Molecular Evolution of a Developmental Pathway: Phylogenetic Analyses of Transforming Growth Factor-β Family Ligands, Receptors and Smad Signal Transducers

Stuart J. Newfeld*,†, Robert G. Wisotzkey‡ and Sudhir Kumar*

*Department of Biology, Arizona State University, Tempe, Arizona 85287-1501, †Graduate Program in Molecular and Cellular Biology, Arizona State University, Tempe, Arizona 85287-1144 and ‡AXYS Pharmaceuticals, Inc., South San Francisco, California 94080

ABSTRACT

Intercellular signaling by transforming growth factor-β (TGF-β) proteins coordinates developmental decisions in many organisms. A receptor complex and Smad signal transducers are required for proper responses to TGF-β signals. We have taken a phylogenetic approach to understanding the developmental evolutionary history of TGF-β signaling pathways. We were interested in detecting evolutionary influences among the physically interacting multigene families encoding TGF-β ligands, receptors, and Smads. Our analyses included new ligands and Smads identified from genomic sequence as well as the newest published family members. From an evolutionary perspective we find that (1) TGF-β pathways do not predate the divergence of animals, plants, and fungi; (2) ligands of the TGF-β/activin subfamily likely originated after the divergence of nematodes and arthropods; (3) type I receptors from *Caenorhabditis elegans* are distinct from other receptors and may reflect an ancestral transitional state between type I and type II receptors; and (4) the Smad family appears to be evolving faster than, and independently of, ligands and receptors. From a developmental perspective we find (1) numerous phylogenetic associations not previously detected in each multigene family; (2) that there are unidentified pathway components that discriminate between type I and type II receptors; (3) that there are more Smads to be discovered in Drosophila and mammals; and (4) that the number of C-terminal serines is the best predictor of a Smad's role in TGF-β signal transduction. We discuss these findings with respect to the coevolution of physically interacting genes.

INTERCELLULAR signaling by growth factors is essential for proper pattern formation in metazoan development. Secreted proteins of the transforming growth factor-β (TGF-β) family regulate key developmental events in many organisms. For example, TGF-β family members induce the establishment of the left-right body axis in mammals (Collignon *et al.* 1996), influence body size and tail morphology in nematodes (Savage *et al.* 1996), and regulate the formation of adult appendages in flies (Posakony *et al.* 1991). Extensive sequence analyses of this multigene family have identified significant amino acid conservation across species. Several TGF-β subfamilies have been identified based upon amino acid identities. The largest is the Dpp/BMP subfamily (*decapentaplegic/*bone morphogenetic protein), with members in flies, nematodes, and vertebrates (reviewed in Kingsley 1994). A comparison of Dpp from two insects (*Drosophila melanogaster* and *Schistocerca americana*) with human BMP4 revealed >75% amino acid identity in all pairwise comparisons (Newfeld and Gelbart 1995). This level of sequence conservation is reflected in the ability of these proteins to

function correctly in cross-phyla experiments. Human BMP2 and BMP4 can rescue *dpp* mutant phenotypes when expressed in Drosophila (Padgett *et al.* 1993) and Dpp can induce bone formation in mammalian cell culture experiments (Sampath *et al.* 1993).

The mechanism by which a specific family member elicits a particular developmental response is still unclear. The current working model for TGF-β family signaling pathways is shown in Figure 1 (reviewed in Whitman 1998). Upon secretion, TGF-β proteins are cleaved and a dimer of the C-terminal fragment is the biologically active ligand. The ligand functions through a complex of two related transmembrane receptor serine-threonine kinases. These receptors are also encoded by members of a large multigene family. The type II receptor is the primary factor in determining ligand-binding specificity. Extracellular binding of ligand by the type II receptor leads to the recruitment of an appropriate type I partner. Table 1 shows the best-characterized ligand-receptor associations. The cytoplasmic domain of the type I receptor is then phosphorylated by the type II receptor kinase. This stimulates the kinase activity of the type I receptor. The type I receptor then initiates a cytoplasmic signal transduction cascade by phosphorylating Smad proteins.
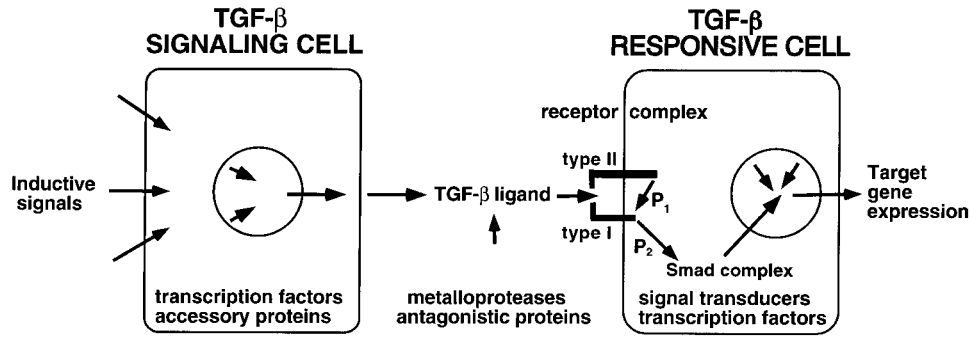
Structurally, the Smad multigene family is character-

Figure 1.—A representative TGF-β family signaling pathway. The cellular location of individual component functions is shown. The hypothetical beginning of the pathway, inductive signals received by the TGF-β signaling cell, may be either TGF-β family or non-TGF-β molecules. In TGF-β-responsive cells, P_1 indicates the phosphorylation of the type I receptor by the type II receptor in response to ligand binding. P_2 indicates the phosphorylation of Smad signal transducers by the type I receptor in response to phosphorylation. Upon phosphorylation, a complex of cytoplasmic Smad proteins translocates to the nucleus. The hypothetical end of the signaling pathway, target gene expression in the TGF-β-responsive cell, reflects transcriptional changes including gene activation and repression. A subset of these changes leads to further TGF-β family signaling (which may be autoregulatory) and/or signaling by non-TGF-β molecules. Subsequent figures correspond to specific pathway components as follows: Figure 2, ligands; Figures 3, 4, 5, and 6, receptors; Figures 7, 8, and 9, Smad signal transducers.

ized by N-terminal (MH1) and C-terminal (MH2) Mad homology domains, which are well conserved across species (reviewed in Derynck *et al.* 1998). Between these domains is a proline-rich region of variable length and sequence. The MH1 domain is required for transcriptional activity and the MH2 domain is involved in forming multi-Smad complexes. Several Smads are missing the MH1 domain. Others lack one or more of the serines of the MH2 domain, which are the target of receptor phosphorylation. Functionally, Smad family proteins play at least three roles in signal transduction. Smads can transduce the signal of a specific ligand, act in the signaling pathways of multiple ligands, or act as antagonists of ligand-dependent signal transduction.

Members of the receptor and Smad multigene families are also highly conserved in distant species. The functional conservation of TGF-β receptors is impressive.

TABLE 1

TGF-β family ligands with known receptor complex components

Ligand	Type II receptor	Type I receptor
Dpp	Punt	Sax
Dpp	Punt	Tkv
BMP2/BMP4	BMPR-II[a]	BMPR-IA (ALK3)[b]
BMP2/BMP4	BMPR-II	BMPR-IB (ALK6)
BMP7	BMPR-II	BMPR-IB (ALK6)
BMP7	BMPR-II	ActR-IA (ALK2)
BMP7	ActR-IIA	BMPR-IB (ALK6)
BMP7	ActR-IIA	ActR-IA (ALK2)
Activin	ActR-IIA	ActR-IA (ALK2)
Activin	ActR-IIA	ActR-IB (ALK4)
Activin	ActR-IIB	ActR-IA (ALK2)
Activin	ActR-IIB	ActR-IB (ALK4)
TGF-β	TBR-II	TBR-I (ALK5)

[a] Vertebrate receptor nomenclature per Massagué *et al.* (1994).

[b] Nomenclature for vertebrate type I receptors per ten Dijke *et al.* (1994a).

For example, a complex containing a type I receptor from *D. melanogaster* and a type II receptor from *C. elegans* binds a human ligand with high affinity (Brummel *et al.* 1994). Smad signal transducers can also function correctly across species. Drosophila Mad mimicked BMP2 and BMP4 signals in mesoderm induction in Xenopus embryos (Newfeld *et al.* 1996). The Xenopus Smad with the highest degree of amino acid identity to Drosophila Mad transduces BMP2 and BMP4 mesoderm-inducing signals normally (Graff *et al.* 1996). Many of the factors influencing the evolutionary diversification and functional conservation of these large multigene families are unidentified.

In this report, we utilize a phylogenetic approach to test hypotheses that address evolutionary and developmental questions about TGF-β signaling pathways. Our focus is on the amino acid sequence relationships within and between the multigene families encoding ligands, receptors, and Smads. For example, we tested the hypothesis that the duplication and divergence of ligands drives the duplication and divergence of receptors and Smads. We included the newest family members in the analyses, many of whom have not been experimentally examined, to provide clues to our fellow investigators. The relationships we identify shed new light on developmental functions for individual family members. For example, our data suggest a number of new hypotheses about the mechanisms underlying the signaling specificity of receptors.

MATERIALS AND METHODS

Database searches and amino acid sequence predictions: As of October 15, 1998, all available full-length amino acid sequences of TGF-β family ligands, receptors, and Smad proteins were obtained by BLAST. For published sequences we used the National Center for Biotechnology Information web site (www.ncbi.nlm.nih.gov/BLAST/). For additional Drosophila sequences we used the Berkeley Drosophila Genome Project and the European Drosophila Genome Project web

pages (fruitfly.berkeley.edu/blast/ and edgp.ebi.ac.uk/www-blast.html/). For additional *C. elegans* sequences we used an AXYS Pharmaceuticals BLAST server. To obtain new *C. elegans* and Drosophila amino acid sequences the GeneScan (Burge and Karlin 1997) and GeneFinder programs (Solovyev and Salamov 1997) were used to predict open reading frames from genomic sequence. Specific sequences, accession numbers, and evidence supporting our use of an outgroup are given in the respective figure legend.

Sequence alignments and phylogenetic analyses: Amino acid sequences were aligned using CLUSTAL V (Higgins *et al.* 1992). Gaps in this alignment were modified to minimize the number of mutations required to explain all differences between the sequences. The maximum-likelihood method in PUZZLE (Strimmer and von Haeseler 1997) was used to identify orthologous sequences from different vertebrate species. A single representative of each vertebrate orthology group was used in the analyses. Evolutionary divergences (number of amino acid substitutions per site) were estimated by the Poisson correction distance to account for multiple substitutions at the same site. Phylogenetic analyses were conducted using (1) neighbor joining (Saitou and Nei 1987) in MEGA (Kumar *et al.* 1993); (2) maximum parsimony (100 replications of the random sequence addition and TBR search algorithms) in PAUP* (Swofford 1998); and (3) the Fitch-Margoliash algorithm (Fitch and Margoliash 1970) in PHYLIP (Felsenstein 1993). The degree of confidence for each branchpoint was obtained by the bootstrap method (1000 replications; Felsenstein 1985). We also conducted an interior branch test based on the minimum evolution principle. In this test, the confidence probability (CP) that the length of a given branch is not equal to zero is estimated by using the Z-test (Rzhetsky and Nei 1993; Dopazo 1994).

RESULTS

Phylogenetic trees generated by the neighbor joining, maximum parsimony, and Fitch-Margoliash methods were largely congruent. As expected (Nei *et al.* 1998), some differences were found for weakly supported nodes. The neighbor-joining trees are presented. For a given branchpoint, the indicated bootstrap value is the percentage of replicates in which the branch is reconstructed. We have used nodes with ≥75% bootstrap support in drawing inferences for two reasons. First, the bootstrap method is conservative (Hillis and Bull 1993; Sitnikova 1996). Second, the amount of data available for inferring relationships within a multigene family is limited by the length of the gene as compared to studies of species evolution where many genes can be used to address the same problem. We also report the CP value for important nodes with bootstrap values <75% to support our proposed relationships. In these cases, we rely on a CP value above >75% to support the branchpoint. Other nodes are preliminary.

Phylogenetic analysis of ligands: Figure 2 shows a phylogeny of TGF-β ligands. Several of the ligands are open reading frame predictions from sequences submitted by genome projects. These are indicated in Figure 2 by clone numbers rather than gene names. For example, we predict a new family member in Drosophila from genomic clone DS07149.

In its overall topology, our analysis agrees with previous studies (*e.g.*, Burt and Law 1994). Figure 2 displays two large subfamilies (CP = 92), the TGF-β/activin subfamily (cluster A) and the Dpp/BMP subfamily (cluster B). Within cluster B, our topology also agrees with others (*e.g.*, Miya *et al.* 1997), showing a distinction between two clusters of sequences (CP = 75). These are the 60A/BMP cluster (cluster C) and the Dpp/BMP cluster (cluster D). The demonstrated functional interchangeability of Dpp with BMP2 and BMP4 in flies and mammalian cells (Padgett *et al.* 1993; Sampath *et al.* 1993) provides a biological foundation for cluster D. No similar experiments have been reported for cluster C.
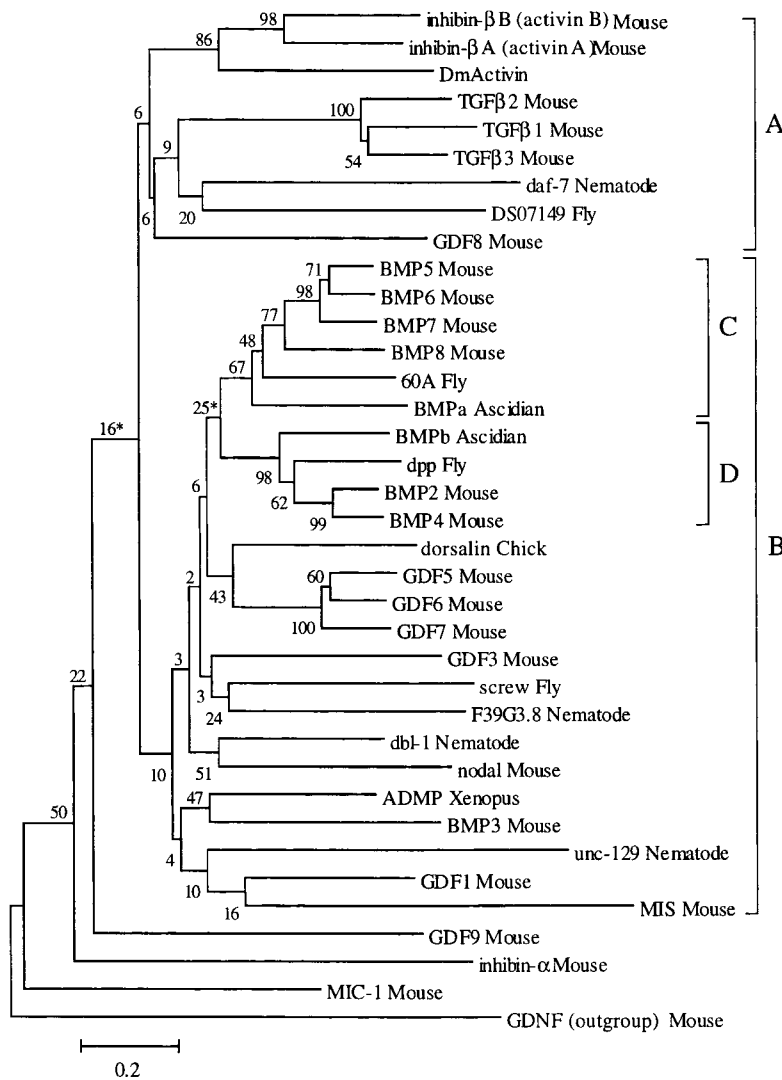
Phylogenetic analysis of type I receptors: Figure 3 shows a phylogeny of type I receptors. The rate of amino acid substitution for type I receptors is roughly the same as for the ligands. The inferred relationships within and between clusters A, B, and C are well supported. Interestingly, both *C. elegans* type I receptors, C32D5.2 (Sma-6; Krishna *et al.* 1999) and Daf-1, form distinct lineages.

The strength of the clustering of the Drosophila Dpp receptor Sax with human ALK1 and ALK2 (Figure 3, cluster A) was unexpected. ALK1 has recently been shown to be a TGF-β receptor (P. ten Dijke, unpublished data) and ALK2 is an activin receptor (ten Dijke *et al.* 1994b). Yet Dpp is in the Dpp/BMP subfamily of ligands and not the TGF-β/activin subfamily. Cluster A was previously noted by Brummel *et al.* (1994) and several others at the time that Dpp receptors were first identified. However, in view of the prevailing notion that there were no TGF-β/activin subfamily ligands in invertebrates this result was not followed up by those investigators. Cluster A raises new questions. Could Sax also be a receptor for the recently identified Drosophila TGF-β/activin subfamily ligand (Kutty *et al.* 1998), and if so, is that its primary function?

The strength of the relationships in cluster B was also unexpected. A relationship between DmATR-1 and ALK4 and ALK5 was previously shown and each of these receptors can bind activin *in vitro* (Wrana *et al.* 1994). However, the supposed type II partner of DmATR-1, DmATR-II, which also binds activin *in vitro*, was subsequently shown to be encoded by the *punt* gene and to act as a Dpp receptor *in vivo* (Ruberte *et al.* 1995). Thus, the cluster C relationship was not followed up by those investigators. Questions about the *in vivo* role of DmATR-1 in signaling for Dpp/BMP or TGF-β/activin subfamily ligands are unanswered. The analysis of mutations in the *baboon* gene that encodes DmATR-1 (Brummel *et al.* 1999) will likely prove very informative.

The clustering of the other Dpp type I receptor Thickveins (Tkv) with the two vertebrate BMP receptors (ALK3 and ALK6; ten Dijke *et al.* 1994a) has not been reported previously (Figure 3, cluster C). Though this seems logical on the basis of the clustering of their respective ligands (Dpp and BMP2 and BMP4), Tkv had

Figure 2.—Phylogenetic relationship of TGF-β family ligands. The criteria for including a sequence in this analysis were (1) if a mammalian sequence was available for a group of likely vertebrate homologs [*e.g.*, nodal is present in mouse (Zhou *et al.* 1993), Xenopus (Jones *et al.* 1995), and chick (Levin *et al.* 1995)] and orthologs [*e.g.*, there are two additional nodal-related sequences in Xenopus (Jones *et al.* 1995; Smith *et al.* 1995)] we used the mouse sequence; (2) any distinctive vertebrate sequences without a mammalian counterpart (*e.g.*, dorsalin and ADMP) were included; and (3) all invertebrate sequences were included. The alignment was performed using the C-terminal ligand region only. For the analysis, the ligand was defined by the first invariant cysteine and the stop codon. The length of the alignment was 113 amino acids. Glial cell line-derived neurotrophic factor (GDNF) was chosen to root the tree because it shares only the pattern of cysteines with other TGF-β members (Lin *et al.* 1993) and uses a novel receptor (Jing *et al.* 1996). The number at each branchpoint represents the relative incidence of that particular relationship (in percent) during bootstrap resampling using 1000 replicates. Branch lengths are drawn to scale, and a scale bar is shown, based upon the number of amino acid substitutions per site between the two sequences. Clusters of sequences showing very strong relationships are given letter designations. An asterisk indicates that the CP value for that node is reported in the text. The accession numbers are as follows: TGFβ1, SWISS-PROT P04202; TGFβ2, SWISS-PROT P27090; TGFβ3, PIR A41397; GDF1, SWISS-PROT P20863; GDF3, SWISS-PROT Q07104; GDF5, PRF 2009388A; GDF6, SWISS-PROT P43028; GDF7, SWISS-PROT P43029; GDF8, SWISS-PROT O08689; GDF9, SWISS-PROT Q07105; BMP8, SWISS-PROT P34821; BMP7, SWISS-PROT P23359; BMP6, SWISS-PROT P20722; BMP5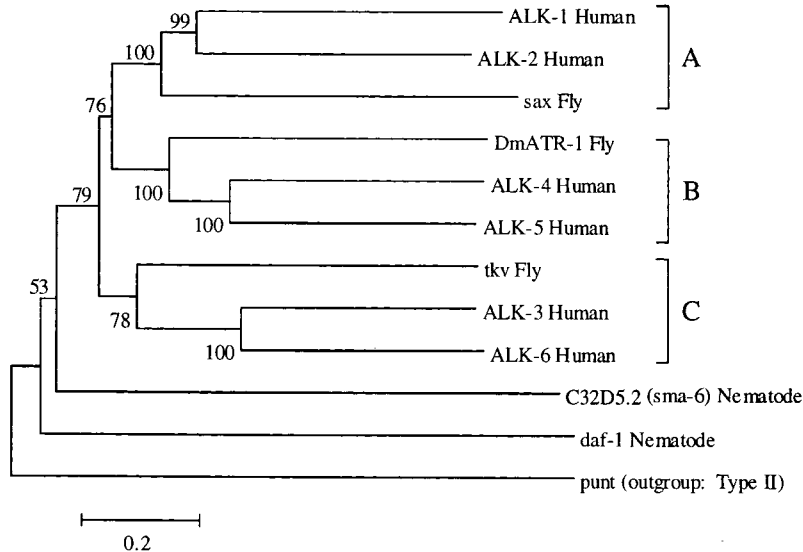, SWISS-PROT P49003; BMP4, SWISS-PROT P21275; BMP3, SWISS-PROT P97737; BMP2, SWISS-PROT P21274;BMPa, DDBJ D83183; BMPb, DDBJ D85464; MIS, SWISS-PROT P27106; MIC-1, GenBank AF019770; inhibin-βB (activin B), SWISS-PROT Q04999; inhibin-βA (activin A), SWISS-PROT Q04998; inhibin-α, SWISS-PROT Q04997; nodal, SWISS-PROT P43021; GDNF, SWISS-PROT P48540; dorsalin, PIR A40735; ADMP, GenBank U22155; Dpp, SWISS-PROT P07713; 60A, SWISS-PROT P27091; Screw, SWISS-PROT P54631; Dactivin, GenBank AF054822; DS07149, our GenScan/GeneFinder prediction from genomic sequence, GenBank AC004120; Daf-7, GenBank U80953; Dbl-1, GenBank AF004395; Unc-129, GenBank AF029887; F39G3.8, GenBank AF016424.

been reported as dissimilar to any other type I receptor (Ruberte *et al.* 1995). On a larger scale, cluster A in Figure 3 is more closely related to cluster B than to cluster C, perhaps reflecting the fact that ALK1 and ALK5 bind TGF-β and ALK2 and ALK4 bind activin (ten Dijke *et al.* 1994b). This suggests that Sax may have a dual role, binding ligands from the Dpp/BMP and TGF-β/activin subfamilies.

Phylogenetic analysis of type II receptors: Figure 4 shows a phylogeny of type II receptors. The rate of amino acid substitution is roughly the same for ligands and type I receptors. The evolutionary relationships within clusters A and B are well supported. Again a receptor from *C. elegans* (Daf-4) forms a distinct lineage.

Within cluster A, a new type II receptor from Drosophila called Wishful thinking (Wit; M. O'Connor, unpub-

lished data) clusters strongly with BMPR-II, suggesting that it may signal for one of the Drosophila Dpp/BMP subfamily ligands (Dpp, 60A, or Screw). The close relationship between these two receptors and MIST-II was unexpected. The relationship in Figure 4, cluster B, between Punt (previously known as DmATR-II) and ActR-IIA/ActR-IIB (activin receptors) has been widely reported. Punt was shown to bind activin *in vitro* (Childs *et al.* 1993). However, as noted above, Punt acts as a Dpp receptor *in vivo* and the sequence similarity between Punt and ActR-IIA and ActR-IIB was not followed up at that time. Now that a Drosophila activin-like ligand has been identified, it seems possible that Punt has a dual role. Punt may bind ligands from the Dpp/BMP and TGF-β/activin subfamilies.

Phylogenetic analysis of the cytoplasmic domain of all

Figure 3.—Phylogenetic relationship of TGF-β family type I receptors. The criteria were (1) if a human sequence was available for a group of likely vertebrate homologs [*e.g.*, ALK3/BMPR-IA is present in human (ten Dijke *et al.* 1994a), mouse (Koenig *et al.* 1994), Xenopus (Graff *et al.* 1994), and chick (Zou *et al.* 1997)] we utilized the human sequence; and (2) all invertebrate sequences were included. The length of the alignment was 812 amino acids. The type II receptor Punt was chosen to root the tree. The accession numbers are as follows: ALK1, SWISS-PROT P37023; ALK2, SWISS-PROT Q04771; ALK3, SWISS-PROT P36894; ALK4, SWISS-PROT P36896; ALK5, SWISS-PROT P36897; ALK6, GenBank U89326; DmATR-1a, GenBank U04692; DmATR-1b, M. O'Connor, unpublished data; Sax, PIR I45712; Tkv, GenBank L33475; Daf-1, SWISS-PROT P20792; C32D5.2 (Sma-6; Krishna *et al.* 1999), SWISS-PROT Q09488.

receptors: Given that the two ends of the receptors have completely distinct environments and roles in TGF-β pathways (ligand binding *vs.* kinase signaling), we wondered if there were any difference in the evolutionary forces affecting the extracellular and the cytoplasmic domains of the receptors. To address this issue we conducted an analysis of each domain alone. The rate of amino acid substitution in the cytoplasmic domain is roughly the same as the entire receptor. The topology of the cytoplasmic tree is identical to the trees of type I receptors and type II receptors with minor differences in bootstrap values. Note that Figure 5, clusters A, B, and C, are the same as Figure 3, clusters A, B, and C, and that Figure 5, clusters D and E, are the same as Figure 4, clusters A and B.

The overall picture gained from this analysis is that the type I and type II receptors, both transmembrane serine-threonine kinases, form a single large lineage. The tree does not bifurcate into two monophyletic groups representing each receptor type. Within our tree the type I receptors form a group excluding type II

receptors. This likely reflects the distinct roles of the receptors (type I, phosphorylate Smads; and type II, phosphorylate type I receptors). The relative placement of the receptors in our tree suggests that type I receptors diverged from an ancestral type II receptor. The *C. elegans* type I receptors fall at the boundary of type I and type II receptors, perhaps representing a transitional receptor type.

Phylogenetic analysis of the extracellular domain of all receptors: In general, the phylogeny of the extracellular domain (Figure 6) has low bootstrap values on many of the branchpoints. This suggests that there is more divergence between the receptor sequences on the ligand-binding side than on the signaling side. The rate of amino acid substitution in the extracellular domain is roughly the same as for the entire receptor. With two exceptions, the overall topology of the relevant portion of the extracellular tree (top half of Figure 6) is congruent with the tree of type I receptors (Figure 3). One exception is that Sax breaks the monophyly of ALK1 and ALK2 (Figure 6, cluster A). However, this
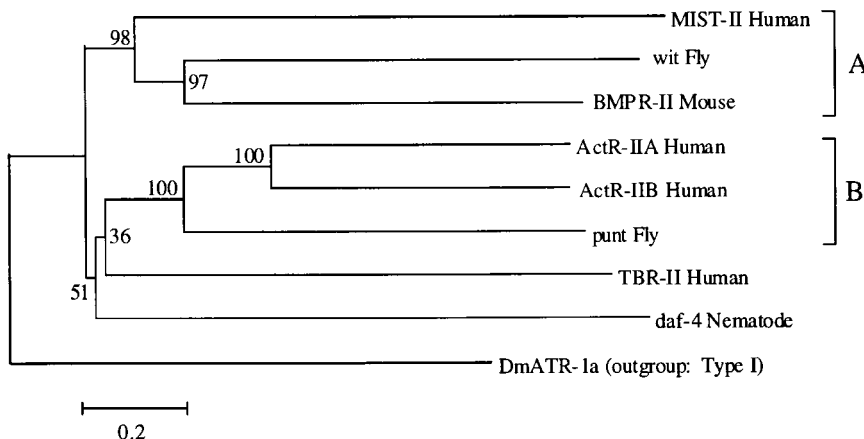

Figure 4.—Phylogenetic relationship of TGF-β family type II receptors. The criteria were (1) if a mammalian sequence was available for a group of likely vertebrate homologs [*e.g.*, BMPR-II is present in human (Liu *et al.* 1995), mouse (Suzuki *et al.* 1994), and Xenopus (Frisch and Wright 1998)] we utilized either the human or the mouse sequence; and (2) all invertebrate sequences were included. The length of the alignment was 1080 amino acids. The type I receptor DmATR-1a was chosen to root the tree. The accession numbers are as follows: ACTR-IIA, SWISS-PROT P27037; ACTR-IIB, SWISS-PROT Q13705; BMPR-II, GenBank U78048; MIST-II, PIR JC4335; TBR-II, SWISS-PROT P37173; Punt, GenBank L38495; Daf-4, GenBank L23110; Wit (M. O'Connor, unpublished data).
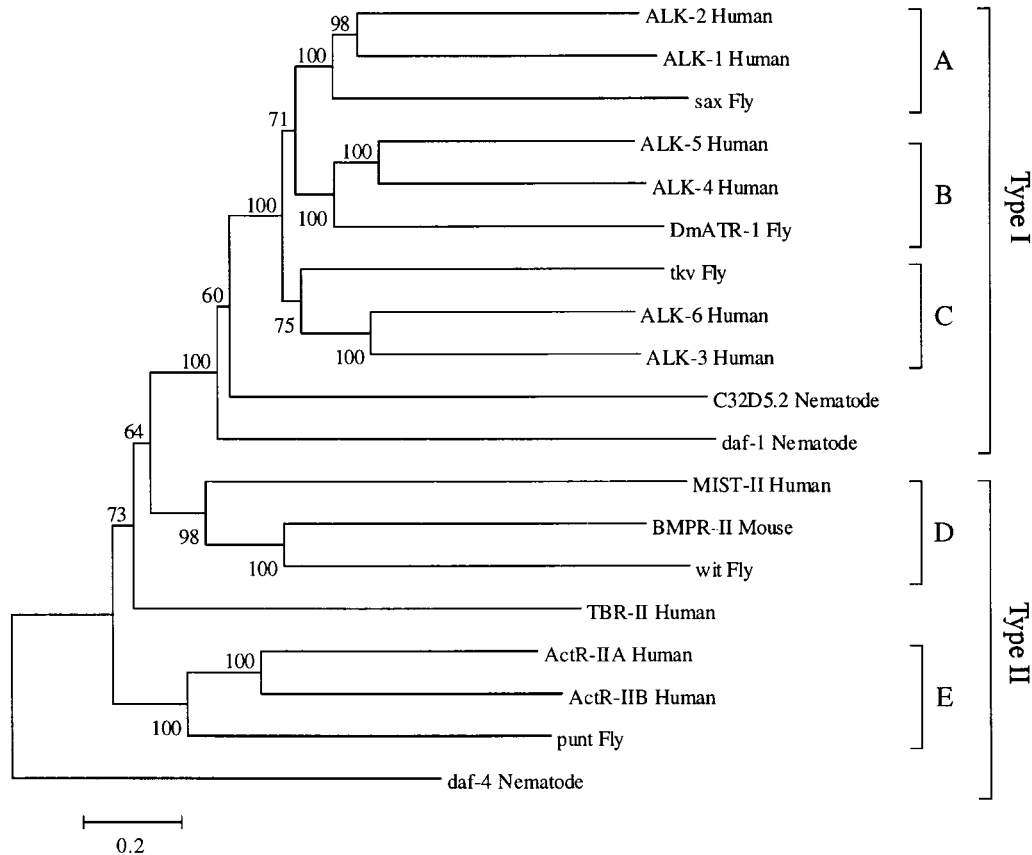
Figure 5.—Phylogenetic relationship of the cytoplasmic domain of all TGF-β family receptors. All type I and type II receptor amino acid sequences from Figures 3 and 4 are included. The alignment was performed using the transmembrane and cytoplasmic regions. The aligned sequences begin at a run of hydrophobic residues and end at the stop codon. The length of the alignment was 900 amino acids. The protein isoforms DmATR-1a and DmATR-1b have identical cytoplasmic domains and are not listed separately. The tree is unrooted.

unique relationship is not strongly supported. Another exception is the secondary connection of the DmATR-1 cluster (Figure 6, cluster B) with the Tkv cluster (Figure 6, cluster C). In the type I tree the DmATR-1 cluster is closest to the Sax cluster. Compare Figure 3 (clusters A and B) with Figure 6 (clusters B and C). While the bootstrap values of these secondary/tertiary clusters are low, the CP values are high (cluster B and C node: CP = 86; cluster A node with cluster B/C: CP = 85).

A number of terminal clusters in the extracellular tree (Figure 6) are different from the tree of type II receptors (Figure 4). In the extracellular tree, Wit is clustered with MIST-II (CP = 91; Figure 6, cluster D) instead of with BMPR-II (Figure 4, cluster A). The two activin receptors and BMPR-II are clustered (CP = 76; Figure 6, cluster E) instead of being distantly related (Figure 4, cluster A vs. cluster B). In Figure 6, Punt is distinct from a node containing clusters D and E rather than branching specifically from the two activin receptors I (Figure 4, cluster B).

Just as in the cytoplasmic tree (Figure 5), the type I receptors form a group of closely related sequences exclusive of the type II receptors. Why do type I and type II receptors that bind the same ligand not cluster together? Note the phylogenetic distance between the Dpp-signaling heteromeric partners Tkv/Punt and Sax/Punt. Perhaps there are factors that bind to only one receptor type that prevent the sequence convergence of type I and type II receptors with a common ligand.

Phylogenetic analysis of Smad signal transducers: Figure 7 shows a phylogeny of Smad signal transducers. Note that the rate of amino acid substitution in the Smad family is roughly 2.5-fold faster than for ligands and receptors. The analysis includes several Smads whose relationships to other family members have not been reported. For example, DSmad2, a new family member in Drosophila (Brummel et al. 1999), clusters with mammalian Smad2 and Smad3 (Figure 7, cluster B). Both mammalian Smads signal for TGF-β/activin subfamily ligands. This result supports the proposal (Brummel et al. 1999) that there is a TGF-β/activin subfamily signaling pathway in Drosophila.

In its overall topology, our tree generally agrees with previous reports demonstrating the existence of three distinct subfamilies (e.g., Wisotzkey et al. 1998). Using Drosophila subfamily members as representatives, these are the Mad subfamily (dedicated to one ligand; Figure 7, cluster E), the Med subfamily (signal for multiple ligands; cluster A), and the Dad subfamily (antagonist; cluster D). These clusters correspond to roles that Smad proteins play in TGF-β signal transduction. However, C. elegans Daf-3, which appears as a highly divergent lineage, may be an exception. Genetically, Daf-3 acts as an antagonist (Patterson et al. 1997) and yet binds DNA like a signal transducer (Thatcher et al. 1999). In addition, we unexpectedly included mammalian Smad8 with cluster C. Smad8 signals for the TGF-β/activin subfamily (Chen et al. 1997), while the other Smads in
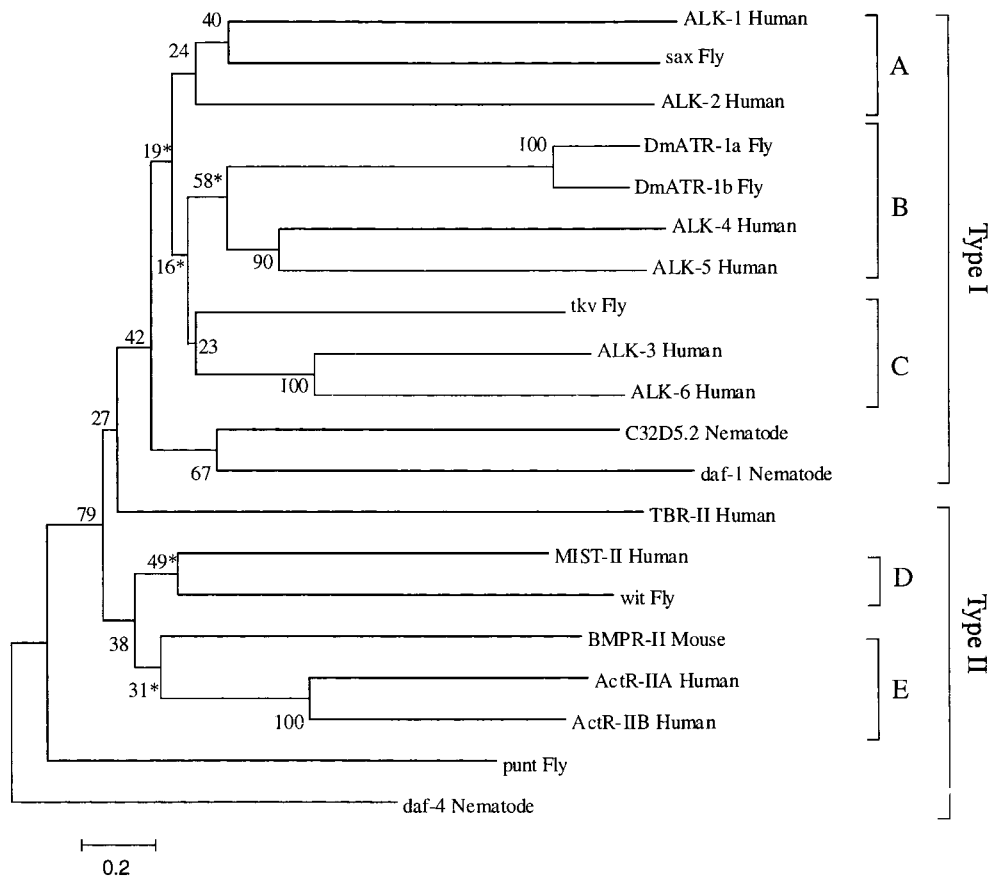
Figure 6.—Phylogenetic relationship of the extracellular domain of all TGF-β family receptors. All type I and type II receptor amino acid sequences from Figures 3 and 4 are included. The alignment does not include the transmembrane domain. The extracellular domain begins at the start codon and ends just before the run of hydrophobic residues. The length of the alignment was 295 amino acids. The Dm-ATR-1a and DmATR-1b protein isoforms contain alternative exons in the extracellular domain (Wrana *et al.* 1994). An asterisk indicates that the CP value for that branchpoint is reported in the text, except for the basal node in cluster B. CP = 75 for this node.

cluster C signal for the Dpp/BMP subfamily (Yingling *et al.* 1996). This suggests that Smads do not cluster according to their ligands.

Other family members derive from open reading frame predictions generated by the recently completed *C. elegans* genome project. Two of these predictions have been connected to genes. These are R05D11.1, which is Daf-8 (D. Riddle, unpublished data), and F01G10.8, which is Daf-14 (J. Thomas, unpublished data). These *C. elegans* Smads do not show close evolutionary relationships to other family members. As unique lineages, both Daf-8 and Daf-14 fall roughly between cluster D and cluster E (Figure 7). We predict that F37D6.a is another family member in *C. elegans.* This open reading frame is mispredicted in the database (see Figure 7 legend). F37D6.a clusters well with antagonist Smads. Thus, *C. elegans* has 7 Smad family members. The fact that there are 8 mammalian Smads and 4 Smads in Drosophila suggests that there are more Smads to be found in these experimental systems. This proposal is supported by comparing the number of mammalian ligands (24) with those in nematodes (4) as shown in Figure 2.

In many new multigene families the identification of the same gene in different species (homologs) is often difficult. This results in the occasional misidentification of genes. We noted one such instance during our analysis involving Xenopus Smad8 (Nakayama *et al.* 1998) and mammalian Smad7 (Imamura *et al.* 1997).

Our maximum-likelihood analysis revealed that Xenopus Smad8 is extremely similar to mammalian Smad7 (data not shown). In addition, both genes share many structural and functional features. These include the absence of the receptor phosphorylated C-terminal serines, ligand-dependent transcription, and an antagonist role (Nakayama *et al.* 1998). Because we found no database references to Xenopus Smad7 as of our cutoff date, we propose that Xenopus Smad8 is the homolog of mammalian Smad7.

Phylogenetic analysis of the conserved domains of Smad signal transducers: The Smad family is characterized by N-terminal (MH1) and C-terminal (MH2) domains that are well conserved between species. Given that the two domains appear to have distinct roles in signal transduction (MH1, DNA binding/transcriptional activation; MH2, Smad complex formation) we wondered if there were any difference in the evolutionary forces affecting the two domains. To address this issue we conducted an analysis of each domain alone. As with the receptors above, we were interested to see whether the phylogenetic relationships among the individual domains showed any differences from each other or from the Smad tree.

The role of the MH2 domain in Smad complex formation is a highly conserved function in all Smad family members (reviewed in Whitman 1998). Our MH2 domain alignment began at the invariant tryptophan
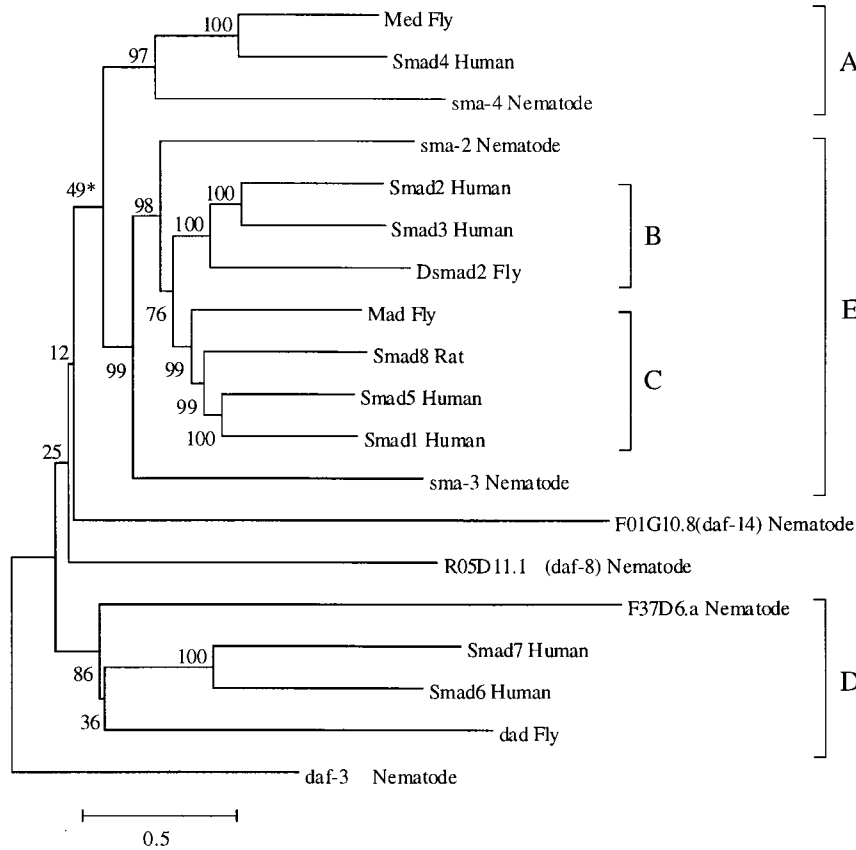
Figure 7.—Phylogenetic relationship of Smad family signal transducers. The criteria were (1) if a human sequence was available for a set of likely vertebrate homologs [*e.g.*, Smad1 is present in human (Hoodless *et al.* 1996), mouse (Yingling *et al.* 1996), and Xenopus (Graff *et al.* 1996)] we used the human sequence; (2) any distinctive vertebrate sequences without a human counterpart [*e.g.*, rat Smad8] were included; and (3) all invertebrate sequences were included. The length of the alignment was 994 amino acids. An asterisk indicates that the CP value for that branchpoint is reported in the text. The tree is unrooted. The accession numbers are as follows: Smad1, PIR S68987; Smad2, GenBank U59911; Smad3, GenBank U76622; Smad4, GenBank U44378; Smad5, GenBank U73825; Smad6, GenBank AF043640; Smad7, GenBank AF015261; Smad8, GenBank AF012347; Mad, SWISS-PROT P42003; Med, GenBank AF027729; Dad, DDBJ AB004232; DSmad2, GenBank AF101386; Sma-2, SWISS-PROT Q02330; Sma-3, SWISS-PROT P45896; Sma-4, SWISS-PROT P45897; Daf-3, GenBank AF005205; R05D11.1 (Daf-8; D. Riddle, unpublished data), EMBL Z75546; F01G10.8 (Daf-14; J. Thomas, unpublished data), EMBL Z81055; F37D6.a, our GenScan/ GeneFinder prediction from genomic sequence. This protein is mispredicted as three proteins (F37D6.6, F37D6.7, F59C6.10) spanning two cosmids (F37D6, EMBL Z75540 and F59C6, EMBL Z79600).

(amino acid 261 in Drosophila Mad) and ended at the stop codon. The only biologically meaningful distinction between Smad sequences that we detect in this region is the number of C-terminal serines (the target of receptor phosphorylation; data not shown). There are two or three serines in Smads dedicated to one ligand (Figure 7, cluster E, plus Daf-8 and Daf-14) and zero or one serine in Smads signaling for multiple ligands and antagonist Smads (Figure 7, clusters A and D, and Daf-3). In comparison with the Smad tree (Figure 7) there are just two minor differences (data not shown). First, Daf-8 moves from a monophyletic lineage between clusters E and D in Figure 7 to between clusters A and E in the MH2 tree. Second, F37D6.a clusters with Dad in the MH2 tree instead of appearing as a divergent antagonist in the Smad tree (Figure 7, cluster D).

The MH1 domain shows wide sequence variation among Smad family members. Figure 8 shows our alignment of this domain using 19 Smad sequences. In the alignment, some portion of the MH1 domain (particularly amino acid numbers 150–160) is recognizable in every sequence except for Daf-14, F37D6.a, and Dad. Smad6 aligns well and is not missing all or part of the MH1 domain as previously reported. The domain is divided into subregions by unique insertions in a num-

ber of Smads. The biological role of these insertions, if any, is unknown.

There are only two absolutely invariant amino acids in the MH1 domain. One is cysteine 45 and the other is proline 153 of the alignment. These amino acids are affected in mutant alleles of Med (Wisotzkey *et al.* 1998) and Smad4 (Thiagalingam *et al.* 1996), respectively. Mutations have been identified in two amino acids that are present in 18 of the 19 sequences. Proline 101 is affected in a Daf-3 allele (Patterson *et al.* 1997) and arginine 117 is affected in alleles of Smad2 (Eppert *et al.* 1996) and Smad4 (Hahn *et al.* 1996). Three additional cysteines (amino acids 114, 137, and 151) are found in 11, 18, and 16 sequences, respectively. The conservation of cysteines capable of disulfide bond formation reinforces the idea that protein complexes containing multiple Smads are essential in TGF-β signal transduction.

A phylogeny generated from the MH1 alignment is shown in Figure 9. Several clusters are present in the Smad and the MH1 trees. For example, Figure 9, clusters A (Med), B (DSmads), and E (signal transducing), are the same as Figure 7, clusters A, B, and E. There are several differences between the trees. In the MH1 tree, Daf-8 and Daf-14 form a cluster and appear more divergent than in the Smad and MH2 trees. Their placement
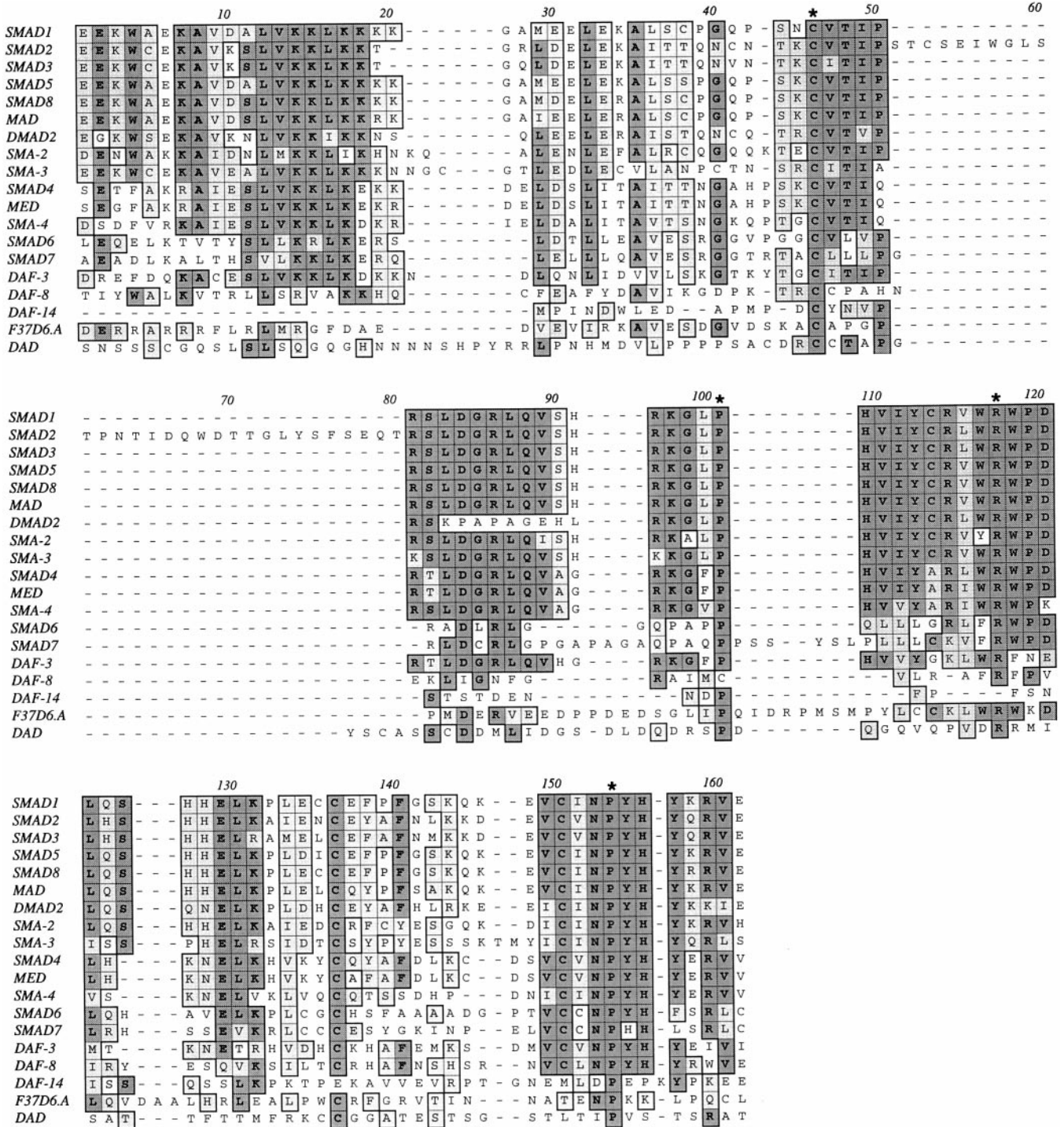
Figure 8.—Alignment of the MH1 domain of all Smad family members. The MH1 domain begins with the conserved aspartic/glutamic acid (39 in Drosophila Mad) and ends at the conserved valine/leucine/isoleucine (144 in Drosophila Mad). Numbers above the alignment begin with the first amino acid and indicate the presence of an amino acid in any sequence. Highlighted amino acids are identical (dark shading and bold print) or similar (light shading and normal print) in a majority (10 of 19) of aligned sequences. Similar amino acids are defined by Higgins *et al.* (1992). An asterisk indicates that a mutation has been identified in that amino acid as follows: *Med* (C[45]), Daf-3 (P[101]), Smad2 and Smad4 (R[117]), and Smad4 (P[153]).

near the outgroup in the MH1 tree likely reflects their lack of (Daf-14) or a minimal (Daf-8) MH1 domain. Interestingly, F37D6.a, which also does not have an obvious MH1 domain, does not move away from its strong

secondary cluster (Figures 7 and 9, cluster D) with the antagonists Smad6 and Smad 7, both of which have recognizable MH1 domains. Daf-3 clusters very strongly with the Med subfamily in the MH1 tree (Figure 9,
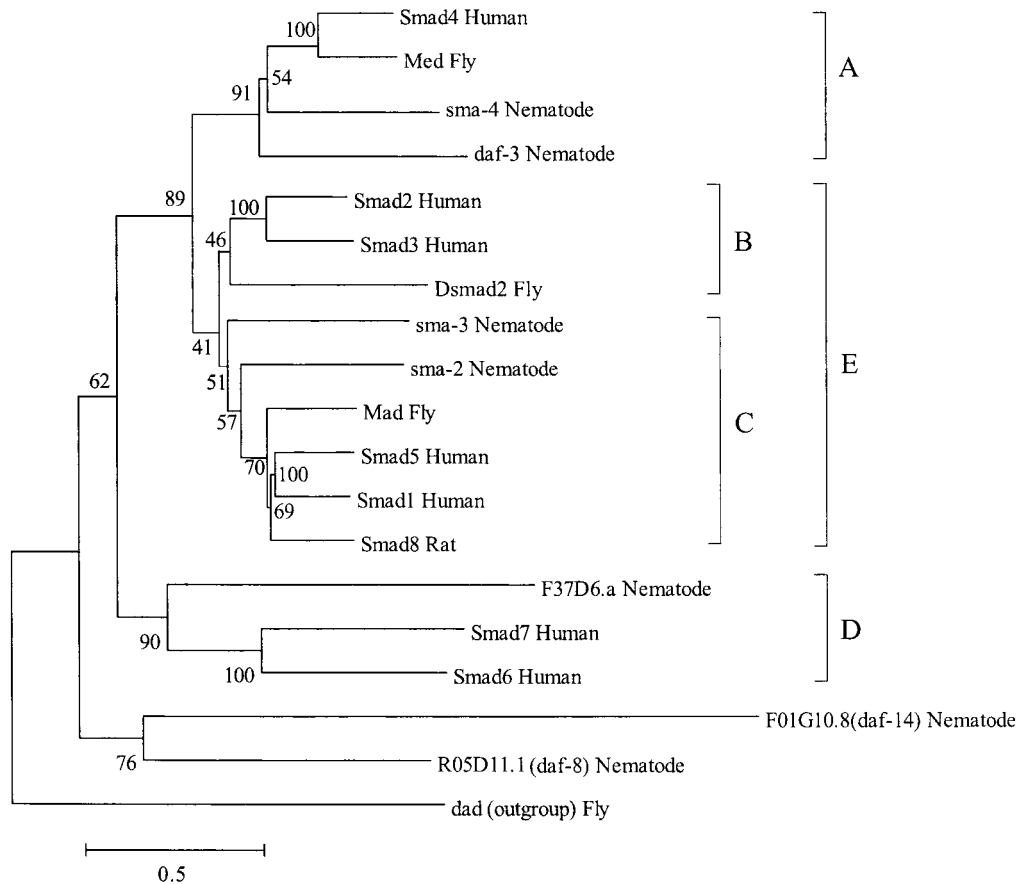
Figure 9.—Phylogenetic analysis of the MH1 domain of all Smad family members. An analysis of all Smad sequences from Figure 8. The tree is rooted with the Drosophila antagonist Dad. According to the MH1 alignment in Figure 8, Dad has no MH1 domain. The length of the alignment was 163 amino acids.

cluster C) instead of as a highly divergent monophyletic lineage. Interestingly, both sma-2 and sma-3 move into cluster C (Mad) in the MH1 tree.

DISCUSSION

Our phylogenetic analyses of three physically interacting multigene families involved in TGF-β signaling provide a number of new insights into the molecular evolution and developmental biology of intercellular communication. One valuable set of observations is derived from our studies of the newest family members, many of whom have not been experimentally examined. These include seven new ligands, two new receptors, and five new Smads. Another set of results with wide relevance are the relationships we detect between sequences not previously connected to each other, such as the clustering of the type I receptors Sax/ALK1/ALK2.

Calibrating our phylogenies for ligands (Figure 2) and Smads (Figure 7) with an arthropod-nematode divergence of 1.1 billion years ago (Wang *et al.* 1999), we were able to roughly date the origin of these two multigene families. Notwithstanding the difference in their amino acid substitution rates, both families appear to originate between 1.2 and 1.4 billion years ago (data not shown). For perspective, Wang *et al.* (1999) dated the divergence of plants, animals, and fungi to 1.6 bil-

lion years ago. From this dating scheme, the TGF-β family likely exists solely in animals.

One important evolutionary question that can be addressed by our data is whether the duplication and divergence of ligands drives the duplication and divergence of receptors and Smads. Our data indicate that the answer to this question is complex. Examination of the recent history of the receptor and Smad families (*e.g.*, terminal clusters) indicates that each appears to be evolving independently or under the influence of factors other than the ligand. Evidence for this comes from three sources: the clustering of receptor extracellular domains by type and not by ligand and the clustering of Smads with different ligands or different roles.

Early in the history of TGF-β signaling, ligand duplication and subsequent sequence diversification may have been a powerful force in shaping the molecular evolution of receptors and Smads. Our data show no convincing evidence for a TGF-β/activin subfamily signaling pathway in *C. elegans.* The bootstrap values for branches placing Daf-7 in the TGF-β/activin subfamily are extremely low (Figure 2) and there are no nematode receptors (Figures 3 and 4) or Smads (Figure 7) in clusters known to signal for TGF-β/activin subfamily ligands. Two explanations are possible. First, the TGF-β/activin subfamily arose after the divergence of nematodes and arthropods. Second, a preexisting TGF-β/activin subfamily signaling pathway was lost in the *C. elegans* lin-

eage. We suggest that TGF-β/activin subfamily signaling arose after the separation of arthropods and nematodes but before the separation of arthropods and vertebrates (1.1 billion *vs.* 950 million years ago; Wang *et al.* 1999). In this window of 150 million years a complete pathway was generated because flies and vertebrates have complete TGF-β/activin subfamily signaling pathways.

A global phylogenetic tree containing a clade of arthropods and nematodes diverging from a common ancestor with vertebrates (Aguinaldo *et al.* 1997) is difficult to reconcile with our proposal for the origin of the TGF-β/activin subfamily. However, Aguinaldo *et al.* (1997) rely upon 18S rDNA sequences from nematodes outside the genus Caenorhabditis and from arthropods outside Drosophila. Our invertebrate data come exclusively from these two genera (except for two ascidian ligands) and are based upon amino acid sequences. We believe that the global phylogeny generated by Wang *et al.* (1999), based upon the amino acid sequences of 75 nuclear genes including Drosophila and *C. elegans* sequences, is a more appropriate reference for our data.

A subset of TGF-β signaling pathways in nematodes may approximate an ancestral TGF-β pathway. Evidence for this hypothesis comes from our analysis of *C. elegans* receptors. Two receptors, Daf-1 (type I; Georgi *et al.* 1990) and Daf-4 (type II; Estevez *et al.* 1993), have been assigned to a receptor type on the basis of *in vitro* studies, while Sma-6 (Krishna *et al.* 1999) has only recently been identified. However, our phylogenetic relationships place all of them in ambiguous positions with regard to receptor type. We show that Daf-1 and Sma-6 are the most divergent type I receptors (Figure 3) and mark the boundary between type I and type II receptors in both the cytoplasmic tree (Figure 5) and the extracellular tree (Figure 6). Daf-4 weakly clusters with TGF-β/activin subfamily receptors in the type II tree (Figure 4) but appears as the most divergent receptor in the cytoplasmic tree (Figure 5) and the extracellular tree (Figure 6). These results suggest that these receptors are not phylogenetically tied (statistically speaking) to a specific receptor type. To test this idea, it would be interesting to see if these receptors are able to signal as homomultimers instead of requiring the "standard" heteromultimeric configuration.

The possibility that Daf-1 and Daf-4 resemble an ancestral "nonspecialized" receptor type led us to compare the literature on other TGF-β family receptors with our data. Perhaps we could identify other receptors that appear "less specialized." Given that receptors have two domains, there are at least three ways in which a receptor can be "nonspecific." A receptor could have relationships with two ligands, two Smads, or bind a single ligand but signal through another ligand's Smad. We identified two such receptors in Drosophila.

The type I receptor Sax signals for Dpp *in vivo*, although its role in Dpp signaling is less significant than Tkv (Singer *et al.* 1996). Sax also binds the human Dpp

homolog BMP2 *in vitro* (Brummel *et al.* 1994). In a clear example of a receptor associating with multiple ligands, Sax also signals *in vivo* for the Drosophila ligands 60A and Screw (Haerry *et al.* 1998; Nguyen *et al.* 1998). None of these Drosophila ligands are in the TGF-β/activin subfamily (Figure 2). However, in all of our receptor trees Sax clusters with ALK1 (a TGF-β receptor) and ALK2 (an activin receptor). The bootstrap value of this association is 100% in the type I tree (Figure 3) and the cytoplasmic tree (Figure 5) but much less in the extracellular tree (Figure 6). Our finding raises the possibility that Sax binds multiple ligands but then signals through TGF-β/activin subfamily Smads. Support for this hypothesis comes from our receptor trees. In the cytoplasmic tree (Figure 5), the Sax cluster forms a secondary cluster with the DmATR-1 cluster. Receptors in the DmATR-1 cluster signal for TGF-β/activin subfamily ligands. In the extracellular tree (Figure 6), the Sax cluster is more divergent, forming a tertiary cluster with the Tkv and DmATR-1 clusters.

A type II receptor with multiple reported ligands is Punt. This receptor binds activin with high affinity *in vitro*, but subsequently was shown to function in Dpp signaling *in vivo.* In the type II tree (Figure 4) and the cytoplasmic tree (Figure 5), Punt clusters with activin receptors with a bootstrap value of 100%. However, in the extracellular tree (Figure 6), Punt appears as a divergent lineage. The data suggest that Punt binds ligands from both subfamilies and is able to signal to both Dpp/BMP and TGF-β/activin subfamily type I receptors. Overall, our data for receptors suggest more "non-specificity" in ligand binding and more flexibility in signaling than had been suspected previously.

Given this potential versatility of individual receptors, how are specific instructive signals transduced? One possibility is that additional pathway components play an important role in ensuring signal specificity. Members of the recently identified SARA family of cytoplasmic proteins (Tsukazaki *et al.* 1998) may ensure that ligand-specific Smads are recruited only to their respective receptor complexes. Alternatively, cell surface proteoglycans such as Dally (implicated in Dpp signaling; Jackson *et al.* 1997) may influence ligand-receptor interactions. A second possibility is suggested by the demonstration that functional receptor complexes are likely heterotetramers or larger units containing multiple type I and multiple type II receptors (Weis-Garcia and Massagué 1996). Receptor complexes containing different stoichiometries of individual receptors or receptor pairs may provide signal specificity through variation in ligand binding or the utilization of distinct constellations of Smads.

The Smad family appears to be evolving independently of the ligands and receptors. Perhaps Smad molecular evolution is instead influenced by interactions with other TGF-β signaling pathway components such as the SARA proteins or transcription factors. Alternatively, Smad family evolution may be driven by interac-

tions with components of other signaling pathways. For example, a recent study in Drosophila (Szuts *et al.* 1998) suggests that Smad molecules can participate in TGF-β and epidermal growth factor signaling pathways.

Smad family members do not cluster by ligand as evidenced by the grouping of Smad8 (TGF-β/activin signaling) with Smad1 and Smad5 (BMP signaling) in the Smad tree (Figure 7). Nor do they cluster by role as evidenced by the grouping of the antagonist Daf-3 with Smads that signal for multiple ligands such as Smad4 in the MH1 tree (Figure 9). A comparison of our Smad tree (Figure 7) and the MH2 tree (data not shown) indicates that the relationships identified in the Smad tree are almost universally reflective of those in the MH2 tree. The MH2 domain facilitates the formation of complexes between Smads. The extensive similarity between the Smad and MH2 trees suggests that the ability to form multi-Smad complexes is the most fundamental feature of Smad function.

In Smads with different roles the most informative sequence difference is the number of C-terminal serines in the MH2 domain. There are two or three serines in Smads dedicated to one ligand and zero or one serine in Smads signaling for multiple ligands and antagonist Smads. The absence of the MH1 domain is not an accurate predictor of Smad function. The *C. elegans* antagonist Daf-3 has an identifiable MH1 domain and one C-terminal serine, while the positively signaling Daf-14 has no MH1 domain and two serines. The function of the MH1 domain appears to be DNA binding/transcriptional activation in both ligand-specific and multiple-ligand Smads. This common function is reflected in the Smad (Figure 7; CP = 95) and MH1 trees (Figure 9). Both trees have secondary clusters of ligand-specific (*e.g.*, Mad; cluster E) and multiple-ligand Smads (*e.g.*, Smad4; cluster A).

In conclusion, we hope that these results will stimulate new experimental directions for both evolutionary and developmental biologists. As new multigene families that participate in TGF-β signaling are discovered we plan to extend our analyses to test hypotheses reported here. We believe this study illustrates the value of applying statistical methods in the analysis of developmentally important signaling pathways.

LITERATURE CITED

Aguinaldo, A., J. Turbeville, L. Linford, M. Rivera, J. Garey *et al.*, 1997   Evidence for a clade of nematodes, arthropods and other moulting animals. Nature **387:** 489–493.

Brummel, T. J., V. Twombly, G. Marques, J. Wrana, S. J. Newfeld *et al.*, 1994   Characterization and relationship of Dpp receptors encoded by the *saxophone* and *thickveins* genes in *Drosophila*. Cell **78:** 251–262.

Brummel, T. J., S. Abdollah, T. Haerry, M. Shimell, J. Merriam *et al.*, 1999   The Drosophila activin receptor Baboon signals through DSmad2 and controls cell proliferation but not patterning during larval development. Genes Dev. **13:** 98–111.

Burge, C., and S. Karlin, 1997   Prediction of complete gene structures in human genomic DNA. J. Mol. Biol. **268:** 78–94.

Burt, D., and A. Law, 1994   Evolution of the TGF-β superfamily. Prog. Growth Factor Res. **5:** 99–118.

Chen, Y., A. Bhushan and W. Vale, 1997   Smad8 mediates the signaling of the receptor serine kinase. Proc. Natl. Acad. Sci. USA **94:** 12938–12943.

Childs, S., J. Wrana, K. Arora, L. Attisano, M. O'Connor *et al.*, 1993   Identification of a *Drosophila* activin receptor. Proc. Natl. Acad. Sci. USA **90:** 9475–9479.

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

Derynck, R., Y. Zhang and X. Feng, 1998   Smads: transcriptional activators of TGF-β responses. Cell **95:** 737–740.

Dopazo, J., 1994   Estimating errors and confidence intervals for branch lengths in phylogenetic trees by a bootstrap approach. J. Mol. Evol. **38:** 300–304.

Eppert, K., S. Scherer, H. Ozcelik, R. Pirone, P. Hoodless *et al.*, 1996   MADR2 maps to 18q21 and encodes a TGF-β regulated MAD-related protein that is functionally mutated in colorectal carcinoma. Cell **86:** 543–552.

Estevez, M., L. Attisano, J. Wrana, P. Albert, J. Massagué *et al.*, 1993   The *daf*-4 gene encodes a BMP receptor controlling *C. elegans* dauer larva development. Nature **365:** 644–649.

Felsenstein, J., 1985   Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:** 783–791.

Felsenstein, J., 1993   PHYLIP version 3.5c. University of Washington, Seattle.

Fitch, W., and E. Margoliash, 1970   The usefulness of amino acid and nucleotide sequences in evolutionary studies. Evol. Biol. **4:** 67–109.

Frisch, A., and C. Wright, 1998   XBMPR-II, a novel Xenopus Type II receptor mediating BMP signaling in embryonic tissues. Development **125:** 431–442.

Georgi, L., P. Albert and D. Riddle, 1990   *daf*-1: a *C. elegans* gene controlling dauer larva development encodes a novel receptor protein kinase. Cell **61:** 635–645.

Graff, J., R. Thies, J. Song, A. Celeste and D. Melton, 1994   Studies with a Xenopus BMP receptor suggest that ventral mesoderm-inducing signals override dorsal signals *in vivo*. Cell **79:** 169–179.

Graff, J., A. Bansal and D. Melton, 1996   Xenopus MAD proteins transduce distinct subsets of signals for the TGF-β family. Cell **85:** 479–487.

Haerry, T., O. Khalsa, M. O'Connor and K. Wharton, 1998   Synergistic signaling by two BMP ligands through the SAX and TKV receptors controls wing growth and patterning in Drosophila. Development **125:** 3977–3987.

Hahn, S., M. Schutte, A. Hoque, C. Moskaluk, L. da Costa *et al.*, 1996   DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. Science **271:** 350–353.

Higgins, D., A. Bleasby and R. Fuchs, 1992   CLUSTAL V: improved software for multiple sequence alignment. Comput. Appl. Biosci. **8:** 189–191.

Hillis, D., and J. Bull, 1993   An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. **42:** 182–192.

Hoodless, P., T. Haerry, S. Abdollah, M. Stapleton, M. O'Connor *et al.*, 1996   MADR1, a *Mad*-related protein that functions in BMP2 signaling pathways. Cell **85:** 489–500.

Imamura, T., M. Takase, A. Nishihara, E. Oeda, J. Hanai *et al.*, 1997   Smad6 inhibits signaling by the TGF-β superfamily. Nature **389:** 622–626.

Jackson, S., H. Nakato, M. Sugiura, A. Jannuzi, R. Oakes *et al.*, 1997   *dally*, a Drosophila glypican controls cellular responses to the TGF-β related morphogen Dpp. Development **124:** 4113–4120.

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 a novel receptor for GDNF. Cell **85:** 1113–1124.

Jones, C., M. Kuehn, B. Hogan, J. Smith and C. Wright, 1995 Nodal-related signals induce axial mesoderm and dorsalize mesoderm in gastrulation. Development **121:** 3651–3662.

Kingsley, D., 1994 The TGF-β superfamily: new members, new receptors, and new genetic tests of function in different organisms. Genes Dev. **8:** 133–146.

Koenig, B., J. Cook, D. Wolsing, J. Ting, J. Tiesman *et al.*, 1994 Characterization and cloning of a receptor for BMP2 and BMP4 from NIH 3T3 cells. Mol. Cell. Biol. **14:** 5961–5974.

Krishna, S., L. Maduzia and R. Padgett, 1999 Specificity of TGF-β signaling is conferred by distinct Type I receptors and their associated Smad proteins in *C. elegans.* Development **126:** 251–260.

Kumar, S., K. Tamura and M. Nei, 1993 *MEGA: Molecular Evolutionary Genetic Analysis.* Pennsylvania State University Press, University Park, PA.

Kutty, G., R. Kutty, W. Samuel, T. Duncan, C. Jaworski *et al.*, 1998 Identification of a new member of the TGF-β family in Drosophila: the first invertebrate activin gene. Biochem. Biophys. Res. Commun. **246:** 644–649.

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

Lin, L., D. Doherty, J. Lile, S. Bektesh and F. Collins, 1993 GDNF: a glial cell line derived neurotrophic factor for midbrain dopaminergic neurons. Science **260:** 1130–1132.

Liu, F., F. Ventura, J. Doody and J. Massagué, 1995 Human type II receptor for BMPs: extension of the two-kinase receptor model to the BMPs. Mol. Cell. Biol. **15:** 3479–3486.

Massagué, J., L. Attisano and J. Wrana, 1994 The TGF-β family and its composite receptors. Trends Cell Biol. **4:** 172–178.

Miya, T., K. Morita, A. Suzuki, N. Ueno and N. Satoh, 1997 Functional analysis of an ascidian homolog of vertebrate BMP2/BMP4 suggests its role in the inhibition of neuralfate specification. Development **124:** 5149–5159.

Nakayama, T., M. Snyder, S. Grewel, K. Tsuneizumi, T. Tabata *et al.*, 1998 Xenopus Smad8 acts downstream of BMP4 to modulate its activity during vertebrate embryonic patterning. Development **125:** 857–867.

Nei, M., S. Kumar and K. Takahashi, 1998 The optimization principle in phylogenetic analysis tends to give incorrect topologies when the number of nucleotides or amino acids used is small. Proc. Natl. Acad. Sci. USA **95:** 12390–12397.

Newfeld, S. J., and W. M. Gelbart, 1995 Identification of two *Drosophila* TGF-β family members in the grasshopper *Schistocerca americana.* J. Mol. Evol. **41:** 155–160.

Newfeld, S. J., E. Chartoff, J. Graff, D. Melton and W. Gelbart, 1996 *Mothers against dpp* encodes a conserved cytoplasmic protein required in DPP/TGF-β responsive cells. Development **122:** 2099–2108.

Nguyen, M., S. Park, G. Marques and K. Arora, 1998 Interpretation of a BMP activity gradient in Drosophila embryos depends on synergystic signaling by two Type I receptors, SAX and TKV. Cell **95:** 495–506.

Padgett, R., J. Wozney and W. Gelbart, 1993 Human BMP sequences confer normal dorsal ventral patterning in the *Drosophila* embryo. Proc. Natl. Acad. Sci. USA **90:** 2905–2909.

Patterson, G., A. Koweek, A. Wong, Y. Liu and G. Ruvkun, 1997 The daf-3 Smad protein antagonizes TGF-β related receptor signaling in the *C. elegans* dauer pathway. Genes Dev. **11:** 2679–2690.

Posakony, L. G., L. A. Raftery and W. M. Gelbart, 1991 Wing formation in Drosophila requires *decapentaplegic* function along the anterior-posterior compartment boundary. Mech. Dev. **33:** 69–82.

Ruberte, E., T. Marty, D. Nellen, M. Affolter and K. Basler, 1995 An absolute requirement for both the Type II and Type I receptors Punt and Thickveins for Dpp signaling *in vivo.* Cell **80:** 889–897.

Rzhetsky, A., and M. Nei, 1993 A simple method for estimating and testing minimum-evolution trees. Mol. Biol. Evol. **9:** 945–967.

Saitou, N., and M. Nei, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425.

Sampath, T., K. Rashka, J. Doctor, R. Tucker and F. M. Hoffmann, 1993 Drosophila TGF-β superfamily proteins induce endochon-

drial bone formation in mammals. Proc. Natl. Acad. Sci. USA **790:** 6004–6008.

Savage, C., P. Das, A. Finelli, S. Townsend, C. Sun *et al.*, 1996 *C. elegans* genes *sma-2, sma-3,* and *sma-4* define a conserved family of TGF-β pathway components. Proc. Natl. Acad. Sci. USA **93:** 790–794.

Singer, M., A. Penton, V. Twombly, F. M. Hoffmann and W. Gelbart, 1996 Signaling through both Type I Dpp receptors is required for anterior-posterior patterning of the entire Drosophila wing. Development **124:** 79–89.

Sitnikova, T., 1996 Bootstrap method of interior-branch test for phylogenetic trees. Mol. Biol. Evol. **13:** 605–611.

Smith, W., R. McKendry, S. Ribisi, Jr. and R. Harland, 1995 A *nodal*-related gene defines a physical and functional domain within the Spemann organizer. Cell **82:** 37–46.

Solovyev, V., and A. Salamov, 1997 The GeneFinder computer tools for analysis of human and model organism genome sequences. Intel. Sys. Mol. Biol. **5:** 294–302.

Strimmer, K., and A. von Haeseler, 1997 Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. Proc. Natl. Acad. Sci. USA **94:** 6815–6819.

Suzuki, A., R. Thies, N. Yamaji, J. Song, J. Wozney *et al.*, 1994 A truncated BMP receptor affects dorsal-ventral patterning in the early Xenopus embryo. Proc. Natl. Acad. Sci. USA **91:** 10255–10259.

Swofford, D. L., 1998 *PAUP*: Phylogenetic Analysis Using Parsimony (and Other Methods).* Sinauer Associates, Sunderland, MA.

Szuts, D., S. Eresh and M. Bienz, 1998 Functional intertwining of Dpp and EGFR signaling during Drosophila endoderm induction. Genes Dev. **12:** 2022–2035.

ten Dijke, P., H. Yamashita, T. Sampath, A. Reddi, M. Estevez *et al.*, 1994a Identification of Type I receptors for Osteogenic Protein1 and BMP4. J. Biol. Chem. **269:** 16985–16988.

ten Dijke, P., H. Yamashita, H. Ichijo, P. Franzen, M. Laiho *et al.*, 1994b Characterization of Type I receptors for TGF-β and activin. Science **264:** 101–104.

Thatcher, J., C. Haun and P. Okkema, 1999 The Daf-3 Smad binds DNA and represses gene expression in the *C. elegans* pharynx. Development **126:** 97–107.

Thiagalingam, S., C. Lengauer, F. Leach, M. Schutte, S. Hahn *et al.*, 1996 Evaluation of candidate tumor suppressor genes in colorectal cancer. Nat. Genet. **13:** 343–346.

Tsukazaki, T., T. Chiang, A. Davison, L. Attisano and J. Wrana, 1998 SARA, a FYVE domain protein that recruits Smad2 to the TGF-β receptor. Cell **95:** 779–791.

Wang, D., S. Kumar and S. B. Hedges, 1999 Divergence time estimates for the early history of animal phyla: the origin of plants, animals and fungi. Proc. R. Soc. Lond. Ser B **266:** 163–171.

Weis-Garcia, F., and J. Massagué, 1996 Complementation between kinase defective and activation defective TGF-β receptors reveals a novel form of receptor cooperativity essential for signaling. EMBO J. **15:** 276–289.

Whitman, M., 1998 Smads and early developmental signaling by the TGF-β superfamily. Genes Dev. **12:** 2445–2462.

Wisotzkey, R. G., A. Mehra, D. Sutherland, L. Dobens, X. Liu *et al.*, 1998 *Medea* is a Drosophila Smad4 homolog that is differentially required to potentiate Dpp responses. Development **125:** 1433–1445.

Wrana, J., H. Tran, L. Attisano, K. Arora, S. Childs *et al.*, 1994 Two distinct transmembrane serine-threonine kinases from Drosophila form an activin receptor complex. Mol. Cell. Biol. **14:** 994–950.

Yingling, J., P. Das, C. Savage, M. Zhang, R. Padgett *et al.*, 1996 Mammalian dwarfins are phosphorylated in response to TGF-β and are implicated in control of cell growth. Proc. Natl. Acad. Sci. USA **93:** 8940–8944.

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

Zou, H., R. Wieser, J. Massagué and L. Niswander, 1997 Distinct roles of Type I BMP receptors in the formation and differentiation of cartilage. Genes Dev. **11:** 2191–2203.

Communicating editor: A. G. Clark