Dev Genes Evol (2010) 220:235–250
DOI 10.1007/s00427-010-0341-5

ORIGINAL ARTICLE

Rapid evolution of a novel signalling mechanism by concerted duplication and divergence of a BMP ligand and its extracellular modulators

Cornelia Fritsch · Robert Lanfear · Robert P. Ray

Received: 4 May 2010 / Accepted: 19 October 2010 / Published online: 18 November 2010

Abstract Gene duplication and divergence is widely considered to be a fundamental mechanism for generating evolutionary novelties. The Bone Morphogenetic Proteins (BMPs) are a diverse family of signalling molecules found in all metazoan genomes that have evolved by duplication and divergence from a small number of ancestral types. In the fruit fly *Drosophila*, there are three BMPs: Decapentaplegic (Dpp) and Glass bottom boat (Gbb), which are the orthologues of vertebrate BMP2/4 and BMP5/6/7/8, respectively, and Screw (Scw), which, at the sequence level, is equally divergent from Dpp and Gbb. It has recently been shown that Scw has arisen from a duplication of Gbb in the lineage leading to higher Diptera. We show that since this duplication event, Gbb has maintained the ancestral BMP5/6/7/8 functionality while Scw has rapidly diverged. The evolution of Scw was accompanied by duplication and divergence of a suite of extracellular regulators that continue to diverge together in the higher Diptera. In addition, Scw has become restricted in its receptor specificity: Gbb proteins can signal through the Type I receptors Thick veins (Tkv) and Saxophone (Sax), while Scw signals through Sax. Thus, in a relatively short span of evolutionary time, the duplication event that gave rise to Scw produced not only a novel ligand but also a novel signalling mode that is functionally distinct from the ancestral Gbb mode. Our results demonstrate the plasticity of the BMP pathway not only in evolving new family members and new functions but also new signalling modes by redeploying key regulators in the pathway.

Keywords Bone morphogenetic protein · *Drosophila* · Signalling · Evolution

Communicated by: P. Simpson

Electronic supplementary material The online version of this article (doi:10.1007/s00427-010-0341-5) contains supplementary material, which is available to authorized users.

C. Fritsch · R. P. Ray (✉)
School of Life Sciences, University of Sussex,
Falmer, Brighton BN1 9QG, United Kingdom
e-mail: r.ray@sussex.ac.uk

R. Lanfear
Centre for Macroevolution and Macroecology, Ecology Evolution and Genetics, Research School of Biology,
Australian National University,
Canberra, ACT, Australia

Introduction

Evidence for gene duplications can be found in the genomes of most organisms in the form of tandem or dispersed arrays of genes that are related in structure and function. Such gene duplications can evolve in three possible ways: (1) *loss*, whereby one of the copies is lost by gene deletion or by degeneration; (2) *subfunctionalization*, whereby the ancestral functions are divided between the duplicated copies by differential regulation of gene expression; and (3) *neofunctionalization*, whereby one copy retains the ancestral function(s) and the other takes up a novel function (Force et al. 1999; Lynch and Conery 2000; Lynch and Force 2000; Zhang 2003). For example, the Hox gene cluster has been amplified by tandem duplications in many lineages and by polyploidization in vertebrates. These amplifications were followed by subfunctionalization of genes within the complex such that expression of individual Hox genes is limited to distinct regions of the body (Lemons and McGinnis 2006). In some lineages, Hox genes have undergone neofunctionaliztion. In the Diptera, the Hox3 orthologue, referred to as *zerknüllt* (*zen*), has

adopted a novel role and is no longer expressed in a segmental pattern but rather functions in extraembryonic tissues. Subsequently, *zen* was duplicated once in the higher Diptera to give rise to *bicoid*, which evolved into the maternally expressed anterior morphogen (McGregor 2005), and then again, in the lineage leading to *Drosophila* to give rise to *zen2*, which is expressed in the same pattern as *zen* but does not have an essential function (Casillas et al. 2006).

The Bone Morphogenetic Proteins (BMPs) are a diverse family of signalling molecules that have also evolved by duplication and divergence. A vertebrate genome typically contains ~20 BMP-related proteins which fall into a number of distinct classes. The two main classes, BMP2/4 and BMP5/6/7/8, are present in Bilateria and in the Cnidaria suggesting they represent the canonical BMPs common to all metazoans. Two additional classes, BMP9/10 and Admp, are widespread in arthropods and vertebrates, but are absent in some lineages suggesting they are not essential elements in the BMP repertoire (Van der Zee et al. 2008). The remaining classes are vertebrate-specific and the evolutionary relationships between them, and to the BMP2/4, BMP5/6/7/8, BMP9/10 and Admp classes, are not clearly resolved in phylogenetic analyses (Herpin et al. 2004; Innis et al. 2000; Katoh and Katoh 2006; Newfeld et al. 1999; Van der Zee et al. 2008).

Invertebrate genomes have fewer BMPs, but, as in the vertebrates, these fall into canonical and non-canonical types. *Drosophila* have three BMP family members: the BMP2/4 orthologue Decapentaplegic (Dpp), the BMP5/6/7/8 orthologue Glass bottom boat (Gbb), and Screw (Scw), the ligand domain of which is equally divergent from that of Dpp and Gbb. Despite their divergent sequences, Scw and Gbb modulate Dpp signalling in a similar fashion. Scw functions in the early embryo where it augments Dpp signalling along the dorsal–ventral axis to produce peak levels of BMP activity at the dorsal midline, thus, specifying cells that will become the extraembryonic amnioserosa (Arora et al. 1994). Scw achieves this focusing of BMP activity by forming heterodimers with Dpp (Shimmi et al. 2005b), interacting with a specific suite of extracellular modulators including the Chordin orthologue Short gastrulation (Sog), the metalloprotease Tolloid (Tld), and the cysteine-rich binding proteins Twisted gastrulation (Tsg) and Shrew (Srw) (Bonds et al. 2007; Holley et al. 1995; Neul and Ferguson 1998; O'Connor et al. 2006; Shimmi and O'Connor 2003), and signalling through the Type I receptor Saxophone (Sax; Neul and Ferguson 1998; Nguyen et al. 1998).

Gbb functions at later stages of development in a variety of different developmental contexts (Ballard et al. 2010; McCabe et al. 2003; Wharton et al. 1999). In the wing, Gbb has long-range and local functions that modulate Dpp activity in a context-dependent fashion. In the larval wing disc, Gbb promotes long-range Dpp signalling to specify pattern elements far from the source of Dpp at the anterior–posterior compartment boundary. In the pupal wing, Gbb augments Dpp signalling locally in the longitudinal veins to achieve peak levels of BMP activity along their length (Ray and Wharton 2001). In the developing crossvein, Gbb focuses Dpp signalling in a manner analogous to the focusing of Dpp activity by Scw in the early embryo and requires a related suite of extracellular modulators including Sog, the metalloprotease Tolkin (Tok), and the cysteine-rich binding protein Crossveinless (Cv) (Serpe et al. 2005; Shimmi et al. 2005a). However, in contrast to the Scw signalling mechanism, Sax function is not required for crossvein specification, so Gbb signalling in this context is presumably transduced by the Type I receptor Thick veins (Tkv; Ray and Wharton 2001). Indeed, while Scw signalling appears to be exclusively transduced by Sax, Gbb can signal through both Type I receptors in a context dependent manner, with some functions requiring only Sax, others only Tkv and still others may require both receptors (Rawson et al. 2003; Ray and Wharton 2001; Singer et al. 1997). Thus, while Gbb and Scw have a similar focusing effect on Dpp signalling, they achieve this effect through interactions with distinct suites of extracellular modulators and receptors.

While Scw and Gbb appear to modulate Dpp signalling in a similar way, the fact that the Scw ligand domain is nearly equally divergent from the BMP2/4 orthologue Dpp and the BMP5/6/7/8 orthologue Gbb raises the question of how this BMP is related to other arthropod BMPs. Indeed, previous studies have provided a number of different explanations for the origin of Scw. Scw was initially placed as an 'orphan' that did not have a direct vertebrate orthologue (Arora et al. 1994). In a subsequent study, Scw was proposed to be the fly orthologue of vertebrate Gdf3 (Newfeld et al. 1999). Recently, a comparison of whole genome sequences concluded that Scw is a paralogue of Gbb that arose from a duplication of Gbb in the lineage leading to the higher Diptera (Van der Zee et al. 2008). This last result raises a number of questions about how the Scw signalling mechanism arose, its relationship to Gbb signalling, and how Scw has come to be so divergent from Gbb.

In this report, we address the relationship between Gbb and Scw by sequence comparisons, phylogenetic analysis, and functional studies. Our results show that in the arthropod lineage, the BMP5/6/7/8 orthologue Gbb has undergone multiple duplications, and that one of these duplication events gave rise to Scw in the lineage leading to the higher Diptera. Using in vivo rescue assays, we demonstrate that although Scw has rapidly diverged from the ancestral Gbb sequence, it retains the ability to function in Gbb-dependent processes in *Drosophila*. In contrast,

while *Drosophila* Gbb has maintained the ancestral Gbb functionality, it cannot replace Scw in the early embryo. This non-reciprocal functional redundancy suggests that Scw has evolved a signalling mechanism that is not compatible with the Gbb ligand. Consistent with this, we find that, like Scw, the suite of extracellular regulators that function with Scw in the early embryo are also novelties of the higher Diptera that have arisen by duplication and divergence of ancestral regulators that interact with Gbb. The evolution of Scw has also resulted in a shift in receptor specificity. In *Drosophila*, Scw signalling is transduced by Sax, while Gbb signalling is transduced by Sax or Tkv. We show that ancestral Gbb can signal through either receptor in *Drosophila*, which implies that the exclusivity of Scw for Sax has evolved with Scw in the lineage leading to the higher Diptera. Using chimeric receptors, we show that this must be due to the changes in Scw and not to changes in Sax. We propose that the evolution of Scw provides a paradigm for the evolutionary potential of BMPs and that a similar mechanism of duplication and divergence may account for the wide range of 'orphan' BMPs that are found in vertebrate genomes.

Materials and methods

Drosophila strains

Drosophila were cultured on standard cornmeal–yeast medium at 25°C. The *gbb* alleles, gbb^{1}, gbb^{4}, and gbb^{D20}, and *Df(2R)S246* have been described (Chen et al. 1998; Wharton et al. 1999). scw^{S12} is a mis-sense mutation (C.F. unpublished) that behaves as a null in trans to *Df(2 L)OD16* (Arora et al. 1994). The maternal effect *sax* alleles, sax^{1} and sax^{2} (from T. Schüpbach), are associated with mis-sense mutations in the kinase domain and behave as dominant antimorphs in conjunction with zygotic loss-of-function *dpp* mutations (Twombly et al. 2009), but as homozygotes, recapitulate the loss-of-function phenotype of *scw*. The nonsense mutation sax^{4} and deficiency *Df(2R)BSC265* are amorphic. cv^{43} (from L. Marsh) is a homozygous viable deletion allele, as described previously (Vilmos et al. 2005). Other strains described in the text were obtained from *Drosophila* stock centres.

Rescue assays

For rescue assays, multiple transgene insertions on the third chromosome were tested for their ability to rescue the genotypes gbb^{D20}*/Df(2R)S246*, scw^{S12}*/Df(2L)OD16*, or sax^{4}*/Df(2R)BSC265* (Table S4). In general, males of the genotype y^{1} w^{1118}*;; Tn** were crossed to y^{1} w^{1118}*; Bl/CyO* females, and the resulting male progeny of the genotype y^{1} w^{1118}*; +/Bl; Tn*/+* were crossed to y^{1} w^{1118}*; Df*/CyO*. From this cross, males of the genotype y^{1} w^{1118}*; Df*/Bl; Tn*/+* were crossed to y^{1} w^{1118}*; mut*/CyO* and the resulting progeny scored. The final cross produces eight progeny classes, four experimental classes with the second chromosome genotypes *mut*/Df**, *mut*/Bl*, *Df*/CyO*, and *Bl/CyO* that carry the transgene, and four control classes with the same genotypes that lack the transgene (see Table S5). The statistic 'percent of expected' was calculated by dividing the number of *mut*/Df* + Tn* flies by the number of *mut*/Bl + Tn* flies and multiplying the quotient by 100. For constructs that did not rescue, we confirmed that the lines tested were expressed at a level comparable to that of the endogenous gene by reverse transcription polymerase chain reaction (RT-PCR; see Figure S1). For rescue assays with gbb^{4}, the same crosses were used with y^{1} w^{1118}*; gbb*/SM6a* replacing the deficiency stock in the second cross, and y^{1} w^{1118}*; (Dp(2;2)DTD48)* gbb^{4} replacing the gbb^{D20} stock in the third cross. To test for epistasis between *cv* and the rescue of *DTD48* gbb^{4} homozygotes by the transgene with Scw being driven by *gbb* cis-regulatory sequences (gbb= Scw, see Results), a stock of the genotype *w* cv^{43}*/FM7h; Dp(2;2)DTD48* gbb^{4}*/CyO; gbb=Scw/+* was constructed which produced males and females of the genotype *w* cv^{43}*; Dp(2;2)DTD48* gbb^{4}*; gbb=Scw/(+)* that could be identified by lack of both balancers and the presence of the w^{+} transgene. For epistasis with *sax*, stocks of the genotype *y w;* sax^{1} gbb^{1}*/CyO; gbb=Scw* and *y w;* sax^{2} gbb^{4}*/CyO; gbb=Scw* were constructed and crossed inter se to produce flies of the genotype *y w;* sax^{1} gbb^{1}/sax^{2} gbb^{4}*; gbb=Scw*. A similar scheme was used for the epistasis experiments with Dm-gbb=AgamGbb1PVLD and Dm-gbb= AgamGbb2PVLD.

Phylogenetic methods

Blast searches The protein sequences of *Drosophila melanogaster* Gbb, Scw, Cv, Tsg, Srw, Tok, Tld, Sax and Tkv were used to search for orthologues in the published genomes of other species using the tblastn function on Ensembl (http://www.ensembl.org/index.html) and FlyBase (http://flybase.org), and against a draft version of the *Glossina morsitans* and *Ceratitis capitata* genomes. Sequence accession numbers can be found in Table S6

Alignment Translated DNA sequences of each gene were aligned by eye using SE-AL (http://evolve.zoo.ox.ac.uk), and regions for which homology could not be confidently inferred were removed. These single-gene alignments form the basis of the DNA and amino acid alignments used for analyses of evolutionary rates and selection (see below). For analyses which included sequences from more than one

gene, single gene alignments were aligned to each other, and regions in which homology between sites of different genes could not be confidently inferred were discarded.

Phylogenetic analysis To determine the evolutionary relationship of paralogous genes, Bayesian analyses were performed on amino acid alignments using MrBayes version 3.1.2 (Huelsenbeck and Ronquist 2001), with the Whelan and Goldman (Whelan and Goldman 2001) amino acid replacement matrix, using random starting trees, with gamma distributed rates across sites (six categories) and the proportion of invariant sites estimated from the data. Each analysis consisted of two independent runs of one chain each. Analyses were run for 6,000,000 generations with samples taken every 500 generations. In order to compare hypotheses for the origin of the Scw gene, the Markov Chain Monte Carlo (MCMC) was constrained to visit only the subset of phylogenetic trees which agreed with a particular hypothesis (see Results). Typically, the initial stages of an MCMC analysis are discarded because they can be unduly influenced by the starting point (Ronquist et al. 2009). This initial phase (known as burn-in) can be determined by examination of the likelihood trace from the MCMC. Analysis of likelihood traces for each run indicated that a burn-in of 1,000,000 generations (yielding a posterior sample of 20,000 trees) was more than sufficient in all cases. Input and output files are available on request. Support for different hypotheses of the phylogenetic placement of Scw was assessed using Bayes factors (Raftery 1996). Harmonic means of likelihoods were calculated from the combined post-burn-in samples of both independent runs for each analysis in Tracer v1.4, using the modifications previously suggested (Suchard et al. 2001).

Comparisons of evolutionary rates Rates of protein evolution were compared by Bayesian analysis of concatenated amino acid alignments of protein pairs for all higher Dipteran species for which we had sequences for both genes. For each analysis, two Bayesian MCMCs were performed—one in which both genes were forced to evolve at the same rate, and one in which the two genes were able to evolve at different rates. In all analyses, the gamma shape parameter and the proportion of invariant sites were estimated separately for each gene. Support for the two-rate model over the one-rate model was assessed using Bayes factors. Since these analyses contained far fewer tips than the phylogenetic analyses (typically just the 12 *Drosophila* species), they were run for 3,000,000 generations with samples taken every 250 generations. Analysis of likelihood traces indicated that a burn-in of 500,000 generations was more than sufficient in all cases, yielding posterior samples of 20,000 trees for each analysis.

Tests for positive selection Evidence for selection was assessed by calculating the likelihood of different models of evolution on DNA alignments using codeml from the PAML4.1 package (Yang 2007). We tested each single-gene alignment to assess whether there was evidence for a class of codons in which the dn/ds ratio is >1, indicative of positive selection. We used both the conservative comparison of models 1a and 2a, and the less conservative comparison of models 7 and 8 in codeml following Yang et al. (2000). Models were compared with Likelihood Ratio Tests (Yang et al. 2000).

P-element constructs and transformation

A 3-kb genomic *SmaI–BamHI* fragment comprising *gbb* coding and regulatory sequences and a 5.1-kb genomic *BamHI–HindIII* fragment comprising *scw* coding and regulatory sequences were used for all rescue experiments with *D. melanogaster*-based *gbb* and *scw* constructs. After addition of restriction sites upstream of the start codon (*SalI*), within the signal peptide (*SpeI*), between the pro-domain and the ligand domain (*MluI* for *gbb* and *BamHI* for *scw*) and at the stop codon (*EcoRI* for *scw*), *gbb* and *scw* coding sequences were exchanged with one another or with PCR-amplified *gbb* or *scw* coding sequences from other species. Genomic rescue constructs of *Drosophila grimshawi gbb* and *scw* and *G. morsitans scw* were PCR-amplified from genomic DNA. A 8.7-kb genomic *EcoRI–SalI* fragment comprising the *Drosophila sax* coding and regulatory sequences was modified by addition of an *NcoI* site at the start codon, and a genomic fragment extending from this *NcoI* site to an endogenous *NheI* site C-terminal of the GS-repeat was replaced by a *NcoI–NheI*-flanked PCR product from the corresponding part of *Anopheles sax*. The constructs were shuttled into the P-element vector pCaSpeR4 (Pirrotta 1988) and injected into y^1 w^{1118} embryos.

Cloning *scw* and *gbb* from other Diptera

Genomic DNA was isolated from adult *Hirtodrosophila duncani*, *Chymomyza procnemis*, *Zaprionus tuberculatus*, *Samoaia leonensis*, and *Scaptodrosophila pattersoni* using the DNeasy tissue kit (QIAGEN) and from *Musca domestica*, *Lucilia caesar*, *Diasemopsis meigenii*, *Ceratitis capitata*, *G. morsitans* and *Megaselia abdita* following standard protocols (Maniatis et al. 1982). Degenerate primers were designed based on conserved sequence stretches of *gbb* or *scw* within the twelve *Drosophila* species (Table S7). The PCR products were either directly sequenced using the same degenerate primers as for the amplification or by cloning the PCR products into pCR2.1-

TOPO (Invitrogen) and sequencing with the M13 universal and reverse primers. Additional flanking sequences were isolated from the *scw* genes of *C. procnemis*, *L. caesar*, and *M. domestica* by inverse PCR using specific primers (Table S8). The resulting PCR fragments were purified and sequenced using the same primers.

RT-RCR

For developmental expression profiling, total RNA was isolated from pupae, ovaries and staged embryos using the QIAGEN RNeasy MiniKit (protocol for animal tissues including on column DNase digestion). RT-PCR was performed on 500 ng of total RNA using the QIAGEN OneStep RT-PCR kit with gene-specific primer pairs (Table S9). For *gbb* and *scw*, we used 25 amplification cycles, while all other reactions where amplified for 35 cycles. For comparison of transgene expression levels with endogenous *gbb* and *scw*, total RNA was isolated from staged embryos or pupae (as above). RT-PCR was performed as above using a common forward primer in the 5'UTR along with two reverse primers, one for the endogenous *scw* (or *gbb*) transcript and one for the transposon (Table S9). All reactions were amplified for 30 cycles.

Results

Scw orthologues can be identified in both branches of the Schizophora

Scw orthologues have been identified in the twelve sequenced *Drosophila* species, but not in the sequenced mosquito genomes, *Anopheles*, *Aedes*, or *Culex* (Culicidae), which has led to the conclusion that Scw arose from a duplication of Gbb in the lineage leading to the higher Diptera (Van der Zee et al. 2008). To address more precisely when Scw appears in this lineage, we used degenerate PCR to identify Scw orthologues in other species. Within the higher Diptera, the Cyclorrhapha are a monophyletic group composed of the Aschiza and the Schizophora, the latter of which is subdivided into two groups, the Calyptratae and Acalyptratae (Yeates and Wiegmann 1999). Using degenerate primers based on highly conserved domains of the Scw proteins in the twelve sequenced *Drosophila* genomes, we amplified fragments of the Scw orthologues from the Acalyptrates *H. duncani*, *C. procnemis*, *Z. tuberculatus*, *S. leonensis*, and *S. pattersoni* (all Drosophilidae), and the stalk-eyed fly *Diasemopsis meigenii* (Diopsidae), and from the Calyptrates *M. domestica* (Muscidae) and *Lucilia ceasar* (Calliphoridae). Our primers could not amplify Scw from the Mediterranean fruitfly *C. capitata* (Tephritidae) or the tsetse fly *G. morsitans* (Glossinidae), but we were able to identify Scw in these species by BLAST against whole genome sequence builds (see Materials and methods). In both of these cases, the failure to recover Scw by PCR was due to the fact that the orthologues in these species are so divergent from the *Drosophila* sequences that they are not recognised by the primers. Indeed, this divergence within the higher Diptera is a hallmark of Scw (see below). Given that these species represent a variety of families from both Calyptratae and Acalyptratae, these results place the origin of Scw basal to the Schizophora (see Fig. 1). As we were unable to amplify Scw from the Aschizans *M. abdita* (Phoridae) or several species of hover fly (Syrphidae) by degenerate PCR, our data do not support the notion that Scw arose basal to the Cyclorrhapha. However, lacking a whole genome sequence for an Aschizan, we do not know whether this is because the genomes do not contain Scw or because the Scw orthologues are too divergent to be amplified by degenerate PCR.

Multiple BMP5/6/7/8 orthologues are widespread in arthropod genomes

The presence of two BMP5/6/7/8 orthologues, Gbb and Scw, in the higher Diptera raises the question of whether this kind of duplication event is unique to this lineage or more widespread. It has previously been reported that *Tribolium* has two Gbb orthologues, Gbb1 and Gbb2, and it was suggested that these arose by a recent duplication event based on the tandem arrangement of the genes on the chromosome and their high degree of sequence conservation (Van der Zee et al. 2008). Systematic screening of other arthropod genomes for BMP-related sequences revealed that the three mosquitoes, *Anopheles gambiae*, *Aedes aegypti* and *Culex pipens*, as well as the jewel wasp, *Nasonia vitripennis*, also have two Gbb-related proteins, while the water flea, *Daphnia pulex*, the louse, *Pediculus humanus*, and the pea aphid, *Acyrthosiphon pisum*, have only one. In species with more than one Gbb, the Gbb proteins vary in how similar they are to one another. In the mosquitoes, the two Gbb genes, *Gbb-60A* and *Gbb-60A2* (here referred to as *Gbb1* and *Gbb2*), are ~80% identical in their ligand domains, but are not direct orthologues of *Tribolium Gbb1* and *Gbb2*: the Gbb1 proteins appear to be orthologous to *Tribolium* Gbb1, showing 75–77% identity in the ligand domain, but the Gbb2 proteins are more related to the Gbb1 proteins of either mosquito species (71% identity) than they are to each other (64% identity) or any other Gbb protein. In *Nasonia,* the two Gbb-related proteins (*Gbb* and *BMP7*-like, here referred to as *Gbb1* and *Gbb2*, respectively) show only 58% sequence identity in their ligand domains. In this species, the Gbb1 protein

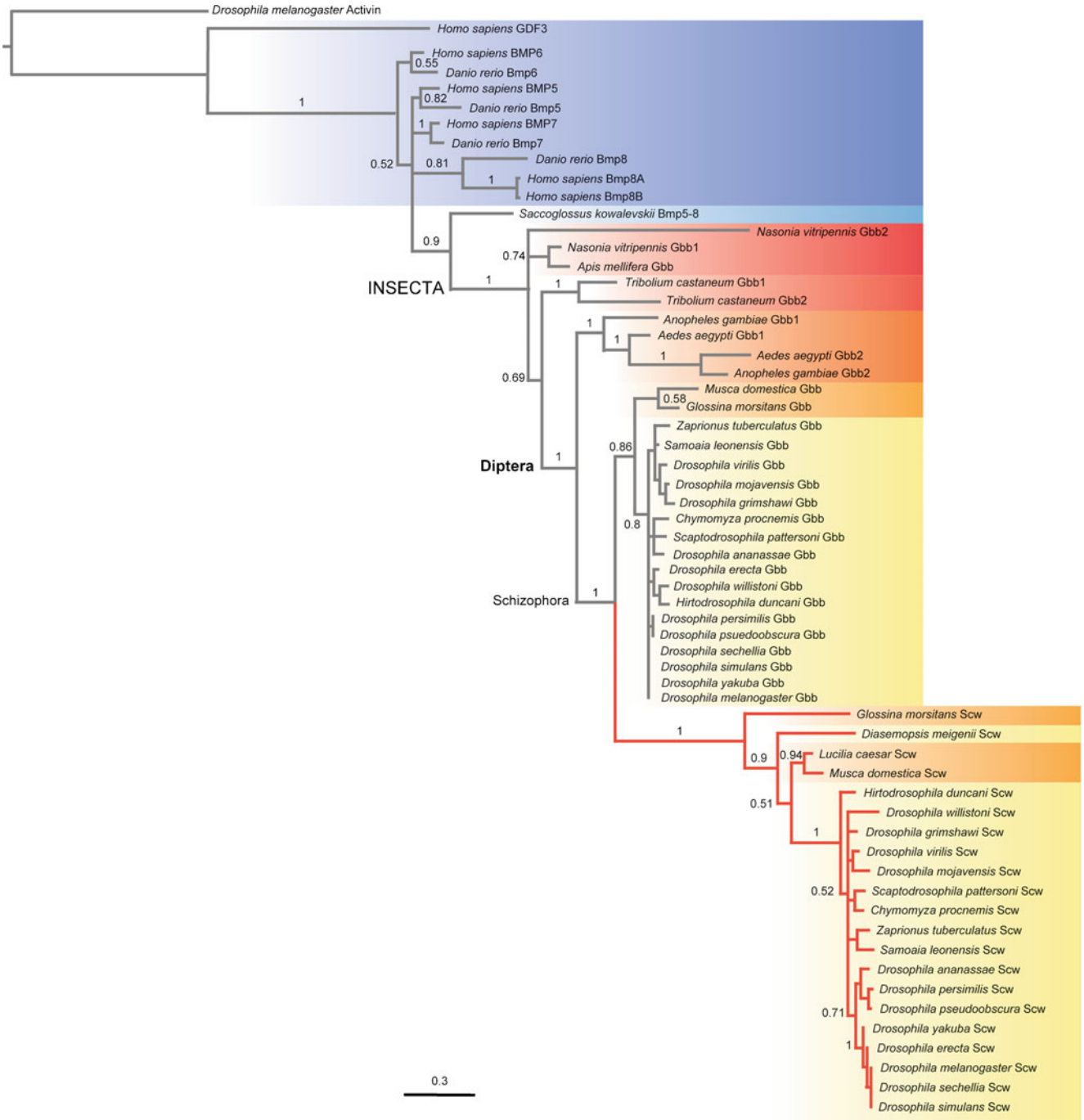

Fig. 1 A phylogeny of the BMP5/6/7/8 class. The phylogenetic tree of the BMP5/6/7/8 class with *Drosophila* Activin as an outgroup as determined from unconstrained and constrained Bayesian analyses. The node uniting the Gbb and Scw proteins in the higher Diptera is strongly supported with a posterior probability of 1.0. This tree places the duplication event that gave rise to Scw after the divergence of the Culiciformae (*orange*) and, as Scw orthologues are found in both Calyptratae (*amber*) and Acalyptratae (*yellow*), basal to the Schizophora. Scw continues to evolve within the higher Diptera as evidenced by the longer branch lengths observed in Scw sequences (*red branches*) compared with Gbb sequences from the same species. Other species groupings are colour-coded as: Coleoptera (*pink*), Hymenoptera (*red*), Hemichordata (*light blue*), and Chordata (*blue*)

appears to be the orthologue of the Gbb1 proteins in *Tribolium*, the mosquitoes, and *Drosophila* (76–86% identity), while the Gbb2 protein is nearly equally divergent from the Gbb1 and Gbb2 proteins in all other species (50–57% identity). Notably, the two Gbbs in *Nasonia* are nearly as divergent from one another as Gbb is from Scw in *Drosophila*. These results suggest that in the Arthropod lineage, there is an 'ancestral' Gbb, which we refer to as

Gbb1, that shows tight sequence conservation in all lineages, and, in some lineages, a second Gbb, which we refer to as Gbb2 (or Scw in the higher Diptera), which has arisen either by independent duplication events or has originated from a single ancestral duplication followed by different degrees of divergence or loss in the various lineages.

Scw arose from a unique duplication in the higher Diptera

The presence of two Gbb-related proteins in the mosquitoes and the higher Diptera raises the possibility that Scw may have evolved by divergence from one of these two ancestral Gbb proteins rather than by a de novo duplication event in the lineage leading to the higher Diptera. To address this question, we performed a Bayesian phylogenetic analysis of all available arthropod Gbb and Scw orthologues. The results of this unconstrained analysis agree with that of Van der Zee et al. (2008) that the duplication event that gave rise to Scw occurred after the split of the mosquitoes and the higher Diptera with a posterior probability of 1.0 on the node uniting higher Dipteran Gbb and Scw sequences (Fig. 1). We then used constrained Bayesian analyses to assess the extent to which this hypothesis should be preferred over other hypotheses for the evolution of Scw. Six hypotheses were compared: (1) the duplication occurred after the separation of the higher Diptera and the mosquitoes (the hypothesis recovered in the unconstrained analysis); (2) the duplication occurred at the base of all Diptera, and Scw is an orthologue of mosquito Gbb1; (3) the duplication occurred at the base of all Diptera, and Scw is an orthologue of mosquito Gbb2; (4) the duplication occurred at the base of all Diptera, and the mosquitoes have lost their *scw* gene and later reduplicated *gbb*; (5) there was a single *gbb* at the base of all Diptera followed by independent duplication events in the mosquitoes and the higher Diptera and (6) the duplication occurred at the base of all insects but Scw was lost independently in different lineages including *Tribolium*, *Nasonia* and the mosquitoes. The results of this analysis show very strong support for hypothesis (1) over all other hypotheses (all other Bayes factors are >8). In particular, the hypothesis that Scw is a direct orthologue of either of the mosquito Gbb proteins yielded the highest Bayes factors (>25), suggesting that these scenarios are extremely unlikely (Table S1). These results very strongly suggest that the duplication that gave rise to Scw occurred after the separation of the mosquitoes and the higher Diptera, and that an independent duplication of Gbb occurred in the mosquito lineage.

Scw continues to evolve rapidly in the higher Diptera

Given that the duplication event that gave rise to Scw occurred after the separation of the mosquito lineage and higher Diptera, subsequent to this event, Gbb and Scw must have diverged rapidly. As a Scw orthologue is present in all Schizophoran species we have analysed, the duplication event that gave rise to Scw occurred between 100 and 250 million years ago (Gaunt and Miles 2002). In this interval of time, the ligand domains of Scw and Gbb have diverged at 47% of their residues to arrive at the sequences currently found in the *D. melanogaster* genome. In comparison, the ligand domains of *Drosophila* Gbb and human BMP5 have only diverged at 27% of their residues in the 555 million years since the last common ancestor of arthropods and vertebrates (Erwin and Davidson 2002). Interestingly, Scw continues to diverge within the higher Diptera. This fact is most obvious from a comparison of the sequence conservation in the ligand domains of the Scw orthologues of *Glossina* and *Drosophila*, which are 54% identical, with that of the Gbb orthologues in the same two species, which are 95% identical. We used Bayesian analysis to quantify the relative rates of evolution of the Gbb and Scw proteins. The results of this analysis show very strong evidence (Bayes factors >10) that the rates of evolution of Gbb and Scw are different, and indicate that Scw is evolving 1.67 times faster than Gbb within the higher Diptera (Fig. 2a; Table S2). These results raise the question of whether the increased rate of divergence is due to positive or relaxed selection. Using two different statistical tests (see Materials and methods), we find no evidence that any of the codons in *scw* are under positive selection ($p>0.05$ for both tests, Table S3), implying that the faster evolution of *scw* is due primarily to relaxed selection and drift.

Gbb and Scw are not reciprocally interchangeable

gbb and *scw* are required for distinct developmental processes, with *scw* functioning exclusively in dorsal–ventral patterning of the early embryo (Arora and Nusslein-Volhard 1992) and *gbb* functioning in a variety of processes during later embryonic, larval, and pupal development (O'Connor et al. 2006). This functional distinction is reflected in the expression profiles of the two genes: *scw* and *gbb* are both ubiquitously expressed, but *scw* expression is confined to the early stages of embryogenesis, while *gbb* expression reaches a peak at the time of gastrulation after which it is maintained for the rest of development (Arora et al. 1994; Wharton et al. 1991). Indeed, RT-PCR of embryos staged at 1-h intervals reveals that *scw* expression peaks at 2–3 h after egg deposition, while *gbb* expression is first detected at low levels in 2–3-h embryos, and reaches its peak at 3–4 h, after which it is maintained throughout embryogenesis (Fig. 3). Given the common origin of Gbb and Scw, this distinct timing of expression raises the possibility that the two proteins are functionally redundant but have been restricted to distinct

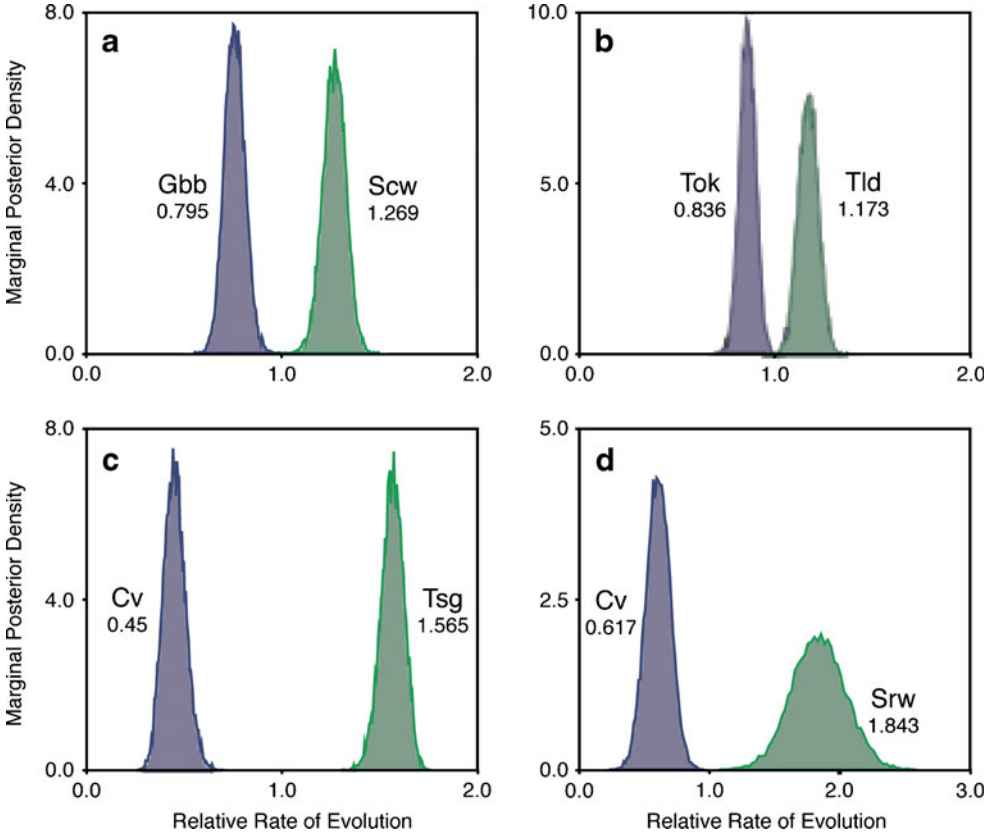

Fig. 2 Scw, Tld, Tsg and Srw evolve faster than their paralogues. Graphs showing the posterior distribution of the relative rate of evolution plotted against the marginal posterior density for each pair of paralogues: Gbb and Scw (**a**), Tok and Tld (**b**), Cv and Tsg (**c**), and Cv and Srw (**d**). In each case, the relative rate of evolution is faster for the paralogue functioning in embryogenesis than for the other paralogue

temporal windows. To address this question, we generated genomic rescue constructs for *gbb* and *scw* and exchanged their protein coding sequences to produce chimeric constructs expressing Gbb under the control of *scw* cis-regulatory sequences (scw=Gbb), or Scw under the control of *gbb* cis-regulatory sequences (gbb=Scw).

To address whether the two proteins are functionally redundant, we tested for the ability of the scw=Gbb construct to rescue *scw* mutants and the gbb=Scw construct to rescue *gbb* mutants. While a control *scw* genomic construct (DmelScw) rescues *scw* mutants with a single copy of the transgene, the scw=Gbb construct does not, even in the presence of two or four copies of the transgene (Table 1, Table S5). Thus, by this rescue assay, Gbb is not able to replace Scw in the early dorsal–ventral patterning function. While these data do not indicate what aspect of the Scw signalling mechanism is incompatible with the Gbb ligand, given that endogenous Gbb protein can be detected in the embryo at this stage, and that the function of the Sax extracellular domain is conserved even in species that lack Scw (see below), the incompatibility must lie downstream of secretion of the active ligand and upstream of receptor binding. Notably, this part of the pathway includes the interactions with the suite of extracellular modulators that are different in the Gbb and Scw signalling mechanisms.

In contrast to the results with the scw=Gbb construct, the gbb=Scw construct can rescue *gbb* mutants. In the

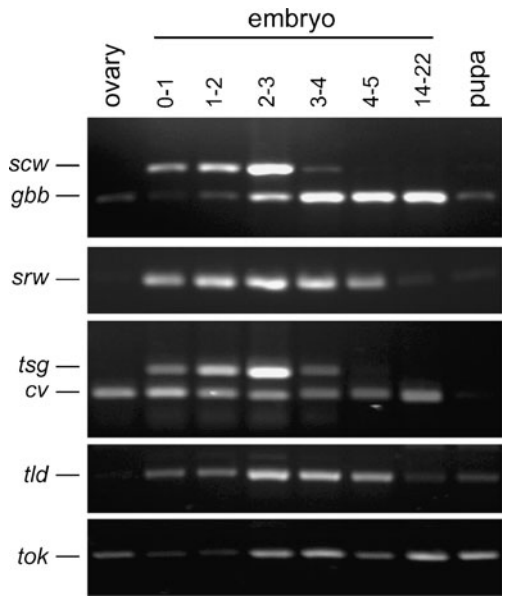

Fig. 3 Expression profiles of derived and ancestral BMP components. Expression profiles of BMP components were determined by RT-PCR on RNA isolated from the stages indicated. The expression profiles of *scw*, *tsg*, *srw*, and *tld* peak at 2–3 h of embryogenesis, and expression of *scw*, *srw*, and *tsg* is confined to early embryogenesis. The ancestral components reach their peak expression after 2–3 h of development, and this expression persists into the later stages

Table 1 Rescue of *gbb*, *scw*, and *sax* mutants with chimeric transgenes and species exchanges

Rescue of *gbb* mutants				
		Progeny[a]		%
Transgene	N	*gbb/Df*	*gbb/Bl*	Exp
DmelGbb-3	512	116	74	**156**
scw=Gbb-34.1.4	755	0	138	**0**
gbb=Scw-20	814	2(85 ph)	151	**1**
scw=Gbb-34+14 (2x)	505	0	107	**0**
gbb=Scw-11+20 (2x)	393	2(74 ph)	85	**2**
Dm-gbb=AgamGbb1PVLD-2.5.1	589	113	100	**113**
Dm-gbb=AgamGbb1PVLD-20.1.4	575	133	90	**147**
Dm-gbb=AgamGbb2PVLD-18.3.1	650	132	80	**165**
DgriGbb-12.3	605	117	112	**104**
DgriGbb-12.2.2	633	136	109	**125**
Rescue of *scw* mutants				
		Progeny[a]		%
Transgene	N	*scw/Df*	*scw/Bl*	Exp
DmelScw-4	682	108	108	**100**
scw=Gbb-34.1.4	764	0	144	**0**
gbb=Scw-20	524	0	129	**0**
scw=Gbb-34+14 (2x)	565	0	117	**0**
gbb=Scw-11+20 (2x)	524	0	91	**0**
DgriScw-3.2	696	131	103	**127**
DgriScw-15.2	742	113	125	**90**
GmorScw-10.1	334	0	58	**0**
GmorScw-F6.1	324	0	46	**0**
Dm-scw=GmorScwPVLD-F22	411	0	92	**0**
Dm-scw=GmorScwPVLD-21	374	0	78	**0**
Dm-scw=GmorScwLD-2	630	0	85	**0**
Dm-scw=GmorScwLD-F3	556	0	108	**0**
Rescue of *sax* mutants				
		Progeny[a]		%
Transgene	*N*	*sax/Df*	*sax/Bl*	Exp
DmelSax-4.1.1	442	61	62	**98**
Dm-sax=AgamSax^{EX}DmelSax^{IN}-F12.1.1f	494	74	84	**88**

[a] Progeny were derived from a cross of *y w; Df*/Bl; Tn[w+]/+* x *y w; mut*/CyO* which yields eight classes, two of which are shown (for the full table, see Table S5). Rescue of *gbb* mutants by gbb=Scw produced a small number of adult escapers and a shift from larval to pharate (ph) lethal that are indicated separately in the table.The statistic 'percent expected' was calculated by dividing the number of *mut*/Df* progeny by the number of *Df*/Bl* progeny and multiplying the quotient by 100. For *gbb*, the heading *gbb/Df* refers to the genotype gbb^{D20}, *pr cn/Df(2R)S246*, and for *scw*, the heading *scw/Df* refers to the genotype scw^{S12}, *b pr cn/Df (2L)OD16*. The heading *sax/Df* refers to sax^{4}, *nub b pr/Df(2R) BSC265*

wing, *gbb* has long-range and local functions: it acts long-range from a focus along the anterior–posterior compartment boundary to influence wing size and specification of the most posterior longitudinal vein L5 (Fig. 4a), and locally from foci along the lengths of the longitudinal veins and crossveins in the process of vein promotion and maintenance (Ray and Wharton 2001). Wings with anterior clones of the null allele gbb^{D20} are smaller than wild type and L5 is truncated at the junction with the posterior crossvein (pcv; Fig. 4a). A single copy of the gbb=Scw transgene can rescue both defects producing a wing of roughly normal size with L5 running to the margin (Fig. 4b). Posterior gbb^{D20} clones lack the pcv and the distal tip of L5 (Ray and Wharton 2001), and a single copy of gbb=Scw partially rescues both defects producing a fragment of the pcv midway between L4 and L5 and an almost complete L5 (Fig. 4c). gbb=Scw also rescues the lethality and wing defects associated with *gbb* hypomorphs. The hypomorphic allele gbb^{4} in trans to gbb^{D20} is almost completely lethal, and the rare escapers have small wings that lack the pcv and the distal quarter of L5 (Fig. 4d). A single copy of the gbb=Scw transgene rescues the lethality of this genotype from 0% to 34% of expected, and the pcv and L5 are partially restored (data not shown), and two copies rescue the pcv and L5 to nearly wild type (Fig. 4e). Rescue is even observed in gbb^{D20} hemizygotes, where a single copy of the transgene

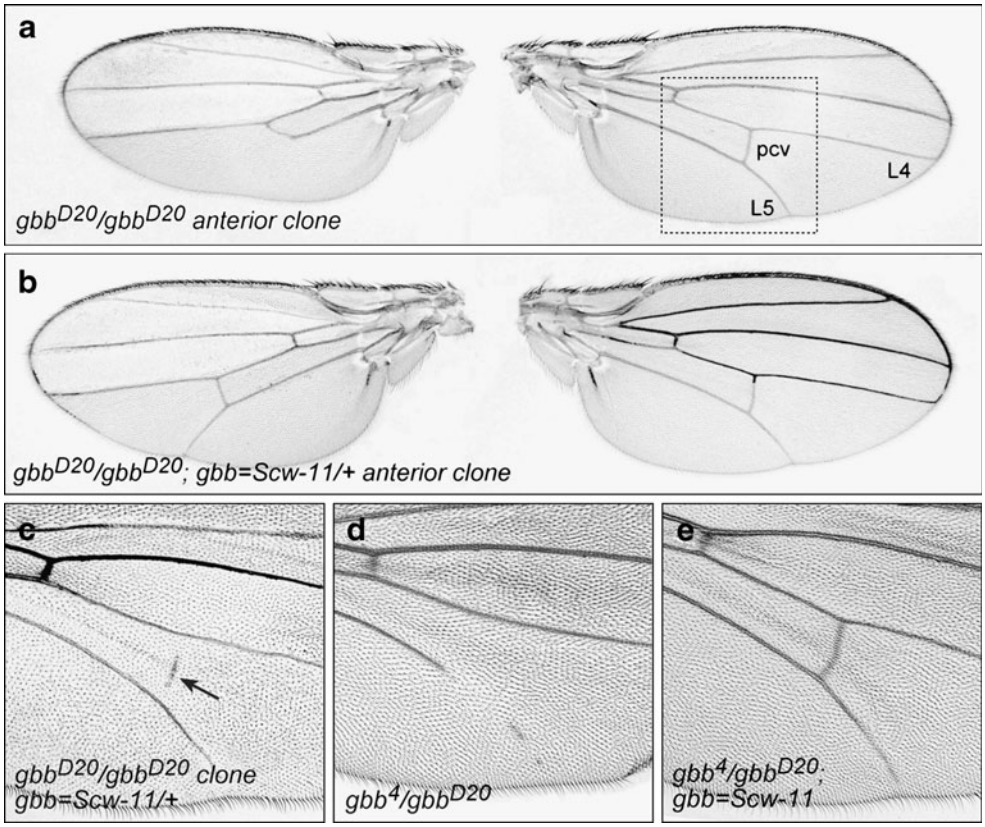

Fig. 4 Rescue of *gbb* mutant phenotypes with a gbb=Scw transgene. Null gbb^{D20} clones in the anterior compartment result in a reduction in wing size and a loss of L5 to the junction with the posterior crossvein (pcv) (**a**, *left side*) shown in comparison with the wild-type wing of the same fly (**a**, *right side*; the *box* indicates the portion of the wing shown in panels **c**–**e**). In the presence of a single copy of the gbb=Scw transgene, both phenotypes are rescued: the wing is nearly-normal in size and L5 reaches the margin (**b**). Null gbb^{D20} clones in the posterior compartment reveal the local functions: they completely lack the pcv and the distal tip of L5 (Ray and Wharton 2001). These phenotypes are partially rescued by a single copy of the gbb=Scw transgene, resulting in a small fragment of crossvein that is detached from the two longitudinal veins (*arrow* in **c**). The gbb=Scw transgene can completely rescue the wing phenotypes associated with hypomorphic *gbb* alleles. The hypomorphic allele gbb^4 in trans to gbb^{D20} is semi-viable and produces a phenotype that is a composite of local and long-range functions lacking the pcv and most of distal L5 (**d**). Two copies of the transgene rescue these phenotypes to wild type (**e**)

rescues the lethality of gbb^{D20}/Df from the larval/pupal stage to pharate, with rare viable escapers, and two copies of the transgene produce the same shift in lethal phase with a higher frequency of escapers (Table 1, Table S5). These results demonstrate that Scw can at least partially compensate for Gbb in all of its essential functions.

The function of Scw is not conserved within the higher Diptera

The non-reciprocal exchangeability of Gbb and Scw suggests that since the duplication event, Scw has evolved a signalling mechanism that is related to, but distinct from, that of Gbb. To assess the functional conservation of Gbb and Scw in different species, we tested whether heterospecific Gbb and Scw proteins can rescue *gbb* and *scw* mutants in *Drosophila*. Within the Drosophilidae, Gbb cis-regulation and function are conserved, as we observe complete rescue of *D. melanogaster gbb* mutants with one copy of a *D. grimshawi* genomic rescue construct (DgriGbb, Table 1, Table S5). To test for rescue with mosquito Gbb, we generated chimeras fusing the *D. melanogaster gbb* cis-regulatory sequences to the coding sequences of *Anopheles* Gbb1 and Gbb2 (Dm-gbb=AgamGbb1PVLD, Dm-gbb=AgamGbb2PVLD). A single copy of either chimera fully rescues *gbb* mutants (Table 1, Table S5), demonstrating that the function of Gbb is conserved across the Diptera. Moreover, chimeras fusing *D. melanogaster* Gbb with the ligand domains of human BMP5, BMP6, or BMP7 also rescue *gbb* mutants in flies (Table S4), thus, Gbb function is conserved both within the Diptera and between arthropods and vertebrates.

In contrast to the broad functional conservation of Gbb, Scw function is not conserved even within the higher Diptera. In the Drosophilidae, Scw cis-regulation and function are conserved, with the *D. grimshawi* genomic construct (DgriScw) fully rescuing *scw* mutants in *D. melanogaster* (Table 1, Table S5). By contrast, a genomic

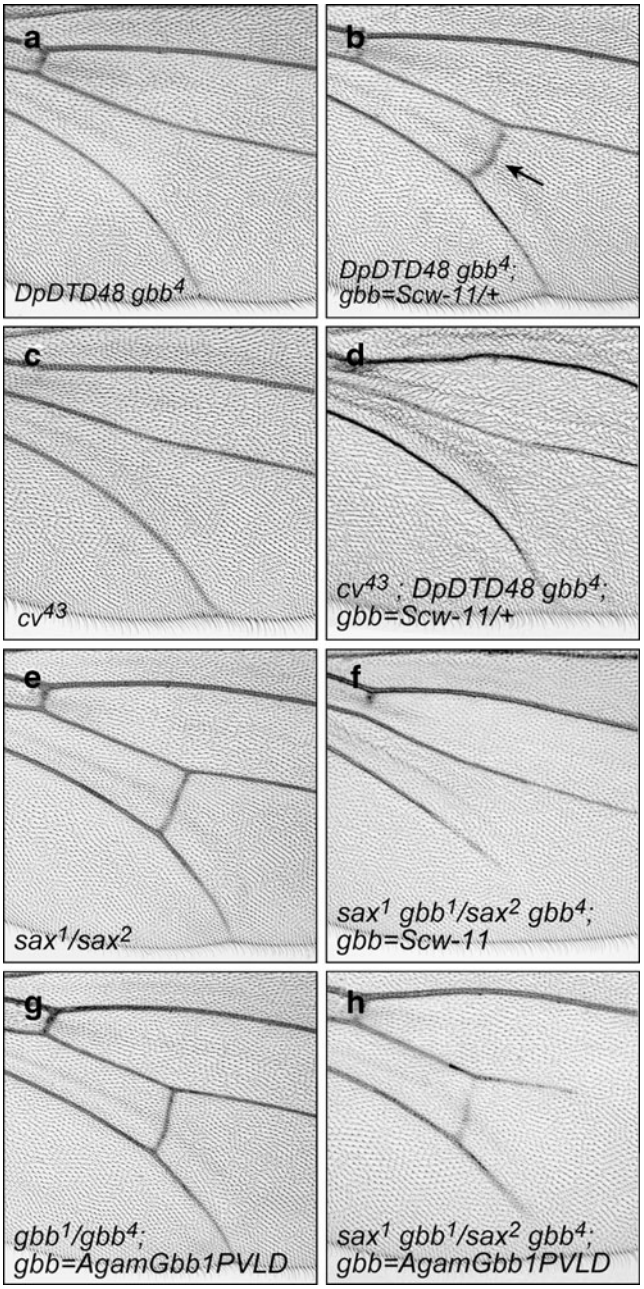

Fig. 5 In the wing Gbb and Scw use the same extracellular modulators but different receptors. In the presence of a duplication of *dpp* (Dp(2;2)DTD48), gbb^4 homzygotes are viable, but completely lack the pcv (**a**), and a single copy of the gbb=Scw transgene rescues this pcv defect (*arrow* in **b**). This rescue of the gbb^4 pcv defect by gbb=Scw requires *crossveinless* (*cv*), as the rescue is completely suppressed in a *cv* mutant background (**c**, **d**). Thus, in this context, Scw signalling requires Cv. In crossvein specification, Gbb signalling does not require the Type I receptor Sax (Ray and Wharton 2001), and consistent with this, sax^1/sax^2 mutants have wild-type crossveins (**e**). By contrast, the rescue of the venation defects of gbb^4/gbb^1 produced by the gbb=Scw transgene is completely suppressed in a sax^1/sax^2 background to the extent that the phenotype resembles that of gbb^1/gbb^4 alone (**f** cf. Fig. 4d). *Anopheles* Gbb1 and Gbb2 can fully rescue *gbb* mutants in *Drosophila* (see Table 1, Table S5), and the wings of these flies have wild-type venation. Similarly, *Anopheles* Gbb1 and Gbb2 can rescue the venation defects of gbb^1/gbb^4 mutants (**g**). This rescue depends, in part, on Sax, as it is suppressed in a sax^1/sax^2 mutant background (**h**). However, since this suppressed phenotype is distinct from that of gbb^1/gbb^4 alone, the *Anopheles* Gbb proteins must also be signalling through Tkv

rescue construct of *G. morsitans* Scw (GmorScw) does not rescue *scw* mutants, nor do chimeras fusing the *D. melanogaster* cis-regulatory sequences to the complete *G. morsitans* Scw protein (Dm-scw=GmorScwPVLD), and the *D. melanogaster* cis-regulatory and pro-domain sequences to the *G. morsitans* ligand domain (Dm-scw= GmorScwLD, Table 1, Table S5). These results show that there are not only differences in cis-regulation but also in protein sequence that interfere with the functional exchange of Scw proteins. Thus, in contrast to Gbb, which is functionally conserved across a range of distantly related species, Scw is not only functionally divergent from Gbb but also from the Scw orthologues of other flies.

A unique signalling mode has evolved with Scw

Given that Gbb cannot rescue *scw* mutants and Scw can only partially rescue the lethality of *gbb* mutants, our data suggest that Gbb and Scw not only differ in their sequences but also in their mode of signalling. Indeed, genetic studies have revealed that Scw and Gbb interact with a distinct suite of extracellular modulators and have different affinities and preferences for the Type I receptors Tkv and Sax. In the early embryo, Scw signalling requires the Chordin orthologue Sog, the metalloprotease Tld, and the cysteine-rich binding proteins Tsg and Srw, and is transduced by the Type I receptor Sax (Bonds et al. 2007; Holley et al. 1995; Neul and Ferguson 1998; Shimmi et al. 2005a). In the pupal wing, Gbb signalling requires Sog, the metalloprotease Tok, and the cysteine-rich binding protein Cv, and is transduced primarily through the Type I receptor Tkv (Ray and Wharton 2001; Serpe et al. 2005; Shimmi et al. 2005a). Like Gbb and Scw, these regulators and receptors are in paralogous pairs: Tld is a paralogue of Tok, Tsg and Srw are paralogues of Cv, and Sax is a paralogue of Tkv.

To address how these modulators have evolved, we traced their evolutionary origin in the Insect lineage. The phylogenetic analyses indicate that, like Scw, the embryonic regulators Tld, Tsg, and Srw arose in the lineage leading to higher Diptera, while the pupal regulators Tok and Cv are found in the Drosophilidae, mosquitoes, and all other sequenced insect genomes, and are thus ancestral. Thus, just as Scw arose from a duplication of Gbb in the lineage leading to the higher Diptera, Tld, Tsg and Srw arose from duplications of the ancestral Tok and Cv genes (Figure S1). Moreover, like Scw, the embryonic regulators all show peaks of expression at 2–3 h after egg deposition (Fig. 3) and are evolving at a faster rate than the ancestral

pupal counterparts (Bayes factors >5, Table S2, Fig. 2b–d). As is the case with Scw, we find no evidence for positive selection in Tld, Tsg or Srw using either of two different tests ($p>0.05$ in all cases, Table S3), suggesting that the faster rates of evolution in these proteins are due primarily to relaxed, rather than positive, selection. Thus, it is not only Scw that was duplicated and rapidly diverged in the lineage leading to the higher Diptera but also the suite of extracellular regulators that are required for its function. Moreover, the fact that these proteins are all evolving at rates faster than their ancestral paralogues suggests that the entire suite of proteins is evolving with Scw to maintain its early embryonic function.

Scw interacts with the Gbb extracellular regulators but uses a different receptor

The fact that gbb=Scw can rescue the pcv in *gbb* mutant wings indicates that Scw is not only able to interact with the early embryonic suite of extracellular modulators but also with the ancestral set that is essential for pcv formation. Consistent with this, the gbb=Scw transgene cannot rescue the crossveinless phenotype of gbb^4 homozygotes in the absence of Cv (Fig. 5a–d). Thus, Scw is not specific to the suite of modulators that have evolved with it to perform the early embryonic function. This result raises the question of whether Scw signalling in the presumptive pcv depends on Sax, as it does in the embryo (Neul and Ferguson 1998), or on Tkv, as Gbb does in the wing (Ray and Wharton 2001). To address this issue, we tested for rescue of the crossvein defect of gbb^1/gbb^4 by gbb=Scw in a *sax* mutant background. The maternal effect *sax* alleles, sax^1 and sax^2, recapitulate the loss-of-function phenotype of zygotic *scw* mutants in the early embryo, which has led to the conclusion that Sax is the primary transducer of Scw signalling in this context (Neul and Ferguson 1998; Nguyen et al. 1998). We used this same test to determine whether the rescue of the *gbb* crossvein defect by gbb=Scw was dependent on *sax*. In an otherwise wild-type background, sax^1 and sax^2 homozygotes or hemizygotes have a wild-type posterior crossvein, demonstrating that these alleles have no effect on pcv specification when Gbb is the ligand (Fig. 5e). By contrast, gbb=Scw cannot rescue the venation defects of gbb^1/gbb^4 mutants in a sax^1/sax^2 background (Fig. 5f, cf. Fig. 4d). Importantly, the phenotype of the $sax^1\ gbb^1/sax^2\ gbb^4$*; gbb=Scw* wings is essentially identical to that of gbb^1/gbb^4 wings (cf. Fig. 4d), indicating that the *sax* mutant background completely blocks the ability for gbb=Scw to rescue *gbb*. Thus, while Gbb signals primarily through Tkv in this context, Scw signals primarily through Sax, the receptor it would normally use in the early embryo.

While *Drosophila* Scw signals through Sax, Gbb can signal through both receptors, raising the question of whether ancestral Gbb proteins are specific for one receptor or the other, or can signal through both. To address this question, we determined the effect of the sax^1 and sax^2 mutations on the rescue of *Drosophila gbb* mutants by *Anopheles* Gbb1 and Gbb2. *Anopheles* Gbb1 and Gbb2 fully rescue the pcv defect of gbb^1/gbb^4 (Fig. 5g). In $sax^1\ gbb^1/sax^2\ gbb^4$ mutants bearing one copy of either Dm-gbb=AgamGbb1PVLD or Dm-gbb=AgamGbb2PVLD, we observe partial rescue of the pcv, loss of the distal tip of L5, and additionally, loss of the distal quarter of L4 (Fig. 5h). Thus, unlike *Drosophila* Gbb, which does not require Sax in this context (cf. Fig. 5e), either *Anopheles* Gbb protein requires Sax to fully rescue the *gbb* mutant phenotype, and thus, must be able to signal through Sax. However, as the *gbb* mutant phenotype is only partially revealed in a *sax* mutant background, the *Anopheles* Gbb proteins must also be able to signal through Tkv. Thus, in the ancestral state, both Sax and Tkv were receptors for Gbb-like ligands, and in the lineage leading to the higher Diptera, Scw has evolved a specificity for Sax.

Changes in Scw are required for its specificity for Sax

Given that the *Anopheles* and *Drosophila* Gbb proteins can signal through Tkv and Sax, while *Drosophila* Scw signals only through Sax, the specificity of Scw for Sax is another novelty of the Scw signalling mechanism. This raises the question of whether this specificity arose from changes in Scw, Sax, or both. Like Scw, Sax has diverged since the separation of the Culiciformae from the lineage leading to the higher Diptera, particularly in its extracellular domain. *Anopheles* and *Drosophila* Sax proteins show 76% amino acid identity in the intracellular domains, but only 42% in their extracellular domains. To address whether the changes in Sax are important for Scw specificity, we generated a chimeric receptor rescue construct fusing the extracellular domain of *Anopheles* Sax with the intracellular domain of *Drosophila* Sax under control of the endogenous *Drosophila* cis-regulatory sequences (sax=AgamSax^{Ex}DmelSax^{IN}). A single copy of either the wild-type 8.7-kb sax genomic rescue construct (DmelSax) or the chimeric construct completely rescues *sax* null mutants in *Drosophila* both for the zygotic requirement and for the maternal requirement (Table 1, Table S5). Thus, the extracellular domain of *Anopheles* Sax can transduce both Gbb and Scw signals in *Drosophila*, indicating that the exclusive specificity of Scw for Sax has resulted from changes in the ligand or extracellular regulators and not in the receptor.

Discussion

Evolution of Scw signalling in the higher Diptera

The BMP5/6/7/8 orthologue Scw arose from a duplication of Gbb in the lineage leading to the higher Diptera. Since this duplication event, Gbb has retained the ancestral sequence and functionality, while Scw has diverged profoundly. Remarkably, the evolution of Scw signalling is not a simple event of ligand duplication and divergence: the duplication event that gave rise to Scw was at some level coordinated with duplications—and in one case, reduplication—of other signalling components and a shift in receptor specificity resulting in a novel signalling mechanism that is specific for its function in early embryogenesis (Fig. 6). Indeed, while Scw can partially rescue mutations in *gbb*, this functional exchangeability is not reciprocal: Gbb cannot rescue *scw* mutations and thus cannot replace Scw in its early embryonic function. Thus, the Scw signalling mechanism in the early embryo is at some level incompatible with Gbb. Curiously, Scw continues to diverge within the higher Diptera, and Scw proteins are not interchangeable even between relatively closely related species. As *scw* coding sequences show no evidence of positive selection, this divergence appears to be due to relaxed selection and drift. The continued divergence, higher evolutionary rates, and lack of any evidence of positive selection of Tld, Tsg and Srw indicate that the whole suite of novel extracellular proteins, including Scw, is evolving as a unit with the only constraint being the functionality of that unit in dorsal–ventral patterning. The implication from these results is that the restriction of Scw function to the early embryo and the duplication of key regulatory proteins were crucial steps in allowing this novel mechanism to diverge from the ancestral Gbb signalling mechanism.

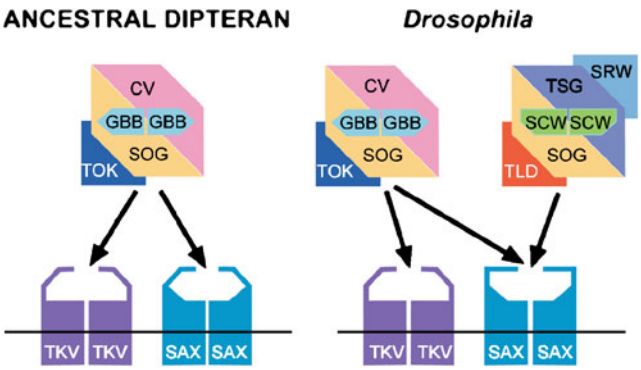

Fig. 6 Evolution of Scw and Gbb signalling in the Diptera. In the ancestral Dipteran, a Gbb-like molecule (or molecules) functions in combination with a suite of extracellular modulators including Sog, Cv and Tok, and signals through both Type I receptors Tkv and Sax. In the lineage leading to the higher Diptera, duplication events gave rise to Scw from Gbb, Tld from Tok, and Tsg and Srw from Cv, generating a new signalling mode that diverged from the ancestral Gbb mode. In *Drosophila*, these two modes have different receptor specificities, with the Gbb mode signalling through Tkv and Sax in a context-dependent fashion, and the Scw mode signalling exclusively through Sax, suggesting that this exclusivity for Sax has emerged as the Scw mode diverged from the Gbb mode

A striking feature of the appearance of the Scw signalling mode is how quickly it evolved. In the 100–250 million years since the duplication event, Scw has diverged to the extent that the sequence of its ligand domain no longer reflects its evolutionary origin, and its signalling mode has diverged to the point that it cannot be substituted by its most closely related BMP family member, Gbb. Moreover, as there is evidence that this mechanism is operating in both branches of the Schizophora, and thus, must have arisen before the separation of these two lineages, the entire mechanism must have evolved in the ~150 million years between the split of the mosquitoes and the higher Diptera but prior to the radiation of the Schizophora. This is evidence for the remarkable plasticity of the BMP pathway to evolve into new signalling mechanisms in a relatively short span of evolutionary time.

The embryonic function of Scw is derived

The mechanistic similarity between the function of Scw in dorsal–ventral patterning and that of Gbb in specification of the pcv has been previously noted (O'Connor et al. 2006), and the demonstration of a common evolutionary origin of Gbb and Scw provides a causal explanation for this similarity: the Scw signalling mechanism arose directly from the ancestral Gbb-based mechanism, thus, they employ similar co-factors to achieve a mechanistically similar outcome (Fig. 6). Nevertheless, the ability to exchange components between the embryonic and crossvein functions is for the most part unidirectional. We have shown that Scw can substitute for Gbb in crossvein specification, and thus, will function to some extent with the ancestral regulators Tok and Cv, but Gbb cannot substitute for Scw in the early embryo, suggesting an incompatibility with the dorsal–ventral patterning system. Similarly, consistent with a previous report (Shimmi et al. 2005a), we have found that Tsg can rescue the crossvein defect of *cv* mutants (data not shown), but the ability for Cv to rescue *tsg* mutants has yet to be conclusively demonstrated. Finally, Tld and Tok appear to be uniquely suited to their particular functions, with neither being able to substitute for the other (Serpe et al. 2005). Thus, with the exception of Tld/Tok, either system is capable of functioning in the context of crossvein specification, but only the derived embryonic regulators function in the context of dorsal–ventral patterning. Thus, it is not the specificity of

the ligands to a particular signalling mechanism, but rather the function itself that is incompatible. In this regard, it must be kept in mind that our experiments assayed for *rescue* of a mutant phenotype and not signalling per se. In our swap experiment, scw=Gbb may well be signalling in the early embryo, but not well enough to rescue *scw* mutants to viability. Thus, the non-reciprocal exchangeability may be due to a difference in stringency and not functionality. Indeed, it is not difficult to image scenarios that might account for this difference. For instance, it is known that embryonic patterning requires synergism between the receptors Tkv and Sax (Neul and Ferguson 1998; Nguyen et al. 1998), and a similar synergism has been reported between the Tkv and Sax orthologues in dorsal–ventral patterning in zebrafish (Little and Mullins 2009). The changes in the Scw signalling system that resulted in its specificity for Sax may have optimised this synergism between Tkv and Sax, leaving Gbb incapable of functioning in this context.

Why does Scw appear in the higher Diptera?

Given the fact that the only function of *Drosophila* Scw is to specify the extraembryonic amnioserosa, it is tempting to speculate that the signalling mechanism may have evolved specifically for the specification of this tissue, which, like Scw, is unique to the higher Diptera. Nematoceran flies such as mosquitoes have two distinct extraembryonic membranes, the amnion and the serosa, while most higher flies, including the Schizophora, have a single extraembryonic tissue, the amnioserosa, that occupies a dorsal position in the early embryo (Rafiqi et al. 2008). The appearance of Scw basal to the Schizophora is consistent with the notion that the signalling mechanism evolved at least in parallel with the evolution of the amnioserosa. However, as the involvement of BMP signalling in early dorsal–ventral patterning is common to all Bilateria, the Scw signalling mechanism has presumably evolved to modulate the existing BMP2/4/Dpp dorsal–ventral patterning mechanism that is common to all metazoans. In this light, the Scw signalling mechanism may have arisen either by subfunctionalization or neofunctionalization. In the former case, the ancestral Gbb would have been involved both in dorsal–ventral patterning and processes in later development, and these two functions were subdivided between the two products of the duplication event. In the latter case, the ancestral Gbb would have only been involved in the later developmental processes, and the duplication event allowed Scw to be recruited to the early embryonic function. Analysis of the function of Gbb paralogues in the lower Diptera or other insect species will resolve this issue. On a broader scale, BMP5/6/7/8 orthologues have been implicated in early dorsal–ventral patterning in zebrafish and *Xenopus*, but not in birds and mammals (Dick et al. 2000), so it is possible that the requirement for a BMP5/6/7/8 orthologue in this process may not be absolute, but rather a lineage-specific adaptation. Thus, in fish and frogs, as in the higher Diptera, a BMP5/6/7/8 has been recruited to modulate the function of the BMP2/4/Dpp dorsal–ventral patterning mechanism to accommodate specific developmental features unique to these lineages.

Scw as a paradigm for the origin of BMP diversity

Our findings about the origin of Scw provide evidence that sequence identity within the ligand domain is not necessarily the most reliable indicator of the relatedness of two BMPs. As such, the evolution of Scw may serve as a paradigm for understanding the diversity of BMP family members. In the vertebrate branch of the BMP family, there are many ‘orphan’ ligands that do not fit into the canonical BMP classes, and given that Scw was once one of these ‘orphans’, some of these vertebrate proteins may have originated from one of the canonical classes by recent duplication and divergence. Indeed, like Scw, vertebrate BMPs appeared in a relatively short span of evolutionary time with the Hemichordates and Cephalochordates having single representatives of the main BMP classes while the fishes have the full repertoire of 20 or more seen in the higher chordates. This expansion included the duplication of members within a class, such as BMP2/4 and BMP56/7/8, and the appearance of new classes such as GDF5/6/7, BMP3/GDF10, and GDF1/3. The example of Scw provides a possible mechanism for this rapid appearance of these novel BMP classes and suggests that some of these classes may be more closely related to the canonical BMP classes than they appear based on ligand domain sequence identity alone.

Vertebrate BMPs also show evidence for duplicated members adopting novel signalling modes. Of the four members of the BMP5/6/7/8 class, BMP8 is the most divergent, both from the other members of the class and from the Gbb proteins in arthropods. In this study, we found that while BMP5, BMP6, and BMP7 rescue Gbb-dependent functions in *Drosophila*, BMP8 does not, and fails to produce dominant phenotypes when mis-expressed at high levels in *Drosophila* tissues (data not shown). Thus, while BMP5, 6, and 7 have retained the ability to function in the context of the *Drosophila* signalling machinery, BMP8 has diverged to the extent that it can no longer do so. The implication is that BMP8 has diverged into a new signalling mode that is to some extent unique to it, and incompatible with the ancestral signalling machinery that works for the other members of the class.

Acknowledgements The authors would like to thank Claire Greer and James Patterson for their contributions to the early stages of this work and to the *Drosophila* Stock Centers, Malaria Research and

Reference Reagent Resource, and the *Drosophila* Genome Resource Center for reagents. We are grateful to Wendy Gibson for providing *Glossina morsitans*, and Hazel Smith for the stalk-eyed flies *Teleopsis dalmanni* and *Diasemopsis meigenii*, Mark Wamalwa for access to the draft version of the *G. morsitans* genome, and Al Handler for access to the draft version of the *C. capitata* genome and Elio Sucena for performing the BLAST searches for Gbb and Scw. We would like to thank Hilary Ashe, Mark Dionne, Florian Maderspacher, and Stefan Thomsen for critical reading of the manuscript and to colleagues in the School of Life Sciences for helpful discussions. During the writing of this manuscript, C.F. was supported by a Wellcome Trust Programme Grant to Juan-Pablo Couso. The research was funded by grants from the Biotechnology and Biological Sciences Research Council (BB/C508050/1) and Medical Research Council (G0500916) to R.P.R.

References

Arora K, Nusslein-Volhard C (1992) Altered mitotic domains reveal fate map changes in *Drosophila* embryos mutant for zygotic dorsoventral patterning genes. Development 114:1003–1024

Arora K, Levine MS, O'Connor MB (1994) The screw gene encodes a ubiquitously expressed member of the TGF-beta family required for specification of dorsal cell fates in the *Drosophila* embryo. Genes Dev 8:2588–2601

Ballard SL, Jarolimova J, Wharton KA (2010) Gbb/BMP signaling is required to maintain energy homeostasis in *Drosophila*. Dev Biol 337:375–385

Bonds M, Sands J, Poulson W, Harvey C, Von Ohlen T (2007) Genetic screen for regulators of ind expression identifies shrew as encoding a novel twisted gastrulation-like protein involved in Dpp signaling. Dev Dyn 236:3524–3531

Casillas S, Negre B, Barbadilla A, Ruiz A (2006) Fast sequence evolution of *Hox* and *Hox*-derived genes in the genus *Drosophila*. BMC Evol Biol 6:106

Chen Y, Riese MJ, Killinger MA, Hoffmann FM (1998) A genetic screen for modifiers of *Drosophila* decapentaplegic signaling identifies mutations in punt, Mothers against dpp and the BMP-7 homologue, 60A. Development 125:1759–1768

Dick A, Hild M, Bauer H, Imai Y, Maifeld H, Schier AF, Talbot WS, Bouwmeester T, Hammerschmidt M (2000) Essential role of Bmp7 (snailhouse) and its prodomain in dorsoventral patterning of the zebrafish embryo. Development 127:343–354

Erwin DH, Davidson EH (2002) The last common bilaterian ancestor. Development 129:3021–3032

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531–1545

Gaunt MW, Miles MA (2002) An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. Mol Biol Evol 19:748–761

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

Holley SA, Jackson PD, Sasai Y, Lu B, De Robertis EM, Hoffmann FM, Ferguson EL (1995) A conserved system for dorsal-ventral patterning in insects and vertebrates involving sog and *chordin*. Nature 376:249–253

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

Innis CA, Shi J, Blundell TL (2000) Evolutionary trace analysis of TGF-beta and related growth factors: implications for site-directed mutagenesis. Protein Eng 13:839–847

Katoh Y, Katoh M (2006) Comparative integromics on BMP/GDF family. Int J Mol Med 17:951–955

Lemons D, McGinnis W (2006) Genomic evolution of Hox gene clusters. Science 313:1918–1922

Little SC, Mullins MC (2009) Bone morphogenetic protein heterodimers assemble heteromeric type I receptor complexes to pattern the dorsoventral axis. Nat Cell Biol 11:637–643

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290:1151–1155

Lynch M, Force A (2000) The probability of duplicate gene preservation by subfunctionalization. Genetics 154:459–473

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. CSHL Press, Cold Spring Harbor

McCabe BD, Marques G, Haghighi AP, Fetter RD, Crotty ML, Haerry TE, Goodman CS, O'Connor MB (2003) The BMP homolog Gbb provides a retrograde signal that regulates synaptic growth at the *Drosophila* neuromuscular junction. Neuron 39:241–254

McGregor AP (2005) How to get ahead: the origin, evolution and function of bicoid. Bioessays 27:904–913

Neul JL, Ferguson EL (1998) Spatially restricted activation of the SAX receptor by SCW modulates DPP/TKV signaling in *Drosophila* dorsal-ventral patterning. Cell 95:483–494

Newfeld SJ, Wisotzkey RG, Kumar S (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

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

O'Connor MB, Umulis D, Othmer HG, Blair SS (2006) Shaping BMP morphogen gradients in the *Drosophila* embryo and pupal wing. Development 133:183–193

Pirrotta V (1988) Vectors for P-element transformation in *Drosophila*. In: Rodriguez RL, Denhardt DT (eds) Vectors: a survey of molecular cloning vectors and their uses. Butterworth, Boston, pp 437–456

Rafiqi AM, Lemke S, Ferguson S, Stauber M, Schmidt-Ott U (2008) Evolutionary origin of the amnioserosa in cyclorrhaphan flies correlates with spatial and temporal expression changes of zen. Proc Natl Acad Sci USA 105:234–239

Raftery AE (1996) Hypothesis testing and model selection. In: Gilks WR, Richardson S, Spiegelhalter DJ (eds) Markov chain Monte Carlo in practice. Chapman & Hall, London, pp 163–188

Rawson JM, Lee M, Kennedy EL, Selleck SB (2003) *Drosophila* neuromuscular synapse assembly and function require the TGF-beta type I receptor saxophone and the transcription factor Mad. J Neurobiol 55:134–150

Ray RP, Wharton KA (2001) Context-dependent relationships between the BMPs *gbb* and *dpp* during development of the *Drosophila* wing imaginal disk. Development 128:3913–3925

Ronquist F, van der Mark P, Huelsenbeck JP (2009) Bayesian phylogenetic analysis using MRBAYES. In: Lemey P, Salemi M, Vandamme A-M (eds) The phylogenetic handbook. Cambridge University Press, Cambridge

Serpe M, Ralston A, Blair SS, O'Connor MB (2005) Matching catalytic activity to developmental function: tolloid-related processes Sog in order to help specify the posterior crossvein in the *Drosophila* wing. Development 132:2645–2656

Shimmi O, O'Connor MB (2003) Physical properties of Tld, Sog, Tsg and Dpp protein interactions are predicted to help create a sharp boundary in Bmp signals during dorsoventral patterning of the *Drosophila* embryo. Development 130:4673–4682

Shimmi O, Ralston A, Blair SS, O'Connor MB (2005a) The *crossveinless* gene encodes a new member of the Twisted gastrulation family of BMP-binding proteins which, with Short gastrulation, promotes BMP signaling in the crossveins of the *Drosophila* wing. Dev Biol 282:70–83

Shimmi O, Umulis D, Othmer H, O'Connor MB (2005b) Facilitated transport of a Dpp/Scw heterodimer by Sog/Tsg leads to robust patterning of the *Drosophila* blastoderm embryo. Cell 120:873–886

Singer MA, Penton A, Twombly V, Hoffmann FM, Gelbart WM (1997) Signaling through both type I DPP receptors is required for anterior-posterior patterning of the entire *Drosophila* wing. Development 124:79–89

Suchard MA, Weiss RE, Sinsheimer JS (2001) Bayesian selection of continuous-time Markov chain evolutionary models. Mol Biol Evol 18:1001–1013

Twombly V, Bangi E, Le V, Malnic B, Singer MA, Wharton KA (2009) Functional analysis of *saxophone*, the Drosophila gene encoding the BMP type I receptor ortholog of human ALK1/ACVRL1 and ACVR1/ALK2. Genetics 183:563–579, 1SI-8SI

Van der Zee M, da Fonseca RN, Roth S (2008) TGFbeta signaling in *Tribolium*: vertebrate-like components in a beetle. Dev Genes Evol 218:203–213

Vilmos P, Sousa-Neves R, Lukacsovich T, Marsh JL (2005) *crossveinless* defines a new family of Twisted-gastrulation-like modulators of bone morphogenetic protein signalling. EMBO Rep 6:262–267

Wharton KA, Thomsen GH, Gelbart WM (1991) Drosophila 60A gene, another transforming growth factor beta family member, is closely related to human bone morphogenetic proteins. Proc Natl Acad Sci USA 88:9214–9218

Wharton KA, Cook JM, Torres-Schumann S, de Castro K, Borod E, Phillips DA (1999) Genetic analysis of the bone morphogenetic protein-related gene, *gbb*, identifies multiple requirements during Drosophila development. Genetics 152:629–640

Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol 18:691–699

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591

Yang Z, Nielsen R, Goldman N, Pedersen AM (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431–449

Yeates DK, Wiegmann BM (1999) Congruence and controversy: toward a higher-level phylogeny of Diptera. Annu Rev Entomol 44:397–428

Zhang J (2003) Evolution by gene duplication: an update. Tends in Ecology and Evolution 18:292–298