Segmentation

Authors: Chad Vosburg, Andrea Lu, Teresa Shippy and Sherry Miller

Introduction

Segmentation is the process by which repeated units of similar groups of cells are created along the anterior-posterior axis of a developing embryo. This is a fundamental property of all arthropods and molecular mechanisms for this process were first elucidated by large scale developmental mutant screens performed by Christiane Nüsslein-Volhard and Eric Wieschaus in the insect model *Drosophila melanogaster* [1]. In fruitflies this complex developmental process begins with cytoplasmic inheritance of mRNAs that are maternally produced and provided to the oocyte. These maternal factors create gradients that provide the coordinates to define the anterior and posterior end of the embryo. Once translated these maternal factors function to activate a group of genes known as gap genes. Gap genes are expressed in broad, well-defined domains in the early embryo and activate the next set of transcription factors, the pair-rule genes. Pair-rule genes are expressed in every other segment of the developing embryo and together they activate the expression of segment polarity genes which are expressed in every segment of the developing embryo. Together, gap, pair-rule and segment polarity genes function to activate homeotic genes, which once activated in a specific region confer the character or identity of the given segment [2,3]. Decades of comparative studies in diverse arthropod species has shown that some aspects of the segmentation pathway are highly conserved while other aspects have undergone evolutionary change [4]. While all the genes discussed in this report were initially identified because of their role in embryonic patterning and segmentation in *Drosophila*, many of them have other important cellular functions during embryogenesis such as pole cell development, neural stem cell maintenance, sex determination, and immune function and thus their presence or absence in *Diaphorina citri* needs to be considered in these additional contexts.

Of the 32 *Drosophila* maternal effect, gap, pair-rule and segment polarity genes examined only 23 were identified in the *D. citri* v3.0 genome. All of the genes reported as absent in this genome have been reported as absent in at least one other insect suggesting our current assembly is likely complete with respect to genes known to be involved in the developmental process of segmentation.

Results and Discussion

Maternal effect genes

One-to-one orthologs were found for the *caudal (cad)*, *dorsal (dl)*, and *nanos (nos)* genes in the *D. citri* v3.0 genome. One TGF-α ligand, homologous to the maternal effect gene *gurken (grk)* and its paralogs *Keren (Krn)* and *spitz (spi)* was identified. Orthologs for *bicoid (bcd)* and *oskar (osk)* were not found (Table 1). Further information on the manual annotation of these genes can be found below.

Maternal Effect		*Drosophila melanogaster*	*Apis mellifera*	*Tribolium castaneum*	*Acyrthosiphon pisum*	*Diaphorina citri*
	caudal	1	1	1	1	1
	dorsal like genes *(dorsal, dorsal related immunity factor)*	2	1	2	2	1
	nanos	1	1	1	4	1
	bicoid	1	0	0	0	0
	oskar	1	0	0	0	0
	tgfα ligand (*gurken, Keren, spitz*)	3	1	1	1	1

Table 1: Maternal Effect Ortholog Number. The Drosophila melanogaster numbers were determined from Flybase. Ortholog numbers for Apis mellifera [5]*, Tribolium castaneum* [6] *and Acyrthosiphon pisum* [7] *are based on genome publications. Diaphorina citri ortholog numbers represent our final manual annotation.*

caudal

One *cad* ortholog has been identified and manually annotated in the *D. citri* genome v3.0. While genome assembly errors are present at this locus the Cad protein sequence in our manual annotation appears complete as PacBio IsoSeq and MCOT transcriptome evidence support our final annotation (Table 2).

dorsal

dl was previously annotated in the *D. citri* genome v1.0 because of its role in innate immunity as the Rel/NFκB transcription factor target of Toll signalling [8]. In this version of the genome (v3.0) a second isoform was annotated (Table 1). Both isoforms are supported with independent *de novo* transcriptome evidence (Table 2). Additionally, *dl* genes in several other insect species have been shown to produce alternative transcripts that result in different protein products [9–11]. Therefore, our findings are consistent with *dl* orthologs in other insects.

nanos

One *nos* ortholog was found in the *D. citri* genome v3.0 (Table 1). Unfortunately, it was only able to be annotated as a partial gene due to genome assembly errors. Initial blasts to the *D. citri* genome using full length Nos protein sequences from *Drosophila*, *Tribolium castaneum*, *Cimex lectularius*, and *Acyrthosiphon pisum* were unsuccessful. However, when a partial Nos protein sequence from *A. pisum* was used as the query sequence the *nos* locus was identified. The expression of the *nos* gene in *D. citri* was confirmed by its presence in a *de novo D. citri* transcriptome (Table 2). The current annotation, while incomplete, does show some of the conserved residues known to be present in Nos insect orthologs [12] and reciprocal blast analysis confirms its identification as Nos.

Group	Gene	D. citri identifier	Gene model		Evidence supporting annotation			
			complete	partial	MCOT	IsoSeq	RNASeq	Ortholog
Maternal Effect								
	caudal	Dcitr06g04620.1.4	x		x	x		
	dorsal	Dcitr02g07710.1.1 Dcitr02g07710.1.2	x		x	x	x	x
	nanos	Dcitr00g11810.1.1		x	x			x
	TGFalpha	Dcitr06g02470.1.1	x		x	x		
Gap								
	tailless	Dcitr01g16840.1.1	x					x
	knirps related 1	Dcitr10g03830.1.1	x			x	x	x
	knirps related 2	Dcitr10g03860.1.1	x		x			x
	Kruppel †		x		x			
	hunchback	Dcitr09g01780.1.1 Dcitr09g01780.1.2	x			x	x	
	huckebein	Dcitr04g11170.1.1	x		x		x	
	orthodenticle	Dcitr04g16960.1.1	x		x		x	x
	empty spiracles	Dcitr09g06330.1.1	x		x			
	cap-n-collar	Dcitr03g12850.1.1	x		x			
	collier	Dcitr03g01400.1.1 Dcitr03g01400.1.2	x		x		x	
Pair-rule								
	paired	Dcitr01g09360.1.1	x		x			x
	odd skipped	Dcitr01g20150.1.1	x		x			
	sloppy paired	Dcitr02g08120.1.1	x		x			
	runt	Dcitr01g07300.1.1	x		x			x
	even-skipped	Dcitr08g10250.1.1	x		x	x		
	hairy	Dcitr02g06890.1.1	x		x	x		x
	odd paired	Dcitr09g07980.1.1	x		x			
Segment polarity								
	gooseberry	Dcitr01g18500.1.1	x		x	x	x	
	engrailed	Dcitr08g03480.1.1 Dcitr08g03480.1.2	x		x		x	
Other related genes								
	Runt related A	Dcitr01g07290.1.1	x		x			x
	Runt related B	Dcitr01g07260.1.1	x		x			x
	lozenge	Dcitr01g07310.1.1		x		x		x
	sister of odd and bowl	Dcitr01g20160.1.1	x		x			
	brother of odd with entrails limited	Dcitr01g11160.1.1	x			x		
	gooseberry neuro	Dcitr01g18520.1.1	x		x	x	x	x

Table 2: Annotated D. citri segmentation genes. Each manually annotated gene has been assigned an OGS3 gene identifier. †denotes genes present in OGS2 but missing from OGS3. For each manually annotated gene it has been denoted as a partial or complete model based on available evidence. Evidence for manual annotation was also recorded. MCOT evidence means a de novo Oases or Trinity model from an independent transcriptome was identified and the sequence from that transcript was used to validate or modify our model. IsoSeq means single reads generated with Pacific Biosciences technology were available and were used to help validate the exon structure of the model. RNASeq means that individually mapped Illumina RNASeq reads were used to help validate or modify our model. Ortholog means ortholog sequences from other insects and information about conserved motifs and domains had to be used to help determine the final annotation.

gurken

In *Drosophila* there are three *TGF-α* ligand paralogs (*grk, Krn, spi*). In other insects, including *Nasonia vitripennis* [13], *Tribolium* [6], honeybee [5] and pea aphid [7], only one *TGF-α* ligand has been identified (Figure 1). In the *D. citri* genome v3.0 we also identified one *TGF-α* gene (Table 1) which had strong PacBio IsoSeq and MCOT transcriptome support (Table 2).

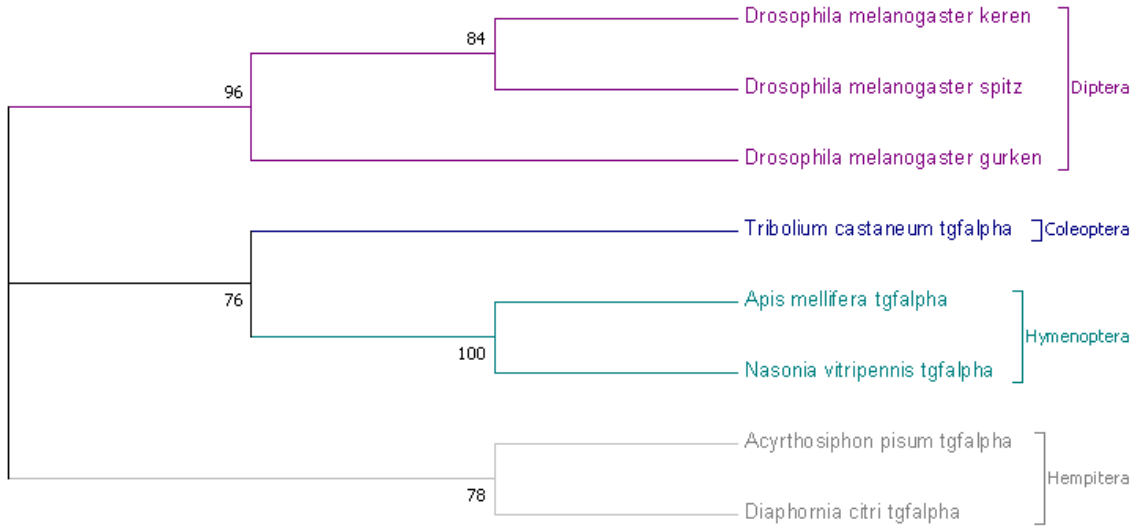

Figure 1: Bootstrap consensus neighbor-joining tree of TGF-alpha homologs. Full length proteins sequences were used in phylogenetic analysis. Colors denote insect orders. Purple: Diptera. Blue: Coleoptera. Green: Hymenoptera. Gray: Hempitera. A single TGF-apha homolog seems to be the norm in insects. In the Drosphila lineage there has been a duplication resulting in three paralogs (gurken, Keren, spitz).

bicoid

A *bcd* ortholog was not identified in the *D. citri* genome v3.0. This result was expected as decades of insect studies [14–17] and genome sequencing projects [5,6,18] have suggested that *bcd* is a dipteran innovation.

oskar

Like *bcd*, *osk* was initially only found in dipterans [12,19,20] and thus it was hypothesized that *osk* also evolved as a novelty in the dipteran lineage. An apparent absence of *osk* orthologs in *Bombyx mori* [21], *Tribolium* [6], *Apis mellifera* [18], *A. pisum* [7], *Rhodnius prolixus* [22] and *Pediculus humanus* [22] supported this hypothesis. However, later findings of *osk* orthologs in the hymenopteran insects *Nasonia* and *Messor pergandei*, called this hypothesis into question [22]. Lynch *et al* suggested that *osk* was an evolutionary novelty of all holometabolous insects (thus not present in hemimetabolous insects) which was then independently lost in some holometabolous insects [22]. Our findings do not contradict this hypothesis as we were unable to find an *osk* ortholog in *D. citri* (Table 1). BLASTs using *Drosophila*, *Aedes aegypti* and *Nasonia* Osk orthologs resulted in no significant hits to the *D. citri* genome, protein database or MCOT transcriptome.

Gap genes

One-to-one orthologs were found for the gap genes *tailless* (*tll*), *Kruppel* (*Kr*), *hunchback* (*hb*) *huckebein* (*hkb*), *empty spiracles* (*ems*), *cap-n-collar* (*cnc*), and *collier* (*col*) in the *D. citri* v3.0 genome. Two *knirps related* (*knrl*) genes were identified and 1 *orthodenticle* (*otd*) homolog was found. *giant* (*gt*) and *buttonhead* (*btd*) appear to be absent in the current *D. citri* genome assembly (Table 3). Further information on the manual annotation of these genes can be found below.

Gap		*Drosophila melanogaster*	*Apis mellifera*	*Tribolium castaneum*	*Acyrthosiphon pisum*	*Diaphorina citri*
	tailless	1	1	1	1	1
	knirps family (*knirps, knirps related, eagle*)	3	3	2	3	2
	giant	1	1	1	0	0
	Kruppel	1	1	1	1	1
	hunchback	1	1	1	1	1
	huckebein	1	1	1	0	1
	orthodentical	1	2	2	1	1
	buttonhead	1	1	1	0	0
	empty spiracles	1	1	1	1	1
	cap-n-collar	1	1	1	1	1
	collier	1	1	1	1	1

Table 3: Gap Gene Ortholog Number. The Drosophila melanogaster numbers were determined from Flybase. Ortholog numbers for Apis mellifera [5]*, Tribolium castaneum* [6] *and Acyrthosiphon pisum* [7] *are based on genome publications. Diaphorina citri ortholog numbers represent our final manual annotation.*

tailless

Tll codes for a protein belonging to the nuclear receptor superfamily. Nuclear receptors, including Tll, are evolutionally conserved in species across metazoans [23]. Analysis of the *D. citri* genome v3.0 suggests that one *tll* ortholog is present in *D. citri* (Table 3). There were no PacBio IsoSeq or *de novo* MCOT models to confirm gene structure and sequence. However, comparisons to *Drosophila*, *A. pisum* and *Bemisia tabaci* Tll orthologs suggests this manual annotation represents the full length protein (Table 2).

knirps

Drosophila has three knirps related genes, *knirps* (*kni*), *knirps related* (*knrl*) and *eagle* (*eg*). While *kni* was the first identified it is actually a paralog of *knrl*, is highly divergent and the most atypical member of the family. Phylogenetic analysis using 28 knirps family genes from 15 arthropods suggests that a single ancestral gene duplicated early in the insect lineage producing the *knrl* and *eg* paralogs (Perl et al). Subsequent duplications in various insect lineages has resulted in most insects having 2 or 3 knirps genes. In the hemipterans examined thus far (*A. pisum*, *R. prolixus*, *P. humanus*) it appears that *knrl* may have duplicated in the hemipteroids resulting in two *knrl* genes and one *eg* gene [24]. In the *D. citri* genome v3.0 w

e were only able to identify two knirps family genes (Table 3). These two genes are located on the same chromosome, about 400 kb apart in the current assembly. Both predicted proteins contain the highly conserved 94 amino acid N terminal domain and the C terminal PIDLS motif commonly found in knirps family members. However, neither contained the GAS motif that is exclusively found in the Eg protein. Due to the lack of this signature motif the resulting *D. citri* annotations were named *knirps related 1* (*knrl1*) and *knirps related 2* (*knrl2*). Despite the lack of GAS motif, it is possible that *D. citri knrl2* is the ortholog of *eg* as our phylogenetic analysis was unable to resolve the relationship of knirps related genes in insects (data `not shown). Interestingly, *knrl1* contains a small exon just 5' of the universally conserved typical first coding exon (Figure 2). This small exon has been shown to be present in some fruitfly, honeybee, pea aphid and human louse *knrl* genes [24]. The conservation of this small exon across diverse insect orders suggests a common origin for this 5' exon in insects.

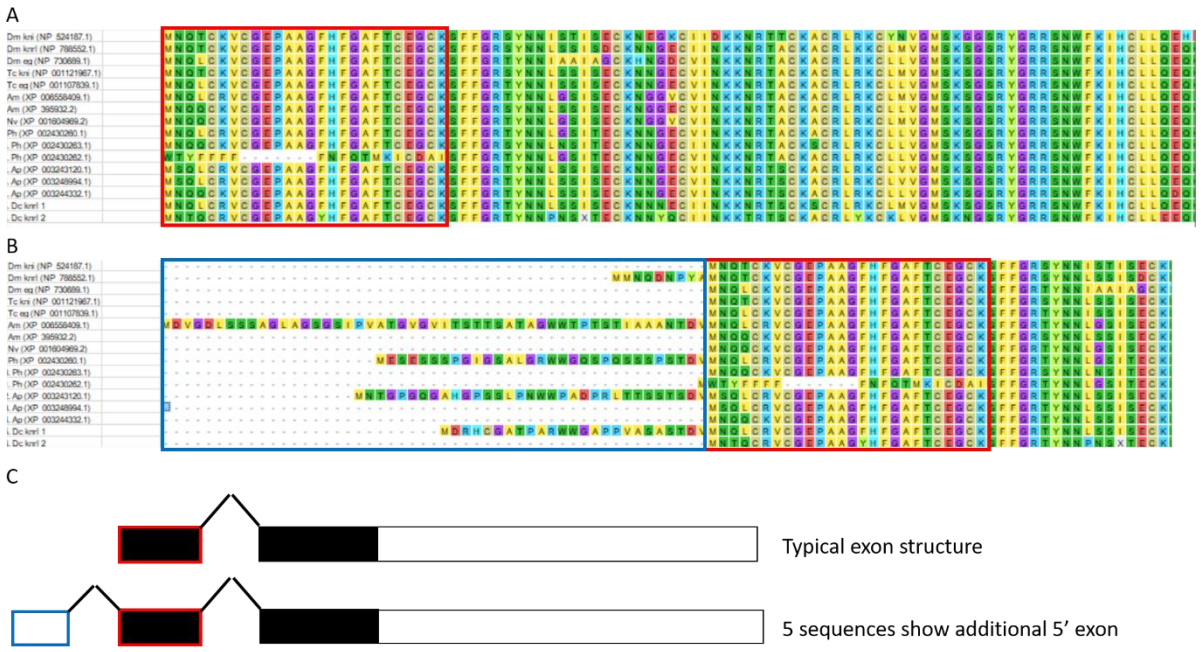

Figure 2: Conservation of knirps related genes. Perl et al reported a 94 amino acid core sequence found in knirps related genes which contains the DNA binding domain [24]. *Panel A shows this conserved sequence. The red box indicates the universally conserved first exon. Muscle alignment was performed in Mega7.0. Named genes have been manually annotated. Unnamed genes are computationally predicted. 7 insect genomes were examined. Dm; Drosophila melanogaster. Tc; Tribolium castaneum. Am; Apis mellifera. Nv; Nasonia vitripennis. Ph; Pediculus humanus. Ap; Acyrthosiphon pisum. Dc; Diaphorina citri. Panel B shows that five of the depicted proteins have sequence upstream of the universally conserved sequence which typically begins in the first exon. In these species the conserved core sequence begins in the 2nd exon. Blue box shows the first exon in these organisms. Red box is the second exon in these organisms. Panel C shows the typical exon structure of knirps family genes with the conserved core sequence (black box) starting in the 1st exon (red box) and extending into the second exon (note: In some organisms additional exons are present after exon 2) and the exon structure of one knirps related gene in Dm, Am, Ph, Ap and Dc which contains an addition exon upstream of the universally conserved 1st exon (blue box).*

giant

Whether *gt* is conserved among hemipterans is unclear. A *gt* ortholog was not found in the *A. pisum* genome [7], however, an ortholog was identified in *R. prolixus* [25]. The *R. prolixus* Gt, while highly diverged from other identified insect giants does have a leucine-zipper domain and appears to function as Gt [25]. Despite the sequence information from *R. prolixus* we were unable to identify a clear *gt* ortholog in the *D. citri* genome v3.0 (Figure 3, Table 3).

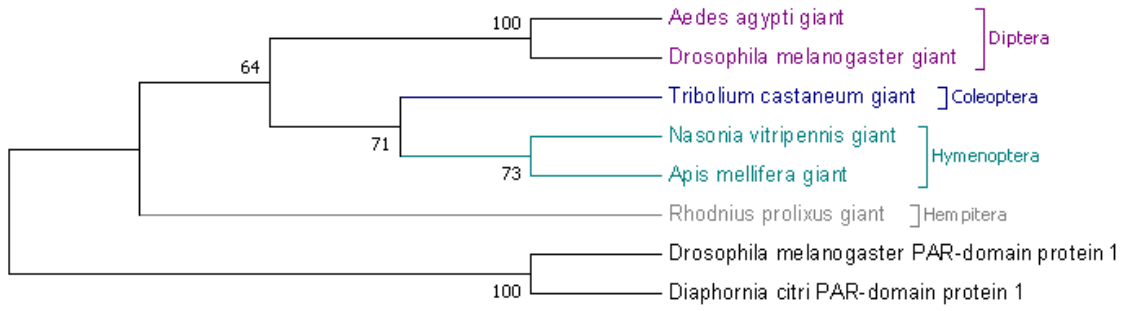

Figure 3: Bootstrap consensus neighbor-joining tree of giant homologs. Full length proteins sequences were used in phylogenetic analysis. Purple: Diptera. Blue: Coleoptera. Green: Hymenoptera. Gray: Hempitera. Black: Par domain proteins as an outgroup. D. citri PAR-domain protein 1 was the top BLAST hit when using insect giant sequences as a query. This protein is clearly not orthologous to other insect giants as it clusters with the Drosophila PAR-domain protein 1 outgroup.

Kruppel

In a variety of insects, including the hemipteran *R. prolixus*, Kruppel (Kr) orthologs have been identified and found to function as a gap gene in the segmentation process [26–34]. In *D. citri* genome v2.0 we identified one *Kr* ortholog (Table 3). The sequence of our annotation was confirmed by sequence from an independent *de novo* transcriptome (Table 2). During the assembly of genome v3.0, several duplicate contigs containing the *Kr* ortholog were removed during deduplication such that *Kr* is currently missing from the v3.0 genome (Table 2).

hunchback

hb plays an important role in anteroposterior patterning in many insects, although its regulation and specific functions differ to some extent [35–38]. *hb* is also important in the developing nervous system in *Drosophila* [39,40], a role which is thought to be conserved in most protostomes [41]. Because Hb has different functions at different life stages, it has been identified as a potential target for insect control strategies [42]. We identified a single copy of *hb* in the *D. citri* genome v3.0 (Table 3). IsoSeq transcripts support the presence of two *hb* isoforms that differ only in 5' UTR sequence (Table 2).

huckebein

In *Drosophila hkb* is a terminal gap gene which plays an important role in head development [43]. However, this function is unlikely to be widely conserved across insects as *hkb* is not expressed in a gap gene pattern in the basal dipteran moth midge *Clogmia albipunctata* [44] or in *Tribolium* [45]. While it's function in *D. citri* is unknown, an ortholog of *hkb* was found in the *D. citri* genome v3.0 and its expression was confirmed via a *de novo* transcriptome (Table 2), despite its reported absence in *A. pisum* [7] (Table 3).

buttonhead

btd is a member of the Sp-family of transcription factors. *Drosophila* has three Sp-family members, *btd*, *Sp1* and *Sp1-like factor for pairing sensitive-silencing* (*Spps*). While initial reports suggested that *btd* might be a *Drosophila* specific duplication [46,47] more recent reports indicate that three Sp members is likely the ancestral state for arthropods and perhaps all metazoans [48]. The insect species *Anopheles*, *Bombyx*, *Apis*, *Nasonia*, *Tribolium*, *Thermobia domestica* and *Folsomia candida*, the non-insect arthropod *Daphnia pulex* and the basal metazoans *Trichoplax adhaerens* and *Nematostella vectensis* all contain three Sp-family genes which cluster into three monophyletic clades (Sp5/btd, Sp1-4 and Sp6-9 clades) [48]. Despite the fact that three Sp-family members appears to be the ancestral state, *btd* is absent from the *A. pisum* genome [7] and repeated efforts to clone *btd* from *Oncopeltus fasiatus* has only resulted in the identification of the two non *btd* Sp genes [48,49]. We too were unable to find a true *btd* ortholog in either the *D. citri* genome v3.0 or in independent *de novo* transcriptomes (Tables 2 & 3). In *D. citri* all evidence suggests there are only two Sp-family members, which are orthologous to *Sp1* and *Spps*.

empty spiracles

ems acts as a head gap gene in Drosophila [50]. It is expressed in the head of a variety of other arthropods, but in a smaller domain [51–53]. In the *D. citri* genome v3.0 we identified one *ems* ortholog (Table 3). This sequence is supported by an independent *D. citri* transcriptome (Table 2).

orthodenticle

Along with *btd* and *ems*, *otd* was identified in *Drosophila* as a head gap gene [50]. While some aspects of the head gap gene patterning system are specific to the *Drosophila* lineage, expression analysis and

loss of function studies in a variety of arthropods have shown that *otd* likely has a conserved role in patterning the arthropod head [53–58]. While *Drosophila* has only one *otd* gene there is evidence that a single ancestral *otd* duplicated before the radiation of arthropods, the result being that most insects examined contain 2 *otd* genes. Functional studies in several insects suggest that *otd-1* is expressed in the early blastoderm and is responsible for patterning the head while *otd-2* is expressed much later in development [55,56]. In *A. pisum* only one *otd* gene has been identified [7]. This *otd* is more similar to *otd-2* in other insects, both in sequence and expression pattern, suggesting there has been a loss of *otd-1* in pea aphids [17]. We too were only able to find one *otd* gene in the *D. citri* v3.0 genome (Table 3). The *D. citri otd* also appears to be more similar to insect *otd-2* (Figure 4) suggesting the gene loss of *otd-1* may have occurred in the lineage leading to sap sucking insects.

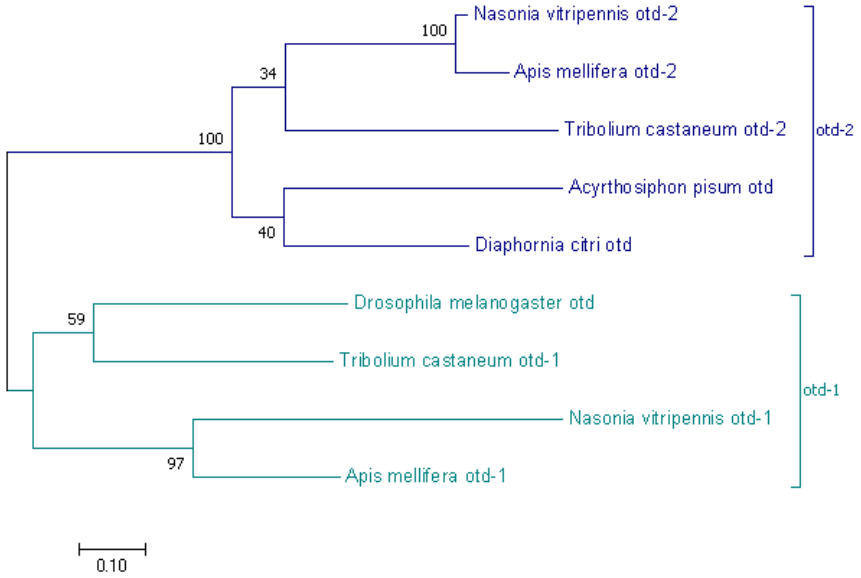

Figure 4: Maximum likelihood tree of otd homologs. Full length proteins sequences were used in phylogenetic analysis. Green: otd-1. Blue: otd-2. Acyrthosiphon pisum and Diaphorina citri have a single otd homolog which appears more similar to insect otd-2 genes.

cap-n-collar

In *Drosophila cnc* is important for patterning of the head [59]. We identified one *cnc* ortholog in the *D. citri* genome v3.0 (Table 3). This annotation has both *de novo* transcriptome and PacBio IsoSeq support (Table 2). In *Drosophila* multiple isoforms for *cnc* have been identified [60]. PacBio IsoSeq data suggest that multiple isoforms also exist in *D. citri*.

collier

In Drosophila, *col* acts downstream of *otd* and *ems* and upstream of *cnc* [61]. One ortholog of the head gap gene *col* was also identified in the *D. citri* v3.0 genome (Table 3). In *Drosophila* multiple isoforms of this gene have been identified [60] Based on independent *de novo* transcriptome evidence (Table 2) two isoforms for this gene have been annotated in the current *D. citri* genome.

Pair-rule genes

One-to-one orthologs were found for all pair-rule genes examined. This includes *paired* (*prd*), *odd skipped* (*odd*), *sloppy paired* (*slp*), *runt* (*run*), *even skipped* (*eve*), *hairy* (*h*) and *odd paired* (*opa*). Further information on the manual annotation of these genes can be found below.

Pair-rule		*Drosophila melanogaster*	*Apis mellifera*	*Tribolium castaneum*	*Acyrthosiphon pisum*	*Diaphorina citri*
	paired	1	1	1	1	1
	odd-skipped	1	1	1	1	1
	sloppy paired	2	1	1	1	1
	runt	1	1	1	1	1
	even-skipped	1	1	1	1	1
	hairy	1	1	1	1	1
	odd paired	1	1	1	1	1

Table 4: Pair-rule Ortholog Number. The Drosophila melanogaster numbers were determined from Flybase. Ortholog numbers for Apis mellifera [5]*, Tribolium castaneum* [6] *and Acyrthosiphon pisum* [7] *are based on genome publications. Diaphorina citri ortholog numbers represent our final manual annotation.*

paired

Prd is a member of the Pax3/7 family of proteins. In *Drosophila* there are three Pax3/7 family genes which are known to play a role in segmentation and neurogenesis, *prd*, *gooseberry* (*gsb*) and *gooseberry neuro* (*gsb-n*). While the number of Pax3/7 genes varies in arthropods, data from insects and arachnids suggest that the Pax3/7 roles in segmentation and neurogenesis are likely to be conserved in all arthropods [62]. In the *D. citri* genome v3.0 we also found three Pax3/7 genes which have been named *D.citri_paired*, *D. citri_gooseberry* and *D. citri_gooseberry neuro* (see below for a discussion of *gsb* and *gsb-n* orthologs) (Tables 4 & 5). Due to the condition of the *D. citri* genome we were only able annotate a partial gene model as *D. citri_paired*. This partial sequence was confirmed with *de novo* MCOT transcript evidence (Table 2). Pax3/7 proteins are defined by the presence of three conserved domains; the paired domain, the paired class homeodomain and a short octapeptide motif typically found between the paired domain and the homeodomain [63]. This partial gene model contains the complete homeodomain but is missing part of the 5' sequence, thus the paired domain is incomplete in this model. A more complete model for paired can be found in the *D. citri* MCOT transcriptome as model MCOT01278.2.CO.

odd skipped

odd is a zinc finger transcription factor. *odd* is closely related to two other zinc finger transcription factors known as *brother of odd with entrails limited* (*bowl*) and *sister of odd and bowl* (*sob*). In *Drosophila* it has been reported that all three genes are found clustered within the genome, with *sob* and *odd* being particularly close [64]. Analysis of insects with sequenced genomes indicates that this clustering may be conserved. In *D. citri* genome v3.0, *odd* and *sob* are actually overlapping one another on opposite strands. The overlap may be an artefact of local misassembly, but it is likely that the genes are closely associated. *bowl* is located on the same chromosome about 25 Mb away. *odd* and *sob* are supported by MCOT transcripts. MCOT models were not available to confirm the annotation of *bowl*, however, PacBio IsoSeq reads support the gene structure of this annotation (Table 2).

sloppy paired

In *Drosophila* there are two *slp* genes involved in segmentation (*slp-1* and *slp-2*)[65]. In more basal insects and other arthopods one *slp* homolog is typically found [66–68]. As expected in *D. citri* one *slp* ortholog was identified (Table 4). The sequence of this annotation has more similarity with *Drosophila* Slp2 than Slp1. However, following the naming convention found in other insects we have named this gene *D. citri_sloppy paired*. Independent *de novo* transcriptome sequence was used to confirm the final Slp sequence (Table 2).

runt

In insects there are four runt domain containing genes 1) *run* 2) *Runt related A (RunxA)* 3) *Runt related B (RunxB)* and 4) *lozenge (lz)* which are typically found clustered on one chromosome. The order and orientation of these four genes is well conserved across insects [69] (Figure 5). In addition to the runt domain, RDPs also contain a C terminal pentapeptide VWRPY sequence. We were able to annotate a full length *runt*, *RunxA* and *RunxB* and a partial *lz* gene. It appears that the cluster is intact, with all four RDP genes identified in their expected order (Figure 5). *lz* appears to be transcribed in the opposite direction compared to other insects, but it is possible that this is due to local misassembly.

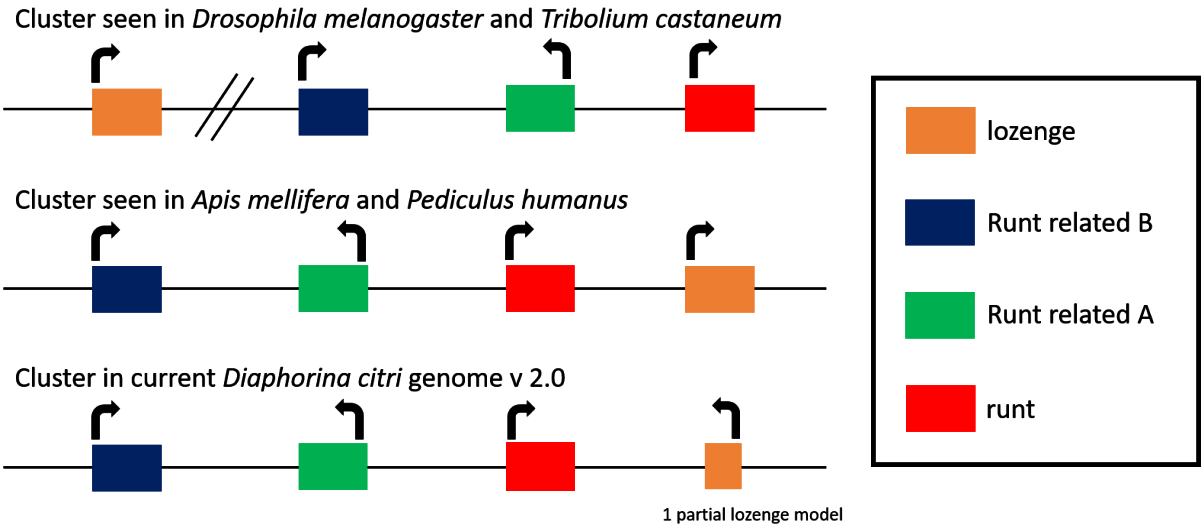

Figure 5: Runt Cluster in Insects. Cluster information from other insects was obtained from [69] *The runt cluster in Drosophila melanogaster and Tribolium castaneum have three genes in a core cluster with lozenge separated, but on the same chromosome, distal to runt. The runt cluster in Apis mellifera and Pediculus humanus have all four genes clustered together with lozenge proximal to runt. The runt cluster in Diaphorina citri most closely follows the pattern seen in Apis mellifera and Pediculus humanus.*

even skipped

One *eve* gene has been identified and manually annotated in the *D. citri* genome v3.0 (Table 4). This annotation is supported by single read PacBio IsoSeq sequence and transcripts from a *de novo* transcriptome (Table 2). In *Drosophila eve* the coding region is contained within two exons and the conserved homeodomain is found in its entirety in the second exon. In amphioxus [70], the polychaete annelid *Capitella teleta* [71], and Tribolium [72] eve genes contain a second intron within the homeodomain resulting in a three exon gene. The homeodomain of the *D. citri eve* gene is split between the second and third coding exons as seen in amphioxus, *C. teleta* and Tribolium.

hairy

In a previous version of the genome a partial gene model for *h* was annotated. With the help of PacBio IsoSeq evidence, the MCOT transcriptome and ortholog sequences we were able to correct the previous gene model so that a full length protein has now been annotated (Table 2) in the *D. citri* genome v3.0.

odd paired

A single copy of the Zic family member *opa* [73] was identified and manually annotated in the *D. citri* genome v3.0 (Table 4). It was also found as a *de novo* model in the MCOT transcriptome (Table 2).

Segment polarity genes

Segment polarity genes include members of the Wnt and Hedgehog signalling pathways. The manual annotation of these genes in the *D. citri* genome v3.0 has been reported in independent reports on signalling pathways. Here we report on the manual annotation of the segment polarity genes *gooseberry* (*gsb*) and *engrailed* (*en*), which are targets of the Wnt and Hedgehog signalling pathways, and the related genes *gooseberry-neuro* (*gsb-n*) and *invected* (inv). One ortholog for the genes *engrailed* (*en*), *gooseberry* (*gsb*) and *gooseberry-neuro* (*gsb-n*) was found in the *D. citri* v3.0 genome. An ortholog for *invected* (*inv*) was not able to be identified in the current genome assembly.

Segment polarity		*Drosophila melanogaster*	*Apis mellifera*	*Tribolium castaneum*	*Acyrthosiphon pisum*	*Diaphorina citri*
	gooseberry	1	1	1	1	1
	engrailed	2	2	2	2	1

Table 5: Segment polarity gene ortholog number. Tribolium castaneum engrailed gene numbers [6,74] *and Acyrthosiphon pisum engrailed and gooseberry gene numbers* [7] *were acquired through genome publications. Ortholog numbers for Drosophila melanogaster, Apis mellifera, Tribolium castaneum gooseberry and gooseberry-neuro and Acythosiphon pisum gooseberry-neuro were determined using NCBI's gene database. Diaphorina citri*

Engrailed family

The engrailed gene family includes two closely related homeodomain-containing transcription factors, *en* and *inv*. In *Drosophila*, they both share a regulatory region and are functionally redundant [75]. One copy of *en* was found in the *D. citri* v3.0 genome (Table 5). The annotated model is supported by *de novo* transcriptome data (Table 2). MCOT model predictions and RNAseq data suggests the presence of two different isoforms (Figure 6). An ortholog for *inv* was not found within the *D. citri* v3.0 genome. The En protein is necessary for the establishment of posterior cell identity within a segment [76]. *en* mutations can cause mild to severe alterations to phenotype and can even prove lethal in developing embryos whereas loss of *inv* has no observable effect on phenotype [75]. Although *inv* appears to be redundant to *en* and could theoretically have been lost through evolution of *D. citri*, many hexapods have both *en* and *inv*. A future version of the *D. citri* genome could reveal the presence of *inv*.

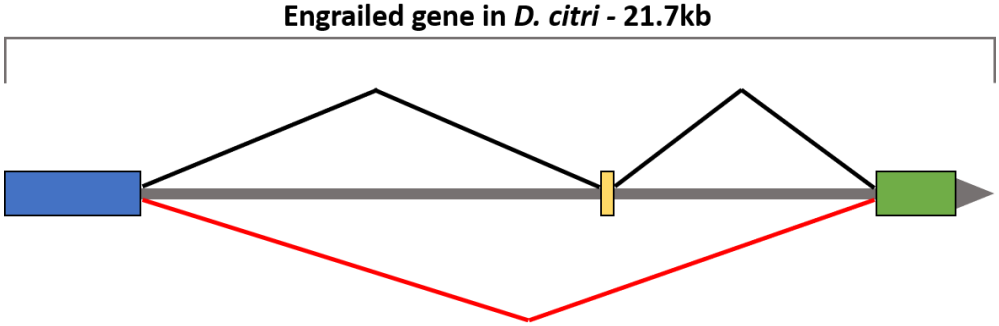

Figure 6: Alternative splicing of engrailed in Diaphorina citri resulting in two different isoforms.

Gooseberry

The Pax3/7 gene *gsb* encodes a transcription factor that is necessary for establishing a wingless-gooseberry autoregulatory loop that is independent of hedgehog signaling [77]. *gsb* expression has also been associated with determining the developmental fate of certain neuroblasts by acting as a selector gene for the homologous *gsb-n* gene [78]. The *gsb* ortholog that has been manually annotated in *D. citri* is characterized by a paired-type homeodomain and a paired domain. The annotation is supported by IsoSeq sequence evidence and *de novo* transcriptome data (Table 2). Along with *prd* and

gsb, *gsb-n* is a Pax3/7 gene. *gsb-n* is found near the homologous *gsb* gene in an area called the gooseberry locus. This two-gene gene cluster appears conserved in insects however, within *D. citri* and other hemipterans, the two genes are transcribed in the same direction whereas in *Drosophila* and *Tribolium* they are transcribed in opposite directions. Phylogenetic analysis alone made it difficult to resolve the true orthology the Pax 3/7 genes (*prd*, *gsb*, *gsb-n*) found in *D. citri* (Figure 7). However, the positional information of *gsb-n* in other insects helped determine proper orthology in the *D. citri* v3.0 genome. The manually annotated model for *gsb-n* is supported by IsoSeq and *de novo* transcriptome data (Table 2).

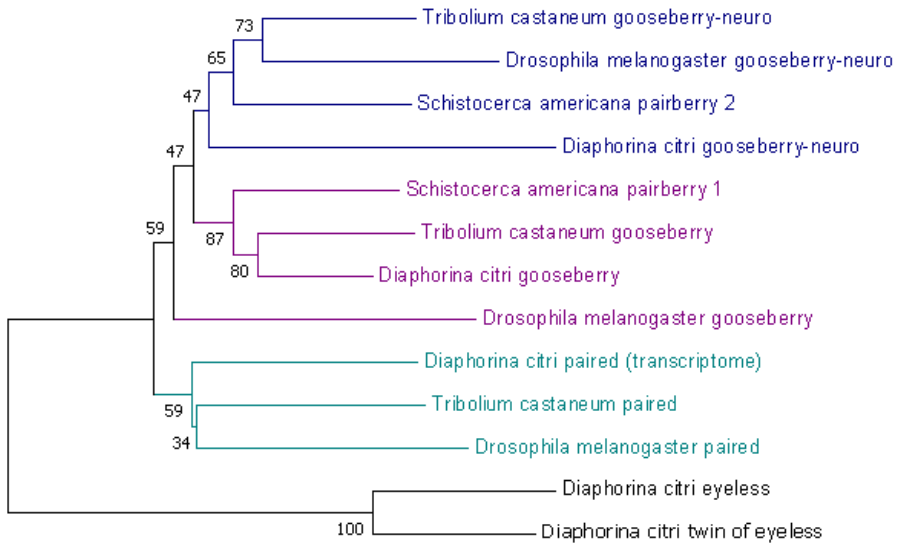

Figure 7: Neighbor-joining tree of gsb-n, gsb and prd homologs. Full length protein sequences were used for phylogenetic analysis except in the case of pairberry 1 and pairberry 2 where only partials were available. In one case (prd) transcriptome sequence was used due to an incomplete genome annotation. D. citri has three Pax 3 genes. We have named them gsb-n, gsb, and prd despite the fact that bootstrap values are low because of their position in the genome. Navy genes show more similarity to gsb-n. Purple sequences show similarity to gsb and teal genes show similarity to prd. Black genes ey and toy are Pax6 genes which are acting as an outgroup for this analysis.

1. Nüsslein-volhard C, Wieschaus E. Mutations affecting segment number and polarity in drosophila. Nature. 1980;287:795–801.

2. Akam M. The molecular basis for metameric pattern in the Drosophila embryo. Development. 1987. p. 1–22.

3. Scott MP, Carroll SB. The segmentation and homeotic gene network in early Drosophila development. Cell. 1987. p. 689–98.

4. Clark E, Peel AD, Akam M. Arthropod segmentation. Dev. Company of Biologists Ltd; 2019.

5. Dearden PK, Wilson MJ, Sablan L, Osborne PW, Havler M, McNaughton E, et al. Patterns of conservation and change in honey bee developmental genes. Genome Res [Internet]. 2006 [cited 2019 Dec 5];16:1376–84. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17065607

6. Richards S, Gibbs RA, Weinstock GM, Brown SJ, Denell R, Beeman RW, et al. The genome of the model beetle and pest Tribolium castaneum. Nature. 2008;452.

7. Shigenobu S, Bickel RD, Brisson JA, Butts T, Chang C, Christiaens O, et al. Comprehensive survey of developmental genes in the pea aphid, Acyrthosiphon pisum: frequent lineage-specific duplications and losses of developmental genes. Insect Mol Biol [Internet]. 2010 [cited 2017 Oct 10];19:47–62. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20482639

8. Saha S, Hosmani PS, Villalobos-Ayala K, Miller S, Shippy T, Flores M, et al. Improved annotation of the insect vector of citrus greening disease: biocuration by a diverse genomics community. Database (Oxford). 2017;2017.

9. Gross I, Georgel P, Oertel-Buchheit P, Schnarr M, Reichhart JM. Dorsal-B, a splice variant of the Drosophila factor Dorsal, is a novel Rel/NF-κB transcriptional activator. Gene. 1999;228:233–42.

10. Sang WS, Kokoza V, Bian G, Cheon HM, Yu JK, Raikhel AS. REL1, a homologue of Drosophila dorsal, regulates toll antifungal immune pathway in the female mosquito Aedes aegypti. J Biol Chem. 2005;280:16499–507.

11. Zou Z, Evans JD, Lu Z, Zhao P, Williams M, Sumathipala N, et al. Comparative genomic analysis of the Tribolium immune system. Genome Biol [Internet]. 2007 [cited 2019 Dec 5];8:R177. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17727709

12. Juhn J, Marinotti O, Calvo E, James AA. Gene structure and expression of nanos (nos) and oskar (osk) orthologues of the vector mosquito, Culex quinquefasciatus. Insect Mol Biol [Internet]. 2008 [cited 2019 Dec 5];17:545–52. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18828840

13. Werren JH, Richards S, Desjardins CA, Niehuis O, Gadau J, Colbourne JK, et al. Functional and evolutionary insights from the genomes of three parasitoid nasonia species. Science (80-). 2010;327:343–8.

14. Saffman EE, Lasko P. Germline development in vertebrates and invertebrates. Cell. Mol. Life Sci. 1999. p. 1141–63.

15. Stauber M, Prell A, Schmidt-Ott U. A single Hox3 gene with composite bicoid and zerknullt expression characteristics in non-Cyclorrhaphan flies. Proc Natl Acad Sci U S A [Internet]. 2002 [cited 2019 Dec 5];99:274–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11773616

16. McGregor AP. How to get ahead: the origin, evolution and function of bicoid. Bioessays [Internet]. 2005 [cited 2019 Dec 5];27:904–13. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16108065

17. Huang T-Y, Cook CE, Davis GK, Shigenobu S, Chen RP-Y, Chang C-C. Anterior development in the parthenogenetic and viviparous form of the pea aphid, Acyrthosiphon pisum: hunchback and orthodenticle expression. Insect Mol Biol [Internet]. 2010 [cited 2019 Dec 5];19 Suppl 2:75–85. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20482641

18. Weinstock GM, Robinson GE, Gibbs RA, Worley KC, Evans JD, Maleszka R, et al. Insights into social insects from the genome of the honeybee Apis mellifera. Nature. 2006;443.

19. Lehmann R, Nüsslein-Volhard C. Abdominal segmentation, pole cell formation, and embryonic polarity require the localized activity of oskar, a maternal gene in drosophila. Cell. 1986;47:141–52.

20. Juhn J, James AA. oskar gene expression in the vector mosquitoes, Anopheles gambiae and Aedes aegypti. Insect Mol Biol [Internet]. 2006 [cited 2019 Dec 5];15:363–72. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16756555

21. Xia Q, Zhou Z, Lu C, Cheng D, Dai F, Li B, et al. A draft sequence for the genome of the domesticated silkworm (Bombyx mori). Science (80-). 2004;306:1937–40.

22. Lynch JA, Ozüak O, Khila A, Abouheif E, Desplan C, Roth S. The phylogenetic origin of oskar coincided with the origin of maternally provisioned germ plasm and pole cells at the base of the Holometabola. PLoS Genet [Internet]. 2011 [cited 2019 Dec 5];7:e1002029. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21552321

23. Gui H, Li ML, Tsai CC. A tale of tailless. Dev. Neurosci. 2011. p. 1–13.

24. Naggan Perl T, Schmid BGM, Schwirz J, Chipman AD. The evolution of the knirps family of transcription factors in arthropods. Mol Biol Evol. 2013;30:1348–57.

25. Lavore A, Pagola L, Esponda-Behrens N, Rivera-Pomar R. The gap gene giant of Rhodnius prolixus is maternally expressed and required for proper head and abdomen formation. Dev Biol [Internet]. 2012 [cited 2019 Dec 5];361:147–55. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21763688

26. Lavore A, Esponda-Behrens N, Pagola L, Rivera-Pomar R. The gap gene Krüppel of Rhodnius prolixus is required for segmentation and for repression of the homeotic gene sex comb-reduced. Dev Biol [Internet]. 2014 [cited 2019 Dec 5];387:121–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24406318

27. Nakao H. Analyses of interactions among pair-rule genes and the gap gene Krüppel in Bombyx segmentation. Dev Biol [Internet]. 2015 [cited 2019 Dec 5];405:149–57. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26102481

28. Crombach A, García-Solache MA, Jaeger J. Evolution of early development in dipterans: reverse-engineering the gap gene network in the moth midge Clogmia albipunctata (Psychodidae). Biosystems [Internet]. 2014 [cited 2019 Dec 5];123:74–85. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24911671

29. Blechert O, Douglas D, Baumgartner S. Conserved function of the Krüppel gap gene in the blowfly Lucilia sericata, despite anterior shift of expression. Insect Mol Biol [Internet]. 2011 [cited 2019 Dec 5];20:257–65. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21166911

30. Wilson MJ, Havler M, Dearden PK. Giant, Krüppel, and caudal act as gap genes with extensive roles in patterning the honeybee embryo. Dev Biol [Internet]. 2010 [cited 2019 Dec 5];339:200–11. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20026025

31. Mito T, Okamoto H, Shinahara W, Shinmyo Y, Miyawaki K, Ohuchi H, et al. Krüppel acts as a gap gene regulating expression of hunchback and even-skipped in the intermediate germ cricket Gryllus

bimaculatus. Dev Biol [Internet]. 2006 [cited 2019 Dec 5];294:471–81. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16616119

32. Cerny AC, Bucher G, Schröder R, Klingler M. Breakdown of abdominal patterning in the Tribolium Kruppel mutant jaws. Development [Internet]. 2005 [cited 2019 Dec 5];132:5353–63. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16280347

33. Liu PZ, Kaufman TC. Krüppel is a gap gene in the intermediate germband insect Oncopeltus fasciatus and is required for development of both blastoderm and germband-derived segments. Development. 2004;131:4567–79.

34. Preiss A, Rosenberg UB, Kienlin A, Seifert E, Jäckle H. Molecular genetics of Krüppel, a gene required for segmentation of the Drosophila embryo. Nature. 1985;313:27–32.

35. Tautz D. Regulation of the Drosophila segmentation gene hunchback by two maternal morphogenetic centres. Nature. 1988;332:281–4.

36. Liu PZ, Kaufman TC. Hunchback is required for suppression of abdominal identity, and for proper germband growth and segmentation in the intermediate germband insect Oncopeltus fasciatus. Development. 2004;131:1515–27.

37. Marques-Souza H, Aranda M, Tautz D. Delimiting the conserved features of hunchback function for the trunk organization of insects. Development [Internet]. 2008 [cited 2019 Dec 5];135:881–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18216167

38. Mao J, Liu C, Zeng F. Hunchback is required for abdominal identity suppression and germband growth in the parthenogenetic embryogenesis of the pea aphid, Acyrthosiphon pisum. Arch Insect Biochem Physiol. 2013;84:209–21.

39. Kambadur R, Koizumi K, Stivers C, Nagle J, Poole SJ, Odenwald WF. Regulation of POU genes by castor and hunchback establishes layered compartments in the Drosophila CNS. Genes Dev. Cold Spring Harbor Laboratory Press; 1998;12:246–60.

40. Novotny T, Eiselt R, Urban J. Hunchback is required for the specification of the early sublineage of neuroblast 7-3 in the Drosophila central nervous system. Development. 2002. p. 1027–36.

41. Pinnell J, Lindeman PS, Colavito S, Lowe C, Savage RM. The divergent roles of the segmentation gene hunchback. Integr Comp Biol [Internet]. 2006 [cited 2019 Dec 5];46:519–32. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21672763

42. Mao J, Zeng F. Feeding-Based RNA Intereference of a Gap Gene Is Lethal to the Pea Aphid, Acyrthosiphon pisum. PLoS One. 2012;7.

43. Weigel D, Jürgens G, Klingler M, Jäckle H. Two gap genes mediate maternal terminal pattern information in Drosophila. Science (80-). 1990;248:495–8.

44. García-Solache M, Jaeger J, Akam M. A systematic analysis of the gap gene system in the moth midge Clogmia albipunctata. Dev Biol [Internet]. 2010 [cited 2019 Dec 5];344:306–18. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20433825

45. Kittelmann S, Ulrich J, Posnien N, Bucher G. Changes in anterior head patterning underlie the evolution of long germ embryogenesis. Dev Biol [Internet]. 2013 [cited 2019 Dec 5];374:174–84. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23201022

46. Wimmer EA, Frommer G, Purnell BA, Jäckle H. buttonhead and D-Sp1: a novel Drosophila gene pair. Mech Dev [Internet]. 1996 [cited 2019 Dec 5];59:53–62. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8892232

47. Schöck F, Purnell BA, Wimmer EA, Jäckle H. Common and diverged functions of the Drosophila gene pair D-Sp1 and buttonhead. Mech Dev. 1999;89:125–32.

48. Schaeper ND, Prpic N-M, Wimmer EA. A clustered set of three Sp-family genes is ancestral in the Metazoa: evidence from sequence analysis, protein domain structure, developmental expression patterns and chromosomal location. BMC Evol Biol [Internet]. 2010 [cited 2019 Dec 5];10:88. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20353601

49. Birkan M, Schaeper ND, Chipman AD. Early patterning and blastodermal fate map of the head in the milkweed bug Oncopeltus fasciatus. Evol Dev [Internet]. [cited 2019 Dec 5];13:436–47. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23016905

50. Cohen SM, Jürgens G. Mediation of Drosophila head development by gap-like segmentation genes. Nature. 1990;346:482–5.

51. Schinko JB, Kreuzer N, Offen N, Posnien N, Wimmer EA, Bucher G. Divergent functions of orthodenticle, empty spiracles and buttonhead in early head patterning of the beetle Tribolium castaneum (Coleoptera). Dev Biol. Academic Press Inc.; 2008;317:600–13.

52. Simonnet F, Célérier ML, Quéinnec E. Orthodenticle and empty spiracles genes are expressed in a segmental pattern in chelicerates. Dev Genes Evol. 2006;216:467–80.

53. Hunnekuhl VS, Akam M. Formation and subdivision of the head field in the centipede Strigamia maritima, as revealed by the expression of head gap gene orthologues and hedgehog dynamics. Evodevo [Internet]. BioMed Central Ltd.; 2017 [cited 2019 Dec 5];8:18. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29075435

54. Janssen R, Budd GE, Damen WGM. Gene expression suggests conserved mechanisms patterning the heads of insects and myriapods. Dev Biol [Internet]. 2011 [cited 2019 Dec 5];357:64–72. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21658375

55. Li Y, Brown SJ, Hausdorf B, Tautz D, Denell RE, Finkelstein R. Two orthodenticle-related genes in the short-germ beetle Tribolium castaneum. Dev Genes Evol [Internet]. 1996 [cited 2019 Dec 5];206:35–45. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24173395

56. Lynch JA, Brent AE, Leaf DS, Pultz MA, Desplan C. Localized maternal orthodenticle patterns anterior and posterior in the long germ wasp Nasonia. Nature [Internet]. 2006 [cited 2019 Dec 5];439:728–32. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16467838

57. Browne WE, Schmid BGM, Wimmer EA, Martindale MQ. Expression of otd orthologs in the amphipod crustacean, Parhyale hawaiensis. Dev Genes Evol [Internet]. 2006 [cited 2019 Dec 5];216:581–95. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16773341

58. Pechmann M, McGregor AP, Schwager EE, Feitosa NM, Damen WGM. Dynamic gene expression is required for anterior regionalization in a spider. Proc Natl Acad Sci U S A [Internet]. 2009 [cited 2019 Dec 5];106:1468–72. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19147844

59. Mohler J, Mahaffey JW, Deutsch E, Vani K. Control of Drosophila head segment identity by the bZIP homeotic gene cnc. Development. 1995;121:237–47.

60. Flybase [Internet]. [cited 2019 Dec 2]. Available from: http://flybase.org

61. Crozatier M, Valle D, Dubois L, Ibnsouda S, Vincent A. Head versus trunk patterning in the Drosophila embryo; collier requirement for formation of the intercalary segment. Development. 1999;126:4385–94.

62. Davis GK, D'Alessio JA, Patel NH. Pax3/7 genes reveal conservation and divergence in the arthropod segmentation hierarchy. Dev Biol [Internet]. 2005 [cited 2019 Dec 5];285:169–84. Available

from: http://www.ncbi.nlm.nih.gov/pubmed/16083872

63. Balczarek KA, Lai ZC, Kumar S. Evolution of functional diversification of the paired box (Pax) DNA-binding domains. Mol Biol Evol [Internet]. 1997 [cited 2019 Dec 5];14:829–42. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9254921

64. Hart MC, Wang L, Coulter DE. Comparison of the structure and expression of odd-skipped and two related genes that encode a new family of zinc finger proteins in Drosophila. Genetics [Internet]. 1996 [cited 2019 Dec 5];144:171–82. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8878683

65. Grossniklaus U, Pearson RK, Gehring WJ. The Drosophila sloppy paired locus encodes two proteins involved in segmentation that show homology to mammalian transcription factors. Genes Dev. 1992;6:1030–51.

66. Damen WGM, Janssen R, Prpic N-M. Pair rule gene orthologs in spider segmentation. Evol Dev [Internet]. [cited 2019 Dec 5];7:618–28. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16336415

67. Choe CP, Miller SC, Brown SJ. A pair-rule gene circuit defines segments sequentially in the short-term insect Tribolium castaneum. Proc Natl Acad Sci U S A. 2006;103:6560–4.

68. Reding K, Chen M, Lu Y, Cheatle Jarvela AM, Pick L. Shifting roles of Drosophila pair-rule gene orthologs: segmental expression and function in the milkweed bug Oncopeltus fasciatus. Development [Internet]. 2019 [cited 2019 Dec 5];146. Available from: http://www.ncbi.nlm.nih.gov/pubmed/31444220

69. Duncan EJ, Wilson MJ, Smith JM, Dearden PK. Evolutionary origin and genomic organisation of runt-domain containing genes in arthropods. BMC Genomics [Internet]. 2008 [cited 2019 Dec 5];9:558. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19032778

70. Ferrier DE, Minguillón C, Cebrián C, Garcia-Fernàndez J. Amphioxus Evx genes: implications for the evolution of the Midbrain-Hindbrain Boundary and the chordate tailbud. Dev Biol [Internet]. 2001 [cited 2019 Dec 5];237:270–81. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11543613

71. Seaver EC, Yamaguchi E, Richards GS, Meyer NP. Expression of the pair-rule gene homologs runt, Pax3/7, even-skipped-1 and even-skipped-2 during larval and juvenile development of the polychaete annelid Capitella teleta does not support a role in segmentation. Evodevo [Internet]. 2012 [cited 2019 Dec 5];3:8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22510249

72. Brown SJ, Parrish JK, Beeman RW, Denell RE. Molecular characterization and embryonic expression of the even-skipped ortholog of Tribolium castaneum. Mech Dev. 1997;61:165–73.

73. Hursh DA, Stultz BG. Odd-Paired: The drosophila Zic gene. Adv Exp Med Biol. Springer New York LLC; 2018. p. 41–58.

74. Peel AD, Telford MJ, Akam M. The evolution of hexapod engrailed-family genes: evidence for conservation and concerted evolution. Proceedings Biol Sci [Internet]. 2006 [cited 2019 Dec 5];273:1733–42. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16790405

75. Gustavson E, Goldsborough AS, Ali Z, Kornberg TB. The Drosophila engrailed and invected Genes: Partners in regulation, expression and function. Genetics. 1996;142:893–906.

76. Kornberg T, Sidén I, O'Farrell P, Simon M. The engrailed locus of drosophila: In situ localization of transcripts reveals compartment-specific expression. Cell. 1985;40:45–53.

77. Li X, Gutjahr T, Noll M. Separable regulatory elements mediate the establishment and maintenance of cell states by the Drosophila segment-polarity gene gooseberry. EMBO J [Internet]. 1993 [cited 2019 Dec 5];12:1427–36. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8096813

78. Zhang Y, Ungar A, Fresquez C, Holmgren R. Ectopic expression of either the Drosophila gooseberry-distal or proximal gene causes alterations of cell fate in the epidermis and central nervous system. Development. 1994;120:1151–61.