Article:
Skelly, J, Pushparajan, C, Duncan, EJ orcid.org/0000-0002-1841-504X et al. (1 more author) (2019) Evolution of the Torso activation cassette, a pathway required for terminal patterning and moulting. Insect Molecular Biology, 28 (3). pp. 392-408. ISSN 0962-1075

https://doi.org/10.1111/imb.12560

Evolution of the Torso activation cassette, a pathway required for terminal patterning and moulting

John Skelly[1], Charlotte Pushparajan[1], Elizabeth Duncan[2], Peter Dearden[1*]

1
Laboratory for Evolution and Development
Genomics Aotearoa
Biochemistry Department
University of Otago
P.O. Box 56
Dunedin
Aotearoa-New Zealand

2
School of Biology
Faculty of Biological Sciences
University of Leeds
Leeds LS2 9JT
United Kingdom

*To whom correspondence should be addressed at peter.dearden@otago.ac.nz

Keywords

Terminal patterning; ecdysis; genome evolution; moulting;

Abstract

Embryonic terminal patterning and moulting are critical developmental processes in insects. In *Drosophila* and *Tribolium* both of these processes are regulated by the Torso-activation cassette (TAC). The TAC consists of a common receptor, Torso, ligands Trunk and PTTH, and the spatially restricted protein Torso-like, with combinations of these elements acting mechanistically to activate the receptor in different developmental contexts. In order to

trace the evolutionary history of the TAC we determined the presence or absence of TAC components in the genomes of arthropods. Our analyses reveal that Torso, Trunk and PTTH are evolutionarily labile components of the TAC with multiple individual or combined losses occurring in the arthropod lineages leading to and within the insects. These losses are often correlated, with both ligands and receptor missing from the genome of the same species. We determine that the *PTTH* gene evolved in the common ancestor of hemiptera and holometabola, and is missing from the genomes of a number of species with experimentally demonstrated PTTH activity, implying another molecule may be involved in ecdysis in these species. In contrast, the *torso-like* gene is a common component of pancrustacean genomes.

Introduction

One of the earliest and most important steps in insect embryogenesis is the specification of the embryonic axes. In *Drosophila melanogaster*, axis formation is accompanied by specification of the embryonic termini, structures which will go on to form the head and very posterior of the adult. The process of terminal patterning is dependent on the receptor tyrosine kinase Torso, the ligand Trunk, and a third protein called Torso-like, together dubbed the 'Torso-activation cassette' (TAC) (Duncan et al., 2014).

The receptor tyrosine kinase Torso was first identified in *Drosophila* via genetic screens as acting in terminal patterning in the just-laid embryo (Klingler et al. 1988). Torso signals via a highly conserved canonical MAP kinase pathway leading to expression of the terminal gap genes *tailless* and *huckebein* (Li 2005; Treisman 1996). *Drosophila*-Torso is activated by a ligand, Trunk (Casanova et al. 1995; Schüpbach and Wieschaus 1986), which is a small, extracellular, cysteine knot protein (Casanova et al. 1995). Trunk is cleaved before it can interact with Torso, and this is carried out intracellularly by the proprotein convertases Furin 1 and Furin 2 (Johnson et al. 2015). Neither Torso nor Trunk are spatially localised in the oocyte or early embryo (Casanova et al. 1995; Casanova and Struhl 1989; Sprenger et al. 1989) and restriction of signalling to the embryonic termini is thought to be controlled by the membrane-attack complex/ perforin (MACPF) domain protein, Torso-like (Savant-Bhonsale and Montell 1993a). *Torso-like* was identified in genetic screens as causing a similar phenotype to mutations in *torso* (Stevens et al, 1990) and named as such. Torso-like

is expressed in specialised follicle cells adjacent to the termini of the mature oocyte (Savant-Bhonsale and Montell 1993b). *Torso-like* protein is placed in the vitelline membrane (Mineo et al. 2015), and may be involved in processing (Casali and Casanova 2001) or secretion (Johnson et al. 2015) of Trunk acting to facilitate or enhance Trunk-mediated signalling through Torso, or may act directly on Torso itself (Amarnath et al. 2017). In the beetle *Tribolium castaneum*, Torso is required to specify the anterior serosa and posterior growth zone (Schoppmeier and Schröder 2005), a modification of terminal patterning seen in *Drosophila*.

In addition to their roles in embryogenesis, members of the TAC also act in post-embryonic development and immunity (Forbes-Beadle et al. 2016). For example, Torso also has a critical role in controlling the timing of moulting in larval *Drosophila* (Rewitz et al. 2009). Torso in this role is the receptor for prothoracicotropic hormone (PTTH), a cysteine knot protein similar to Trunk (Chung et al. 1994) considered part of the TAC (Duncan et al., 2014). PTTH was the first insect hormone identified (Kopec 1922; Wigglesworth 1934a) and has been extensively studied for its role in regulating moulting in the Lepidoptera (Fain and Riddiford 1976; Meola and Adkisson 1977; Nagasawa et al. 1984) and Hemiptera (Garcia et al. 1990; Wigglesworth 1934a; Wigglesworth 1934b). In *Drosophila,* PTTH is expressed by paired neurons and released to activate Torso in the prothoracic gland (Rewitz et al. 2009), which stimulates MAP kinase signalling leading to the release of ecdysteroids (McBrayer et al. 2007), and moulting. PTTH, like Trunk, appears to require cleavage for its activity (Sauman and Reppert 1996b). Unlike Trunk in *Drosophila*, PTTH signalling through Torso does not require Torso-like for its action (Grillo et al. 2012; Johnson et al. 2013), and although Torso-like is expressed by the prothoracic gland, and regulates developmental timing, this occurs through a pathway independent of PTTH and Torso (Johnson et al. 2013).

Trunk and PTTH are both cysteine knot proteins, and are part of a larger gene family that includes vertebrate *noggin* and invertebrate *noggin-like* genes (Duncan et al. 2013). The latter share sequence similarity with vertebrate *noggin* genes (Duncan et al. 2013; Molina et al. 2011) like *noggin*, two of these genes have been shown to encode inhibitors of BMP signalling (Molina et al. 2011). Expression of *noggin-like* genes in insects has only been

investigated in the pea aphid (*Acyrthosiphon pisum*) (Duncan et al. 2013) and may indicate a role for these proteins in neurogenesis.

As both regulation of moulting and terminal patterning are essential developmental functions, it came as some surprise that the honeybee (*Apis mellifera*) genome contained neither orthologs of *trunk* nor *torso* (Dearden et al. 2006; The Honey Bee Genome Sequencing Consortium 2006); and that terminal patterning is carried out through a different pathway in this species (Wilson and Dearden 2009). The honeybee genome also does not encode an ortholog of *PTTH* (Duncan et al. 2013), despite immunoreactivity for PTTH being reported in the developing larval brain (Simões et al. 1997) and the adult protocerebrum (Závodská et al. 2003). In another hymenopteran, the jewel wasp (*Nasonia vitripennis*), *PTTH* and *torso* are present in the genome, but not *trunk* (Werren et al. 2010), and terminal patterning occurs through a pathway that differs from both *Drosophila* and honeybee (Lynch et al. 2006). Studies in two hemipterans (an order of hemimetabolous insects which do not undergo complete metamorphosis) indicate that embryonic terminal patterning differs from that seen in *Drosophila* (Bickel et al. 2013; Duncan et al. 2013; Weisbrod et al. 2013). In the pea aphid, expression of Torso is restricted to paired lateral expression domains within the prothoracic segment (Duncan et al. 2013), suggesting a possible conserved role for Torso signalling in regulation of ecdysteroid release from the prothoracic gland. Independent of genome sequence, PTTH immunoreactivity (Sauman and Reppert 1996a) has been detected in many insects, including deeply branching groups such as the Odonata and Ephemeroptera (Závodská et al. 2003), and PTTH activity, as measured by the ability to stimulate ecdysteroid release from the prothoracic gland, have been detected in the Blattodea (Hiragaki et al. 2009; Richter 1992; Richter et al. 1999) suggesting that this may be the conserved function for the TAC in insects.

To test this hypothesis and investigate the evolutionary dynamics of the TAC, here we investigate 126 arthropod genomes for components of the TAC and provide the first comprehensive examination of the evolutionary history of this developmental pathway.

Results

The sequencing and assembly of genomes of multiple arthropod species, led by the i5K consortium (i5K Consortium 2013), allows a detailed examination of the presence or absence of genes and pathways during arthropod evolution. This analysis is complicated by the incomplete nature of some of these draft genomes; the absence of a gene from a single genome could reflect either real absence of the gene, or a sequencing or assembly error. To address this 'incomplete genome' problem, in this study we based any claims that a gene/pathway has been lost on phylogenetic groups of genomes, arguing that the loss of a single sequence from a number of independently sequenced genomes is more likely to indicate the evolutionary loss of that gene rather than sequencing or assembly errors. Presence or absence of TAC genes is indicated in Figure 1. Genomes used in this analysis are listed (with abbreviations) in supplementary table 1 with peptide databases used in supplementary table 2.

The receptor Torso

Torso, the receptor in the Torso-activation cassette (Duncan et al. 2014), is a receptor tyrosine kinase (RTK), similar to the large numbers of RTKs present in animal genomes. Determining the phylogenetic relationships of RTKs is challenging as there are many of these receptors encoded in insect genomes. In *Drosophila*, over 20 RTKs have been identified from 10 receptor families (Schartl et al. 2015; Sopko and Perrimon 2013). We have used Hidden Markov-Model (HMM) techniques (Eddy 2001) to identify the receptors (Supplementary data file 3) in arthropod genomes and then phylogenetic analysis to classify Torso receptors in arthropod genomes to determine presence or absence (Figure 2 and supplementary data file 3 and supplementary data file 4). The three RTKs most closely related to Torso, Ret, Heartless and Breathless from *Drosophila melanogaster* and *Apis mellifera* were used as an outgroup in constructing the phylogeny (Figure 2 and supplementary figure 5).

To understand the evolutionary history of these receptors outside the arthropods, we used BLAST searches with Torso proteins from selected species (*Nasonia vitripennis, Drosophila melanogaster, Tribolium casteneum, Bombyx mori* and *Lottia gigantea)* to identify Torso receptors encoded in the genomes of 9 lophotrochozoans (*Capitella teleta, Hellobdella robusta, Octopus bimaculoides, Crassostrea gigas, Pinctada fucata martensii, Notospermus geniculatus, Phoronis australis, Lingula anatina*

Pomacea canaliculata), and five ecdysozoans (*Priapulus caudatus*, *Ramazzottius varieornatus*, *Hypsibius dujardini*, *Caenorhabditis elegans* and *Strongyloides ratti*). This less accurate but faster method allowed us to rapidly determine the origins of the Torso receptor. *Torso* is found in seven of the ten Lophotrochozoan genomes examined, and two of the five Ecdysozoa genomes examined, indicating the origin of *torso* is older than the divergence of lophotrochozoans and Ecdysozoans (supplementary figure 6).

In arthropods, a clear *torso* ortholog was found in the genome of just one of nine chelicerates examined, *Ixodes scapularis*. The myriapod *Strigamia martima* also has an identifiable *torso* ortholog. The *Ixodes* and *Strigamia torso* genes are not included in Figure 2, but shown in supplementary figure 5. In the five crustacean species we examined, no *torso* orthologs could be found.

Torso receptors are encoded in the sequenced genomes of a dipluran, a mayfly and a dragonfly (*Catajapyx aquilonaris*, *Ephemera danica* and *Ladona fulva),* a termite (*Zootermopsis nevadensis*), a cockroach (*Blattella germanica),* and a thrip (*Frankliniella occidentalis)*.

Torso was found in the genomes of all hemipterans except *Diaphorina citri*, *Pachypsylia venusta* and *Oncopeltus fasciatus*. *Torso* genes were not found in the human body louse (*Pediculus humanus*). It is not clear if the apparent absence from the *Pediculus humanus* genome is due to genome assembly, but there is independent evidence from *Oncopeltus fasciatus* where repeated attempts to clone *torso* by degenerate PCR (Weisbrod et al. 2013) supports the absence of *torso* in this species. That two psyllid species have no *torso* genes in their genomes additionally supports this as real loss rather than genome assembly issues.

Torso is found in Hymenoptera, the most basally branching group of the Holometabola (Hedges et al. 2015; Whiting 2002), but is not ubiquitous in these species. Nine of the 24 hymenopteran species we have investigated have no *torso* ortholog, with many of these losses following phylogenetic patterns. This pattern is consistent with *torso* being lost at least once in the Apidae bees (~95 million years ago,(Cardinal and Danforth 2013)). All

coleopteran, dipteran and lepidopteran genomes investigated uniformly encode Torso receptors.

The ligands, Trunk, PTTH and Noggin-like

Genes encoding *trunk* and *PTTH* are difficult to identify in genome sequences as they are short and poorly conserved. Using HMM methods (Eddy 2001) (supplementary data file 7), we have scanned predicted protein sequences, and then identified them as trunk or PTTH using Bayesian techniques (Ronquist and Huelsenbeck 2003) (Figure 3, supplementary data file 8 and supplementary figure 9).

Noggin-like molecules have been identified in non-arthropod animals previously (Duncan et al. 2013; Molina et al. 2011). To determine the origin of Trunk, we used BLAST searches with Trunk protein from selected species (*Nasonia vitripennis, Drosophila melanogaster, Tribolium casteneum, Bombyx mori* and *Lottia gigantea)* to find traces of *trunk* genes in 9 Lophotrochozoa and 5 Ecdysozoa genomes (species as above). Using this fast method, we identified *trunk* genes in five out of ten Lophotrochozoa genomes (including *Lottia*, and two out of five Ecdysozoa genomes. This approach likely produces false negatives, implying that *trunk* may be present in many Lophotrochozoa and Ecdysozoa genomes and evolved before their divergence.

In Arthropods, using our more stringent approaches, a *trunk* sequence was identified in the genomes of five out of nine chelicerates, including *Ixodes scapularis*, in which *torso* was also identified; and *trunk* and *torso* have previously been identified in the myriapod *Strigamia maritima* (Chipman et al. 2014). Neither *trunk* nor *PTTH* was found in any crustacean genomes.

PTTH and *trunk* are not found in the sequenced genomes of Ephemeroptera, Odonata, Blattodea or Isoptera, indicating at least two independent losses in these lineages (Misof et al. 2014). Given experimental evidence for PTTH activity in the Blattodea (Hiragaki et al. 2009; Richter 1992; Richter et al. 1999), it seems likely that another molecule in these species has PTTH-like activity.

No *trunk* genes were identified in the Hemiptera, but seven out of the nine genomes investigated have *PTTH* orthologs. This is consistent with PTTH activity detected in *Rhodnius prolixus* (Garcia et al. 1990; Wigglesworth 1934b), and supported by evidence that knock-downs of *PTTH* in *Nilaparvata lugens* show reduced ecdysteriodogenesis (Chen et al. 2017). *Trunk* and *PTTH* are common components of holometabolous insect genomes. In Diptera, *PTTH* was found in every species investigated, while *trunk* was found in all but tsetse fly (*Glossina*) and fruit fly (*Dacinae*) species.

No lepidopteran genomes investigated had *trunk* sequences, but all had *PTTH,* implying *trunk* was lost early in the lepidopteran lineage. In the closely related Trichoptera, *trunk* was present, but not *PTTH*.

In the Coleoptera, orthologs for *trunk* and *PTTH* were found in *Tribolium castaneum* and *Anoplophora glabripennis*, while *Agrilus planipennis* and *Leptinotarsa decemlineata* had only *trunk* orthologs. *Onthophagus taurus* has only an identifiable *PTTH* ortholog. Neither *trunk* nor *PTTH* were identified in the genome of *Dendroctonus ponderosae.* The presence of *torso* in the genome of *Dendroctonus ponderosae* implies that the apparent absence of *trunk* and *PTTH* are likely due to incomplete genome sequencing or annotation. The conservation patterns of *PTTH* and *trunk* in the Coleoptera imply multiple loss events (two apparent losses of *trunk* and three apparent losses of *PTTH*) if absence is not due to incomplete genomes or annotations.

Of the 24 species of Hymenoptera included in the analysis, a *trunk* ortholog was identified only in the sawfly species (*Athalia rosa and Neodiprion lecontei*), indicating that *trunk* was lost early in the hymenopteran lineage prior to the diversification of wasps, ants and bees (~200 mya (Misof et al. 2014)). Twelve of the 24 hymenopteran species investigated do not have an identifiable *PTTH* ortholog, indicating that *PTTH* was lost in the lineages that gave rise to the Apidae (~50 mya). Species in which both *torso* and *PTTH* were not identifiable include *Trichogramma pretiosum*, *Lasioglossum albipes*, *Apis florea, Apis dorsata*, *Apis mellifera*, *Melipona quadrifasciata*, *Bombus impatiens* and *Bombus terrestris*. All but *Trichogramma pretiosum,* are bees.

Noggin-like genes (Molina et al. 2011) appear to be the ancestors of *trunk* and *PTTH* (Duncan et al. 2013), linking these molecules and Noggin, an extracellular regulator of BMP signalling found in many animal genomes but thought to have been lost in insects. *Noggin-like* genes are found in the genomes of hemipterans, basal branching insects, 'Crustacea', myriapods, and chelicerates, as well as non-arthropod Ecdysozoa (Duncan et al. 2013) and Lophotrochozoa (Molina et al. 2011). *Noggin-like* was found in the genomes of every hemipteran species investigated except for *Diaphorina citri*. *Noggin-like* genes are absent from holometabolous insect genomes, implying that these genes were lost from the lineage leading to holometabolous insects ~350 mya.

Torso-like

Torso-like is found in all insect and crustacean genomes surveyed, implying an origin in the common ancestor of the pancrustacea. While it is possible that *torso-like* is older than this (indeed, we have only examined a single myriapod genome), it is not present in multiple chelicerate genomes, implying it did not evolve in the common ancestor of arthopods (Figure 4, supplementary file 10, supplementary figure 11). *Torso-like* has been duplicated several times within lineages of the Pancrustacea including in the Hemiptera, where there are apparent lineage-specific duplications of the gene in the pea aphid (*Acyrthosiphon pisum*) (Bickel et al. 2013; Duncan et al. 2013; Shigenobu et al. 2010), the kissing bug (*Rhodnius prolixus*) (Mesquita et al. 2015) and the water strider (*Gerris buenoi*)(Armisen et al. 2018). In this study, we have also identified a duplication of *torso-like* within the Lepidoptera, occurring prior to the diversification of moths and butterflies 250 mya, with two copies of this gene maintained in the genomes of all sequenced lepidopteran species.

Evolution of the TAC

Ancestral state reconstruction (Figure 5 and 6, supplementary figures 12-15), our phylogenetic analysis (Figures 2-4), and BLAST searches in lophotrochozoan and ecdysozoan genomes, indicate that the receptor tyrosine kinase Torso, and the ligand Trunk, have arisen in the ancestor of Ecdysozoa and Lophotrochozoa.

Torso-like and *PTTH* appear to be innovations at the base of the pancrustacean and insect lineages, respectively. *Torso-like* is present in the genomes of all the Pancrustacea, and is

duplicated in some insects. *PTTH*/*trunk* are less evolutionarily labile in the Holometabola, perhaps related to a key role in metamorphosis, with the only losses of both *PTTH* and *trunk* occurring in the Apidae, and the loss of *trunk* in the Lepidoptera.

Discussion

Origins of the TAC.

The presence/absence data we present here are complex but imply that while many TAC components have deep evolutionary histories in the arthropods. *Trunk* and *torso* appear in a range of Ecdysozoan and Lophotrochozoan genomes, supporting the notion that these are ancient molecules that are often lost in genomes, rather than an example of parallel evolution. The loss of components appears to be the major theme in TAC evolution (Figure 7). For example, *Noggin-like* is an ancient molecule present in many lophotrochozoan and ecdysozoan genomes, but is completely absent from the Holometabola. While it seems that *trunk* evolved from *noggin-like* in the common ancestor of Ecdysozoa and Lophotrochozoa, both have then been lost in 'Crustacea' and some basal insects. *PTTH* appears to have evolved from *trunk*, in the common ancestor of Hemiptera and Holometabola, but is lost in some holometabolous groups.

Torso-like presents perhaps the simplest evolutionary history in this group, with an evolutionary origin in the pancrustacean ancestor. While *torso-like* is well conserved in the Pancrustacea, there is no clear ortholog of this gene in the rest of the Ecdysozoa. We have detected MACPF domains in some arthropod genomes (data not shown), especially spiders, but these are not *torso-like* orthologs.

Patterns of absence in insect TAC

While the broad pattern of presence or absence of components of the Torso-activation cassette is as described, apparent loss of some or all of these components is common. Some of these apparent losses can be attributed to incomplete genomes, but good evidence for genuine loss can be found. In most cases (excepting the chelicerates and basal insects) the loss of both ligands is predictive of the loss of the receptor. This implies that loss of either the receptor, or the ligands renders the other non-functional, and thus dispensable. This further implies that these are the only ligands that bind to and activate this receptor and

that over the course of arthropod evolution, at least, this receptor, and these ligands, have developed no other activities.

Perhaps the most obvious group where the TAC has been lost is within the Hymenoptera, where *trunk* is missing from all species but the basally branching sawflies (*Athalia* and *Neodiprion*), and *PTTH* is missing from seven bee genomes, while *torso* is missing from eight. This pattern is consistent with one loss of *trunk* and *torso* within the Hymenoptera, and three independent losses of *PTTH* in the bees. Given the high quality of the honeybee genome (Elsik et al. 2014; The Honey Bee Genome Sequencing Consortium 2006) it seems likely that the loss of *trunk*, *torso* and *PTTH* in bees is not an artefact of genome sequencing, despite reported PTTH immunoreactivity in these species (Závodská et al. 2003).

Trunk is also absent from the Lepidoptera, though present in the closely related Trichoptera. All species of Lepidoptera have *PTTH* and *torso*, but no *trunk*, implying loss of *trunk* early in the evolution of this order. No studies have yet been undertaken to determine the role of Torso signalling in terminal patterning of lepidopteran embryos.

Trunk appears to be lost from the genomes of some species of Diptera, especially in the Dacinae, its absence from this whole group supporting the view that this is evolutionary loss rather than an artefact.

Terminal patterning and moulting

TAC components regulate two developmental processes, terminal patterning and control of moulting. While these processes may be broadly conserved it is clear that the genes underlying them in *Drosophila* are not. In canonical terminal patterning, Torso-like interacts with Trunk and Torso leading to activation of MAPK in the terminal regions of the embryo (Duncan et al. 2014). In many species, *trunk* and/or *torso* are not present in the genome. In some of these species, studies have indicated that terminal patterning occurs in other ways. In *Drosophila*, Torso-like has other functions, including a role in dorso-ventral patterning (Mineo et al. 2017) and gastrulation (Johnson et al. 2017) The non-terminal patterning phenotypes produced in Torso-like null mutants are reminiscent of those caused by *Oncopeltus* RNAi (Weisbrod et al. 2013) and over-expression of aphid *torso-like* in

Drosophila (Duncan et al. 2013), implying these functions, rather than terminal patterning, may be more conserved.

In moulting control in *Drosophila*, PTTH is released by prothoracic neurons in the brain, leading to the ecdysteroid release from the prothoracic gland required for moulting (Rewitz et al. 2009). PTTH has been well-studied in this role in the Lepidoptera (Gu et al. 2011; Meola and Adkisson 1977; Nagasawa et al. 1984) and Hemiptera (Garcia et al. 1990; Wigglesworth 1934a; Wigglesworth 1934b) (Chen et al. 2017), and PTTH immunoreactivity (Hiragaki et al. 2009; Simões et al. 1997; Závodská et al. 2003) and activity (Richter 1992; Richter et al. 1999) has been assayed in deeply branching insect groups, implying deep conservation of PTTH and its activity. Despite this, the *PTTH* gene is not conserved in all insects, being missing from whole groups of moulting insects (such as bees) and indeed evolving after the evolution of many of the deep branching insects that show PTTH immunoreactivity and activity.

The failure to identify *PTTH* genes in groups, such as honeybees and dragonflies, that have been reported to have immunoreactivity for PTTH (Simões et al. 1997; Závodská et al. 2003) is puzzling, but probably due to difficulties with cross-reacting antibodies. One of these studies, for example, used an antibody raised against a short peptide based on a lepidopteran PTTH sequence to survey expression in multiple insects (Závodská et al. 2003). This peptide is not encoded in most *PTTH* genes we have identified (data not shown), and tBLASTn (Altschul et al. 1990) searches of the honeybee genome, perhaps the best, non-drosophilid, insect genome, indicate no homologies. While this antibody stains patches of cells in the honeybee brain, there is little data indicating that it indeed cross-reacts with a PTTH protein in honeybee or other insects.

Rather more problematic is the assignment of PTTH activity (Richter 1992; Richter et al. 1999) to insects such as *Periplaneta americanum*, a member of the Blattodea, in which no sequenced species has a *PTTH* gene. In these experiments PTTH is not directly assayed, but inferred through the stimulation of ecdysteriod release from isolated prothoracic glands through co-culturing with brain tissue. While a PTTH effect is measurable, this does not

require a classic PTTH molecule. It seems likely that some other molecule, perhaps another cysteine knot protein, carries the PTTH activity in deeply branching insects.

One potential candidate for this PTTH activity in the apparent absence of *PTTH* genes are Noggin-like proteins. As Trunk and PTTH evolved from noggin-like ancestors, is it possible that these Noggin-like proteins took over roles previously played by either Trunk or PTTH in groups that had lost these genes? While this possibility exists, the sequence of *noggin-like* genes is similar enough to ancestral *noggin* genes to suggest that they may instead act in BMP signalling (Molina et al. 2011), rather than the MAPK processes regulated by *trunk* and *PTTH*. Furthermore, Noggin-like proteins are not present in the genomes of bees implying they cannot replace PTTH in these groups. Certainly, our data implies that the control of moulting in bees, and more basally branching insects, may differ from that of most insects. Understanding how moulting is controlled in these species will give us a fuller understanding of the evolution of this vital process.

Understanding the function of Trunk in hemipteran species and Trunk-like in chelicerates will give us a better understanding of the evolution of terminal patterning. For example, does Trunk ancestrally play a role in terminal patterning? Or does Trunk have other roles, which have been co-opted for terminal patterning in some holometabolous insects?

Conclusions

Terminal patterning and moulting control are critical processes in insect development. In *Drosophila*, these are regulated by genes in the TAC, comprising *torso*, *trunk*, *torso-like* and *PTTH*. This set of genes has, however, a complex evolutionary history, with multiple losses of ancestral genes, evolution of new gene variants and the potential for convergent evolution. Understanding the functions of this signalling process in a variety of arthropods is required if we are to understand the evolution and assembly of this modified signalling process, as well as the biological processes it regulates.

Experimental procedures

Identification of trunk/PTTH genes.

Peptide databases (supplementary table 1) were searched with hmmsearch from the HMMER package (version 3.1b2; (Eddy 2001)), using a hidden Markōv model constructed by (Duncan et al. 2013) from *Drosophila melanogaster trunk* and *Bombyx mori PTTH* with an E-value cut-off of 0.1. Identified sequences were trimmed using the *trunk*/*PTTH* HMM model using hmmtrim from the HMMER suite of programs (version 3.1b2; (Eddy 2001)).

Trimmed sequences were aligned using ClustalX (version 2.1 ; (Thompson et al. 1994)). Aligned sequences were used to construct initial neighbor-joining trees using Quicktree (version 1.1; (Howe et al. 2002)) with a bootstrap of 10, 000, 000. When multiple sequences from a single species were identified in the BLAST analysis and these clustered together in the neighbor-joining tree (bootstrap confidence values >90%), the sequences were aligned with ClustalX (Thompson et al. 1994) to evaluate whether they were splice variants. Only the longest splice variants were retained for subsequent analysis.

The genome sequences (supplementary table 1) of included species in which either *PTTH* or *trunk* were not identified, were BLAST searched using tBLASTn (Altschul et al. 1990), using either partial or complete peptide sequences from the most closely related species, to identify *trunk* or *PTTH* orthologs which were not annotated as peptide sequences.

The NCBI arthropod peptide database was subsequently BLAST (Altschul et al. 1990) searched using peptide sequences from nucleotide sequences that had been identified in the initial BLAST search in order to establish whether there were any *trunk* or *PTTH* orthologs in other peptide databases for the same species, and to confirm the identity of hits.

Augustus (Stanke and Waack 2003) was used to create *de novo* models from scaffolds identified in the BLAST search as containing a potential *PTTH*/*trunk* gene. Augustus was used to create *de novo* gene models for missing or partial models based on the gene prediction model of the most closely related species available (version 2.5.5; (Stanke and Waack 2003)). Parameters in Augustus were set to both strands (forward and backward strands searched), partial gene model (allowed for the prediction of incomplete genes at the sequence boundaries) and single stranded predictions (predicted genes independently on each strand, allows for overlapping genes on opposite strands). Models were visualised

using Artemis (Rutherford et al. 2000). Peptide models from tBLASTn and the Augustus model were compared using NCBI BLASTp on non-redundant arthropod peptide database. The model with the highest similarity, measured in the E-value of the most closely related identified peptide was considered the most complete and was retained for use by downstream analysis (Altschul et al. 1990).

All identified orthologs were aligned using ClustalX (version 2.1; (Thompson et al. 1994)). trimAl was used to remove the "gappiest" regions of the alignments using the gt 50 parameter (version 1.2rev59; (Capella-Gutiérrez et al. 2009)).

Identification of torso-like genes

The peptide databases of species were BLASTp searched with the *Drosophila melanogaster* Torso-like protein (FBpp0083563) and identified sequences aligned using ClustalX (version 2.1; (Thompson et al. 1994)). The alignment was used to construct a neighbor joining tree using Quicktree (version 1.1; (Howe et al. 2002)) with a bootstrap of 10000000.

Species in which multiple sequences were identified in the BLAST analysis, which also clustered together in the neighbor joining tree with a high bootstrap value (>90%), were aligned to evaluate whether they are splice variants. Splice variants were defined as sequences that were identical in overlapping regions. Only the longest variant was retained whilst probable splice variants were removed. Curated sequences were aligned using ClustalX (version 2.1; (Thompson et al. 1994)).

Identification of torso genes.

The HMM model for *torso* was constructed using *torso* orthologs from *Drosophila* (NP_476762.1), *N. vitripennis* (XP_016840977.1), *T. castaneum* (NP_001034536.1) and *B. mori* (NP_001164049.1). The peptide sequences were aligned using ClustalX (version 2.1; (Thompson et al. 1994)) and trimmed with trimAl using the strict parameter (Capella-Gutiérrez et al. 2009). HMMER (Eddy 2001) hmmbuild was used to construct a hidden Markōv model. The peptide databases of a number of species were searched using the previously constructed Torso HMM model (supplementary file 3) using hmmsearch (Eddy

2001) with an E-value parameter of 0.00001. hmmalign from the HMMER package was used to trim the resulting sequences using the *torso* HMM model. Sequences shorter than 200 amino acids were removed using minimumseq (Cock et al. 2009). Clustal Omega (version 1.2.1; (Sievers and Higgins 2014)) was used to align the resulting sequences. trimAl (Capella-Gutiérrez et al. 2009) was used to remove the gappiest sequence regions from the alignment using the strict parameter.

RapidNJ (Simonsen et al. 2010) was used to construct neighbor joining trees. The *torso* sub-tree was identified through well-supported (by bootstrap) association with *torso* orthologs from *A. pisum* (ACYPI005080), *Drosophila* (NP_476762.1), *T. castaneum* (EFA07525.1) and *B. mori* (XP_012549743.1). Hmmsearch was used on individual species' genomes in which an identifiable *torso* ortholog in the primary search using the torso HMM was not found, with an E-value parameter of 0.00001. These results were combined with previously identified *Torso* orthologs. Identified proteins were trimmed using hmmtrim from HMMER (version3.1b2; (Eddy 2001)), aligned using Clustal Omega (version 1.2.1; (Sievers and Higgins 2014)), and neighbor joining trees constructed for each species (Simonsen et al. 2010) in order to identify sequences that clustered with previously identified *torso* orthologs with high support.

The nucleotide sequence databases of species that did not have an identifiable *torso* ortholog were BLAST searched with a minimum of 4 *torso* orthologs from the closest available species using BlastX (Altschul et al. 1990) with an E-value cutoff of 0.00001. The corresponding peptide sequence of the best BLAST hits were BLAST searched using NCBI non-redundant protein database limited to arthropods. Blast results were used to assess the plausibility of the BLAST hit being a *torso* ortholog. If a *torso* ortholog from the same species from an alternative database was identified, this was used in further analysis. If no model was available for the species of interest, then Augustus (version 3.3 (Stanke and Waack 2003)) was used to predict potential *torso* orthologs. The species model used to predict *torso* was the closest available species (i.e. *Drosophila melanogaster, Nasonia vitripennis, Acyrthosiphon pisum* and *Tribolium castaneum*). Other parameters used in the running of

Augustus (version 3.3 (Stanke and Waack 2003)) include: both strands, partial gene model and single stranded predictions. Predicted models were BLASTed against the NCBI non-redundant peptide database, to assess similarity to Torso proteins.

Estimation of final trees.

Final trees were estimated using Bayesian methods using MrBAYES ((Ronquist and Huelsenbeck 2003); version 3.2.6). The Monte Carlo Markov Chain search was run using four chains over 1,000,000 generations with trees sampled every 1000 generations.

The first 25% of trees were discarded as 'burn-in'. To estimate the phylogeny for *trunk*/*PTTH* sequences, the WAG model (Whelan and Goldman 2001) of amino-acid replacement was used after experiments with mixed models, indicating it was the most appropriate. For *torso-like* sequences, MrBAYES was run using 10 chains, overs 5,000,000 generations, sampled every 1000 generations with a 'burn-in' of 25%. The analysis was carried out using the JTT model (Jones et al. 1992) and resulted in a phylogeny in which the standard deviation of the posterior probability was less than 0.01.

Constructing cladograms

The higher-level phylogenetic relationships between included arthropod species in the cladogram is based on the phylogenetic data from (Misof et al. 2014). The phylogenetic relationships between the included species were based on the best current phylogenetic data available for the Chelicerates (Bond et al. 2014; Fernández and Giribet 2015; Jeyaprakash and Hoy 2009), 'Crustacea'(Aleshin et al. 2009; Braga et al. 1999; Koenemann et al. 2010; Mathers et al. 2013; Oakley et al. 2012), Hemiptera (Cryan and Urban 2012), Hymenoptera (Branstetter et al. 2017; Cardinal et al. 2010), Coleoptera (Hunt et al. 2007; Mckenna et al. 2015), Lepidoptera (Kawahara and Breinholt 2014; Mitter et al. 2017) and Diptera (Chu et al. 2016; Dyer et al. 2008; Gibson et al. 2010; Giribet and Edgecombe 2012; Kutty et al. 2010; Sallum et al. 2002). The cladogram was constructed in Mesquite (version 3.2; (Maddison and Maddison 2001)). Mesquite was also used to make mirror trees and to reconstruct ancestral states through parsimony.

Acknowledgements: This work was funded by a New Zealand Ministry of Business, Innovation and Employment Smart Ideas grant to P.K.D. The authors would like to thank P.M. Dearden for critical reading of this manuscript. The authors are grateful for the efforts of the i5K consortium and the Baylor College of Medicine Human Genome Sequencing Center, which provided genome sequences and annotation. The authors would also like to thank two anonymous reviewers for their extremely helpful comments on a previous draft of this manuscript. The authors declare no conflicts of interest.

Author contributions: J.S. carried out the bioinformatics analyses, C.P. helped draft and edit the manuscript, EJD co-supervised this work and edited the manuscript, PKD co-supervised the work and drafted the manuscript.

References

Aleshin, V, Mikhailov, K, Konstantinova, A, Nikitin, M, Rusin, LY, Buinova, D, Kedrova, O and Petrov, N (2009) On the phylogenetic position of insects in the Pancrustacea clade. *Molecular biology* **43**: 804-818.

Altschul, SF, Gish, W, Miller, W, Myers, EW and Lipman, DJ (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403-10.

Amarnath, S, Stevens, LM and Stein, DS (2017) Reconstitution of Torso signaling in cultured cells suggests a role for both Trunk and Torso-like in receptor activation. *Development* **144**: 677-686.

Armisen, D, Rajakumar, R, Friedrich, M, Benoit, JB, Robertson, HM, Panfilio, KA, Ahn, S-J, Poelchau, MF, Chao, H and Dinh, H (2018) The genome of the water strider Gerris buenoi reveals expansions of gene repertoires associated with adaptations to life on the water. *bioRxiv*: 242230.

Bickel, RD, Cleveland, HC, Barkas, J, Jeschke, CC, Raz, AA, Stern, DL and Davis, GK (2013) The pea aphid uses a version of the terminal system during oviparous, but not viviparous, development. *EvoDevo* **4**: 10.

Bond, JE, Garrison, NL, Hamilton, CA, Godwin, RL, Hedin, M and Agnarsson, I (2014) Phylogenomics resolves a spider backbone phylogeny and rejects a prevailing paradigm for orb web evolution. *Current Biology* **24**: 1765-1771.

Braga, E, Zardoya, R, Meyer, A and Yen, J (1999) Mitochondrial and nuclear rRNA based copepod phylogeny with emphasis on the Euchaetidae (Calanoida). *Marine Biology* **133**: 79-90.

Branstetter, MG, Danforth, BN, Pitts, JP, Faircloth, BC, Ward, PS, Buffington, ML, Gates, MW, Kula, RR and Brady, SG (2017) Phylogenomic insights into the evolution of stinging wasps and the origins of ants and bees. *Current Biology* **27**: 1019-1025.

Capella-Gutiérrez, S, Silla-Martínez, JM and Gabaldón, T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972-1973.

Cardinal, S and Danforth, BN (2013) Bees diversified in the age of eudicots. In: *Book Bees diversified in the age of eudicots* (Editor, ed.^eds.). Vol. 280, pp. 20122686. The Royal Society, City.

Cardinal, S, Straka, J and Danforth, BN (2010) Comprehensive phylogeny of apid bees reveals the evolutionary origins and antiquity of cleptoparasitism. *Proceedings of the national Academy of Sciences* **107**: 16207-16211.

Casali, A and Casanova, J (2001) The spatial control of Torso RTK activation: a C-terminal fragment of the Trunk protein acts as a signal for Torso receptor in the Drosophila embryo. *Development* **128**: 1709-1715.

Casanova, J, Furriols, M, McCormick, CA and Struhl, G (1995) Similarities between trunk and spätzle, putative extracellular ligands specifying body pattern in Drosophila. *Genes & Development* **9**: 2539-2544.

Casanova, J and Struhl, G (1989) Localized surface activity of torso, a receptor tyrosine kinase, specifies terminal body pattern in Drosophila. *Genes & development* **3**: 2025-2038.

Chen, X, Li, G-Q, Wan, P-J and Fu, Q (2017) Efficient RNA interference for three neuronally-expressed genes in Nilaparvata lugens (Stål)(Hemiptera: Delphacidae). *Journal of Asia-Pacific Entomology* **20**: 513-519.

Chipman, AD, Ferrier, DE, Brena, C, Qu, J, Hughes, DS, Schröder, R, Torres-Oliva, M, Znassi, N, Jiang, H and Almeida, FC (2014) The first myriapod genome sequence reveals conservative arthropod gene content and genome organisation in the centipede Strigamia maritima. *PLoS biology* **12**: e1002005.

Chu, H, Li, C, Guo, X, Zhang, H, Luo, P, Wu, Z, Wang, G and Zhao, T (2016) The phylogenetic relationships of known mosquito (Diptera: Culicidae) mitogenomes. *Mitochondrial DNA Part A*: 1-5.

Chung, KW, Oh, HS, Kim, AJ, Won, YJ, Kim, NK, Namkoong, Y, Kim, K and Lee, C (1994) Isolation of small prothoracicotropic hormone-like gene in Drosophila melanogaster. *Korean J. Zool* **37**: 12-18.

Cock, PJ, Antao, T, Chang, JT, Chapman, BA, Cox, CJ, Dalke, A, Friedberg, I, Hamelryck, T, Kauff, F and Wilczynski, B (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**: 1422-1423.

Cryan, JR and Urban, JM (2012) Higher-level phylogeny of the insect order Hemiptera: is Auchenorrhyncha really paraphyletic? *Systematic Entomology* **37**: 7-21.

Dearden, PK, Wilson, MJ, Sablan, L, Osborne, PW, Havler, M, McNaughton, E, Kimura, K, Milshina, NV, Hasselmann, M, Gemp, T, Schioett, M, Brown, SJ, Elsik, CG, Holland, PW, Kadowaki, T and Beye, M (2006) Patterns of conservation and change in honey bee developmental genes. *Genome Research* **16**: 1376-1384.

Duncan, EJ, Benton, MA and Dearden, PK (2013) Canonical terminal patterning is an evolutionary novelty. *Dev Biol*.

Duncan, EJ, Johnson, TK, Whisstock, JC, Warr, C, G. and Dearden, PK (2014) Capturing embryonic development from metamorphosis: How did the terminal patterning signalling pathway of Drosophila evolve? *Current Opinions in Insect science* **1**: 45-51.

Dyer, N, Lawton, S, Ravel, S, Choi, K, Lehane, M, Robinson, A, Okedi, L, Hall, M, Solano, P and Donnelly, M (2008) Molecular phylogenetics of tsetse flies (Diptera: Glossinidae) based on mitochondrial (COI, 16S, ND2) and nuclear ribosomal DNA sequences, with an emphasis on the palpalis group. *Molecular phylogenetics and evolution* **49**: 227-239.

Eddy, SR (2001) HMMER: Profile hidden Markov models for biological sequence analysis.

Elsik, CG, Worley, KC, Bennett, AK, Beye, M, Camara, F, Childers, CP, de Graaf, DC, Debyser, G, Deng, J and Devreese, B (2014) Finding the missing honey bee genes: lessons learned from a genome upgrade. *BMC genomics* **15**: 86.

Fain, MJ and Riddiford, LM (1976) Reassessment of the critical periods for prothoracicotropic hormone and juvenile hormone secretion in the larval molt of the tobacco hornworm Manduca sexta. *General and comparative endocrinology* **30**: 131-141.

Fernández, R and Giribet, G (2015) Unnoticed in the tropics: phylogenomic resolution of the poorly known arachnid order Ricinulei (Arachnida). *Royal Society open science* **2**: 150065.

Forbes-Beadle, L, Crossman, T, Johnson, TK, Burke, R, Warr, CG and Whisstock, JC (2016) Development of the Cellular Immune System of Drosophila Requires the Membrane Attack Complex/Perforin-Like Protein Torso-Like. *Genetics* **204**: 675-681.

Garcia, E, Luz, N, Azambuja, P and Rembold, H (1990) Azadirachtin depresses the release of prothoracicotropic hormone in Rhodnius prolixus larvae: evidence from head transplantations. *Journal of Insect Physiology* **36**: 679-682.

Gibson, JF, Skevington, JH and Kelso, S (2010) Placement of Conopidae (Diptera) within Schizophora based on mtDNA and nrDNA gene regions. *Molecular Phylogenetics and Evolution* **56**: 91-103.

Giribet, G and Edgecombe, GD (2012) Reevaluating the arthropod tree of life. *Annual review of entomology* **57**: 167-186.

Grillo, M, Furriols, M, De Miguel, C, Franch-Marro, X and Casanova, J (2012) Conserved and divergent elements in Torso RTK activation in Drosophila development. *Scientific reports* **2**: 762.

Gu, S-H, Young, S-C, Lin, J-L and Lin, P-L (2011) Involvement of PI3K/Akt signaling in PTTH-stimulated ecdysteroidogenesis by prothoracic glands of the silkworm, Bombyx mori. *Insect biochemistry and molecular biology* **41**: 197-202.

Hedges, SB, Marin, J, Suleski, M, Paymer, M and Kumar, S (2015) Tree of life reveals clock-like speciation and diversification. *Mol Biol Evol* **32**: 835-45.

Hiragaki, S, Uno, T and Takeda, M (2009) Putative regulatory mechanism of prothoracicotropic hormone (PTTH) secretion in the American cockroach, Periplaneta americana as inferred from co-localization of Rab8, PTTH, and protein kinase C in neurosecretory cells. *Cell and tissue research* **335**: 607-615.

Howe, K, Bateman, A and Durbin, R (2002) QuickTree: building huge Neighbour-Joining trees of protein sequences. *Bioinformatics* **18**: 1546-1547.

Hunt, T, Bergsten, J, Levkanicova, Z, Papadopoulou, A, John, OS, Wild, R, Hammond, PM, Ahrens, D, Balke, M and Caterino, MS (2007) A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. *Science* **318**: 1913-1916.

i5K Consortium (2013) The i5K Initiative: Advancing Arthropod Genomics for Knowledge, Human Health, Agriculture, and the Environment. *Journal of Heredity* **104**: 595.

Jeyaprakash, A and Hoy, MA (2009) First divergence time estimate of spiders, scorpions, mites and ticks (subphylum: Chelicerata) inferred from mitochondrial phylogeny. *Experimental and Applied Acarology* **47**: 1-18.

Johnson, TK, Crossman, T, Foote, KA, Henstridge, MA, Saligari, MJ, Beadle, LF, Herr, A, Whisstock, JC and Warr, CG (2013) Torso-like functions independently of Torso to regulate Drosophila growth and developmental timing. *Proceedings of the National Academy of Sciences* **110**: 14688-14692.

Johnson, TK, Henstridge, MA, Herr, A, Moore, KA, Whisstock, JC and Warr, CG (2015) Torso-like mediates extracellular accumulation of Furin-cleaved Trunk to pattern the Drosophila embryo termini. *Nature communications* **6**.

Johnson, TK, Moore, KA, Whisstock, JC and Warr, CG (2017) Maternal Torso-Like Coordinates Tissue Folding During Drosophila Gastrulation. *Genetics*: genetics. 117.200576.

Jones, DT, Taylor, WR and Thornton, JM (1992) The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* **8**: 275-82.

Kawahara, AY and Breinholt, JW (2014) Phylogenomics provides strong evidence for relationships of butterflies and moths. In: *Book Phylogenomics provides strong evidence for relationships of butterflies and moths* (Editor, ed.^eds.). Vol. 281, pp. 20140970. The Royal Society, City.

Klingler, M, Erdélyi, M, Szabad, J and Nüsslein-Volhard, C (1988) Function of torso in determining the terminal anlagen of the Drosophila embryo. *Nature* **335**: 275-277.

Koenemann, S, Jenner, RA, Hoenemann, M, Stemme, T and von Reumont, BM (2010) Arthropod phylogeny revisited, with a focus on crustacean relationships. *Arthropod Structure & Development* **39**: 88-110.

Kopec, S (1922) Studies on the necessity of the brain for the inception of insect metamorphosis. *The Biological Bulletin* **42**: 323-342.

Kutty, SN, Pape, T, Wiegmann, BM and Meier, R (2010) Molecular phylogeny of the Calyptratae (Diptera: Cyclorrhapha) with an emphasis on the superfamily Oestroidea and the position of Mystacinobiidae and McAlpine's fly. *Systematic Entomology* **35**: 614-635.

Li, WX (2005) Functions and mechanisms of receptor tyrosine kinase Torso signaling: lessons from Drosophila embryonic terminal development. *Dev Dyn* **232**: 656-72.

Lynch, JA, Olesnicky, EC and Desplan, C (2006) Regulation and function of tailless in the long germ wasp Nasonia vitripennis. *Dev Genes Evol* **216**: 493-8.

Maddison, WP and Maddison, DR (2001) Mesquite: a modular system for evolutionary analysis.

Mathers, TC, Hammond, RL, Jenner, RA, Hänfling, B and Gomez, A (2013) Multiple global radiations in tadpole shrimps challenge the concept of 'living fossils'. *PeerJ* **1**: e62.

McBrayer, Z, Ono, H, Shimell, M, Parvy, J-P, Beckstead, RB, Warren, JT, Thummel, CS, Dauphin-Villemant, C, Gilbert, LI and O'Connor, MB (2007) Prothoracicotropic hormone regulates developmental timing and body size in Drosophila. *Developmental cell* **13**: 857-871.

Mckenna, DD, Wild, AL, Kanda, K, Bellamy, CL, Beutel, RG, Caterino, MS, Farnum, CW, Hawks, DC, Ivie, MA and Jameson, ML (2015) The beetle tree of life reveals that Coleoptera survived end-Permian mass extinction to diversify during the Cretaceous terrestrial revolution. *Systematic Entomology* **40**: 835-880.

Meola, RW and Adkisson, PL (1977) Release of prothoracicotropic hormone and potentiation of developmental ability during diapause in the bollworm, Heliothis zea. *Journal of Insect Physiology* **23**: 683-688.

Mesquita, RD, Vionette-Amaral, RJ, Lowenberger, C, Rivera-Pomar, R, Monteiro, FA, Minx, P, Spieth, J, Carvalho, AB, Panzera, F and Lawson, D (2015) Genome of Rhodnius prolixus, an insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite infection. *Proceedings of the National Academy of Sciences* **112**: 14936-14941.

Mineo, A, Furriols, M and Casanova, J (2015) Accumulation of the Drosophila Torso-like protein at the BLASToderm plasma membrane suggests that it translocates from the eggshell. *Development* **142**: 1299-1304.

Mineo, A, Furriols, M and Casanova, J (2017) Transfer of dorsoventral and terminal information from the ovary to the embryo by a common group of eggshell proteins in Drosophila. *Genetics* **205**: 1529-1536.

Misof, B, Liu, S, Meusemann, K, Peters, RS, Donath, A, Mayer, C, Frandsen, PB, Ware, J, Flouri, T and Beutel, RG (2014) Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**: 763-767.

Mitter, C, Davis, DR and Cummings, MP (2017) Phylogeny and Evolution of Lepidoptera. *Annual review of entomology* **62**: 265-283.

Molina, MD, Neto, A, Maeso, I, Gómez-Skarmeta, JL, Saló, E and Cebrià, F (2011) Noggin and noggin-like genes control dorsoventral axis regeneration in planarians. *Current Biology* **21**: 300-305.

Nagasawa, H, Kataoka, H, Hori, Y, Isogai, A, Tamura, S, Suzuki, A, Guo, F, Zhong, X, Mizoguchi, A and Fujishita, M (1984) Isolation and some characterization of the prothoracicotropic hormone from Bombyx mori. *General and comparative endocrinology* **53**: 143-152.

Oakley, TH, Wolfe, JM, Lindgren, AR and Zaharoff, AK (2012) Phylotranscriptomics to bring the understudied into the fold: monophyletic ostracoda, fossil placement, and pancrustacean phylogeny. *Molecular biology and evolution* **30**: 215-233.

Rewitz, KF, Yamanaka, N, Gilbert, LI and O'Connor, MB (2009) The insect neuropeptide PTTH activates receptor tyrosine kinase torso to initiate metamorphosis. *Science* **326**: 1403-5.

Richter, K (1992) Prothoracicotropic activity in the brain of the cockroach Periplaneta americana. *Journal of insect physiology* **38**: 349-355.

Richter, K, Peschke, E and Peschke, D (1999) Effect of melatonin on the release of prothoracicotropic hormone from the brain of Periplaneta americana (Blattodea: Blattidae). *European Journal of Entomology* **96**: 341-346.

Ronquist, F and Huelsenbeck, JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572-4.

Rutherford, K, Parkhill, J, Crook, J, Horsnell, T, Rice, P, Rajandream, M-A and Barrell, B (2000) Artemis: sequence visualization and annotation. *Bioinformatics* **16**: 944-945.

Sallum, M, Schultz, T, Foster, P, Aronstein, K, Wirtz, R and Wilkerson, R (2002) Phylogeny of Anophelinae (Diptera: Culicidae) based on nuclear ribosomal and mitochondrial DNA sequences. *Systematic Entomology* **27**: 361-382.

Sauman, I and Reppert, S (1996a) Molecular characterization of prothoracicotropic hormone (PTTH) from the giant silkmoth Antheraea pernyi: developmental appearance of PTTH-expressing cells and relationship to circadian clock cells in central brain. *Developmental biology* **178**: 418.

Sauman, I and Reppert, SM (1996b) Molecular characterization of prothoracicotropic hormone (PTTH) from the giant silkmothAntheraea pernyi: developmental appearance of PTTH-expressing cells and relationship to circadian clock cells in central brain. *Developmental biology* **178**: 418-429.

Savant-Bhonsale, S and Montell, DJ (1993a) torso-like encodes the localized determinant of Drosophila terminal pattern formation. *Genes Dev* **7**: 2548-55.

Savant-Bhonsale, S and Montell, DJ (1993b) torso-like encodes the localized determinant of Drosophila terminal pattern formation. *Genes & development* **7**: 2548-2555.

Schartl, M, Volff, J-N and Brunet, F (2015) Evolution of receptor tyrosine kinases. In: *Receptor Tyrosine Kinases: Structure, Functions and Role in Human Disease*. pp. 17-36. Springer.

Schoppmeier, M and Schröder, R (2005) Maternal torso signaling controls body axis elongation in a short germ insect. *Current Biology* **15**: 2131-2136.

Schüpbach, T and Wieschaus, E (1986) Maternal-effect mutations altering the anterior-posterior pattern of the Drosophila embryo. *Development genes and evolution* **195**: 302-317.

Shigenobu, S, Bickel, R, Brisson, J, Butts, T, Chang, C, Christiaens, O, Davis, G, Duncan, E, Ferrier, D and Iga, M (2010) Comprehensive survey of developmental genes in the pea aphid, Acyrthosiphon pisum: frequent lineage-specific duplications and losses of developmental genes. *Insect Molecular Biology* **19**: 47-62.

Sievers, F and Higgins, DG (2014) Clustal Omega, accurate alignment of very large numbers of sequences. *Multiple sequence alignment methods*: 105-116.

Simões, ZP, Boleli, I and Hartfelder, K (1997) Occurrence of a prothoracicotropic hormone-like peptide in the developing nervous system of the honey bee (Apis mellifera L). *Apidologie* **28**: 399-409.

Simonsen, M, Mailund, T and Pedersen, CN (2010) Inference of large phylogenies using neighbour-joining. In: *Book Inference of large phylogenies using neighbour-joining* (Editor, ed.^eds.). Vol., pp. 334-344. Springer, City.

Sopko, R and Perrimon, N (2013) Receptor tyrosine kinases in Drosophila development. *Cold Spring Harb Perspect Biol* **5**.

Sprenger, F, Stevens, LM and Nüsslein-Volhard, C (1989) The Drosophila gene torso encodes a putative receptor tyrosine kinase. *Nature* **338**: 478-483.

Stanke, M and Waack, S (2003) Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19**: ii215-ii225.

The Honey Bee Genome Sequencing Consortium (2006) Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* **443**: 931-949.

Thompson, JD, Higgins, DG and Gibson, TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**: 4673-4680.

Treisman, R (1996) Regulation of transcription by MAP kinase cascades. *Current opinion in cell biology* **8**: 205-215.
Weisbrod, A, Cohen, M and Chipman, AD (2013) Evolution of the insect terminal patterning system—Insights from the milkweed bug, Oncopeltus fasciatus. *Developmental biology* **380**: 125-131.
Werren, JH, Richards, S, Desjardins, CA, Niehuis, O, Gadau, J, Colbourne, JK and Group, NGW (2010) Functional and evolutionary insights from the genomes of three parasitoid Nasonia species. *Science* **327**: 343-348.
Whelan, S and Goldman, N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* **18**: 691-9.
Whiting, MF (2002) Phylogeny of the holometabolous insect orders: molecular evidence. *Zoologica scripta* **31**: 3-15.
Wigglesworth, VB (1934a) Factors controlling moulting and metamorphosis in an insect. *Nature* **133**: 725-726.
Wigglesworth, VB (1934b) Memoirs: The physiology of ecdysis in Rhodnius prolixus (Hemiptera). II. Factors controlling moulting and 'metamorphosis'. *Journal of Cell Science* **2**: 191-222.
Wilson, MJ and Dearden, PK (2009) Tailless patterning functions are conserved in the honeybee even in the absence of Torso signaling. *Dev Biol* **335**: 276-87.
Závodská, R, Šauman, I and Sehnal, F (2003) Distribution of PER protein, pigment-dispersing hormone, prothoracicotropic hormone, and eclosion hormone in the cephalic nervous system of insects. *Journal of Biological Rhythms* **18**: 106-122.

Figure legends

Figure 1) Presence and absence of TAC components in arthropod genomes. Using stringent approaches to be as sure as possible about presence or absence we have marked each gene as present, absent or duplicated in each species. Where 'no-data' is listed we have not found a clear orthologue of the gene being assessed, but the integrity of that genome is such that we are not prepared to rule out the presence of the gene.

Figure 2) Bayesian phylogeny of identified Torso orthologs from arthropod species and a lophotrochozoan. Trees were estimated using MrBAYES run using ten chains over 50,000,000 generations using the WAG model. Branches that included multiple species from the same genus were collapsed in order to simplify the tree. Species abbreviations are shown in Figure 1.

Figure 3) Bayesian phylogeny of Trunk, PTTH and Noggin-like orthologs from arthropod species. Trees were estimated using MrBAYES run using ten chains over 5,000,000,000

generations using the WAG model. Branches that included multiple species from the same Genus were collapsed in order to simplify the tree. Species abbreviations are shown in Figure 1.

Figure 4) Bayesian phylogeny of Torso-like orthologs from arthropod species. Trees were estimated using MrBAYES run using four chains over 5,000,000 generations using the JTT model. Branches that included multiple species from the same genus were collapsed in order to simplify the tree. Species abbreviations are shown in Figure 1.

Figure 5) Cladogram of ancestral states of Torso vs Trunk reconstructed through parsimony for arthropod species. The cladogram emphasises the relationship between loss of the torso receptor and loss of its ligands, especially in the apidae clade (marked by red dotted lines). Reconstruction was performed using Mesquite.

Figure 6) Cladogram of ancestral states of Torso vs PTTH reconstructed through parsimony for arthropod species. The cladogram emphasises the relationship between loss of the torso receptor and loss of its ligands, especially in the apidae clade (marked by red dotted lines). Reconstruction was performed using Mesquite.

Figure 7) Cladogram of the proposed evolution of the TAC in protostomes. Evolutionary innovations are marked on the cladogram in blue on the left, variation within each clade is shown in red on the right.

Supplementary figures

Supplementary table **1)** Presence and absence of TAC components from genomes examined using stringent techniques

Supplementary table **2)** Peptide databases used in this analysis.

Supplementary data file **3)** Torso HMM model constructed using torso orthologs from *Drosophila* (NP_476762.1), *N. vitripennis* (XP_016840977.1), *T. castaneum* (NP_001034536.1) and *B. mori* (NP_001164049.1).

Supplementary data file **4)** Alignment of identified torso orthologs, trimmed using trimmal, using ClustalX.

Supplementary figure **5)** Bayesian phylogeny of identified torso orthologs. Trees were estimated using MrBAYES run using ten chains over 50,000,000 generations using the WAG model. Species abbreviations are shown in Figure 1.

Supplementary table **6)** Detection of TAC components in selected lophotrochozoan and ecdysozoan genomes

Supplementary data file **7)** Trunk and PTTH Hidden Markov Models.

Supplementary data file **8)** Alignment of the identified trunk, PTTH and noggin-like ortholog, which were trimmed using trimmal, using ClustalX.

Supplementary figure **9)** Bayesian phylogeny of trunk, PTTH and noggin-like orthologs. Trees were estimated using MrBAYES run using ten chains over 5,000,000,000 generations using the WAG model. Species abbreviations are shown in Figure 1.

Supplementary data file **10)** Alignment of the identified torso-like orthologs using ClustalX.

Supplementary figure **11)** Bayesian phylogeny of Torso-like. Trees were estimated using MrBAYES run using four chains over 5,000,000 generations using the JTT model. Species abbreviations are shown in Figure 1.

Supplementary figure **12)** Cladogram ancestral states parsimony Torso vs Trunk reconstructed through parsimony. Reconstruction was performed using Mesquite.

Supplementary figure **13)** Cladogram ancestral states parsimony Torso vs PTTH reconstructed through parsimony. Reconstruction was performed using Mesquite.

Supplementary figure **14)** Cladogram ancestral states parsimony Torso vs Torso-like reconstructed through parsimony. Reconstruction was performed using Mesquite.

Supplementary figure **15)** Cladogram ancestral states parsimony Torso vs Noggin-like reconstructed through parsimony. Reconstruction was performed using Mesquite.