Annotation of segmentation pathway genes in the Asian citrus psyllid, *Diaphorina citri*

Sherry Miller^{1,2}, Teresa D. Shippy¹, Prashant S. Hosmani³, Mirella Flores-Gonzalez³, Lukas A. Mueller³, Wayne B. Hunter⁴, Susan J. Brown¹, Tom D'Elia⁵ and Surya Saha^{3,6,*}

¹ Division of Biology, Kansas State University, Manhattan, KS 66506, USA

² Allen County Community College, Burlingame, KS 66413, USA

³ Boyce Thompson Institute, Ithaca, NY 14853, USA

⁴ USDA-ARS, U.S. Horticultural Research Laboratory, Fort Pierce, FL 34945, USA

⁵ Indian River State College, Fort Pierce, FL 34981, USA

⁶ Animal and Comparative Biomedical Sciences, University of Arizona, Tucson, AZ 85721, USA

ABSTRACT

Insects have a segmented body plan that is established during embryogenesis when the anterior–posterior (A–P) axis is divided into repeated units by a cascade of gene expression. The cascade is initiated by protein gradients created by translation of maternally provided mRNAs, localized at the anterior and posterior poles of the embryo. Combinations of these proteins activate specific gap genes to divide the embryo into distinct regions along the anterior–posterior axis. Gap genes then activate pair-rule genes, which are usually expressed in parts of every other segment. The pair-rule genes, in turn, activate expression of segment polarity genes in a portion of each segment. The segmentation genes are generally conserved among insects, although there is considerable variation in how they are deployed. We annotated 25 segmentation gene homologs in the Asian citrus psyllid, *Diaphorina citri*. Most of the genes expected to be present in *D. citri* based on their phylogenetic distribution in other insects were identified and annotated. Two exceptions were *eagle* and *invected*, which are present in at least some hemipterans, but were not found in *D. citri*. Many of the segmentation pathway genes are likely to be essential for *D. citri* development, and thus they may be useful targets for gene-based pest control methods.

Submitted: 24 December 2020
Accepted: 02 July 2021
Published: 08 July 2021

Subjects Genetics and Genomics, Animal Genetics, Bioinformatics

* Corresponding author. E-mail: suryasaha@cornell.edu

Published by GigaScience Press.

Preprint submitted at <https://doi.org/10.1101/2020.12.24.424320>

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Gigabyte, 2021, 1–12

DATA DESCRIPTION

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. The molecular mechanisms involved in this process were first elucidated by large-scale developmental mutant screens in the insect model *Drosophila melanogaster* [1–5]. In *Drosophila*, segmentation begins with cytoplasmic inheritance of mRNAs that are maternally produced and provided to the oocyte. The products of these maternal-effect genes create gradients that define positional information within the embryo and 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. Together, they activate the

Figure 1. Protocol for psyllid genome community curation [14]. <https://www.protocols.io/widgets/doi?uri=dx.doi.org/10.17504/protocols.io.bniimcce>

expression of segment polarity genes, which are expressed in every segment of the developing embryo. Comparative studies in diverse arthropod species have shown that some aspects of the segmentation pathway are highly conserved, while other aspects have undergone evolutionary change [6]. The hemipteran insects that have been examined seem to employ a particularly divergent method of segmentation. Most strikingly, the pair-rule genes, which are usually considered to be the most conserved portion of the segmentation pathway among insects, have lost their pair-rule expression and function and are expressed segmentally in at least some hemipterans [7–9].

Context

As part of a community genome annotation project, we have annotated 25 homologs of *Drosophila* segmentation pathway genes in the genome of the hemipteran agricultural pest, *Diaphorina citri* (NCBI:txid121845), also known as the Asian citrus psyllid. A few segmentation genes were identified in a previous version of the *D. citri* genome [10]. However, many of those genes were incomplete because of genome assembly errors. The current genome version (v3) is much higher quality [11], allowing us to annotate genes with much higher accuracy and confidence. Except for *eagle* and *invected*, which appear to be missing from the *D. citri* v3 genome, we annotated all of the segmentation genes expected to be present in *D. citri*. Our annotations pave the way for future work aimed at understanding the expression and function of these genes during *D. citri* segmentation and the identification of essential genes that could be used as insect control targets.

METHODS

Segmentation gene orthologs in *D. citri* genome v3 [11] were identified by BLAST (RRID:SCR_004870) search and confirmed by reciprocal BLAST. Manual annotation was performed in Apollo 2.1.0 (RRID:SCR_001936) [12] using available evidence, such as RNA-Seq reads, Iso-Seq transcripts and *de novo*-assembled transcripts. Multiple alignment and phylogenetic analysis were done in MEGA X (RRID:SCR_000667) [13]. More details of the annotation process are described in a [protocols.io](https://www.protocols.io) protocol (Figure 1) [14].

Table 1. Annotated *D. citri* segmentation gene orthologs.

Group	Gene/Isoform	OGSv3 ID	Complete model	Evidence supporting annotation			
				MCOT	Iso-Seq	RNA-Seq	Ortholog
Maternal Effect							
	<i>caudal</i>	Dcitr06g04620.1.4	X	MCOT08625.0.OO	X		
	<i>dorsal</i>	Dcitr02g07710.1.1	X	MCOT04961.1.CO	X	X	X
		Dcitr02g07710.1.2		MCOT04961.3.CO			
	<i>nanos</i>	Dcitr00g11810.1.1		MCOT03924.0.CT			X
	<i>TGFalpha</i>	Dcitr06g02470.1.1	X	MCOT20343.1.CO	X		
Gap							
	<i>tailless</i>	Dcitr01g16840.1.1	X				X
	<i>knirps related 1</i>	Dcitr10g03830.1.1	X		X	X	X
	<i>knirps related 2</i>	Dcitr10g03860.1.1	X	MCOT03238.1.CT			X
	<i>knirps related 3</i>	Dcitr10g03875.1.1	X			X	X
	<i>Kruppel</i> [†]	Dcitr00g04150.1.1		MCOT03100.0.CO		X	X
		Dcitr00g04150.1.2					
	<i>hunchback</i>	Dcitr09g01780.1.1	X		X	X	
		Dcitr09g01780.1.2					
	<i>huckebein</i>	Dcitr04g11170.1.1	X	MCOT23091.2.CT		X	
	<i>orthodenticle</i>	Dcitr04g16960.1.1	X	MCOT05861.1.CO		X	X
	<i>empty spiracles</i>	Dcitr09g06330.1.1	X	MCOT04329.0.CO			
	<i>cap-n-collar</i>	Dcitr03g12850.1.1	X	MCOT01547.1.CT			
	<i>collier</i>	Dcitr03g01400.1.1	X	MCOT16351.0.CO		X	
		Dcitr03g01400.1.2		MCOT09939.0.OO			
Pair-rule							
	<i>paired</i>	Dcitr01g09360.1.1	X	MCOT01278.2.CO			X
	<i>odd skipped</i>	Dcitr01g20150.1.1	X	MCOT15683.0.CO			
	<i>sloppy paired</i>	Dcitr02g08120.1.1	X	MCOT19895.0.CT			
	<i>runt</i>	Dcitr01g07300.1.1	X	MCOT14329.0.CT			X
	<i>even skipped</i>	Dcitr08g10250.1.1	X	MCOT05103.0.CT	X		
	<i>hairy</i>	Dcitr02g06890.1.1	X	MCOT00149.0.CT	X		X
	<i>odd paired</i>	Dcitr09g07980.1.1	X	MCOT12397.2.CT			
				MCOT03041.1.CT			
Segment polarity							
	<i>gooseberry</i>	Dcitr01g18500.1.1	X	MCOT03447.1.CO	X	X	
	<i>engrailed</i>	Dcitr08g03480.1.1	X	MCOT21494.0.CO		X	
		Dcitr08g03480.1.2		MCOT01333.0.CT			
Other related genes							
	<i>Runt related A</i>	Dcitr01g07290.1.1	X	MCOT01714.2.CT			X
	<i>Runt related B</i>	Dcitr01g07260.1.1	X	MCOT22437.2.CT			X
	<i>lozenge</i>	Dcitr01g07310.1.1			X		X
	<i>sister of odd and bowl</i>	Dcitr01g20160.1.1	X	MCOT11314.0.MO			
	<i>brother of odd with entrails limited</i>	Dcitr01g11160.1.1	X		X		
	<i>gooseberry neuro</i>	Dcitr01g18520.1.1	X	MCOT03447.2.CT	X	X	X

Each manually annotated gene has been assigned an OGSv3 gene identifier. Genes marked as complete models have a complete coding region. Evidence used for manual annotation was also recorded. Maker, Cufflinks, Oases Trinity (MCOT) [10] and Iso-Seq [11] indicate use of these genome-independent transcriptomes to validate the annotation. RNA-Seq indicates use of mapped RNA-Seq reads, while orthologs indicate comparison with related proteins was essential for the annotation. Additional information on specific evidence types is included in [14]. [†] denotes genes that were complete in OGSv2 but are partial in OGSv3.

RESULTS AND DISCUSSION

We searched the *D. citri* v3 genome for orthologs of genes known to be involved in segmentation in *Drosophila* [15] (Table 1). We then used available evidence to manually annotate the genes that were present [11, 14] (Tables 1, 2). Most manual annotations were straightforward and performed using our workflow (see Methods), so only those requiring additional explanation are described in detail here.

Table 2. Segmentation gene ortholog number.

	<i>Drosophila melanogaster</i>	<i>Apis mellifera</i>	<i>Tribolium castaneum</i>	<i>Acyrtosiphon pisum</i>	<i>Diaphorina citri</i>
Maternal Effect					
<i>caudal</i>	1	1	1	1	1
<i>dorsal</i> family (<i>dorsal</i> , <i>Dorsal-related immunity factor</i>)	2	1	2	2	1
<i>nanos</i>	1	1	1	4	1
<i>bicoid</i>	1	0	0	0	0
<i>oskar</i>	1	0	0	0	0
TGF α ligand (<i>gurken</i> , <i>Keren</i> , <i>spitz</i>)	3	1	1	1	1
Gap					
<i>tailless</i>	1	1	1	1	1
<i>knirps</i> family (<i>knirps</i> , <i>knirps-like</i> , <i>eagle</i>)	3	3	2	3	3
<i>giant</i>	1	1	1	0	0
<i>Kruppel</i>	1	1	1	1	1
<i>hunchback</i>	1	1	1	1	1
<i>huckebein</i>	1	1	1	0	1
<i>orthodenticle</i>	1	2	2	1	1
<i>buttonhead</i>	1	1	1	0	0
<i>empty spiracles</i>	1	1	1	1	1
<i>cap-n-collar</i>	1	1	1	1	1
<i>collier</i>	1	1	1	1	1
Pair-rule					
<i>paired</i>	1	1	1	1	1
<i>odd skipped</i>	1	1	1	1	1
<i>sloppy paired 1/sloppy paired 2</i>	2	1	1	1	1
<i>runt</i>	1	1	1	1	1
<i>even skipped</i>	1	1	1	1	1
<i>hairy</i>	1	1	1	1	1
<i>odd paired</i>	1	1	1	1	1
Segment polarity					
<i>gooseberry</i>	1	1	1	1	1
<i>engrailed</i> family (<i>engrailed</i> , <i>invected</i>)	2	2	2	2	1

The *Drosophila melanogaster* numbers were determined from Flybase (RRID:SCR_006549) [16]. Ortholog numbers for *Apis mellifera* [17], *Tribolium castaneum* [18] and *Acyrtosiphon pisum* [19] are based on genome publications or NCBI (RRID:SCR_006472). *Diaphorina citri* ortholog numbers represent our final manual annotation.

Maternal effect genes

One-to-one orthologs of *caudal* (*cad*), *dorsal* (*dl*), and *nanos* (*nos*) were found in the *D. citri* v3 genome. *dl* was previously annotated in the *D. citri* genome v1.1 because of its role in innate immunity [10]. Here, we annotated a second isoform of *dl* (Table 1).

We also identified a single TGF- α ligand-encoding gene in *D. citri*. In *Drosophila*, there are three TGF- α ligand paralogs (*grk*, *Krn*, *spi*), but in many other insects, only one TGF- α ligand has been identified [17–21]. Orthologs for *bicoid* (*bcd*) and *oskar* (*osk*) were not found in *D. citri* (Table 2), which is consistent with the previously described phylogenetic distribution of these genes [22].

Gap genes

One-to-one orthologs of the gap genes *tailless* (*tll*), *Kruppel* (*Kr*), *hunchback* (*hb*), *huckebein* (*hkb*), *empty spiracles* (*ems*), *cap-n-collar* (*cnc*), and *collier* (*col*) were identified and annotated in the *D. citri* v3 genome (Tables 1, 2). Except for *hkb*, which was reported

missing in *Acyrtosiphon pisum* (*A. pisum*) [19], conservation of these genes was expected (Table 2). Two *knirps-related* (*knrl*) genes and two *orthodenticle* (*otd*) homologs were annotated and are discussed in more detail below. *giant* (*gt*) and *buttonhead* (*btd*) appear to be absent in the v3 *D. citri* genome assembly (Table 2).

Phylogenetics of the *knirps* family suggests that a single ancestral gene duplicated early in the insect lineage, producing two paralogs called *knirps-related* (*knrl*) and *eagle* (*eg*) [23]. Subsequent duplications have occurred in various insect lineages. A duplication in the lineage leading to *Drosophila* resulted in the paralogs *knirps* and *knirps-like* (also called *knirps-related*) [23]. A separate duplication of *knrl* seems to have occurred in the hemipteroid lineage, leading to three *knirps* family genes (two *knrl* and one *eg*) in most hemipterans [23]. In the *D. citri* genome v3 we identified three potential *knirps* family genes (Tables 1, 2), one of which was annotated as *knirps* in *D. citri* genome v1.1 [10]. These three *knirps* family genes are all located on the same chromosome, within a 900-kilobase pair (Kb) region. All three predicted proteins contain the highly conserved 94-amino acid N-terminal domain and the C-terminal PIDLS motif commonly found in *knirps* family members [23]. However, none of them contain the GASS-domain motif that is unique to the *Eg* protein [23]. Owing to the lack of this signature Eagle motif, the resulting *D. citri* annotations were named *knirps-related 1* (*knrl1*), *knirps-related 2* (*knrl2*) and *knirps-related 3* (*knrl3*). Despite the lack of GASS domain, it is possible that one of these genes is the ortholog of *eg* but has lost the characteristic motif. Interestingly, *D. citri knrl1* has a small exon just 5' of the highly conserved coding exon that is the first exon in most *knirps* family genes (Figure 2). Similar gene structure has been reported for one *knirps* family gene each in *D. melanogaster*, the honeybee *Apis mellifera*, *A. pisum*, and the human louse *Pediculus* [23]. All but *D. melanogaster* share small stretches of sequence identity in the amino acid sequence encoded by this additional exon, suggesting that the 5' exon might have been present in a common ancestor. This model suggests the duplication and acquisition of an additional exon by one paralog early in the insect lineage. The paralog containing the additional exon appears to have been lost in holometabolous insects sometime after the divergence of the Hymenoptera from the rest of the Holometabola.

Most insects have two *otd* genes. However, in both *Drosophila* and *A. pisum* only one *otd* gene has been identified. *Drosophila* is missing the *otd-2* ortholog, while *otd-1* has apparently been lost in pea aphids [24]. In the *D. citri* v3 genome, we found two *otd* genes adjacent to one another on chromosome 4 (Table 1). Phylogenetic analysis suggests that one of these genes is an *otd-1* ortholog, while the other is an *otd-2* ortholog (Figure 3). The genomic clustering of *otd-1* and *otd-2* has also been noted in other insects and crustaceans where their genomic location has been examined [25], providing further support for the identification of the *D. citri* genes as *otd-1* and *otd-2*. The presence of *otd-1* in *D. citri* suggests that *otd-1* was lost in the pea aphid lineage after divergence of the psyllid and aphid lineages.

gt is conserved in some, but not all, hemipterans. No *gt* ortholog was found in the *A. pisum* genome [19]; however, apparent *gt* orthologs are present in *Rhodnius prolixus*, *Cimex lectularius*, *Halyomorpha halys*, *Oncopeltus fasciatus* and *Bemisia tabaci* [26–30]. RNA interference (RNAi) studies indicate that the *R. prolixus* and *O. fasciatus* *gt* orthologs both function as gap genes [26, 29]. We performed BLAST searches of *D. citri* genome v3 with all of these hemipteran *gt* orthologs, but were unable to identify a *D. citri* *gt* ortholog (Table 2).

Figure 2. Multiple alignment of the N-terminal region of Knirps family proteins. The red box denotes the sequence encoded by a highly conserved exon [23]. Species represented are *Drosophila melanogaster* (Dm), *Tribolium castaneum* (Tc), *Apis mellifera* (Am), *Nasonia vitripennis* (Nv), *Pediculus humanus* (Ph), *Acyrtosiphon pisum* (Ap) and *Diaphorina citri* (Dc). Named proteins have been manually annotated, while those with only an accession number are computationally predicted. Five of the *knirps* family genes (one each in Dm, Am, Ph, Ap and Dc) have sequence upstream of the universally conserved sequence that typically begins in the first exon [23, and this work]. In these proteins the conserved core sequence (red box) begins in the second exon. Four of the five proteins with an additional 5' exon (Am, Ph, Ap and Dc) share a small region of sequence identity at the C-terminal end of the sequence encoded by this exon.

Figure 3. Neighbor joining tree of Otd homologs from *Drosophila melanogaster* (Dm), *Tribolium castaneum* (Tc), *Apis mellifera* (Am), and *Diaphorina citri* (Dc). *D. citri* proteins are marked with black circles.

btd is a member of the Sp family of transcription factors. Recent reports indicate that the ancestral state for arthropods—and perhaps all metazoans—is likely to be the presence of three Sp members [31]. These three Sp family genes cluster into three monophyletic clades (Sp5/*btd*, Sp1-4/(Sp-pps) and Sp6-9 (Sp1)) [31]. Even though the ancestral state appears to be the presence of three Sp family members, *btd* is absent from the *A. pisum* genome [19]. Furthermore, repeated efforts to clone *btd* from *Oncopeltus fasciatus* have only resulted in the identification of the two non-*btd* Sp genes. This suggests that *btd* may have been lost in the lineage leading to hemipterans. We were also unable to find a true *btd* ortholog in either the *D. citri* genome v3, or in independent *de novo* transcriptomes (Table 2). Two Sp family members were found that appear to be orthologous to *Sp1* and *Sp6*, but these were not annotated.

Pair-rule genes

One-to-one orthologs were found for all pair-rule genes examined, including *paired* (*prd*), *odd skipped* (*odd*), *sloppy paired* (*slp*), *runt* (*run*), *even skipped* (*eve*), *hairy* (*h*) and *odd paired*

(*opa*) in the *D. citri* genome v3 (Tables 1, 2). A partial copy of *h* had been annotated in a previous genome version [10]. Three of the pair-rule genes we annotated have closely related paralogs and required additional analysis before gene identities could be assigned.

Prd is a member of the Pax3/7 family of proteins. In *Drosophila* there are three Pax3/7 family genes, which are known to be involved 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 roles of Pax3/7 in segmentation and neurogenesis are likely to be conserved in all arthropods [32]. In the *D. citri* genome v3, we also found three Pax3/7 genes, which we named *paired*, *gooseberry* and *gooseberry-neuro* based on reciprocal BLAST analysis and genomic location (Tables 1, 2). The *gsb* and *gsb-n* orthologs are discussed in more detail in the segment polarity gene section.

odd is a zinc finger transcription factor with three close relatives known as *brother of odd with entrails limited* (*bowl*), *sister of odd and bowl* (*sob*) and *drumstick* (*drum*). All four genes are in a conserved cluster in *Tribolium castaneum* and *D. melanogaster* [33]. In *D. citri* genome v3, *drum*, *odd* and *sob* are all located within a 400-Kb region, with *odd* and *sob* overlapping one another on opposite strands. It remains unclear whether the overlap is correct or results from misassembly, but the genes are almost certainly located very close together. *D. citri bowl* is located on the same chromosome about 20 megabase pairs (Mb) away. Separation of *bowl* from the rest of the cluster has also been observed in *Anopheles gambiae* [34].

Insects have four *runt* domain-containing genes: *run*, *Runt-related A* (*RunxA*), *Runt-related B* (*RunxB*) and *lozenge* (*lz*). All four genes are typically found in a cluster and their order and orientation is well conserved across insects [35] (Figure 4). We were able to annotate full length models for all four genes in the *D. citri* genome. It appears that the cluster is intact, with all four genes identified in their expected order within a 300-Kb region (Figure 4).

Segment polarity genes

Many segment polarity genes are members of the Wnt and Hedgehog signaling pathways. Manual annotation of the Wnt pathway genes in the *D. citri* genome v3 is described in a separate report [36]. Here, we report the manual annotation of the segment polarity genes *gooseberry* (*gsb*) and *engrailed* (*en*) (Table 1). *gsb* and *en* each have a tightly linked paralog in many insects [37–39]. Surprisingly, we were unable to find the *en* paralog *invected* (*inv*) in the current genome assembly or the *de novo* transcriptome. However, we did find and annotate the *gsb* paralog *gooseberry-neuro* (*gsb-n*) in its expected position adjacent to *gsb* (Table 1). This positional information helped verify the identity of *gsb-n*, since phylogenetic analysis was inconclusive.

CONCLUSION

We searched for orthologs of 33 *Drosophila* segmentation genes in the *D. citri* v3 genome and identified and annotated 25 homologous genes. We were unable to find orthologs for 10 of the *Drosophila* genes, while *D. citri* has one segmentation gene (*otd-2*) whose ortholog has been lost in *Drosophila*. Most of the absences, except *eagle* and *invected*, were expected, based on the known phylogenetic distribution of the genes. While all the genes discussed in this report were initially identified because of their role in embryonic patterning and

Figure 4. Runt domain (RD) cluster in representative insects. Cluster information from other insects was obtained from [35]. The RD clusters in *D. melanogaster* and *T. castaneum* have three genes in a core cluster, with *lozenge* (*lz*) separated from the cluster, but on the same chromosome distal to *runt* (*run*). The RD clusters in *A. mellifera* and *P. humanus* have all four genes clustered together with *lz* proximal to *run*. The RD cluster in *D. citri* most closely follows the pattern seen in *A. mellifera* and *P. humanus*. *D. citri* *lz* appears to be transcribed in the opposite direction compared to other insects, but it is possible that this is attributed to local misassembly. The orientation of *D. citri* *Runx-related B* (*RunxB*) is uncertain, since there are tandem artifactual duplicates that are on opposite strands. We chose the *RunxB* copy closest to *Runx-related A* (*RunxA*) for annotation. Future assembly improvements may help resolve the gene orientation in this cluster.

segmentation in *Drosophila*, many have other important functions, such as pole cell development, neural stem cell maintenance, sex determination, and immune function. Analysis of expression patterns and gene function will be required to determine which of these genes are involved in *D. citri* segmentation.

DATA VALIDATION AND QUALITY CONTROL

Orthology of annotated genes was verified by reciprocal blasts and phylogenetic analysis. Completeness of gene models (defined as a complete coding region) was assessed by comparing the encoded protein to orthologous proteins. Evidence used for validation of gene models is shown in Table 1.

REUSE POTENTIAL

This manual curation was carried out as a part of the *D. citri* community annotation project [40], with a goal to annotate gene families related to immune response, metabolism and other major functions [36, 41–43]. As scientists search for ways to control the spread of Huanglongbing, understanding the development pathways of its vector, *D. citri*, may provide insights into essential genes that could be targeted by pest control methods. The availability of accurate gene models will facilitate the design of experiments aimed at understanding the expression and function of these genes.

DATA AVAILABILITY

The gene models will be part of an updated official gene set (OGS) for *D. citri* that will be submitted to NCBI. The OGS (v3) will also be publicly available for download, BLAST analysis and expression profiling on Citrusgreening.org and the Citrus Greening Expression Network [40]. The *D. citri* genome assembly (v3), OGS (v3) and transcriptomes are accessible on Citrusgreening.org and NCBI. Accession numbers for genes used in multiple

Table 3. Orthologs used in alignments and phylogenetic analysis.

Species	Accession	Name in NCBI	Name in Tree
<i>Drosophila melanogaster</i>	NP_524187.1	knirps, isoform A	Dm Knirps
<i>Drosophila melanogaster</i>	NP_788552.1	knirps-like, isoform A	Dm Knirps-like
<i>Drosophila melanogaster</i>	NP_730689.1	eagle, isoform B	Dm Eagle
<i>Tribolium castaneum</i>	NP_001121967.1	knirps	Tc Knirps
<i>Tribolium castaneum</i>	NP_001107839.1	eagle	Tc Eagle
<i>Apis mellifera</i>	XP_006558409.1	nuclear receptor subfamily 1 group D member 2 isoform X1	Am XP_006558409.1
<i>Apis mellifera</i>	XP_006559124.1	knirps-related protein	Am XP_395932.2
<i>Nasonia vitripennis</i>	XP_001604969.2	protein embryonic gonad	Nv XP_001604969.2
<i>Pediculus humanus corporis</i>	XP_002430263.1	conserved hypothetical protein	Phc XP_002430263.1
<i>Pediculus humanus corporis</i>	XP_002430262.1	conserved hypothetical protein	Phc XP_002430262.1
<i>Pediculus humanus corporis</i>	XP_002430260.1	knirps related protein, putative	Phc XP_002430260.1
<i>Acyrtosiphon pisum</i>	XP_003243120.1	knirps-related protein-like	Ap XP_003243120.1
<i>Acyrtosiphon pisum</i>	XP_003248994.1	protein embryonic gonad-like	Ap XP_003248994.1
<i>Acyrtosiphon pisum</i>	XP_003244332.1	protein doublesex*	Ap XP_003244332.1
<i>Drosophila melanogaster</i>	NP_001259345.1	ocelliless, isoform G	Dm Otd
<i>Tribolium castaneum</i>	NP_001034513.1	orthodenticle-1	Tc Otd-1
<i>Tribolium castaneum</i>	KYB28866.1	orthodenticle-2	Tc Otd-2
<i>Apis mellifera</i>	AER27696.1	orthodenticle 1	Am Otd-1
<i>Apis mellifera</i>	AER27697.1	orthodenticle 2	Am Otd-2

Species, accession number, full name and abbreviated name are provided for all orthologs used in multiple alignments and phylogenetic trees [18, 21, 23, 45–50]. The asterisk (*) denotes a gene that appears to be incorrectly named *doublesex* in GenBank, but is actually a *knirps* family gene.

alignments or phylogenetic trees are provided in Table 3, and all additional data is available via the *GigaScience* GigaDB repository [44].

EDITOR'S NOTE

This article is one of a series of Data Releases crediting the outputs of a student-focused and community-driven manual annotation project curating gene models and, if required, correcting assembly anomalies, for the *Diaphorina citri* genome project [11].

DECLARATIONS LIST OF ABBREVIATIONS

Am: *Apis mellifera*; Ap: *Acyrtosiphon pisum*; CGEN: Citrus Greening Expression Network; Dc: *Diaphorina citri*; Dm: *Drosophila melanogaster*; Kb: kilobase pair; MCOT: Maker, Cufflinks, Oases and Trinity; NCBI: National Center for Biotechnology Information; Nv: *Nasonia vitripennis*; OGS: official gene set; Ph: *Pediculus humanus*; RD: Runt domain; RNA-Seq: RNA sequencing; Tc: *Tribolium castaneum*.

ETHICAL APPROVAL

Not applicable.

CONSENT FOR PUBLICATION

Not applicable.

COMPETING INTERESTS

The authors declare that they have no competing interests.

FUNDING

This work was supported by USDA-NIFA grant 2015- 70016-23028, HSI 1300394, 2020-70029-33199 and an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant number P20GM103418.

AUTHOR CONTRIBUTIONS

WBH, SJB, TD and LAM conceptualized the study; TDS, SJB, TD and SS supervised the study; SJB, TD, SS and LAM were involved in project administration; SM and TDS conducted the investigation; PH, MF-G and SS contributed to software development; TDS, SS, PH and MF-G developed the methodology; SJB, TD, WBH and LAM acquired funding; SM and TDS wrote the original draft; SJB, WBH and SS reviewed and edited the draft.

ACKNOWLEDGEMENTS

We thank Will Tank for technical assistance and helpful discussions.

REFERENCES

- 1 Nüsslein-Volhard C, Wieschaus E, Mutations affecting segment number and polarity in *Drosophila*. *Nature*, 1980; **287**: 795–801.
- 2 Jürgens G, Wieschaus E, Nüsslein-Volhard C, Kluding H, Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster* - II. Zygotic loci on the third chromosome. *Wilhelm Roux Arch. Dev. Biol.*, 1984; **193**: 283–295.
- 3 Wieschaus E, Nüsslein-Volhard C, Jürgens G, Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster* - III. Zygotic loci on the X-chromosome and fourth chromosome. *Wilhelm Roux Arch. Dev. Biol.*, 1984; **193**: 296–307.
- 4 Nüsslein-Volhard C, Wieschaus E, Kluding H, Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster* - I. Zygotic loci on the second chromosome. *Wilhelm Roux Arch. Dev. Biol.*, 1984; **193**: 267–282.
- 5 Wieschaus E, Nüsslein-Volhard C, The Heidelberg screen for pattern mutants of *Drosophila*: a personal account. *Annu. Rev. Cell Dev. Biol.*, 2016; **32**: 1–46.
- 6 Clark E, Peel AD, Akam M, Arthropod segmentation. *Development*, 2019; **146**: dev170480.
- 7 Hernandez J, Pick L, Reding K, Oncopeltus-like gene expression patterns in *Murgantia histrionica*, a new hemipteran model system, suggest ancient regulatory network divergence. *Evodevo*, 2020; **11**: 9.
- 8 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*, 2019; **146**: dev181453.
- 9 Auman T, Chipman AD, Growth zone segmentation in the milkweed bug *Oncopeltus fasciatus* sheds light on the evolution of insect segmentation. *BMC Evol. Biol.*, 2018; **18**: 178.
- 10 Saha S et al. Improved annotation of the insect vector of citrus greening disease: biocuration by a diverse genomics community. *Database*, 2017; **2017**: bax032.
- 11 Hosmani P et al. Chromosomal length reference assembly for *Diaphorina citri* using single-molecule sequencing and Hi-C proximity ligation with manually curated genes in developmental, structural and immune pathways. bioRxiv. Cold Spring Harbor Laboratory; 2019; 869685. <https://doi.org/10.1101/869685>.
- 12 Dunn NA et al. Apollo: democratizing genome annotation. *PLoS Comput. Biol.*, 2019; **15**: e1006790.
- 13 Kumar S, Stecher G, Li M, Knyaz C, Tamura K, MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.*, 2018; **35**: 1547–1549.
- 14 Shippy TD et al. Annotating genes in *Diaphorina citri* genome version 3. protocols.io. 2020; <https://dx.doi.org/10.17504/protocols.io.bniiimcce>.
- 15 Peel AD, Chipman AD, Akam M, Arthropod segmentation: beyond the *Drosophila* paradigm. *Nat. Rev. Genet.*, 2005; **6**: 905–916.

- 16 Larkin A et al. FlyBase: updates to the *Drosophila melanogaster* knowledge base. *Nucleic Acids Res.*, 2021; **49**: D899–D907.
- 17 Dearden PK et al. Patterns of conservation and change in honey bee developmental genes. *Genome Res.*, 2006; **16**: 1376–1384.
- 18 Richards S et al. Tribolium Genome Sequencing Consortium. The genome of the model beetle and pest *Tribolium castaneum*. *Nature*, 2008; **452**: 949–955.
- 19 Shigenobu S et al. Comprehensive survey of developmental genes in the pea aphid, *Acyrtosiphon pisum*: frequent lineage-specific duplications and losses of developmental genes. *Insect Mol. Biol.*, 2010; **19**: 47–62.
- 20 Stein RA, Staros JV., Insights into the evolution of the ErbB receptor family and their ligands from sequence analysis. *BMC Evol. Biol.*, 2006; **6**: 79.
- 21 Werren JH et al. Functional and evolutionary insights from the genomes of three parasitoid nasonia species. *Science*, 2010; **327**: 343–348.
- 22 Rosenberg MI, Lynch JA, Desplan C, Heads and tails: evolution of antero-posterior patterning in insects. *Biochim. Biophys. Acta*, 2009; **1789**: 333–342.
- 23 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**: 1348–1357.
- 24 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, *Acyrtosiphon pisum*: hunchback and orthodenticle expression. *Insect Mol. Biol.*, 2010; **19**: 75–85.
- 25 Browne WE, Schmid BGM, Wimmer EA, Martindale MQ, Expression of otd orthologs in the amphipod crustacean, *Parhyale hawaiiensis*. *Dev. Genes Evol.*, 2006; **216**: 581–595.
- 26 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.*, 2012; **361**: 147–155.
- 27 Benoit JB et al. Unique features of a global human ectoparasite identified through sequencing of the bed bug genome. *Nat. Commun.*, 2016; **7**: 10165.
- 28 Sparks ME et al. Brown marmorated stink bug, *Halyomorpha halys* (Stål), genome: putative underpinnings of polyphagy, insecticide resistance potential and biology of a top worldwide pest. *BMC Genomics*, 2020; **21**: 227.
- 29 Liu PZ, Patel NH, Giant is a bona fide gap gene in the intermediate germband insect, *Oncopeltus fasciatus*. *Development*, 2010; **137**: 835–844.
- 30 Chen W et al. The draft genome of whitefly *Bemisia tabaci* MEAM1, a global crop pest, provides novel insights into virus transmission, host adaptation, and insecticide resistance. *BMC Biol.*, 2016; **14**: 110.
- 31 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.*, 2010; **10**: 88.
- 32 Davis GK, D'Alessio JA, Patel NH, Pax3/7 genes reveal conservation and divergence in the arthropod segmentation hierarchy. *Dev. Biol.*, 2005; **285**: 169–184.
- 33 Linz DM, Tomoyasu Y, RNAi screening of developmental toolkit genes: a search for novel wing genes in the red flour beetle, *Tribolium castaneum*. *Dev. Genes Evol.*, 2015; **225**: 11–22.
- 34 Johansen KA, Green RB, Iwaki DD, Hernandez JB, Lengyel JA, The Drm-Bowl-Lin relief-of-repression hierarchy controls fore- and hindgut patterning and morphogenesis. *Mech. Dev.*, 2003; **120**: 1139–1151.
- 35 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.
- 36 Vosburg C et al. Utilizing a chromosomal-length genome assembly to annotate the Wnt signaling pathway in the Asian citrus psyllid, *Diaphorina citri*. *Gigabyte*, 2021; **1**: 1–15, <https://doi.org/10.46471/gigabyte.21>.
- 37 Peel AD, Telford MJ, Akam M, The evolution of hexapod engrailed-family genes: evidence for conservation and concerted evolution. *Proc. Biol. Sci.*, 2006; **273**: 1733–1742.
- 38 Bopp D, Burri M, Baumgartner S, Frigerio G, Noll M, Conservation of a large protein domain in the segmentation gene paired and in functionally related genes of *Drosophila*. *Cell*, 1986; **47**: 1033–1040.

- 39 Osborne PW, Dearden PK. Expression of Pax group III genes in the honeybee (*Apis mellifera*). *Dev. Genes Evol.*, 2005; 215: 499–508.
- 40 Flores-Gonzalez M et al. Citrusgreening.org: An open access and integrated systems biology portal for the Huanglongbing (HLB) disease complex. bioRxiv. Cold Spring Harbor Laboratory; 2019; <https://doi.org/10.1101/868364>.
- 41 Massimino C et al. Annotation of yellow genes in *Diaphorina citri*, the vector for Huanglongbing disease. *Gigabyte*, 2021; 1: 1–15, <https://doi.org/10.46471/gigabyte.20>.
- 42 Miller S et al. Annotation of chitin biosynthesis genes in *Diaphorina citri*, the Asian citrus psyllid. *Gigabyte*, 2021; 1: 1–12, <https://doi.org/10.46471/gigabyte.23>.
- 43 Miller S et al. In silico characterization of chitin deacetylase genes in the *Diaphorina citri* genome. *Gigabyte*, 2021; 1: 1–12, <https://doi.org/10.46471/gigabyte.25>.
- 44 Miller S et al. Supporting data for “Annotation of segmentation pathway genes in the Asian citrus psyllid, *Diaphorina citri*”. GigaScience Database. 2021; <http://dx.doi.org/10.5524/100902>.
- 45 Adams MD et al. The genome sequence of *Drosophila melanogaster*. *Science*, 2000; 287: 2185–2195.
- 46 Cerny AC, Grossmann D, Bucher G, Klingler M., The Tribolium ortholog of knirps and knirps-related is crucial for head segmentation but plays a minor role during abdominal patterning. *Dev. Biol.*, 2008; 321: 284–294.
- 47 Bucher G, Farzana L, Brown SJ, Klingler M. Anterior localization of maternal mRNAs in a short germ insect lacking bicoid. *Evol. Dev.*, 2005; 7: 142–149.
- 48 Elsik CG et al. Finding the missing honey bee genes: lessons learned from a genome upgrade. *BMC Genomics*, 2014; 15: 86.
- 49 Kirkness EF et al. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proc. Natl. Acad. Sci. USA*, 2010; 107: 12168–12173.
- 50 International Aphid Genomics Consortium. Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS Biol.*, 2010; 8: e1000313.