

RESEARCH ARTICLE

Open Access



Origin and evolution of the enhancer of split complex

Peter K. Dearden

Abstract

Background: The Enhancer of split complex is an unusual gene complex found in Arthropod genomes. Where known this complex of genes is often regulated by Notch cell signalling and is critically important for neurogenesis. The Enhancer of split complex is made up of two different classes of genes, basic helix-loop-helix-orange domain transcription factors and bearded class genes. The association of these genes has been detected in the genomes of insects and crustaceans.

Results: Tracing the evolution of the Enhancer of split complex in recently sequenced Arthropod genomes indicates that enhancer of split basic helix-loop-helix orange domain genes arose before the common ancestor of insects and Crustacea, and before the formation of the complex. Throughout insect and crustacean evolution, a four-gene cluster has been present with lineage specific gene losses and duplications. The complex can be found in the vast majority of genomes, but appears to be missing from the genomes of chalcid wasps, raising questions as to how they carry out neurogenesis in the absence of these crucial genes.

Conclusions: The enhancer of split complex arose in the common ancestor of Crustacea and insects, probably through the linkage of a basic helix-loop-helix orange domain gene and a bearded class gene. The complex has been maintained, with variations, throughout insect and crustacean evolution indicating some function of the complex, such as coordinate regulation, may maintain its structure through evolutionary time.

Background

Evolutionary conserved complexes of genes are rare in insect genomes [1] while relatively common in vertebrates [2–4]. The best characterised is the Hox complex, found in the genomes of widely diverse animals, in which inter-linked and coordinate gene expression appear to stabilise the genomic structure of the complex over evolutionary time [5]. Complexes of genes that remain intact over wide evolutionary distances in insect genomes presumably have similar features to the Hox complex, such as coordinate regulation, that maintain their genomic structure, while the genome is rearranged around them.

The Enhancer of Split Complex (E(spl)-C) is an unusual and conserved complex of genes first identified in *Drosophila melanogaster*. This complex differs from most in that the genes it contains encode two completely different sorts of proteins: bHLH-Orange domain transcription

factors [6, 7] (bHLHO), and Bearded class proteins [8, 9] (Brd). The association between these two types of genes has been found in the genomes of both insects and crustaceans [10], implying that this complex first formed through the association of these genes rather than the more usual gene duplication.

A limited survey of insect and crustacean genomes has shown that the E(spl) complex is ancestrally made up of three bHLHO encoding genes (*bHLH1*, *bHLH2* and *her*) and a single Brd-class gene [10]. The structure of the complex is modified in some insects, *Drosophila* being an example, where two bHLH-orange domain genes are absent from the complex and there are seven copies of the remaining one, the Brd-class gene, *ma*, has been duplicated and there are a range of unrelated genes inserted in the complex [6, 9–16].

The E(spl)-C was first identified as a modifier of the *Notch* mutant *Split* [17]. Subsequent studies have shown that the bHLH proteins encoded by E(spl)-C act as effectors of Notch cell signalling [13, 18]. During neurogenesis in *Drosophila*, presumptive neuroblasts signal to

Correspondence: peter.dearden@otago.ac.nz
Genetics Otago and Gravidia (National Centre for Growth and Development),
Biochemistry Department, University of Otago, Dunedin, Aotearoa, New
Zealand

surrounding cells in proneural clusters through the Notch cell-signalling pathway (Reviewed in [19–21]). This pathway leads to expression of E(spl)-C bHLH genes [18], which act to repress the expression [22], and function [23], of proneural genes, a set of transcription factors that promote neural cell fate [24, 25]. Presumptive neuroblasts thus signal surrounding cells to block their differentiation as neuroblasts through the activation of Notch cell signalling, and the expression of the E(spl) bHLHO domain proteins [26–29]. E(spl) bHLHO domain proteins encode a c-terminal WRPW motif which recruits the transcriptional repressor Groucho, which in turn acts to attenuate gene expression by promoting RNA polymerase pausing, and causes local histone deacetylation [30]. E(spl) bHLHO domain proteins thus target transcriptional repression to proneural genes, and proneural gene targets.

Brd class proteins from the E(spl)-C, in contrast, antagonise Notch signalling by interacting with Neuralised [31], an E3 ubiquitin ligase, stimulating the degradation of a Notch ligand, Delta [32]. In *Drosophila*, brd-class proteins particularly act to pattern adult sensory precursor formation [8, 9]. Bearded class genes encode small proteins with amphipathic alpha helices, and little sequence conservation [9].

The genes contained within the E(spl)-C are regulated in a range of ways. While individual enhancer elements for some of the genes have been identified [33, 34], the entire complex appears also to be activated by Su(H) [9, 35–40], a transcription factor usually regulated by Notch signalling [41]. This Notch responsiveness is also found in a crustacean, *Daphnia* [42]. Individual transcripts are repressed, in *Drosophila*, by miRNAs binding to 3'UTR located sites (named GY, Brd and K- boxes) [15, 43, 44]. There is also evidence for coordinate regulation of the E(Spl)-C by cohesin [45], a chromatin structure regulating protein that coats the E(spl)-C in cells and represses expression, and repression by Polycomb group proteins [46]. The genomics DNA containing the E(spl)-C also is structured in three dimensions in cells such that the chromatin of the complex interacts with itself forming an isolated domain, but does not interact with flanking regions [46]. This self-interactive structure implies the complex is regulated in a coordinated manner [46].

E(Spl)-C bHLH proteins are closely related to other clades of insect bHLH-orange domain proteins, including *clockwork orange* (*cwo*), *Similar to Deadpan*, (*Side*), *Hairy and E(spl) related with a YRPW domain* (*Hey*), *hairy* (*h*), *deadpan* (*dpn*). These genes have multiple roles in insects. *Hairy* is a regulator of segmentation [47, 48], acting as a pair-rule gene in many arthropods [49–51]. *Deadpan* [52] has multiple roles in *Drosophila* development, such as dosage compensation [53] and regulating neuroblast proliferation through responding to Notch signalling [54].

Hey has been shown to regulate neuron fate determination [55]. *Cwo* (clockwork orange) has a role in regulating the circadian clock [56–59]. *Side* has no identified function in *Drosophila*. In non-arthropod animals, closely related genes, the HES genes (hairy-enhancer of split genes) act, in most situations, as effectors of Notch signalling [60, 61] though some these genes do have non-Notch related roles [62, 63]. Given this close phylogenetic relationship, and the role in Notch signalling conserved between arthropods and vertebrates, it seems likely that Notch responsiveness, and a role as an effector of Notch signalling, may be an ancestral function for this group of transcription factors. Indeed expression of *E(spl)-mδ* produces a neurogenic effect when mis-expressed in *Xenopus* embryos [64].

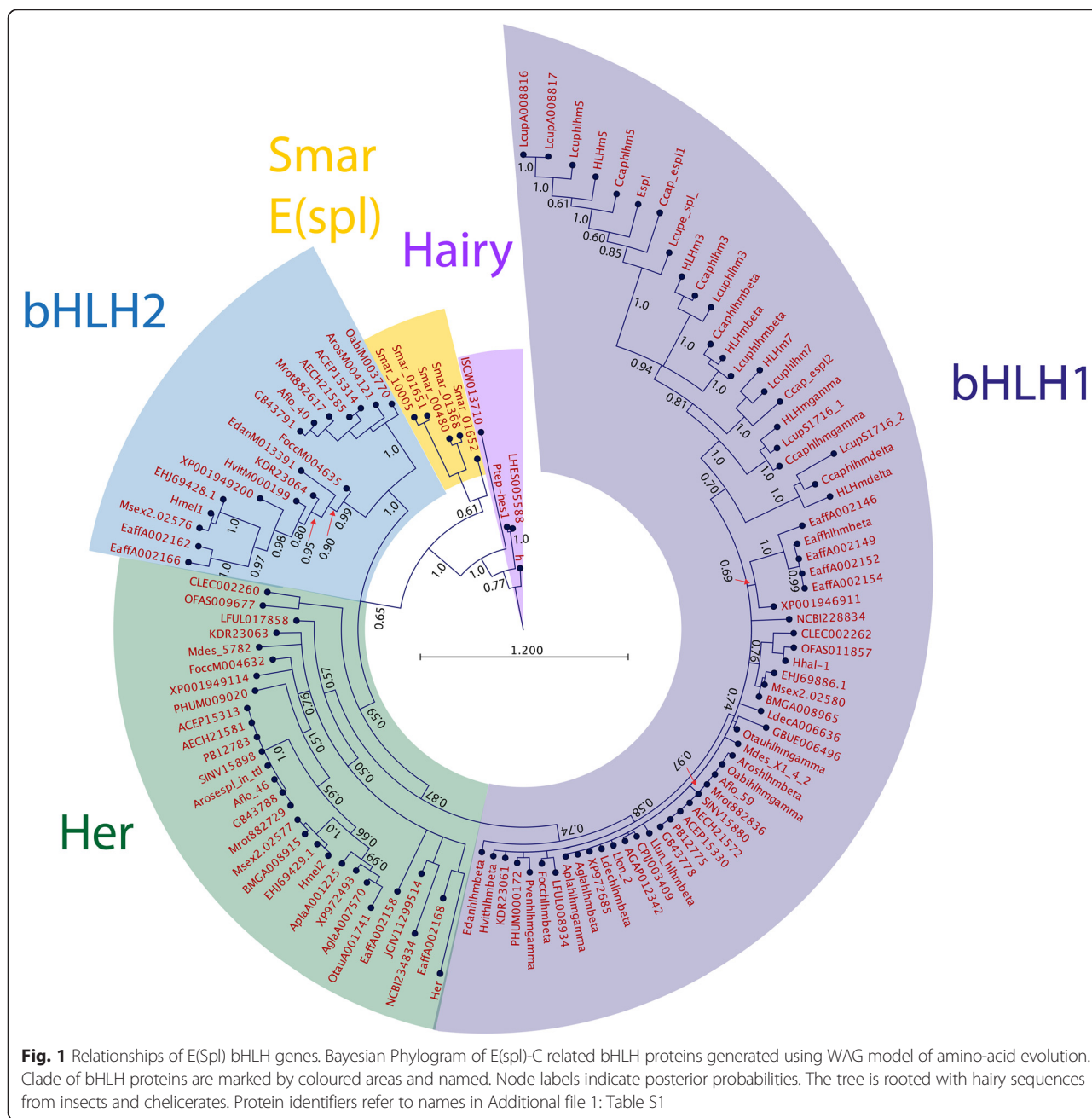
The unusual nature of the E(spl)-C, containing two types of genes, and its potential coordinate regulation, make understanding the dynamics of its evolution important. Here I examine the structure and relationships of the E(spl)-C from arthropod and onychophoran genomes, particularly those provided by the i5K consortium [65, 66]. The i5K consortium provides high quality Arthropod genome sequences that allow both the identification of genes, and examination of genome structure. This dataset allows us to trace both the origins and subsequent evolution of the E(spl)-C in arthropods.

Results and discussion

Phylogeny of all isolated E(spl) bHLH genes recovers three major clades

Searching for E(spl)-like bHLH sequences in arthropod genomes identifies large numbers of sequences with similarity to E(spl) bHLHs, and other clades of bHLHO domain sequences. In most species of insects and Crustacea, multiple bHLHO sequences can be identified with strong similarity to E(Spl)-C bHLHs, and these genes are often found adjacent to each other in contigs, linkage groups or genome scaffolds. Aligning the protein products of these genes and reconstructing their relationships using Bayesian phylogenetics [67] identifies the three major clades of E(spl) bHLHO proteins as first described in Duncan and Dearden [10] (*her*, bHLH1 and bHLH2) (Fig. 1). The analysis also indicates that the *Strigamia maritima* (centipede) genome contains a number E(Spl) bHLH proteins, but these do not fall into the three classes of E(spl) bHLH proteins found in crustacean and insects. Orthologues of non E(Spl) bHLHO proteins robustly fall outside the E(Spl) bHLH clades (Protein identifiers refer to names in Additional file 1: Table S1).

Branch-lengths within the E(spl) bHLH clades are short, indicating strong conservation of sequence, but these branch lengths increase between bHLH1 proteins from Diptera, implying more sequence divergence in that group. The assignment of E(spl) bHLH genes from

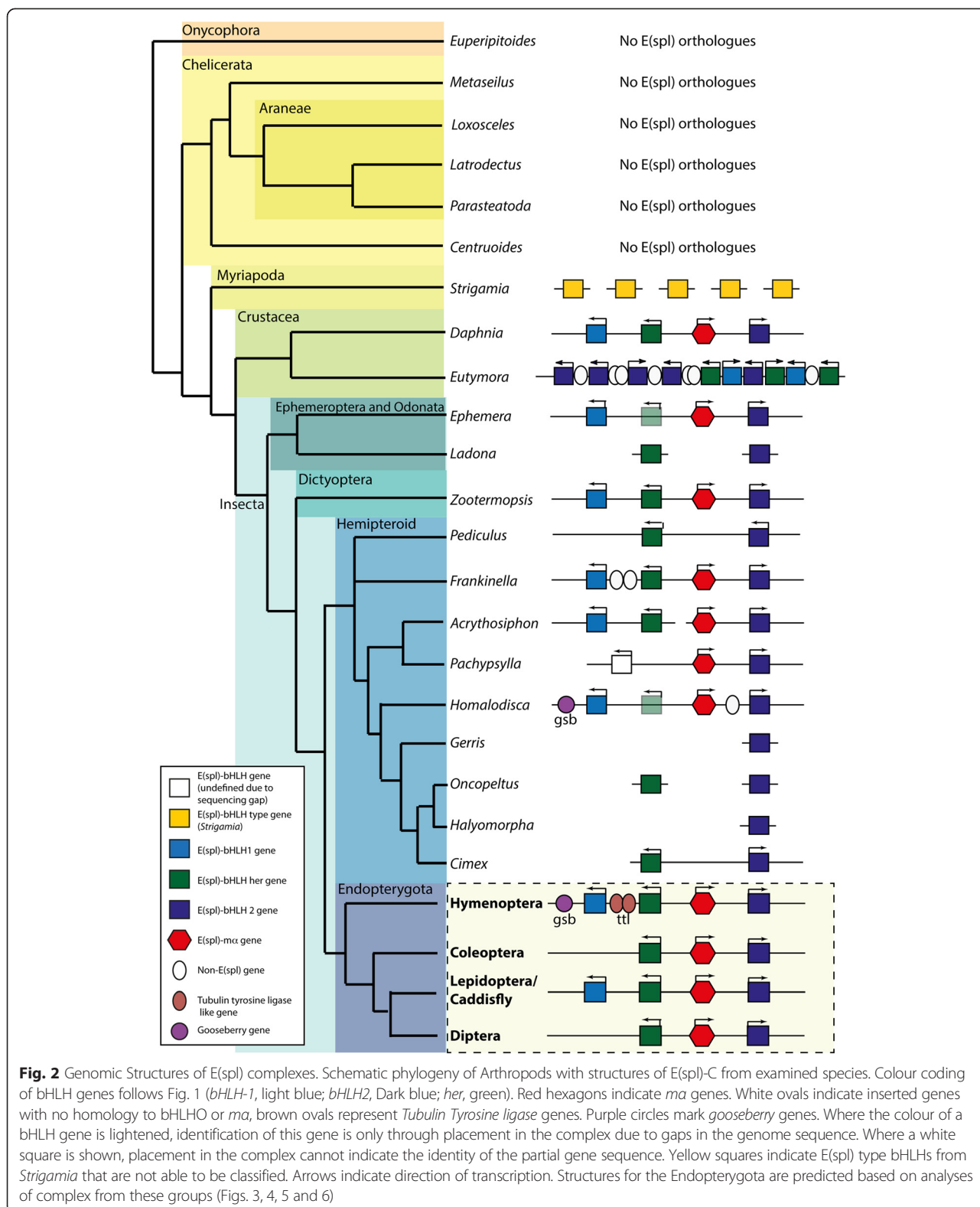


crustacean and insect genomes to these three clades of allows us to interpret the genomic structure of the E(spl)-C in crustacean and insect genomes.

Genomic analysis of E(spl) genes indicates clusters and conserved structure

Duncan and Dearden [10] identified a four gene E(spl) complex as the ancestral state for insects and Crustacea. This four-gene cluster (*bHLH-2*, *her*, *ma* and *bHLH-1*) was identified in the genome of the crustacean *Daphnia* and in sequenced insect genomes [10]. I have expanded this analysis, using the i5k consortium data, to identify

E(spl)-Cs within Arthropod and Onychophoran genomes and then used phylogenetic analyses (Fig. 1) to categorise those genes. This analysis indicates that the bHLH complex is a component of the vast majority of insect genomes, but there are clade and species-specific losses and expansions. In Chelicerates, and the partial genome of an onychophoran (Velvet worm, Non arthropod Ecdysozoan, basal to arthropods), we can find no evidence for bHLHO proteins that are orthologous (by reciprocal blast) to E(spl) bHLHO proteins (Fig. 2). A myriapod genome [68] (*Strigamia*) encodes 5 E(spl) proteins, but these are not co-located in the genome. In two



crustacean genomes (*Daphnia* and *Eurytemora*), the E(spl)-C is more apparent. In *Daphnia* (Water Flea), as described previously, a four-gene complex is present, with

an identifiable Brd-class gene, *ma*, embedded within it. *Ma* genes are often difficult to identify by Blast alone because their sequence evolves rapidly. In all species in

which *ma* genes could be identified by tblastn [69] of the genome, they were found in the E(spl)-C except in *Drosophila*, where a complex of Brd-class genes lies outside the E(spl)-C [31]. Given the difficulties in identifying these genes, it is possible that other members of the Brd-class are present outside the E(spl)-C in other species.

In *Eurytemora*, a copepod crustacean, the E(spl)-C is modified with duplications of all three E(spl) bHLH genes (two *bHLH-1*, three *her* and 5 *bHLH2*). Alongside this, no *ma* gene can be identified.

In insects, a four-gene complex can also be identified in deeply branching clades (Fig. 2). In a mayfly (*Ephemera*) and a termite (*Zootermopsis*), a four-gene complex, identical to *Daphnia*, is present.

In the large hemipteroid clade of insects, examples of a four-gene complex are relatively common (*Frankliniella* (Western flower thrips), *Acrythosiphon* (Pea Aphid, though this is split between two contigs) [10], and *Homalodisca* (Glassy-winged sharpshooter)). Despite this conservation, gene-loss has occurred in some Hemipteroid lineage, with species such as *Cimex* (bedbug) and *Oncopeltus* (milkweed bug) having only two bHLH genes (*her* and *bHLH2*) and species such as *Gerris* (Waterstrider) and *Halyomorpha* (Brown marmorated stink bug) having a single E(spl) gene (*bHLH2*). The patterns of change in the complex imply that such losses are lineage or species specific, and that

the selective pressure to maintain the full E(spl)-C is somewhat reduced in this assemblage. In species without the full complex of genes, *ma* is invariably missing, though its fast sequence evolution means that it could be located elsewhere in the genome.

Within Endopterygote lineages, the E(spl)-C appears more stable. In the Hymenoptera (Fig. 3), the classic four-gene complex is present in most species, though in *Orussus* (Parasitic woodwasp) one gene (*her*) is missing, and in Chalcidid wasps (*Trichogramma*, *Copidosoma* and *Nasonia*) no complex is detectable (see later). With these exceptions the E(spl)-C (including flanking and unrelated genes inserted into the complex (see later)) is completely conserved in gene complement, order and orientation.

In Coleoptera, by contrast, the complex is stable, but only contains three genes (*her*, *ma* and *bHLH2*) (Fig. 4). An orthologue of *bHLH1* is not present in any of the coleopteran genome examined. In *Leptinotarsa* (Colorado potato beetle), *her* and *ma* are missing, and *bHLH2* has been duplicated. The fact that *bHLH1* is absent from all Coleopteran species examined indicates that it was likely lost early in Coleopteran evolution.

The basal four-gene complex can be found in the genomes of the Lepidopterans (Fig. 5) *Manduca* (Tobacco Hornworm), and *Bombyx* [10] (Silk moth), and possibly in *Danaeus* (Monarch Butterfly) though in this genome

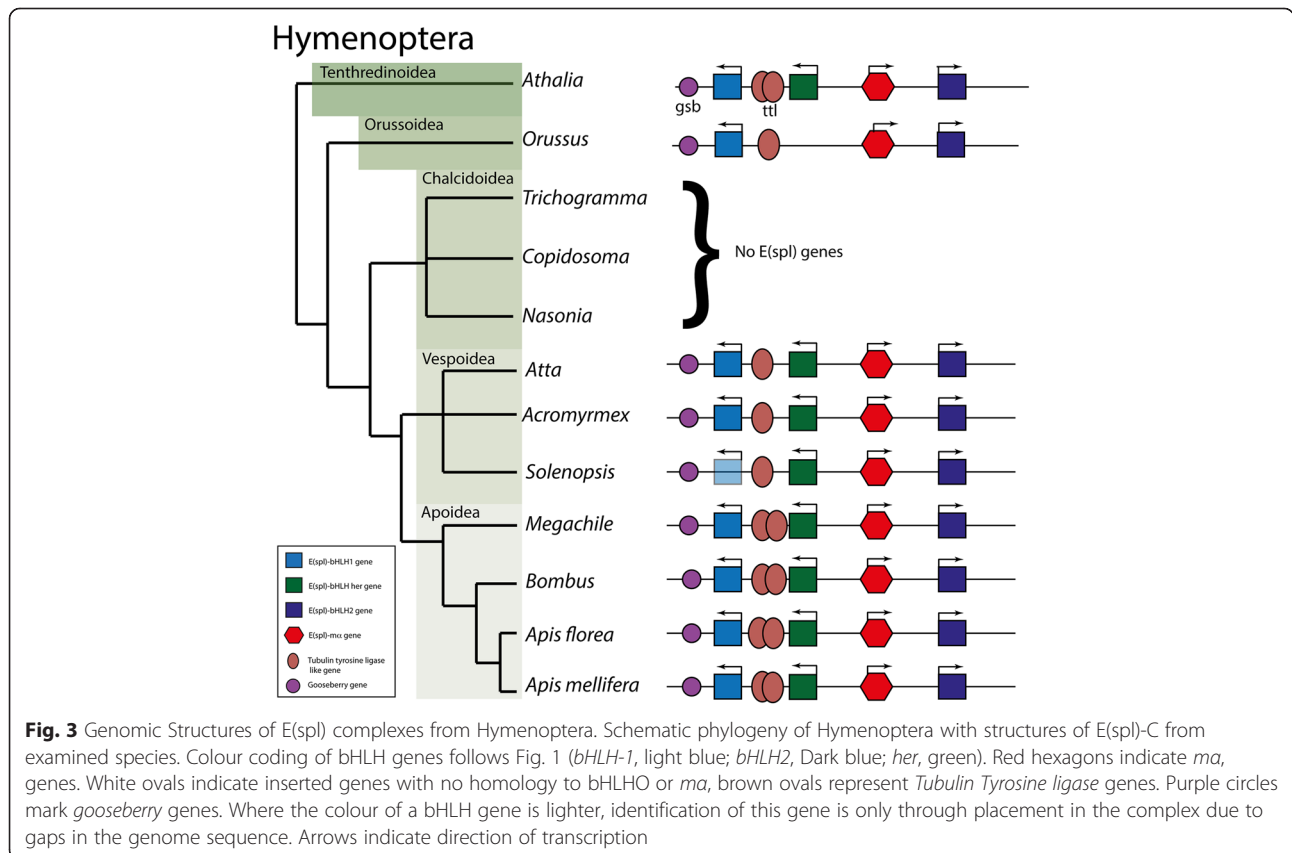
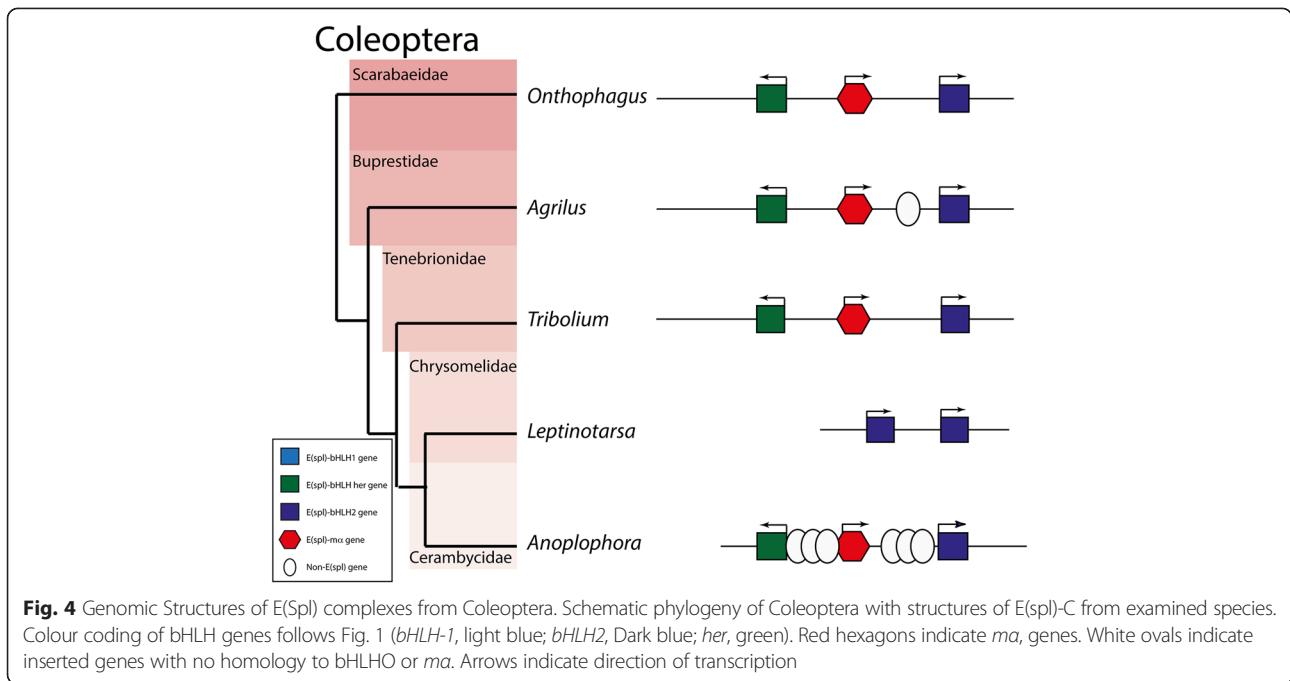


Fig. 3 Genomic Structures of E(spl) complexes from Hymenoptera. Schematic phylogeny of Hymenoptera with structures of E(spl)-C from examined species. Colour coding of bHLH genes follows Fig. 1 (*bHLH-1*, light blue; *bHLH2*, Dark blue; *her*, green). Red hexagons indicate *ma* genes. White ovals indicate inserted genes with no homology to bHLHO or *ma*, brown ovals represent *Tubulin Tyrosine ligase* genes. Purple circles mark *gooseberry* genes. Where the colour of a bHLH gene is lighter, identification of this gene is only through placement in the complex due to gaps in the genome sequence. Arrows indicate direction of transcription



the complex is split across two contigs. The complex is reduced to two genes (*bHLH1* and *her*) in *Heliconius* (Postman Butterfly). This is the only insect species of the 42 studied that does not have a *bHLH2* gene, raising the possibility that this may be a genome sequencing error. In a species of Caddisfly (*Limnephilus*, sister group to the Lepidoptera) only *bHLH2* can be found. It is unclear if this is species specific or lost in the entire lineage.

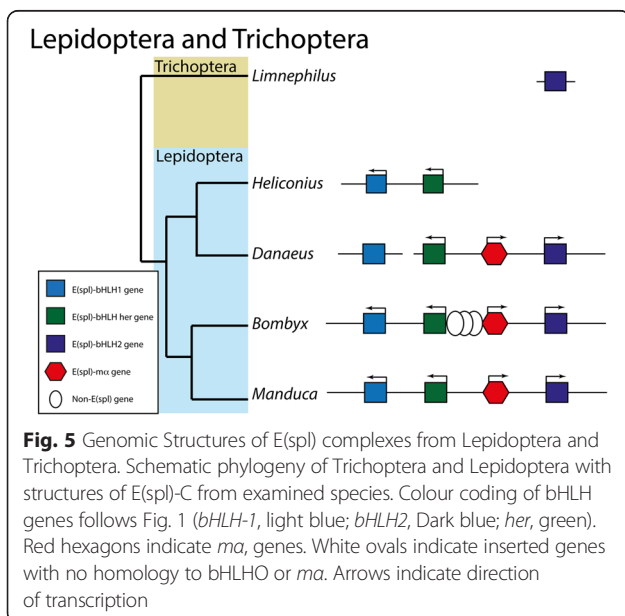
In Diptera (Fig. 6), *bHLH1* is absent from all genomes examined. In the Cuculidae (*Anopheles*, *Aedes* and *Culex*),

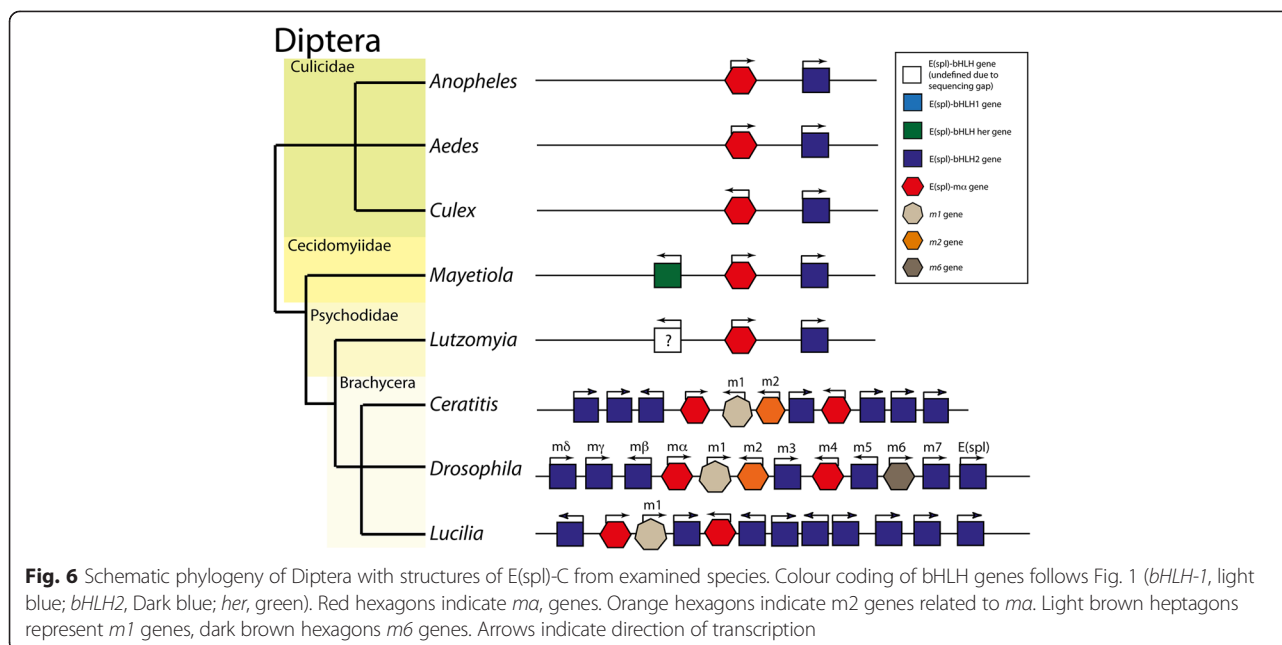
her is also absent from the genome and *ma* and *bHLH2* make up the complex. In the Brachycera, the complex is expanded, with multiple copies of *bHLH2*; 7 in *Ceratitis* (Medfly), 7 in *Drosophila* species and 9 in *Lucilia* (Common Green Bottle fly). *Ma* is also expanded, with two copies in each genome. In *Drosophila*, *her* is present in the genome, though not linked to the E(spl)-C. In *Lucilia* and *Ceratitis*, no *her* ortholog is present.

Her is present in the E(spl)-C of *Mayetiola* (Hessian fly), a more deeply branching fly, and there is an unidentifiable gene (due to a gap in genome sequence), in the same position in the fragmentary genome of *Lutzomyia* (Blackfly). These patterns imply that at least a three gene complex (*her*, *ma* and *bHLH2*) was present in the common ancestor of Diptera, but that gene loss (Cuculidae) and expansion of the complex (Brachycera) have extensively modified the E(spl)-C in this group.

In most of the complexes I have identified, the orientation of genes with respect to the direction of transcription of *ma* is conserved. *bHLH-1* and *her* are transcribed on the opposite strand to *ma*, and *bHLH2* on the same strand. The only variations to this pattern are in Diptera (Fig. 6) where the multiple copies of *bHLH2* in Brachycera are transcribed from either strand, and in *Culex* (Mosquito), where *ma* and *bHLH2* are transcribed from opposite strands.

These data imply that the origins of the E(spl)-C lie in the pan-crustacean clade. I can find no evidence for E(spl)-C bHLH genes in chelicerates (4 genomes), the most basally branching clade of arthropods, nor the onychophora, the closest non-arthropod ecdysozoan group. While this analysis is not conclusive as to the presence of





these genes in onychophorans (due to the partial nature of the genome sequence), their absence in all chelicerates examined is best explained by absence of these genes in this lineage, but could plausibly be due to gene loss.

The patterns of conservation, gene loss and expansion indicate that the E(spl)-C has a history of conservation of an ancestral four-gene structure, with gene-loss in some lineages, usually not affecting *bHLH2*, and expansion via gene duplication in Brachyceran flies, and the copepod *Eurytemora*. These expansions are difficult to explain, as their patterns of expansion are different. In *Eurytemora*, all bHLH genes are expanded, while *ma* is missing. In Brachycera, *bHLH1* and *her* are missing from the complex, with an expansion of *bHLH2* and *mα*. Given there are many species with two, or only one class of E(Spl) gene, these expansions are not best explained as a way to replace missing members of the complex, but may be related to complexity of gene regulation, or pattern formation, required from the complex.

Insertions into the E(spl)-C

In *Drosophila melanogaster* (Fig. 6), 2 non bHLH/*ma* genes are found in the E(spl)-C. These genes, *m1* (encoding a Kazal-type protease inhibitor) and *m6* (encoding a protein with a Myelin proteolipid protein PLP), do not produce Notch signalling-like defects when mutant, though *m6*, like the other E(spl)-C genes, is regulated by Notch signalling [70]. The insertion of *m1* and *m6* in the complex are conserved in other Drosophilids [10, 71] but only *m1* is conserved in the Brachyceran flies *Ceratitis* and *Lucilia*. Flanking the *Drosophila* complex is another Notch related

gene, *groucho*, which encodes a protein that interacts with E(spl)-bHLH genes to suppress gene expression [30, 72]. This gene does not flank the E(spl)-C outside *Drosophila* species.

Unrelated genes are inserted into the E(spl)-C in many insect species but these insertions are most often not conserved between species. The exception to this are a set of tubulin tyrosine ligase genes inserted between *bHLH1* and *her* in Hymenopteran genomes (Fig. 3). One or two of these genes are present in this location in all Hymenopteran examined that have an E(spl)-C. The maintenance of this insertion over 250 million years of evolutionary time, and the expression pattern of one of these genes in Honeybees [10] implies that these tyrosine-tubulin ligase genes may be regulated by Notch signalling.

The stability of the E(spl)-C in hymenopteran genomes extends to flanking genes. All Hymenopteran complexes are also flanked by a gene named *gooseberry* (Fig. 3). *Gooseberry* is a paired-box containing transcription factor that has been shown to have roles in patterning the nervous system and cuticle in a number of insect species [73–79] including hymenoptera [80, 81]. This gene is also found flanking the E(spl)-C in *Homoladisca* (Fig. 2), implying the association of *gooseberry* and E(spl)-C may date from the common ancestor of Endopterygota and the hemipteroid assemblage.

Chalcid Wasps have lost the E(spl) complex

While the E(spl)-C in the hymenoptera is highly conserved and stereotyped, I can find no evidence for E(spl)-C bHLH genes in the genomes of three chalcid wasps. These species have a full complement of other

bHLHO domain genes (Fig. 7), but no E(spl)-C orthologues are present in these genomes, either distributed, or in a complex. That this pattern of loss is present in three genomes, one of which is the well-sequenced (both genomes and transcriptomes) *Nasonia* genome, implies that loss of the complex in this group is not a sequencing error, but that Chalcid wasps have lost their E(spl)-C. These are the only group yet found in insects that do not have identifiable E(spl)-C genes. All of these wasp's genomes encode the core components of the Notch signalling pathways, as well as other direct targets of Notch signalling (eg *glass*, *sugarless* etc.) (data not shown). This deficit is specific, therefore, to the genes of the E(spl)-C. That the E(spl)-C is missing from the genomes of these wasps raises important

questions as to how cell specification in the nervous system of these animals is achieved.

Interestingly, studies of the effect of *Nasonia* venom on their fly hosts indicates that E(spl)-C genes are upregulated in the host in response to the venom [82], possibly to trigger developmental arrest. Is it possible that the evolution of resistance to their own venom has necessitated the deletion of the E(spl)-C from the genomes of wasps that use this mechanism?

Origins of the e(spl) complex

E(spl) bHLHO proteins are related to a broad family of BHLHO proteins found in animal genomes. In arthropods, 5 major families (Hairy/deadpan, Side, clockwork

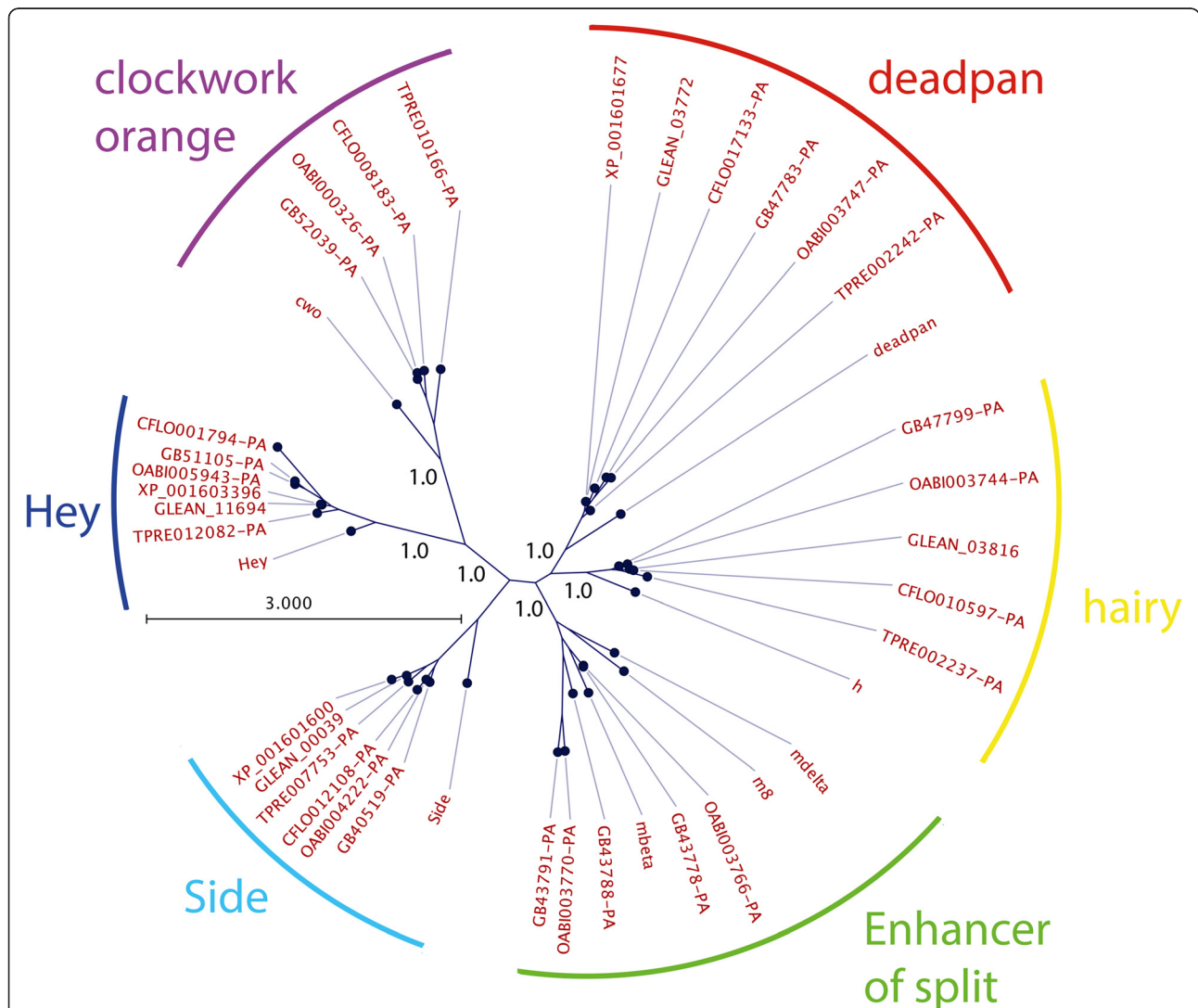


Fig. 7 Relationships of bHLHO proteins from Chalcid wasps and other insects. Bayesian phylogram of representative hymenopteran bHLH-orange domain proteins reconstructed using the WAG model of protein evolution. While Chalcid wasps (genes identified by the prefixes, CFL (*Copidosoma*), TRE (*Trichogramma*) and XP_ or GLEAN (*Nasonia*)) have members of the hairy, deadpan, clockwork orange, Hey and Side families, they do not have proteins related to E(spl)-C bHLH proteins from *Apis* (GB identifiers, *Orussus* (woodwasp, OAB identifiers), or *Drosophila* (gene names)

The E(spl)-like genes from *Strigamia* do not form a cluster in the genome, and the proteins encoded by these genes form a clade separate to the three crustacean/insect E(spl)-C bHLH clades. As most phylogenetic examinations of the placement of myriapods within the arthropods indicates that they are the sister group to crustaceans and insects, the origins of E(spl)-c bHLH proteins must lie somewhere after the separation of the lineage leading to chelicerates, but before the last common ancestor of insects/crustacea and myriapods. E(spl)-C bHLHO proteins thus pre-date the origin of the E(spl)-C, which is present only in the genomes of insects and crustacea.

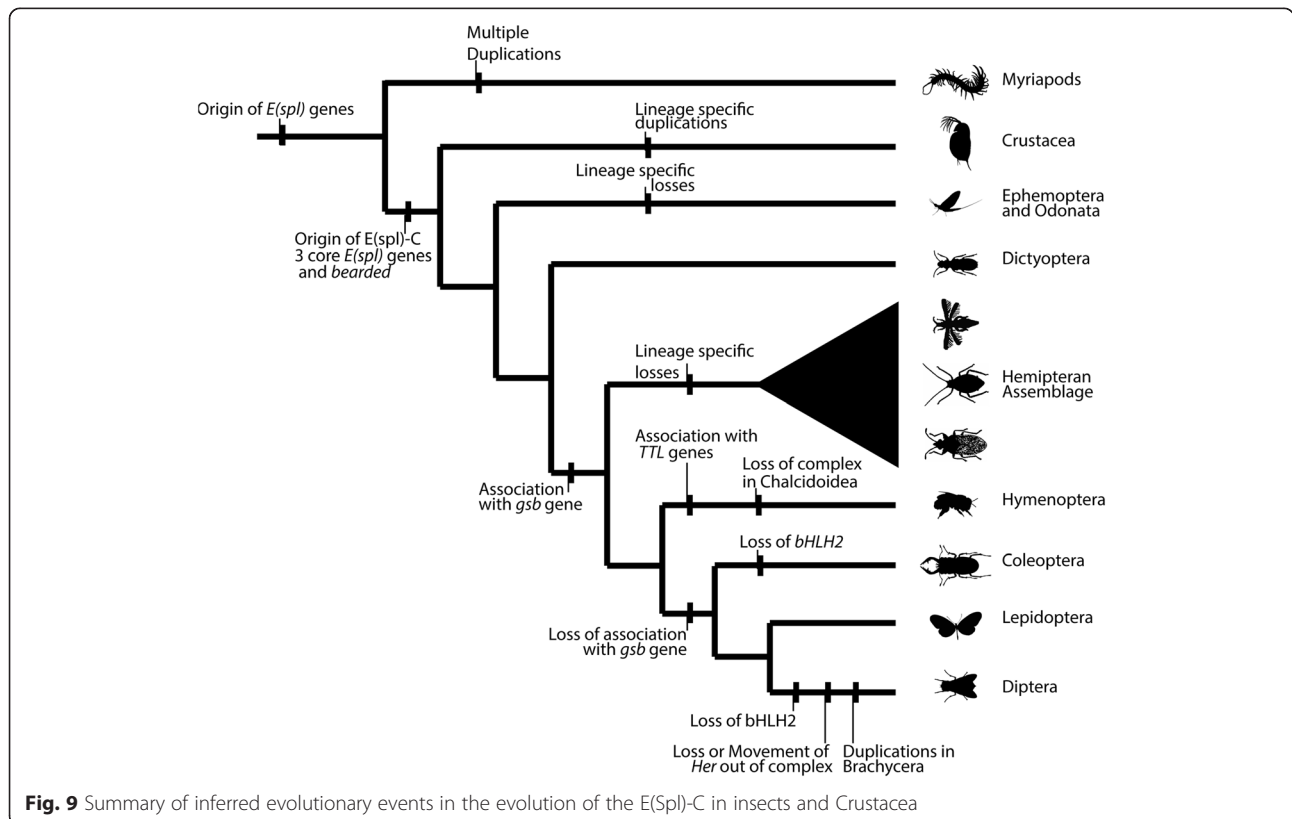
These differences in the organisation of the E(spl)-C genes and complex mirror, to some extent, differences between these clades in the presence of the neural stem cells, neuroblasts, that the E(spl)-C regulates. In chelicerates and myriapods there is no evidence for cells similar to the neural stem cells that arise out of proneural clusters and repress their neighbours through Notch signalling and the E(spl)-C in *Drosophila* [83–87]. Neither of these groups have an E(spl)-C in our analyses, while Crustacea and insects, which do have identifiable neuroblasts, do. There is some evidence that Myriapods, which have no E(spl)-C but do have E(Spl) bHLHO proteins, may have specialised neural precursors in groups of cells specified to become neural [83]. Understanding how neural cells are specified in these groups, and how this is related to

Notch signalling, will allow us to determine if the formation of the E(spl)-C is linked to the evolution of neuroblasts.

Conclusions

The genomes of arthropods contain few evolutionary conserved gene complexes [1], the most well known being the Hox [5], runt [88] and E(spl)-C [10]. The E(spl)-C is restricted to Crustacea and insects, but the key bHLH genes arose before the formation of the complex. The complex appears to have become assembled in the lineage leading to insects and crustaceans, possibly through the association of the bHLH genes (with a long evolutionary history of Notch responsiveness) with the *ma* Brd-class gene (Fig. 9). Presumably the formation of this complex gave some advantage in the regulation or expression of these genes, cementing the structure of the complex. The regulation of this complex in *Drosophila* through chromatin conformation regulators [45, 46], and the suggestion of coordinate regulation [46], may provide an explanation for the conservation of the complex through 540 million years of arthropod evolution. This complex has remained stable in insect genomes, but while gene-loss and duplication has reshaped it in some lineages, its complete absence has only been detected in a group of Chalcidoid wasps.

The pattern of evolution of the E(spl)-C implies some regulatory reason for the conservation of its structure,



perhaps on a par with the well-described coordinate regulation of the Hox complex [5]. Examining the expression and function of these genes in species with variations of the complex, and in deeply branching groups such as myriapods, will provide insight into the reasons behind the conservation of this remarkable gene complex.

Methods

Gene identification

BHLHO domain genes and Brd-class genes were identified in arthropod genomes using Blast [69], with orthology assigned using a reciprocal blast best-hit approach. Coding sequences were either extracted from gene prediction sets, or, if such predictions were absent or erroneous, using FGENESH [89] on contigs identified as containing bHLHO sequences using tblastn [69]. Predicted proteins were generated and aligned using CLC Genomics Workbench (<http://www.clcbio.com>).

Genomic analyses were carried out using CLC Genomics Workbench to visualise the placement of bHLHO genes on scaffolds and contigs. Predicted proteins encoded in these regions were analysed with blastp [69], in the first instance, and HMMER [90] (to identify Brd-class encoding genes).

Phylogenetics

All phylogenetics were carried out using MrBayes [67] using the WAG model of protein [91] which proved to be the most appropriate model after testing using mixed models. Monte-Carlo Markov chains were run for 1000000 generations with the initial 25 % of trees discarded as burn-in. Consensus trees were visualised with Dendroscope [92] or CLC Genomics Workbench.

Availability of data and materials

All sequences and accession numbers (either derived from Genbank or the Baylor i5K pilot project (<https://www.hgsc.bcm.edu/i5k-pilot-project-summary>)) are available in Additional file 2.

Ethics

This manuscript describes analysis of publically available genome sequences and annotation (used with permission) and thus did not require animal ethics approval.

Additional files

Additional file 1: Table of bHLH Orange domain genes used in this analysis. (XLSX 35 kb)

Additional file 2: Alignments used for phylogenetic analyses. (TXT 111 kb)

Competing interests

The author declares that he has no competing interests.

Authors' contributions

All work described was carried out by PKD.

Acknowledgements

The author would like to acknowledge P.M. Dearden for critical reading of this manuscript. The author is grateful for the efforts of the i5K consortium and the Baylor College of Medicine Human Genome Sequencing Center who provided genome sequences and annotation. Hemipteran taxon images in Fig. 9 are from Joseph Hughes, used under a creative commons license (<https://creativecommons.org/licenses/by-nc-sa/3.0/>).

Received: 18 June 2015 Accepted: 12 September 2015

Published online: 18 September 2015

References

- Zdobnov EM, Bork P. Quantification of insect genome divergence. *Trends Genet.* 2007;23(1):16–20.
- Barbazuk WB, Korf I, Kadavi C, Heyen J, Tate S, Wun E, et al. The syntenic relationship of the zebrafish and human genomes. *Genome Res.* 2000;10(9):1351–8.
- Kohn M, Hogel J, Vogel W, Minich P, Kehrer-Sawatzki H, Graves JA, et al. Reconstruction of a 450-My-old ancestral vertebrate protokaryotype. *Trends Genet.* 2006;22(4):203–10.
- Kikuta H, Laplante M, Navratilova P, Komisarczuk AZ, Engstrom PG, Fredman D, et al. Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. *Genome Res.* 2007;17(5):545–55.
- Hughes CL, Kaufman TC. Hox genes and the evolution of the arthropod body plan. *Evol Dev.* 2002;4(6):459–99.
- Knust E, Schrons H, Grawe F, Campos-Ortega JA. Seven genes of the Enhancer of split complex of *Drosophila melanogaster* encode helix-loop-helix proteins. *Genetics.* 1992;132(2):505–18.
- Delidakis C, Monastiriotti M, Magadi SS. E(spl): genetic, developmental, and evolutionary aspects of a group of invertebrate Hes proteins with close ties to Notch signaling. *Curr Top Dev Biol.* 2014;110:217–62.
- Apidianakis Y, Nagel AC, Chalkiadaki A, Preiss A, Delidakis C. Overexpression of the m4 and malpha genes of the E(spl)-complex antagonizes notch mediated lateral inhibition. *Mech Dev.* 1999;86(1–2):39–50.
- Lai EC, Bodner R, Posakony JW. The enhancer of split complex of *Drosophila* includes four Notch-regulated members of the bearded gene family. *Development.* 2000;127(16):3441–55.
- Duncan EJ, Dearden PK. Evolution of a genomic regulatory domain: the role of gene co-option and gene duplication in the Enhancer of split complex. *Genome Res.* 2010;20(7):917–28.
- Schrons H, Knust E, Campos-Ortega JA. Toward a functional dissection of the Enhancer-of-split gene complex. *J Neurogenet.* 1991;7:145.
- Delidakis C, Artavanis-Tsakonas S. The Enhancer of split [E(spl)] locus of *Drosophila* encodes seven independent helix-loop-helix proteins. *Proc Natl Acad Sci U S A.* 1992;89(18):8731–5.
- Schrons H, Knust E, Campos-Ortega JA. The Enhancer of split complex and adjacent genes in the 96 F region of *Drosophila melanogaster* are required for segregation of neural and epidermal progenitor cells. *Genetics.* 1992;132(2):481–503.
- Oellers N, Tietze K, Knust E. The genes of the Enhancer-of-split complex of *Drosophila melanogaster* encode regulatory helix-loop-helix proteins. *J Neurogenet.* 1993;8:242.
- Lai EC, Posakony JW. The Bearded box, a novel 3' UTR sequence motif, mediates negative post-transcriptional regulation of Bearded and Enhancer of split Complex gene expression. *Development.* 1997;124(23):4847–56.
- Leviten MW, Lai EC, Posakony JW. The *Drosophila* gene Bearded encodes a novel small protein and shares 3' UTR sequence motifs with multiple Enhancer of split complex genes. *Development.* 1997;124(20):4039–51.
- Welshons WJ. Dosage experiments with split mutants in the presence of an enhancer of split. *Drosophila Inf Serv.* 1956;30:157–8.
- Jennings B, Preiss A, Delidakis C, Bray S. The Notch signalling pathway is required for Enhancer of split bHLH protein expression during neurogenesis in the *Drosophila* embryo. *Development.* 1994;120(12):3537–48.
- Knust E. Cell fate choice during early neurogenesis in *Drosophila melanogaster*. *Perspect Dev Neurobiol.* 1994;2(2):141–9.
- Bray SJ. Expression and function of Enhancer of split bHLH proteins during *Drosophila* neurogenesis. *Perspect Dev Neurobiol.* 1997;4(4):313–23.

21. Hartenstein V, Wodarz A. Initial neurogenesis in *Drosophila*. *Wiley Interdiscipl Rev Dev Biol*. 2013;2(5):701–21.
22. Nakao K, Campos-Ortega JA. Persistent expression of genes of the enhancer of split complex suppresses neural development in *Drosophila*. *Neuron*. 1996;16(2):275–86.
23. Oellers N, Dehio M, Knust E. bHLH proteins encoded by the Enhancer of split complex of *Drosophila* negatively interfere with transcriptional activation mediated by proneural genes. *Mol Gen Genet*. 1994;244(5):465–73.
24. Ruiz-Gomez M, Ghysen A. The expression and role of a proneural gene, *achaete*, in the development of the larval nervous system of *Drosophila*. *EMBO J*. 1993;12(3):1121–30.
25. Tepass U, Hartenstein V. Neurogenic and proneural genes control cell fate specification in the *Drosophila* endoderm. *Development*. 1995;121(2):393–405.
26. Lecourtois M, Schweisguth F. The neurogenic suppressor of hairless DNA-binding protein mediates the transcriptional activation of the enhancer of split complex genes triggered by Notch signaling. *Genes Dev*. 1995;9(21):2598–608.
27. Gigliani F, Longo F, Gaddini L, Battaglia PA. Interactions among the bHLH domains of the proteins encoded by the Enhancer of split and *achaete-scute* gene complexes of *Drosophila*. *Mol Gen Genet*. 1996;251(6):628–34.
28. Heitzler P, Bourouis M, Ruel L, Carteret C, Simpson P. Genes of the Enhancer of split and *achaete-scute* complexes are required for a regulatory loop between Notch and Delta during lateral signalling in *Drosophila*. *Development*. 1996;122(1):161–71.
29. Jennings BH, Tyler DM, Bray SJ. Target specificities of *Drosophila* enhancer of split basic helix-loop-helix proteins. *Mol Cell Biol*. 1999;19(7):4600–10.
30. Kaul A, Schuster E, Jennings BH. The Groucho co-repressor is primarily recruited to local target sites in active chromatin to attenuate transcription. *PLoS Genet*. 2014;10(8):e1004595.
31. Lai EC, Bodner R, Kavalier J, Freschi G, Posakony JW. Antagonism of notch signaling activity by members of a novel protein family encoded by the bearded and enhancer of split gene complexes. *Development*. 2000;127(2):291–306.
32. Pavlopoulos E, Pitsouli C, Klueg KM, Muskavitch MA, Moschonas NK, Delidakis C. *neuralized* Encodes a peripheral membrane protein involved in delta signaling and endocytosis. *Dev Cell*. 2001;1(6):807–16.
33. Preiss A, de Celis J, Jennings B, Wech I, Wurmlich E, Delidakis C, et al. Distinct expression of individual Enhancer of split genes. *J Neurogenet*. 1997;11:187–8.
34. Wech I, Bray S, Delidakis C, Preiss A. Distinct expression patterns of different enhancer of split bHLH genes during embryogenesis of *Drosophila melanogaster*. *Dev Genes Evol*. 1999;209(6):370–5.
35. Eastman DS, Slee R, Skoufos E, Bangalore L, Bray S, Delidakis C. Synergy between suppressor of Hairless and Notch in regulation of Enhancer of split γ and δ expression. *Mol Cell Biol*. 1997;17(9):5620–8.
36. Cave JW, Xia L, Caudy M. Differential regulation of transcription through distinct Suppressor of Hairless DNA binding site architectures during Notch signaling in proneural clusters. *Mol Cell Biol*. 2011;31(1):22–9.
37. Nellesen DT, Lai EC, Posakony JW. Discrete enhancer elements mediate selective responsiveness of enhancer of split complex genes to common transcriptional activators. *Dev Biol*. 1999;213(1):33–53.
38. Maeder ML, Polansky BJ, Robson BE, Eastman DA. Phylogenetic footprinting analysis in the upstream regulatory regions of the *Drosophila* enhancer of split genes. *Genetics*. 2007;177(3):1377–94.
39. Bailey AM, Posakony JW. Suppressor of hairless directly activates transcription of enhancer of split complex genes in response to Notch receptor activity. *Genes Dev*. 1995;9(21):2609–22.
40. Furukawa T, Kobayakawa Y, Tamura K, Kimura K, Kawaichi M, Tanimura T, et al. Suppressor of hairless, the *Drosophila* homologue of RBP-J kappa, transactivates the neurogenic gene *E(spl)m8*. *Idengaku Zasshi*. 1995;70(4):505–24.
41. Fortini ME, Artavanis-Tsakonas S. The suppressor of Hairless protein participates in Notch receptor signaling. *Cell*. 1994;79(2):273–82.
42. Ungerer P, Eriksson BJ, Stollewerk A. Unravelling the evolution of neural stem cells in arthropods: notch signalling in neural stem cell development in the crustacean *Daphnia magna*. *Dev Biol*. 2012;371(2):302–11.
43. Lai EC, Burks C, Posakony JW. The K box, a conserved 3' UTR sequence motif, negatively regulates accumulation of enhancer of split complex transcripts. *Development*. 1998;125(20):4077–88.
44. Lai EC, Tam B, Rubin GM. Pervasive regulation of *Drosophila* Notch target genes by GY-box-, Brd-box-, and K-box-class microRNAs. *Genes Dev*. 2005;19(9):1067–80.
45. Schaaf CA, Misulovin Z, Sahota G, Siddiqui AM, Schwartz YB, Kahn TG, et al. Regulation of the *Drosophila* Enhancer of split and invected-engrailed gene complexes by sister chromatid cohesion proteins. *PLoS One*. 2009;4(7):e6202.
46. Schaaf CA, Misulovin Z, Gause M, Koenig A, Dorsett D. The *Drosophila* enhancer of split gene complex: architecture and coordinate regulation by notch, cohesin, and polycomb group proteins. *G3*. 2013;3(10):1785–94.
47. Ingham PW, Pinchin SM, Howard KR, Ish-Horowitz D. Genetic analysis of the hairy locus of *Drosophila melanogaster*. *Genetics*. 1985;111:463–86.
48. Rushlow CA, Hogan A, Pinchin SM, Howe KM, Lardelli M, Ish-Horowitz D. The *Drosophila* hairy protein acts in both segmentation and bristle patterning and shows homology to N-myc. *EMBO J*. 1989;8(10):3095–104.
49. Sommer RJ, Tautz D. Involvement of an orthologue of the *Drosophila* pair-rule gene hairy in segment formation of the short germ-band embryo of *Tribolium* (Coleoptera). *Nature*. 1993;361(6411):448–50.
50. Damen WG, Weller M, Tautz D. Expression patterns of *hairy*, *even-skipped*, and *run1* in the spider *Cupiennius salei* imply that these genes were segmentation genes in a basal arthropod. *Proc Natl Acad Sci U S A*. 2000;97(9):4515–9.
51. Wilson MJ, Dearden PK. Pair-rule gene orthologues have unexpected maternal roles in the honeybee (*Apis mellifera*). *PLoS One*. 2012;7(9):e46490.
52. Bier E, Vaessin H, Younger-Shepherd S, Jan LY, Jan YN. *deadpan*, an essential pan-neural gene in *Drosophila*, encodes a helix-loop-helix protein similar to the hairy gene product. *Genes Dev*. 1992;6:2137–51.
53. Younger-Shepherd S, Vassin H, Bier E, Jan LY, Jan YN. *deadpan*, an essential pan-neural gene encoding an HLH protein, acts as a denominator in *Drosophila* sex determination. *Cell*. 1992;70:911–22.
54. San-Juan BP, Baonza A. The bHLH factor *deadpan* is a direct target of Notch signaling and regulates neuroblast self-renewal in *Drosophila*. *Dev Biol*. 2011;352(1):70–82.
55. Monastiriotti M, Giagtzoglou N, Koumbanakis KA, Zacharioudaki E, Deligiannaki M, Wech I, et al. *Drosophila* Hey is a target of Notch in asymmetric divisions during embryonic and larval neurogenesis. *Development*. 2010;137(2):191–201.
56. Kadener S, Stoleru D, McDonald M, Nawatwan P, Rosbash M. Clockwork Orange is a transcriptional repressor and a new *Drosophila* circadian pacemaker component. *Genes Dev*. 2007;21(13):1675–86.
57. Lim C, Chung BY, Pitman JL, McGill JJ, Pradhan S, Lee J, et al. Clockwork orange encodes a transcriptional repressor important for circadian-clock amplitude in *Drosophila*. *Curr Biol*. 2007;17(12):1082–9.
58. Matsumoto A, Ukai-Tadenuma M, Yamada RG, Houl J, Uno KD, Kasukawa T, et al. A functional genomics strategy reveals clockwork orange as a transcriptional regulator in the *Drosophila* circadian clock. *Genes Dev*. 2007;21(13):1687–700.
59. Richier B, Michard-Vanhee C, Lamouroux A, Papin C, Rouyer F. The clockwork orange *Drosophila* protein functions as both an activator and a repressor of clock gene expression. *J Biol Rhythm*. 2008;23(2):103–16.
60. Kobayashi T, Kageyama R. Expression dynamics and functions of Hes factors in development and diseases. *Curr Top Dev Biol*. 2014;110:263–83.
61. Gazave E, Guillou A, Balavoine G. History of a prolific family: the Hes/Hey-related genes of the annelid *Platynereis*. *EvoDevo*. 2014;5:29.
62. San Juan BP, Andrade-Zapata I, Baonza A. The bHLH factors Dpn and members of the E(spl) complex mediate the function of Notch signalling regulating cell proliferation during wing disc development. *Biol Open*. 2012;1(7):667–76.
63. Geling A, Plessy C, Rastegar S, Strahle U, Bally-Cuif L. Her5 acts as a prepattern factor that blocks neurogenin1 and coe2 expression upstream of Notch to inhibit neurogenesis at the midbrain-hindbrain boundary. *Development*. 2004;131(9):1993–2006.
64. Tannahill D, Bray S, Harris WA. A *Drosophila* E(spl) gene is “neurogenic” in *Xenopus*: a green fluorescent protein study. *Dev Biol*. 1995;168(2):694–7.
65. i KC. The i5K Initiative: advancing arthropod genomics for knowledge, human health, agriculture, and the environment. *J Hered*. 2013;104(5):595–600.
66. Poelchau M, Childers C, Moore G, Tsavatapalli V, Evans J, Lee CY, et al. The i5K Workspace@NAL-enabling genomic data access, visualization and curation of arthropod genomes. *Nucleic Acids Res*. 2015;43(Database issue):D714–719.
67. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 2003;19(12):1572–4.
68. Chipman AD, Ferrier DE, Brena C, Qu J, Hughes DS, Schroder R, et al. The first myriapod genome sequence reveals conservative arthropod gene content and genome organisation in the centipede *Strigamia maritima*. *PLoS Biol*. 2014;12(11):e1002005.

69. Altschul S, Gish W, Miller W, Myers E, Lipman D. Basic local alignment search tool. *J Mol Biol.* 1990;215(3):403–10.
70. Wurmbach E, Wech I, Preiss A. The Enhancer of split complex of *Drosophila melanogaster* harbors three classes of Notch responsive genes. *Mech Dev.* 1999;80(2):171–80.
71. Schlatter R, Maier D. The Enhancer of split and Achaete-Scute complexes of *Drosophilids* derived from simple ur-complexes preserved in mosquito and honeybee. *BMC Evol Biol.* 2005;5:67.
72. Chen G, Courey AJ. Groucho/TLE family proteins and transcriptional repression. *Gene.* 2000;249(1–2):1–16.
73. Gutjahr T, Patel NH, Li X, Goodman CS, Noll M. Analysis of the gooseberry locus in *Drosophila* embryos: gooseberry determines the cuticular pattern and activates gooseberry neuro. *Development.* 1993;118(1):21–31.
74. Bhat KM. The patched signaling pathway mediates repression of gooseberry allowing neuroblast specification by wingless during *Drosophila* neurogenesis. *Development.* 1996;122(9):2921–32.
75. Duman-Scheel M, Li X, Orlov I, Noll M, Patel NH. Genetic separation of the neural and cuticular patterning functions of gooseberry. *Development.* 1997;124(15):2855–65.
76. Xue L, Li X, Noll M. Multiple protein functions of paired in *Drosophila* development and their conservation in the Gooseberry and Pax3 homologs. *Development.* 2001;128(3):395–405.
77. Liu W, Xue L. Functional conservation of the *Drosophila* gooseberry gene and its evolutionary alleles. *PLoS One.* 2012;7(1):e30980.
78. Savard J, Marques-Souza H, Aranda M, Tautz D. A segmentation gene in *tribolium* produces a polycistronic mRNA that codes for multiple conserved peptides. *Cell.* 2006;126(3):559–69.
79. Davis GK, Jaramillo CA, Patel NH. Pax group III genes and the evolution of insect pair-rule patterning. *Development.* 2001;128(18):3445–58.
80. Osborne PW, Dearden PK. Expression of Pax group III genes in the honeybee (*Apis mellifera*). *Dev Genes Evol.* 2005;215(10):499–508.
81. Keller RG, Desplan C, Rosenberg MI. Identification and characterization of *Nasonia* Pax genes. *Insect Mol Biol.* 2010;19 Suppl 1:109–20.
82. Martinson EO, Wheeler D, Wright J, Mrinalini, Siebert AL, Werren JH. *Nasonia vitripennis* venom causes targeted gene expression changes in its fly host. *Mol Ecol.* 2014;23(23):5918–30.
83. Dove H, Stollewerk A. Comparative analysis of neurogenesis in the myriapod *Glomeris marginata* (Diplopoda) suggests more similarities to chelicerates than to insects. *Development.* 2003;130(10):2161–71.
84. Brenneis G, Stollewerk A, Scholtz G. Embryonic neurogenesis in *Pseudopallene* sp. (Arthropoda, Pycnogonida) includes two subsequent phases with similarities to different arthropod groups. *Evo Devo.* 2013;4(1):32.
85. Stollewerk A, Chipman AD. Neurogenesis in myriapods and chelicerates and its importance for understanding arthropod relationships. *Integr Comp Biol.* 2006;46(2):195–206.
86. Kadner D, Stollewerk A. Neurogenesis in the chilopod *Lithobius forficatus* suggests more similarities to chelicerates than to insects. *Dev Genes Evol.* 2004;214(8):367–79.
87. Stollewerk A, Weller M, Tautz D. Neurogenesis in the spider *Cupiennius salei*. *Development.* 2001;128(14):2673–88.
88. Duncan EJ, Wilson MJ, Smith JM, Dearden PK. Evolutionary origin and genomic organisation of runt-domain containing genes in arthropods. *BMC Genomics.* 2008;9:558.
89. Solovyev V, Kosarev P, Seledsov I, Vorobyev D. Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol.* 2006;7 Suppl 1:S10 11–12.
90. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 2011;39(Web Server issue):W29–37.
91. Whelan S, Goldman N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol.* 2001;18(5):691–9.
92. Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, Rupp R. Dendroscope: an interactive viewer for large phylogenetic trees. *BMC Bioinformatics.* 2007;8:460.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

