONRKGRS -ONRKGRS -ONRKGRS -ONRKGRS -ONRKGRS -ONSNELN SWLCD-PR 	
DmSCW DmGBB MmBmp8a MmBmp6 MmBmp6 MmBmp7 CeDBL-1 MmBmp10 MmGdf2 CeTIG-2 DmDPP MmGdf2 MmBmp4 MmBmp4 MmBmp4 MmBmp4 MmBdf1 MmGdf5 MmGdf6 MmGdf6 MmGdf7 DmSCW DmGBB	179 - S 228 - T 172 - D 283 - D 228 - D 283 - D 228 - D 283 - D 228 - D 153 - 157 - N 155 - 157 - N 156 - 157 - N 156 - 158 - C 157 - V 121 - D 158 - C 157 - V 121 - D 158 - C 157 - D 158 - C 158 - C 159 -	QSFL         DVS	SEALE VN           NTTS;           NTTG;           TTRS;           XWWA;           XTVD;           YTYG;           2DIYG;           2DIS;           XTNA;           XSVD;           YTYG;           2DIS;           XTVNQ;           XLVNQ;           XLVNQ;           XLVNQ;           XLVNQ;           XSV PGPQGQ;           XSV PGPQGQ;           XSV PGL           XTD	β6           SQR M         E           NY         E           SQR M         E           NY         V           SQR M         E           SQR M         V	SPEPD	YYMLINKGLQ HENLVKSKDN EWLLNHHKD SWLLNHHKD SWLLNHHKD HUVVYPQHN HUVVYPQHN HUVVYPQHN HUVVYPQHN HUVVPRINT SWLKSGPST CWVRADST SWSHLQLST SWSHL SWSHLQLST SWSHLQLST SWSHLQLST SWSHLQLST SWSHL SWSHLQLST SWSHLST SWSHL SWSHLQLST SWSHLQLST SWSHLQLST SWSHL SWSHLQLST SWSHLD SWSHLQLST SWSHLQLST SWSHLQLST SWSHLQLST SWSHLQLST SWSHLQLST SWSHLQLST SWSHLQLST SWSHLQLST SWSHLQLST SWSHLQLST SWSHLQLST SWSHLQLST SWSHLQLST SWSHLQLST SWSHLD	β7 RNEPRISG7 IGIRIYUETA-D IGIRIYUETA-D IGIRIYUET MGLOUSUVTR-D IGIOCAETG-D IGIOCAETG-D IGIOCYETL-D IDIPEOVEIEETQS HQUEHIESRONQA NKNKEVTVQS LPUTARAST YGLUEVRTVRSLK HGFVDEVAHLEENP YGALEVTHLHQTR -VIRAFRCQQQ HISWNFTCTKD VRAA VRAA VRAA VRAA VRAA
DmSCW DmGBB MmBmp8b MmBmp6 MmBmp6 MmBmp7 CeDBL-1 MmBmp10 CeDBL-1 MmBmp10 MmBmp15 MmBmp2 MmBmp2 MmBmp2 MmBmp4 MmGdf3 MmGdf3 MmGdf3 MmGdf5 MmGdf5 MmGdf5 MmGdf7 DmSCW DmGBB MmBmp8a	179 S 228 T 172 C 228 D 228 D 228 D 153 - 153 - 181 T 156 - 181 T 155 P 181 F 185 V 190 I 155 P 181 F 185 V 191 V 192 V	QSFL     DVS	SEAME WVN           NTTS;           NTRS;           TTRS;           XXTQA;           XXYQA;           XXYQA;           XXY	ADLR IL RR 	SPEPD	YWULINKGLQ EEULVKSKDN EEULVKSKDN EWULNHHKD EWULNPHN EWUNPQNN EWUNPQNN EWUNPQNN EWUNPQNN EWUNPQNN EWUNPQNN EWTQGHSSEF EWSSES - TSSERSI- - MNRKGRS- - MNRKGRS- - TSSERSI- - MNRKGRS- - TSSERSI- - MNRKGRS- - MNRKGRS- - OPNKGRS- - QPUKAS- - QPUKAS- - QPUKAS- - SA- - S	
DmSCW DmGBB MmBmp8a MmBmp6 MmBmp5 MmBmp7 CeDBL-1 MmBmp10 MmGdf2 CeTIG-2 DmDPP MmGdf2 MmBmp4 MmBmp4 MmBmp4 MmGdf5 MmGdf1 MmGdf5 MmGdf5 MmGdf6 MmGdf6 MmGdf7 DmSCW DmSCB DmSCW DmSCW	179 - S 228 - T 172 - D 283 - D 228 - D 228 - D 228 - D 228 - D 228 - D 153 - 177 - M 155 - D 228 - D 155 - D 156 - S 215 - D 156 - S 216 - D 217 266 - D 211 211	QSFL         DVS	SEAME WVN           NTTS;           NTTG;           TTRS;           XWWA;           XTWA;           XTWA;           XTWA;           SSVD;           YMWA;           SSVD;           YMS;           TTVS;           TVS;           TVS;           TVS;           RLWNQ;           RLWNQ;           RLWNQ;           RLW	ABLR         L         RR          β6         SQR         E         N           SQR         E         N         S         N           SQR         E         N         S         N         N           SQR         E         N         S         N	SPEPD	YWULINKGLQ HEGLVKSKDN FRULNHHKD SWLLNHHKD SWLLNHHKD SWLLNHKD HGVVTPQHN HGVVTPQHN HGVVTPQHN HGVVTPQHN HGVVTPGHT HGVVTPGHT HGVTGGFT HGVSGGFT HGVSGGFT HGVSGGFT HGVSGGFT HGTGG HGTGG HGTGG HGTGG HGTGG HGTGG HGTGG HGTGG HGTGG HGTGG HGTGG HGTGG HGTGG HGTGG HGTGG HGTGG HGTGG HGTGG HGTG HGTGG HGTG HG	
DmSCW DmGBB MmBmp8b MmBmp6 MmBmp6 MmBmp7 CeDBL-1 MmBmp10 MmBmp10 MmBmp10 MmBmp12 MmBmp2 MmBmp2 MmBmp2 MmBmp4 MmBmp4 MmGdf1 MmGdf1 MmGdf1 MmGdf5 MmGdf5 MmGdf5 MmGdf5 MmGdf7 DmSCW DmGBB MmBmp8a MmBmp8a MmBmp8a	179 S 228 T 172 D 228 D 228 D 228 D 153 - 153 - 155 P 181 T 156 - 185 P 181 F 186 V 190 I 185 P 181 F 185 V 192 - 176 V 192 V	QSFL         DVS	SEAME WVN           NTTS;           NTRS;           TTRS;           TTRS;           XXWA;           XXWA;           XXWA;           XXWA;           XXWA;           XXQA;           XX		SPEPD		
DmSCW DmGBB MmBmp8b MmBmp6 MmBmp6 MmBmp7 CeDBL-1 MmBmp10 MmBmp7 CeDBL-1 MmBmp2 CeTIG-2 DmDPP MmGdf9 MmBmp4 MmBmp4 MmBmp4 MmBmp5 MmGdf5 MmGdf5 MmGdf5 DmSCW DmSCW DmSCW DmSCW DmSCW DmSCW DmSCW DmBmp8a MmBmp8 MmBmp6	179 - S 228 - T 172 - D 228 - D 217 - D 218 - D 217 - D 218 - D 217 - D 228	QSFLEDV&S           β5           MEPDS           MEPDS           JLFFUDI           LLFFUDI           LFFUDI           LFFUDI           LFFUDI	SEALE VVN           NTTS;           NTRS;           TTRS;           XKTQA;           RVWA;           XKTQA;           STQA;           TTRS;           XKTQA;           STQA;           ST	ADLR IL RR 	SPEPD		β7 RRNEPRISI-G HGIYIGAHAVNR-P LGRRYVETA-D LGRRYVETA-D LGRQYVETR-D LGRQYVETR-D LGRQSVTC-D IDIPEDVEIEETQS HQUEHESRQNQA NKNKEVTVQS LPHVTARAST YGLUEVRTVRSLK HGFVUEVAHLEENP YGANEVTHLHQTR -VRRRFMCQQQ HISNFTCTND RIADRA COCEDIEAW -QICEDIRA VKIQKLRFKR KATAHSSHHSKR VRAPRA VRAPTAR VRAPTAR VRAPTAR VRAPTAR
DmSCW DmGBB MmBmp8b MmBmp5 MmBmp5 MmBmp7 CeDBL-1 MmBmp10 MmGdf2 CeTIG-2 DmDPP MmBmp10 MmBmp10 MmBmp3 MmBmp3 MmSDF3 MmSdf5 MmGdf5 MmGdf5 MmGdf5 MmGdf5 MmGdf7 DmSCW DmSBB MmBmp8a MmBmp8a MmBmp8 MmBmp5	179 S 228 T 172 D 228 T 283 D 203 D 153 - 155 P 181 T 155 P 181 T 155 P 181 T 155 P 181 T 155 P 181 T 260 P 202 - 176 T 217 266 211 217 222 267 222 2 227 222 267 222 27 222 27 27 27 27 27 27 27 27 27 27 27 27 27 2	QSFL DV& S 	SEAREWVN           NNTTG           NTRS           QTIRS           QTIRS           QTIRS           QTIRS           QTIRS           QTIRS           QTIRS           QTIRS           QTI	ABLR IL RR 	SPEPD		
DmSCW DmGBB MmBmp8b MmBmp6 MmBmp6 MmBmp7 CeDBL-1 MmBmp10 MmBmp10 MmBmp2 MmBmp2 MmBmp2 MmBmp4 MmBmp15 MmBmp4 MmBmp4 MmSdf1 MmSdf5 MmSdf5 MmSdf6 MmSmg8a MmBmp8a MmBmp8a MmBmp8 MmBmp6 MmBm6 MmBm6 MmBm6 MmBm6 MmBm6 MmBm6 MmBm6 MmBm6 MmBm6 MmBm6 MmBm6 MmBm6 M	179° S 228° T 172° D 228° D 228° D 228° D 228° D 228° D 177° M 181° T 155° F 181° T 155° F 181° C 155° T 155° T 155° T 155° T 157° V 190° T 155° T 155° T 157° V 190° T 155° T 157° V 157° V 15	QSFLEDV&S           β5           MEPDS           MEPDS           JLFFNDL           JLFFNDL           JLFFNDT	SEAME WVN           NTTS;           NTTG;           TTRS;           XKTQA;           RXQA;           RXQA;           RXQA;           RXQA;           RXQA;           RXQA;           RY	ADLR IL RR 	SPEPD	YWWLINKGLQ HEWLVKSKDN FWLLNHHKD SWLLNHHKD WWVYPQHN HWVYPQHN HWVYPQHN HWVYPQHN HWVYPQHN HWVYPQHN HWVYPGHN HWVYPGHN HWVYPGHN HWVTQGHS HQUSSNL HWTTQGHS HQUSSNRLKN -ONRKGRS -ONRS -ONRKGRS -ONRKGRS -ONRKGRS -ONRKGRS -ONRKGRS -ONRKGRS -ONRKGRS -ONRKGRS -ONRS -ONRKGRS -ONRKGRS -ONRKGRS -ONRKGRS -ONRKGRS -ONRKGRS -ONRKGRS -ONRKGRS -ONRKGRS -ONRKGRS -ONRKGRS -ONRKGRS -ONRKGRS -ONR -ONRKGRS -ONR -ONRKGRS -ONR -ONRKGRS -ONR -ONR -ONRKGRS -ONR -ONR -ONR -ONR -ONR -ONR -ONR -ONR	β7 RRNEPRISI-G HGIYIGAHAVNR-P LGIRRYVETA-D LGIRRYVETA-D LGIRRYVETR-D LGIQISVTR-D LGIQISVTC-D DIPEDVEIEETQS HQUEHESRQNQA NKNKEVTVQS LPHVTARAST YGLIVEVRTVRSLK HGFVUEVAHLEENP YGLAFVTHLHQTR -VIRRFMCQQQ HISNFTCTKD VRIA LDHHETLVKED-R -ADEKCVSSRAEKC -QUCCEIRAV NO α5 VKIQKLRFKR KATAHSSHHRSKR VRAPR VRAFTAR LRSVR
DmSCW DmGBB MmBmp8b MmBmp5 MmBmp5 MmBmp7 CeDBL-1 MmBmp10 MmGdf2 CeTIG-2 DmDPP MmBmp4 MmBmp15 MmGdf3 MmBmp4 MmGdf6 MmGdf6 MmGdf6 MmGdf6 MmBmg4a MmBmp8a MmBmp8 MmBmp5 MmBmp5 MmBmp5 MmBmp5	179 S 228 T 172 D 228 C 223 D 233 D 153 155 181 T 155 P 181 T 155 P 181 T 155 P 185 C 187 C	QSFL DV& S 	SEAME WVN           NTTG           VNTTG           TTRS           VNTRS           VNTRS           VNTRS           VIT	ABLR         L         RR          β6         SQR         E         N           SQR         E         N         S           SQR         E         N         S         N           SQR         E         N         S         N           SQR         S         S         N         N           SQR         N         S         S         N           NTSORES         D         N         N         N           SQR         S         D         N         N         N           SQR         S         D         N	SPEPD	YMLINKGLQ YMLINKGLQ HEOLVKSKDN EWLLNHHKD HEOVVTPQHN HEOVVTPQHN HEOVVTPQHN HEOVVTPQHN HEOVTPQHN HEOVTPQHN HEOVTPQHN HEOVTPQHN HEOVTPQHN HEOVSGEPST HE	
DmSCW DmGBB MmBmp8b MmBmp6 MmBmp6 MmBmp7 CeDBL-1 MmBmp1 CeDBL-1 MmBmp12 MmBmp2 MmBmp2 MmBmp4 MmBmp4 MmBmp4 MmBmp4 MmBmf6 MmBmf6 MmBmf6 MmBmf6 MmBmf6 MmBmg8a MmBmp8a MmBmp5 MmBmp7 CeDBL-1 MmBmp7	179° S 228° T 172° D 228° D 228° D 228° D 228° D 177° M 153° S 181° T 155° F 181° T 155° F 181° F 182° F 181° F 181° F 181° F 181° F 181° F 181° F 182° F 181° F 181° F 181° F 181° F 182° F 181° F 182° F 181° F 181° F 182° F 181° F 182° F 181° F 182° F 18	$\begin{array}{c} \mbox{QSFL} \mbox{B} \mbox{QSFL} \mbox{B} \mbox{S} \mbox{S}$	SEAME WVN           NTTS;           NTTG;           TTRS;           XKTQA;           RVWA;           SSVD;           TTWA;           SSVD;           TTWA;           SSVD;           TTWA;           SSVD;           TTWA;           SS PGPQGQ;           SSV PGPQGQ;           SSV PGPQGQ;           SSV PGL	ADLR IL RR 	SPEPD	YYWLINKGLQ YYWLINKGLQ EEGLVKSKDN FWLLNHHKD FWLLNHHKD EWLLNHHKD EWLNHKD EWLNHKD EWLNHKD EWLNHKD EWLST EW	β7 RRNEPRISI-G HGIYIGAHAVNR-P LGIRRYVETA-D LGIRRYVETA-D LGIRRYVETR-D LGIRSYVETR-D LGIQSSVTL-D TDIPEDVEIEETQS HQUEHESRQNQA NKNKEWTVQS LPIVTARAST YGLLWERTVRSLK HGFVIEVAHLEENP YGLAEVTHLHQTR -VRERFMCQQQ HISVNFTCTKD RAA VRIAR VRIAR VRIAR VRIAR VRIAR VKIQKLRFKR KATAHSSHHRSKR VRAPRTAR
DmSCW DmGBB MmBmp8b MmBmp5 MmBmp5 MmBmp7 CeDBL-1 MmBmp10 MmBmp15 MmBmp10 MmGdf2 CeTIG-2 DmDPP MmBmp4 MmBmp15 MmGdf9 MmSdf5 MmGdf9 MmGdf7 DmSCW DmGBB MmBmp8a MmBmp8a MmBmp8 MmBmp5 MmBmp7 CeDBL-1 MmBmp10 MmGmf12	179 - S 228 - T 172 - D 228 - D 228 - D 228 - D 228 - D 228 - D 228 - D 153 177 - M 155 - F 156 - S 157 - V 190 - T 156 - S 157 - V 157 - V 158 - C 157 - V 157	QSFLEDV&S           β5           MEPUSS           MEPUSS           UFFUDI           UFFUDI	SEAME WVN           NTTS;           NTTG;           TTRS;           XWWA;           XTWA;           XTWA;           XTWA;           SSVD;           YFYG;           DDS;           TTVS;           TVNQ;           RLWNQ;           SVPGPQGQ;           SVP-GL];           RABPS PLG;           I:e no β9	ADLR IL RR 	SPEPD	YMLINKGLQ HEGLVKSKDN FWLLNHHKD SWLLNHHKD SWLLNHHKD SWLLNHKD HGVVTPQHN HGVVTPQHN HGVVTPQHN HGVVTPQHN HGVTRSF HGVKSGPST CWVRASCH HGVTCGHT HGVTCGHT HGVTCGHT HGVTCGHT HGVTCGHT HGVTCGHT HGVTCGHT HGVTCGHT HGVTCGHT HGVTCGHT HGT HGT HGT HGT HGT HGT HGT HGT HGT	
DmSCW DmGBB MmBmp8b MmBmp6 MmBmp7 CeDBL-1 MmBmp17 CeDBL-1 MmBmp10 MmBmp17 CeDBL-1 MmBmp10 MmBmp3 MmBmp4 MmBmp4 MmBmp4 MmBmp4 MmBmf4 MmBmf4 MmBmf4 MmBmf4 MmBmf4 MmBmf4 MmBmf4 MmBmf4 MmBmp5 MmBmp7 CeDBL-1 CEDBL-1 CEDBL-1 CEDBL-1 CEDBL-1 CEDBL-1 CEDBP	179 S 228 T 172 C 228 D 228 D 228 D 228 D 172 C 228 D 153 - 181 T 155 P 181 T 155 P 181 F 185 C 180 V 190 I 155 P 181 T 155 V 191 V 192 V 192 V 192 V 192 V 193 V 192 V 193 V 192 V 193 V 193 V 192 V 193 V 193 V 193 V 194 V 195 V 19	$\begin{array}{c} \mbox{QSFL} \mbox{B} \mbox{QSFL} \mbox{B} \mbox{S} \mbox{S}$	SEAME WVN           NTTS;           NTTG;           TTRS;           XXTQA;           XXVQA;           XXV	ADLR LARRAL 	SPEPD		
DmSCW DmGBB MmBmp8b MmBmp7 MmBmp7 CeDBL-1 MmBmp7 CeDBL-1 MmBmp10 MmBmp2 MmBmp2 MmBmp4 MmBmp4 MmBmp4 MmBmp4 MmBmp4 MmBmp5 MmGdf9 MmGdf7 DmSCW DmGB7 DmSCW DmGB7 DmSCW DmGB7 CeDBL-1 MmBmp6 MmBmp5 CeDL-1 MmBmp7 CeDBL-1	179 - S 228 - T 172 - D 228 - D 53 177 - M 172 - D 228 - D 153 177 - D 217 - D 218 - D 218 - D 217 - D 218 - D 218 - D 218 - D 218 - D 218 - D 217 - D 218 - D 218 - D 219	$\begin{array}{c} \mbox{QSFL} \mbox{B} \mbox{QSFL} \mbox{B} \mbox{S} \mbox{S}$	SEALE VVN           NTTS;           NTRS;           TTRS;           XKTQA;           RVWA;           SCVD;           TTS;           TTS;           TTS;           TTS;           TTNQ;           RLNQ;           RL	ADLR IL RR 	SPEPD	YYMLINKGLQ HEGLVKSKDN FWLLNHHKD SWLLNHHKD SWLLNHHKD SWLLNHKD HGVVYPQHN HWVYPQHN HWVYPQHN HWVYPGHN HWVYPGHN HWVYPGHN HWVYPGHN HWVYGSGPST HWVKSGPST HWVKSGPST HWFRGS- HWTTGCHY HWTTGCHY HWTTGCHY HWTTGCHY HWTTGCHY HWTTGCHY HWTTGCHY HWTTGCHY HWTTGCHY HWTGCHY HWTGCHY HWTGCHY HWTGCHY HWFRASSPU H HWFRASSPU H HTFRASSP H H H H H H H H H H H H H H H H H H	β7 RRNEPRISG7 IGIRPYETA-D IGIRPYETA-D IGIRPYETA-D IGIRPYET GIGUSYVT IGIOSYET IGIOSYET IDIPEDYEIEETQS HQUEHESRQNQA NKNKEVTVQS IPYTARAST YGLIEWRTVRSLK HGFVEVAHLEENP YGANEVTHLHQTR -VRURRFMCQQC HISWRTTCTND HSWRTTCTND KIQCIEIRA -QCCEIRA NG CS VKIQKLRFKR KATAHSSHHRSKR VRAPR VRAPTAR VRAPTAR SSVRKK-R SSVRKK-R BTK BTK
DmSCW DmGBB MmBmp8b MmBmp6 MmBmp7 CeDBL-1 MmBmp10 MmGdf2 CeTIG-2 DmDPP MmBmp10 MmGdf3 MmBmp4 MmGdf3 MmBmp4 MmGdf3 MmSdf5 MmGdf1 MmGdf5 MmGdf5 MmGdf5 MmGdf7 DmSCW DmGBB MmBmp8a MmBmp8a MmBmp8b MmBmp7 CeDBL-1 MmBmp10 MmGdf2 CeTIG-2 DmDPP	179 S 228 T 172 C 228 D 228 D 228 D 228 D 172 C 228 D 153 - 153 - 153 - 180 V 190 I 155 P 181 F 180 V 191 V 192 - 185 V 190 V 191 V 192 V 192 V 193 V 192 V 193 V 193 V 193 V 194 V 195 V 19	$\begin{array}{c} \mbox{QSFL} \mbox{B} \mbox{QSFL} \mbox{B} \mbox{S} \mbox{S}$	SEAME WVN           NTTS;           NTTG;           TTRS;           XXQA;           XXTQA;           XXYQA;           XXYQA;           XXYQA;           XXY	ADLR IL RR 	SPEPD		
DmSCW DmGBB MmBmp8a MmBmp6 MmBmp6 MmBmp7 CeDBL-1 MmBmp10 MmBmp7 CeDBL-1 MmBmp2 MmBmp4 MmBmp4 MmBmp4 MmBmp4 MmBmp4 MmBdf4 MmBmp4 MmBdf4 MmBdf4 MmBdf4 MmBdf4 MmBdf4 MmBmg6 MmBmp6 MmBm6 MmBmp6 MmBm6 MmBm6 MmBm6 MmBm6 MmBm6 MmBm6 MmBm6 MmBm6 MmBm6 MmB	179 - S 228 - T 172 - D 228 - D 359 - S 359 - S 359 - S 359 - S 177 - M 181 - T 155 - F 155 - F 155 - F 157 - V 190 - T 155 - F 157 - V 190 - T 155 - F 157 - V 190 - T 155 - F 157 - V 157	QSFLEDVSS           β5           MEPDSS           MEPDSS           MEPDSS           MEPDSS           UFFUDI           FODSS	SEALE VVN           NTTS;           NTRS;           TTRS;           XKTQA;           RVWA;           XKTQA;           XKTQA;           XKTQA;           XKTQA;           XKTQA;           XKT	ADLR IL RR 	SPEPD	YYMLINKGLQ HEGLVKSKDN FWLLNHHKD SWLLNHHKD SWLLNHHKD SWLLNHKD IGUVYPQHN IGUVYPQHN IGUVYPQHN IGWYNGSGPST GWVRADST GWVRADST GWVRADST CWVGSGPST GWVRADST CWVGSGPST GWVRADST CWVGSGPST GWVRADST CWVGSGPST CWSSNRLKN - TSSERSI - KN-SA- - CVGSSNRLKN GWGSGPST GWSSNRLKN - SA- - CVGSSNRLKN - GPGRGPLLI IGFGRGPLI IGFGGGGLG IGFGGGGLG IGFGGGGGG IGFGGGGGGGGGG	β7 RRNEPRISI-G HGIYIGHAWNR-P LGRRYVETA-D LGRRYVETA-D LGRQYVETR-D LGRQYVETR-D LGRQSVT TDIPEDVEIEETQS HQUEHESRQNQA NKNKEVTVQS LPHVTARAST YGLLEVRTVRSK HGFVIEVAHLEENP YGRAFVTHLHQTR -VRRRFMCQQQ HSNFTCTKD HSNFTCTKD HSNFTCTKD KATAHSSHHSKR VKIQKLRFKR KATAHSSHHSKR VRAPRTAR VRAPRTAR SSVRRK-R DKEQ-KEELNE GTK-ET-RLELKE PPKKRSRSASTT KARSIRDVSGG HPLHKREKR
DmSCW DmGBB MmBmp8b MmBmp56 MmBmp5 MmBmp7 CeDBL-1 MmBmp10 MmGdf2 CeTIG-2 DmDPP MmBmp4 MmBmp10 MmGdf3 MmNoda1 MmGdf5 MmGdf5 MmGdf5 MmGdf5 MmGdf5 MmGdf5 MmGdf7 DmSCW DmGBB MmBmp8a MmBmp8b MmBmp5 MmBmp7 CeDBL-1 MmBmp10 MmBmp10 MmGdf2 CeTIG-2 DmDPP MmBmp10 MmGdf2 CeTIG-2 DmPP MmBmp10	179 S 228 T 172 D 228 D 228 D 228 D 228 D 228 D 228 D 153 - 153 - 153 - 155 P 181 T 155 P 181 S 180 V 190 I 155 P 181 S 180 V 191 I 192 - 177 M 181 T 195 V 192 - 177 M 192 - 177 M 193 V 193 V 192 - 177 M 193 V 192 - 177 M 193 V 192 - 177 V 193 V 193 V 193 V 193 V 193 V 194 V 195 V 197 V 19	QSFL         DVS	SEAME WVN           NTTTS;           NTTRS;           TTRS;           XXTQA;           XXY	ABLR         LB RR          β6         SORGE EM           SORGE EM         BO           DYVGWEIN         BD           SDE GUIVE         BD           SEE COUE         BD           NTRE DE         SEE COUE           SDDPTUTSONES         SED           VUTRE DE         MARA           SET COUE         SEE COUE           SEE COUE         SEE COUE           SEE COUE         SEE COUE           SEE COUE         SEE COUE           SEE COUE         SEE C	SPEPD		
DmSCW DmGBB MmBmp8a MmBmp6 MmBmp6 MmBmp7 CeDBL-1 MmBmp1 CeDBL-1 MmBmp1 CeDBL-1 MmBmp4 MmBmp4 MmBmp4 MmBmp4 MmSdf9 MmSdf9 MmSdf9 MmSdf9 MmSdf9 MmSdf9 MmSdf9 MmSdf9 MmSdf9 MmSdf9 MmSdf9 MmSdf9 MmSdf9 MmSdf9 MmSdf9 MmSdf9 MmSmp4 MmBmp4 MmBmp4 MmBmp4 MmBmp4	179° S 228° T 172° D 228° D 228° D 228° D 228° D 177° M 153° S 359° S 177° M 155° F 153° C 155° F 153° C 155° C 157° V 190° T 155° F 192° C 157° V 190° T 155° F 192° C 157° V 190° T 155° C 190° V 190° T 176° T 190° V 190° T 176° T 190° V 190° T 176° T 190° V 190° T 176° T 121° T 222° C 191° C 192° C 222° C 220° C 22	QSFLEDVSS           β5           MEPDSS           MEPDSS           MEPDSS           UFFUDI           UFFUDI	SEAME WVN           NTTS;           NTTG;           TTRS;           XKTQA;           RXQA;           RX PGPQGQ;           CV P-SQVT;           SX PGPQGQ;           CV P-SQVT;           SX PGPQGQ;           CV P-SQVT;           SX PGPQGQ;           CT D-PQGP;           AAAE -PLG           SI I no β9	ADLR IL RR 	SPEPD	YYWLINKGLQ HEGLVKSKDN FWLLNHHKD SWLLNHHKD SWLLNHHKD SWLLNHHKD SWLLNHKD SWLLNF SWLNHKD SWLLNF SWLKSGPST SWLKSGPST SWLKSPGN SWLKS SWLKSPGN SWLKSPGN SWLKSPGN SWLKSPGN SWLKSPGN SWLKSPGN SWLKSPGN SWLKSPGN SWLKSPGN SWLKSPGN SWLKSPGN SWLKS SWLKSPGN SWLK	β7 RRNEPRISI-G HGIYIGAHAVNR-P LGIRBYVETA-D LGIRBYVETA-D LGIRBYVETR-D LGIQOSVTR-D LGIQOSVTD DIPEDVEIEETQS HQUEHESRQNQA NKNKEVTVQS LPIVTARAST YGLIENTVRSLK HGFVIEVAHLEENP YGRAEVTHLHQTR -VIRBRMCQQQ HISNFTCTKD VRIARAV NG C5 VKIQKLRFKR KATAHSSHHRSKR VRAPR VRAPRTAR VRAPRTAR LRSVR LRSVR SSVRKK-R LRSVR SSVRKSRSSASTT KARSIRDVSGG HPLHKREKR
DmSCW DmGBB MmBmp8b MmBmp5 MmBmp5 MmBmp7 CeDBL-1 MmBmp10 MmGdf2 CeTIG-2 DmDPP MmBmp1 MmBmp15 MmGdf3 MmNodal MmGdf3 MmNodal MmGdf5 MmGdf5 MmGdf5 MmGdf5 MmGdf5 MmGdf5 MmGdf5 MmGdf5 MmGdf5 MmBmp4 DmSCW DmGBB MmBmp8b MmBmp5 MmBmp1 MmBmp1 MmBmp1 MmBmp1 MmBmp1 Starter MmBmp1 Starter MmBmp2 MmBmp5 MmBm	179 S 228 T 172 D 228 D 203 D 203 D 203 D 203 D 203 D 153 - 153 - 155 D 155 D 181 T 156 - 181 T 181 T 181 T 185 D 180 V 190 T 190 T 190 T 190 T 192 - 197 T 197 T 195 T 203 Z 203 C 203 C 205 C 20	QSFL         DVS	SEAME EVVN         NTTG         VNTTG         TTRS         VNTTG         TTRS         VNTTG         TTRS         VNT	ABLR         L	SPEPD		
DmSCW DmGBB MmBmp8a MmBmp6 MmBmp6 MmBmp7 CeDBL-1 MmBmp7 CeDBL-1 MmBmp7 CeDBL-1 MmBmp7 SMmBmp7 CeDBL-1 MmBmp4 MmGdf3 MmBmp4 MmBdf5 MmBmg6 MmBmg6 MmBmg8a MmBmg8 MmBmp6 MmBmp7 CeDBL-1 MmBmp7 CeDBL-1 MmBmp7 CeDBL-1 MmBmp5 MmBmp5 MmBmp5 MmBmp5 MmBmp5 MmBmp5 MmBmp5 MmBmp5 MmBmp5 SMmBmp4 MmBmp4 MmBmp4 MmBmp4 SMmBmp5 SMmBmp5 SMmBmp4 SMmBmp4 SMMBmp4	179° S 228° T 172° D 228° D 228° D 228° D 228° D 228° D 177° M 170° D 228° D 173° D 228° D 173° D 17	QSFL         DVS	SEAME WVN           NTTS;           NTTG;           TTRS;           XKTQA;           RVWA;           SSVD;           TTWA;           SSVD;           VUS;           TTWA;           SSVD;           VU	ADLR IL RR 	SPEPD	YYWLINKGLQ HEGLVKSKDN FWLLNHHKD SWLLNHHKD SWLLNHHKD SWLLNHHKD SWLLNHKS SWLLNPQNN GWVRNSFF GWVRADST GWVRADST GWVRADST GWVRADST GWVRADST GWVRADST GWVRADST GWVRADST GWVRADST GWVRADST GWVRASS SWLQLST GWVRASS SWLQLST GWVRASS SWLQLST GWTGGHN GWTGGHN GWTGGHN GWTGGHN GWFGASS SWLCD-PR SWLCD SWLCD-PR SWLCD-PR SWLCD SW	
DmSCW DmGBB MmBmp6 MmBmp5 MmBmp5 MmBmp7 CeDBL-1 MmBmp10 MmGdf2 CeTIG-2 DmDPP MmBmp4 MmBmp15 MmGdf1 MmGdf5 MmGdf5 MmGdf7 DmSCW DmGBB MmBmp8a MmBmp8a MmBmp8 MmBmp5 MmBmp7 CeDBL-1 MmBmp10 MmBmp7 CeDBL-1 MmBmp10 MmBmp2 MmBmp2 MmBmp4 MmBmp15 MmBmp15 MmBmp15 MmBmp15 MmBmp15 MmSdf1 MmBmp15 MmBmp15 MmBmp15 MmBmp15 MmBmp15 MmBmp15 MmBmp15 MmBmp15 MmBmp15 MmBmp15 MmBmp15 MmBmp15 MmBmp16 Mm	179 S 228 T 172 D 228 D 223 D 233 D 153 233 D 155 T 155 T 155 T 181 T 156 157 V 190 T 181 T 156 157 V 190 T 190 T 191 T 192 195 196 195 196 195	QSFL         DVS	SEAME WVN           NTTS;           NTTG;           TTRS;           XWWA;           XKTQA;           XKTQA;           XKTQA;           XKTQA;           XKTQA;           XKT	ABLR         L         RR	SPEPD	YMLINKGLQ HEGLVKSKDN FRULNHRKD SWLLNHRKD SWLLNHRKD SWLLNHRKD SWLLNHRKD SWLLNFRKD SWLLNPQNN ARTNRISFF RWQKSGPST CWVRADST SWLKSPQRN AWT7GGHT AWT7GGHT SWLKD-FR	
DmSCW DmGBB MmBmp8b MmBmp6 MmBmp7 CeDBL-1 MmBmp17 CeDBL-1 MmBmp10 MmBmp15 MmBmp17 CeDBL-1 MmBmp15 MmBmp4 MmBmp3 MmSDF3 MmSdf1 MmGdf3 MmSdf4 MmBmp4 MmBmp5 MmBmp6 MmBmp3 CeDBL-1 CEDBL-1 CEDBL-	179 S 228 T 172 C 228 D 228 D 228 D 228 D 228 D 228 D 177 M 171 C 228 D 228 D 153 - 153 - 181 T 155 P 181 F 155 P 181 F 155 P 181 F 157 V 192 - 177 M 190 T 197 V 197 V 247 V 242 V 191 V 195 V 242 V 191 V 195 V 242 V 191 V 242 V 195 V 195 V 195 V 195 V 195 V 195 V 197 V 19	QSFL     DVS     S	SEAME WVN           NTTS;           NTTG;           TTRS;           XXTQA;           XXYQA;           XXYQA;           XXYQA;           XXY	ADLR IL RR 	SPEPD	YYMLINKGLQ YYMLINKGLQ ERLUNKSKDN FWLLNHHKD FWLLNHHKD FWLINHKD FWUNPQNN FWUNPQNN FWUNPQNN FWUNPQNN FWUNPGNN FWUNPGNN FWUNGHTN FWINGHT	
DmSCW DmGBB MmBmp8b MmBmp5 MmBmp5 MmBmp7 CeDBL-1 MmBmp10 MmBmp7 CeDBL-1 MmBmp10 MmBmp4 MmBmp4 MmBmp4 MmGdf5 MmGdf5 MmGdf6 MmBmp5 MmBmp6 MmBmp6 MmBmp6 MmBmp6 MmBmp6 MmBmp6 MmBmp7 CeDBL-1 MmBmp4 MmBmp4 MmBmp4 MmBmp4 MmBmp4 MmBmp4 MmBmp4 MmBmp4 MmBmp4 MmBmp5 MmBmp4 MmBmp4 MmBmp4 MmBmp5 MmBmp4 MmBmp5 MmBmp4 MmBmp4 MmBmp4 MmBmp4 MmBmp5 MmBmp4 MmBmp5 MmBmp4 MmBmp5 MmBmp4 MmBmp5 MmBmp4 MmBmp5 MmBmp4 MmBmp5 MmBmp4 MmBmp5 MmBmp4 MmBmp5 MmBmp4 MmBmp5 MmBmp4 MmBmp5 MmBmp4 MmBmp5 MmBmp5 MmBmp4 MmBmp5 MmBmp5 MmBmp5 MmBmp4 MmBmp5 MmBmb	179 - S 228 - T 172 - D 228 - D 359 - S - S - S - S - S - S - S - S	QSFLEDV&S       β5       MEPDS       MEPDS       JLFFUDI       LLFFUDI       LLFFUDI       TRUDT       YLUDT       TRUDT       SSKSGSSKI       MMSGR       TRUDT	SEALE VVN           NTTS;           NTTG;           TTRS;           XWWA;           XTQA;           XT V	ADLR IL RR 	SPEPD	YYNLINKGLQ HEGLVKSKDN FWELNHHKD SWLLNHHKD SWLLNHHKD SWLLNHKD IGVVTPQHN IGVVTPQHN IGVVTPQHN IGVVNPQNN GRTNRISFF GWCKSGPST GWCKASPQRN CWVRAST CWSSNLQLST CWSSNLQLST CWCKSPQRN GWTCGGHY GWTCGGHY CWSSNELKN CWSSNE	β7           RRNEPRISG           HGIYIGAHAVNR-P           LGIRBYVETA-D           LGIRBYVETA-D           LGIRBYVETA-D           LGIRSYVETR-D           LGOCSYVTR-D           LGOCSYVTR-D           LGOCSYVTR-D           LGOCSYVTR-D           LGOCSYVTR-D           LGOCSYVTR-D           LGOCSYVTR-D           LGOCSYVTRD           LGUENTYRS           HGYVEY           LGOCSYVTRD           LGOCSYVTSEQ           LGUENTYRS           HGFVVEVSIK           HGFVVEVSIK           HGFVVEVSIK           HGFVEVSIK           HGFVEVSIK           HGFVEVSIK           LGTRYS           VIRDRAMCQQO           HESVETTRO           VRAPRA
DmSCW DmGBB MmBmp8b MmBmp55 MmBmp7 CeDBL-1 MmBmp15 MmBmp7 CeDBL-1 MmBmp10 MmBmp15 MmBmp4 MmBmp15 MmGdf1 MmBdf2 CeTIG-2 DmSCW DmCPP MmBmp4 MmBmp4 MmBmf4 MmBmp5 MmBmp5 MmBmp5 MmBmp5 MmBmp7 CeDBL-1 MmBmp10 MmBmp15 MmBmp15 MmBmp4 MmBmp15 MmBmp4 MmBmp15 MmBmp4 MmBmp15 MmBmp4 MmBmp15 MmBmp4 MmBmp15 MmBmp4 MmBmp15 MmBmp4 MmBmp15 MmBmp4 MmBmp15 MmGdf1 MmGdf2 MmGdf1 MmGdf1 MmGdf1 MmGdf1 MmGdf1 MmGdf1 MmGdf1 MmGdf1 MmGdf1 MmGdf1 MmGdf1 MmGdf1 MmGdf1 MmGdf1 MmGdf1 MmGdf1 MmGdf1 MmGdf1 MmGdf1 MmGdf2 MmBmp5 MmGdf1 MmBmp5 MmGdf1 MmBmp5	179 S 228 T 172 D 228 D 228 D 228 D 228 D 228 D 228 D 153 - 153 - 153 - 155 P 181 T 155 P 181 T 155 P 181 T 155 P 181 T 157 V 121 P 220 T 177 M 181 T 156 - 177 M 181 T 157 V 190 T 192 - 177 M 192 - 177 M 192 - 177 M 193 S 180 V 193 S 180 V 192 - 177 M 193 S 193 S 19	QSFL DVS S βS	SEAME WVN           NTTS;           NTTG;           TTRS;           XXWA;           XXTWA;           XXTWA;           XXTWA;           XXTQA;           XXTQA;           XXT	AbLR         L	SPEPD		
DmSCW DmGBB MmBmp8a MmBmp8a MmBmp6 MmBmp6 MmBmp7 CeDBL-1 MmBmp17 CeDBL-1 MmBmp10 CeDBL-1 MmBmp4 MmBmp4 MmBmp4 MmBmp4 MmBmp4 MmBmp4 MmSdf9 MmBmp6 MmBmp1 Kall MmBmp6 MmBmp1 Kall MmBmp1 Kall MmBmp6 MmBmp6 MmBmp6 MmBmp1 Kall MmBmp1 MmBmp	179 S 228 T 172 C 228 C 177 M 150 C 153 C 153 C 155 C 153 C 155 C 157 C 155 C 157 C 15	QSFLEDVS       β5       MEPDS       MEPDS       MEPDS       JLFFUDI       LIFFUDI       JLFFUDI       FODS       SSKSGS       SSKSGS       GRUP       FLUDT       FDDD       FLUDT       ALLDS       FRINKER       GRUP       FLUDA       AND       SCUSTFAAG       GLINNPRAAG       S       S       S       GUSUNPKLAG       S       GOUSUNPKLAG       S       GOUSUNPKLAG       S       GOUSUNPKLAG       S       S       S   <	SEAME WVN           NTTS;           NTTG;           TTRS;           RVWA;           RXTWA;           RXTWA;           RXTQA;           RYRG;           RY	ADLR IL RR 	SPEPD	YYNLINKGLQ HEGLVKSKDN FWLLNHHKD SWLLNHHKD SWLLNHHKD SWLLNHHKD SWLLNHKD SWLLNHKD SWLLNPQNN GWVNADST GWVRADST GWVRADST GWVRADST GWVRADST GWVRADST GWVRADST GWVRADST GWVRADST GWVRADST GWVRADST GWVRADST GWVRADST GWVRADST GWVRADST GWVRADST GWVRADST GWVRADST GWRADST GWRADST GWRADST GWRADST GWFRANSP GWFRANSP GWFRANSP GWFRANSP GWFRANSP GWFRANSP GFRANQS	β7 RRNEPRISI-G HGIYIGAHAVNR-P LGIRRYVETA-D LGIRRYVETA-D LGIRRYVETR-D LGIRGYVETR-D LGIQOSVTR-D TDIPEDVEIEETQS HQUEHESRQNQA NKNKEVTVQS LFVVTARAST YGLIENERFMCQQQ HISNFTCTKD VRAR LDHHETLVKED-R -ADEKCVSSRAEKC -QUCBIRAV NO α5 VKIQKLRFKR KATAHSSHHRSKR VRAPR VRAPRTAR LRSVR SSVRKK-R LRSVR SSVRKK-R LRSVR RATAHSSHHRSKR VRAPRTAR
DmSCW DmGBB MmBmp6 MmBmp5 MmBmp5 MmBmp7 CeDBL-1 MmBmp10 MmGdf2 CeTIG-2 DmDPP MmBmp10 MmBmp15 MmGdf3 MmNodal MmGdf5 MmBmp5 MmBmp5 MmBmp5 MmBmp5 MmBmp5 MmBmp7 CeDBL-1 MmBmp10 MmBmp7 CeDBL-1 MmBmp10 MmBmp7 MmBmp15 MmBmp15 MmBmp15 MmBmp4 MmBmp15 MmBmp4 MmBmp15 MmBmp4 MmBmp4 MmBmp15 MmSdf9 MmBmp4 MmBmp4 MmBmp4 MmBmp4 MmBmp5 MmBmp4 MmBmp4 MmBmp4 MmBmp5 MmBmp4 MmBmp4 MmBmp5 MmBmp4 MmBmp4 MmBmp4 MmBmp5 MmBmp4 MmBmp4 MmBmp4 MmBmp5 MmBmp4 MmBm4 MmB	179 S 228 T 172 D 228 D 223 D 230 D 233 D 153 - 234 D 155 - 155 - 155 - 155 - 181 T 156 - 181 T 157 V 190 T 190 T 190 T 190 T 191 T 192 - 192 - 177 M 191 T 192 - 192 - 193 - 192 - 192 - 192 - 193 - 192 - 192 - 192 - 192 - 192 - 193 - 195 - 192 - 192 - 193 - 195 - 192 - 192 - 192 - 193 - 195 - 195 - 195 - 192 - 195 - 198 - 199 - 199 - 195 - 198 - 199 - 199 - 198 - 199 - 198 - 199 - 198 - 199 - 198 - 198 - 198 - 198 - 198 - 198 - 199 - 198 - 19	QSFL     DVS     S	SEAME WVN           NTTS;           NTTG;           TTRS;           TTRS;           XWWA;           XKTQA;           XKTQA;           XKTQA;           XKTQA;           XKTQA;           XKT	ADLR IL RR 	SPEPD		

**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  $\alpha$ 3 is indicated by the underlining in GDF2. The features  $\beta$ 7 and  $\beta$ 9' are conserved in this subfamily, while  $\beta$ 8,  $\beta$ 9 (the distinctive bowtie), and  $\alpha$ 5 of TGF- $\beta$ 1 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- $\alpha$  and this group's not quite significant cluster with Myoglianin was unexpected. The not quite significant cluster of DAF-7 with the four Inhibin- $\beta$  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- $\beta$  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 clusters are expected. (D) Cystine knot nodes ~0.85 are significant clusters are expected, such as Activin with four Inhibin- $\beta$  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- $\beta$  subfamily members Inhibin- $\alpha$ /GDF15. Together these form a fourth cross-subfamily cluster with significant, or just below significant, nodes to the TGF- $\beta$ proteins and five Activin subfamily members (Myoglianin, Myostatin/GDF11, and BMP3/GDF10). This cluster suggests that these prodomains are similar to Inhibin- $\alpha$  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-B subfamily members Inhibin- $\alpha$ /GDF15 could explain the ability of BMP subfamily Nodal/ GDF1 heterodimers to signal via the TGF-B subfamily receptor ActRIIA.

A fifth cross-subfamily cluster in the prodomain is DAF-7 (TGF- $\beta$  subfamily) with heterodimerizing Inhibin- $\beta$  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- $\beta$  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- $\beta$  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- $\beta$  subfamily members Inhibin- $\alpha$ /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- $\alpha$ /GDF15 may explain why Nodal heterodimers with GDF1/GDF3 can signal through ActRIIA.

The placement of Nodal and DBL-1 as solos in the fulllength 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- $\beta$  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- $\beta$  cluster. Alternatively, these two are weakly linked to an Activin + TGF- $\beta$  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- $\beta$  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  $\alpha$ 1, the Latency Lasso and  $\alpha$ 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  $\alpha$ 2.

The  $\alpha 2$  region of TGF- $\beta 1$  contains three conserved I/L/V residues (one universal) and a tyrosine. These are not conserved in the TGF- $\beta$  subfamily or the Activin + TGF- $\beta$  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  $\alpha 2$  of the other subfamilies. In other words, the All family members alignment erases the distinction in the location of  $\alpha 2$  in the Activin and TGF- $\beta$  subfamilies. In  $\beta 1$ , one I/L/V and a phenylalanine present in the Activin + BMP subfamilies are absent in the TGF- $\beta$  subfamily, maintaining the prior distinction in  $\beta 1$  location and conservation.

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

		Assnal Late	encyLass	o	$\alpha 2 $	-β1	$\beta 2$	β3		β4	6	60	4	β7-	β8-	β10-
MmBmp15	35			-•RYMLK	YHR•KN	WRLVKP.	DF-P.	ELIRA	TVVYRH.		KAWT-	EIDITH	CI-QQKL	V · SVIRLR	F•C•.	AFLLL
MmGdf9	55			-•YYMKKI	YKT•N1	VRLESP	• LF-N•	HLLKS	VLLYTL		HRWI-	EIDVTS	LL-QPLV	Γ•IHLSVN	F•C•	PSLIL
MmNodal	1	•MSAHSLRIL	LLQ	-•AYMES	YRD•DI	IRSLQA	•TF-D•	DLVWA	ELRLQL		SGST-	VLEVTK	PLSI	K • RALEKQ	V.AEKC.	NVLML
MmGdf5	92	TV	TPKGQL	G•EYMLSI	YRT•N1	TITSFID	•VF-D•	GLLGA	ELRILR		•SGWE-	VFDIWK	LF-RNF-	- • AQLCLE	LEAW.	KALFL
MmGdf7	50	SALQA-	AAV	G•HFMMSI	YRS • DI	TITGETD	•LF-D•	EVVNA	ELRVLR	•	ARWE-	AFDVTD	AV-QSHRI	R • RKFCLV	LRAV.	RALLV
MmGdf6	53	•-TPPEHGLRQK	DLRRRPE	G•EYMLS	YKT•N7	TITSEVD	·LF-D•	ELVGA	ELRLYR		AGWE-	VFDVWQ	GL-RP	- • KQLCLE:	LRAA.	RALLV
DmAct	314	CTLC · ESIKROILTKL	GLSH-KE	N•TYPID	NHS • QE	EIITEAE •	·SAQN ·	SIRSA	QIHIRI.	IKIWVFQ	• LGWQ-	KFDLTD	TI-REWY	G•EKLRLL	- • CTGC •	PFLVL
CeTIG-3	1			-•SRKHD	YGG •	D	• OYSI •	-ILSA	SLII	ISIVVYE.	LDKY-	HFDISH	LF-HKWMI	K•SDKMIK	IEIT.	
DmSCW	17	TT•LSEQMEMIDIL	DLGD-RE	R•KFLLE	YNE•NS	SILTESS	• TENT •	SLVQA	MLRIYK	FTVSVYR	RGWL-	EFNLTD	TL-RYWL	H . LQRRNE	LRIS.	PFIVG
DmGBB	40	DN•LDVSYEILEFL	GIAE-RE	T•KFLLD	YHR•DI	IMTELN	•WFDV•	YLVMA	ELRIYQ.	FTITVYA	VGWL-	ELNVTE	GL-HEWL	V · DNHGIY	I•-GAH•	PFMIG
MmBmp8a	19	GP•RDMQREILAVL	GLPG-RE	R•LFMLDI	YHA•DI	LVMSEVN	HFDL.	AVTAA	EFRIYK	LHISMFE	EGWL-	VLDITA	AS-DRWL	L•KDLGLR	LYVE.	PFMVT
MmBmp8b	19	LS•RDMQREIREVL	GLPG-RE	R•LFMLDI	YRA•DI	LIMSEVN	HFDL.	AVTAA	EFRIYK.	LHISMFE	EGWL-	VLDITA	AS-DRWL	L•KDLGLR	LYVE.	PFMVG
MmBmp6	52	PQ•REMQKEILSVL	GLPH-RE	R•LFMLD	YNA•DN	IVMSEVN	•KFNL•	AVTAA	EFRVY <mark>K</mark>	FLISIYQ	•EGWL-	EFDITA	TS-NLWV	V • HNMGLQ	LSVV.	PFMVA
MmBmp7	30	DN• REMOREILSIL	GLPH-RE	R•MFMLDI	YNA• DN	IVMSEVN•	RFDL.	AVTAA	EFRIY <mark>K</mark> •	FQITVYQ	EGWL-	VFDITA	TS-NHWV	V • HNLGLO	LSVE.	PFMVA
MmBmp5	27	DN•REIQREILSIL	GLPH-RF	R•LFMLDI	YNA • DN	IVMSEVN	•RFDL•	AVTAA	EFRIYK.	IKISIYQ	•VGWL-	VFDITV	TS-NHWV	I • NNLGLQ	LCAE.	PFMVA
MmBmp3	4	•QGQR-PNLHLP	GLRETEF	S•EHMLWI	YD-•		TF-N•	NILSA	TLYFYV.	IDLSAWI	VSWL-	SKDITQ	LL-RKAK	Q • FLIGFN	I SRA.	R
MmGdf10	17	• PSWSSLPSAAA	GLQGDRD	S•IHMLRI	YE-•		FF-N•	MILTA	AFHFYS.	LHLIFRS	GLWQ-	AKDISS	II-KAARI	R • LLLSAQ	LTGE.	R
DmMyo	175	SVSL•ESIKMHILMRL	NLKKL-P	N•-NIIDN	JFYR•SS	SIYIFPE	•RF-Q•	SYA	TLHLYL	IRVANTT	GQWV-	AVDLKS	LL-GNLG	S • MTQE	IIKG.	PLTVH
MmMstn	35	CNAC•EAIKIQILSKL	RLETA-P	N•RELIDO	YDV		KF-S•		QYNKVV		GINQ-	SIDVKT	VL-QNWLI	K • SNLGIE	IALD.	PFLEV
MmGdf11	56	CPVC•ESIKSQIISKI	RLKEA-P	N•QQILD	HDF		HF-S•		MFTKVL		GHWQ-	SIDFKQ	VL-HSWF1	R • SNWGIE	IAFD.	PFMEL
MmTgfb1	29	CKTI•EAIRGQILSKL	RLASP-P	S•-AVLAI	YNS•		••	D	ISH	QHVELYQ.	PEWL-	SEDVTG	VV-RQ <mark>W</mark> LI	N • GIQGFR	F•CSC-•	PFLLL
MmTgfb2	20	CSTL•EAIRGQILSKL	KLTSP-P	E•-EVISI	YNS•	•	•SF-C•	GYLDA	IPPTFY	QRIELYQ	• GEWL-	SFDVTD	AV-QEWLI	H • RNIGFK	I.CPC	PHLLL
MmTgfb3	23	CTTL•EAIRGQILSKL	RLTSP-F	E•-QVLAI	YNS•	•	••	·		QRIELFQ	AEWL-	SFDVTD	TV-REWLI	L•SNLGLE	I.CPC	PHLIL
MmInha	21	CQG-•AKVKALFLDAL	GPPAM-D	G•HAVGGH	MHR		·VF-R•	Q		QGPPRWA	·			•	-••	PILVL
MmGdf15	33	• SGPESQLNA	D	E LSF	RLHA	•	L•	RALLL	LTP•	AGRGRR-	•			•	-••	
MmGdf1	22	•MGPAAALLQVL	GLPEA-F	R∙−−VMWB	RLFR		• P•	G	•	-SVALWA	GVLL-	RADILG	TA-VAAN	A • CTVRLA	L••	SLLLV
MmGDF3	20	•EFQDSDLLQFL	G <mark>l</mark> eka-f	S•VLRF	<iir< td=""><td>•</td><td>• T •</td><td>G</td><td>•</td><td>•QDRG<mark>VW</mark>G •</td><td>• GQL</td><td>QFNLQG.</td><td>AL-KDWS:</td><td>S•KNLDLH</td><td>LILV.</td><td>SL<mark>L</mark>VV</td></iir<>	•	• T •	G	•	•QDRG <mark>VW</mark> G •	• GQL	QFNLQG.	AL-KDWS:	S•KNLDLH	LILV.	SL <mark>L</mark> VV
CeUNC129	9 2	IVLL • DLINETIRDLL	HFKSSD	N•EHMKNI	YEN•	-VRAIEP	•VF-D•	SIMRA	ELHFYL.	•-CVNEYC	• EEYKV	IWDATK	SVFDSYH:	L • AVFRIT	-•R•	PFLVI
CeDAF-7	30	CIEK•EYLKNEILDQL	NMKE-AF	K•SVY <mark>l</mark> e	(YRD•QI	LVAKEDV	••	DILQA	TLTVSI	•VQVQVYE •	• KS <mark>W</mark> FT	ISPIQG	IFVKAML	D•VALH	PQQ.	
DmMav	249	•TRLKHLVLKGL	GIKK-LI	D•SKYIEY	LSR•N]	[LLHEPL•	••	KIDEA	NVRLML•	•LTLKVYQ•	• TQ <mark>W</mark> I-	EFDVTK	AV-RS <mark>W</mark> LI	N•ENLGIE	- • CDKC •	PVLNI
MmInhba	44	CPSC • EAVKKHILNML	H <mark>l</mark> kk-rf	D•AAILNA	AIRK•SE	EIITEAE •	•HFEI•	VVERA	EVWLFL•	•VTIRLFQ	•STWH-	IFPVSS	SI-QRLL	D•SSLDVR•	-•CEQC•	PFLML
MmInhbb	68	CTSC • EAVKRHILSRL	QLRG-RF	N•AAMVT <i>I</i>	ALRK•SE	EIISFAE •	•YFFV•	FVVQA	SLWLYL	VRVKVYF	•SGWH-	TFPITE	AI-QALFI	E•RRLNLD	-•CDSC•	PFVVV
MmInhbc	39	CPAC • DLAKKSILDKL	HLSQ-RE	I•GALKTA	ALQR•YE	EIISFAD	• EFHF •	EVRQT	RFMFFV•	MNIRVLV	•SGWY-	QLLLGP	EA-QAAC:	S•LTLEL-	-•VP•	PFVAA
MmInhbe	39	CPSC • ELAKQQILEGL	HLTS-RF	R•AALTRA	ALRR•EF	KVISFAT	• TFQL•	-LYHA	RLWL•	LYLRIFR	•SG <mark>W</mark> H-	ALTIPS	SG-LRSE	D•VKLQLE·	-•-FRP•	PFLEL
CeDBL-1	24	HLFL•HATRRGILRKL	GLEH-VF	V•QHMWD1	IYDD•	P	·SL·	EVTKA	TLKLRL•	GNISIYF	• TEWI-	DFDVTA	AF-LRRTI	N•FFIDLPI	E•-VEI•	PLIVF
MmGdf2	56	M•ENMKVDFLRSL	NLSG-IF	S•QYMIDI	YNR•N]	IVRSESV•	•IF•	QITRA	ELRLY <mark>V</mark> •	•GSMVVYD•	•EGWE-	TLEVSS.	AV-KRWVI	R • TTNKNK	-•-LEV•	PFFVV
MmBmp10	53	T•QSMKNEFLKTL	NLSD-IF	V•EYMLEI	YNK•N]	IIRSEKN	•LF•	EVVMA	ELRLYT	RKITIFE	•SEWE-	TFDVTD	AT-RRWQI	K•STHQLE	-•-IHI•	PLLVV
CeTIG-2	42	•DKIGEQIRELF	NIDI-NI	N•	••	H·	•VSHM•	DDGSY	SIRFAK	•LRIHIQG	• DDPTV	VTDVTT	MV-DRWSI	H•STLPIV	TAR.	AFLVI
DmDawdle	226	<u>CPKC</u> • EFVKQQILEKL	RLKE-SE	K•KPIFD0	GMTL•NE	PSMCETF	••	-VSTA	VLWLFK.	•QTIVVSE ·	• DE <mark>W</mark> M-	KIDIEW	PI-KHWI:	S•LSHLIQ	-•CGGC•	PFIVI
MmLeftyl	27	•-II-GSILQQL	QLDQ-PF	V•SHVRT(	2YVA •	KRFSQ	• VFGM •	ELVQA	VLRLFQ.	•ARVTI	• SGWK-	AFDVTE.	AV-NFWQ0	Q•PRQPLL	-•-LQV•	
MmLefty2	2 27	•-VI-SSIIQQL	QLSQ-A	T•THVRS(	QYVA•	KRESQ	• VFGM •	ELVQA	VLRLFQ	ARVTI	• SGWK-	AFDVTE.	AV-NFWQ	Q•PRQPLL	-•-LQV•	
MmAmh	191	VL.DFP-AGAWSGS	GUIL-TI	Q•IDQ <b>I</b> Q-	-AF • DI	-MPEPQ	•LE-T•	ALRGP	LTQASN	GLVNLSD	• TEREP	MPLHGP.	AS-APWA	A • RRVAVE	-•-LQA•.	ALUII
MmGdnf	65	HSL	GHRR-VI	F•EDYPDQ	2FDD		·•				•			•	-••	
DmDPP	223	•EIE-KS	NWKR-PF	K•EPMKKI	YAE•N1	IVRSETH.	•HF-D•	KIIKAA	ELQITR	YQVLVYD	TDTV-	SLDVQP.	AV-DRWL	A•RNYGLL	-•-VEV•	PLLFT
MmBmp2	49	•EFE-LR	GLKQ-RI	T•PYMLDI	YRR•N]	TVRSEHH•	• FF-N•	FLTSA	ELQIFR	INIYE	SQWE-	SFDVTP.	AV-MRWT	r • TNHGFV	-•-VEV•	PLLVT
MmBmp4	54	•DFE-AT	GERR-R	O•DYMRD∎	MRL•N]	IWRSEHH .	·LD-N•	VIIISSA	ELRLER	NIME	<ul> <li>TRME -</li> </ul>	THOWSP.	AM-LRMTI	R • PNYGIA	- • - TEV •	PLLVT

**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- $\beta$ 1 and BMP9 as noted above with the latter shown here by its synonym GDF2. In the "LTBP-Association region", the cysteines in TGF- $\beta$ 1–3 that form a disulfide bond with LTBP are underlined in black. The unexpectedly conserved pairs of cysteines in Activin, Dawdle, the four Inhibin- $\beta$ s, and Myostatin/GDF11 are underlined in red. Note that the first cysteine of each pair aligns with the cysteine of TGF- $\beta$ 1–3. The conserved cysteines of Inhibin- $\alpha$  and DAF-7 are also underlined in red. The two underlines of  $\alpha$ 2 show that TGF- $\beta$ 1 and Inhibin- $\beta$ a have been pulled into alignment by the homogeneity of the Activin + BMP version is conserved, while the TGF- $\beta$ 1 in  $\beta$ 3 and  $\beta$ 7 shows that the Activin + BMP version is conserved, while the TGF- $1\beta$  version is not, although TGF- $\beta$  subfamily members have been pulled into alignment by the homogeneity of the Activin + BMP version is conserved, while the TGF- $1\beta$  version is not, although TGF- $\beta$ 1 and its siblings (CxC) are underlined in black. The unexpectedly conserved pairs of cysteines in Activin, Maverick, Dawdle, Inhibin- $\beta$ a, and Inhibin- $\beta$ b (CxxC) are underlined in red. Note that the first cysteine of the CxxC 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 CxxC.

six having lost the glutamic acid.  $\beta$ 4 conservation is also reduced, with its stretch of seven residues now only three: two I/L/V residues and an F/Y.  $\beta$ 4 is absent in Nodal and seven other BMP subfamily members. The middle region of the arm,  $\beta$ 6 to  $\alpha$ 4 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 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  $\alpha 4$  is still present.

At the 3' end of the arm conservation is also limited.  $\beta7$ , like  $\beta3$ , is conserved in an Activin + BMP pattern that draws in several TGF- $\beta$  subfamily members, although not TGF- $\beta1$ .  $\beta7$  retains two moderately conserved I/L/V residues of three previously.  $\beta10$  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 ( $\alpha1$ , Latency Lasso, and  $\alpha2$ ) plus the central  $\beta6$  to  $\alpha4$  and 3' end of the arm ( $\beta$ 10) are the six best conserved features of the 17 prodomain features in the TGF- $\beta$  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- $\beta$ 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 LTBPs (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- $\beta s$ , Dawdle, and the duplicated pair Myostatin/GDF11. The first cysteine of the pair in each of these eight proteins from the Activin and TGF- $\beta$  subfamilies is aligned with the cysteine of TGF- $\beta 1$ –3. Further, a single cysteine is present in Inhibin- $\alpha$ and DAF-7 that also aligns with the cysteine of TGF- $\beta 1$ –3. A total of 10 proteins in the Activin + TGF- $\beta$  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- $\beta$  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- $\beta$  subfamilies is the order of  $\alpha 3$  and  $\beta 3$  (Activin has  $\beta 3$  first and TGF- $\beta$  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- $\beta$  subfamily.  $\beta 9'$  is distinct between the BMP and Activin + TGF- $\beta$ 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- $\beta$  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 β8 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-βa 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- $\beta$  subfamily members and a single conserved cysteine in two additional members of the TGF- $\beta$  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- $\beta$  subfamily cysteines to be captured in the alignment.

The presence of conserved cysteines in all Inhibin- $\beta$ s and Inhibin- $\alpha$  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- $\beta$  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 receptorassociated Smads have been noted. In flies, the Activin/TGF- $\beta$ 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-B Type II receptor Wit (Lee-Hoeflich et al. 2005). In mammals, Inhibin-B 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 crosssubfamily prodomain similarity. Nodal heterodimerization may serve as an example as its partners GDF1/GDF3 have prodomains that cluster with Inhibin- $\alpha$ /GDF15 suggesting a mechanism for Nodal signaling through ActRIIA.

Overall, we identified six cross-subfamily and two withinsubfamily 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  $\beta 8$  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- $\beta$  proteins (Inhibin- $\beta$ 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- $\beta$  tree. This cluster has strong statistical support. In the same tree Dawdle is adjacent to Inhibin- $\alpha$ , the heterodimerization partner of Activin's Inhibin- $\beta$  relatives. The Dawdle and Inhibin- $\alpha$  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  $\beta$ 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- $\beta$  receptor ActRIIA. This is a cross-subfamily cluster that links the BMP subfamily members GDF3/GDF1 with absolute confidence to the TGF- $\beta$  subfamily members Inhibin- $\alpha$ /GDF15. These four are in a second larger crosssubfamily cluster with the prototype TGF- $\beta$  proteins and five Activin subfamily members. All the proteins in the larger cluster, except GDF3/GDF1, signal through TGF- $\beta$  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- $\beta$  receptors. Nodal heterodimers with GDF1/GDF3 that signal via TGF- $\beta$ 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 crosssubfamily clusters in the All family members prodomain tree: DAF-7 with the Activin subfamily heterodimerizing Inhibin- $\beta$  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  $\beta$ 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- $\beta$  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. Zacchigna *et al.*, 2009 FAM/USP9x a deubiquitinating enzyme essential for TGF- $\beta$  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
- 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 heterotype 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 24: 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 ectodermin/Tif1gamma 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 $\beta$  ligand and the Gfrl 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 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.
- Özüak, 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. Wijayatonge 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 LTBPs 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 (BMPR2) in immortalized gonadotropelike 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