Annotation of Hox cluster and Hox cofactor genes in the Asian citrus psyllid, *Diaphorina citri*, reveals novel features

Teresa D Shippy (ORCID: 0000-0003-2305-7432)¹, Prashant S Hosmani (ORCID: 0000-0001-5722-4118)², Mirella Flores-Gonzalez (ORCID: 0000-0002-7759-1617)², Lukas A Mueller (ORCID: 0000-0001-8640-1750)², Wayne B Hunter (ORCID: 0000-0001-8603-3337)³, Susan J Brown (ORCID: 0000-0002-7984-0445)¹, Tom D'Elia (ORCID: 0000-0003-2601-3128)⁴ and Surya Saha (ORCID: 0000-0002-1160-1413)^{2,5*}

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

² Boyce Thompson Institute, Ithaca, NY 14853, USA

³USDA-ARS, U.S. Horticultural Research Laboratory, Fort Pierce, FL 34945, USA

⁴ Indian River State College, Fort Pierce, FL 34981, USA

⁵ Animal and Comparative Biomedical Sciences, University of Arizona, Tucson, AZ 85721, USA

* suryasaha@cornell.edu

Abstract

Hox genes and their cofactors are essential developmental genes that specify regional identity in animals, including insects. A particularly interesting feature of Hox genes is their conserved arrangement in clusters in the same order in which they specify identity along the anterior-posterior axis. Among insects, breaks in the cluster have been reported in a few species, but these seem to be the exception rather than the rule. We have annotated the ten Hox genes of the Asian citrus psyllid, *Diaphorina citri*, and

determined that there is a split in its Hox cluster between the *Deformed* and *Sex combs* *reduced* genes. This is the first time a break at this position has been observed in an insect Hox cluster. We have also annotated the *D. citri* orthologs of the Hox cofactor genes *homothorax*, *PKNOX* and *extradenticle*. Interestingly, we found an additional copy of *extradenticle* in *D. citri* that appears to be a retrogene. Expression data and sequence conservation suggest that the *extradenticle* retrogene may have retained the original *extradenticle* function and allowed the parental *extradenticle* gene to diverge.

Research Area: Genetics and Genomics

Classifications: Animal Genetics, Bioinformatics

Data Description

Introduction

Hox genes encode transcription factors best known for their role in specifying regional identity along the anterior-posterior (A-P) axis [1]. They were first identified in the fruit fly, *Drosophila melanogaster* [2–4], but are conserved throughout the Metazoa [1]. The arrangement of Hox genes in clusters in an order that parallels their function along the A-P axis — a phenomenon known as spatial colinearity — is also widely conserved [5,6]. Researchers investigating the reasons for this conserved arrangement have discovered several constraints that could be keeping Hox clusters intact. First, the Hox clusters of vertebrates show temporal collinearity (the sequential activation of Hox genes from the 3' end to the 5' end of the cluster) which may be regulated by progressive opening of the chromatin to an active state [7–9]. Therefore, the need to maintain the timing of

Hox gene expression during development could drive Hox cluster conservation. There is also evidence that the presence of shared regulatory elements is a factor in the conservation of Hox clusters. In vertebrates, these include enhancers shared by two Hox genes [10–12] and global enhancer elements outside the cluster that regulate multiple Hox genes [6,13,14]. Others have suggested that the conservation of colinearity may stem from the need to minimize the number of boundaries between active and inactive chromatin to avoid inappropriate activation of genes [15].

In insects, the constraints on Hox clusters appear to be more relaxed than those affecting vertebrate clusters. The most recent common ancestor of insects is believed to have had ten Hox genes: *labial (lab)*, *proboscipedia (pb)*, *zerknüllt (zen)*, *Deformed (Dfd)*, *Sex combs reduced (Scr)*, *fushi tarazu (ftz)*, *Antennapedia (Antp)*, *Ultrabithorax (Ubx)*, *abdominal-A (abd-A)* and *Abdominal-B (Abd-B)*. Ancestrally, these genes were located in a single cluster and all transcribed on the same strand [16–18]. However, splits of the cluster have occurred in several insect lineages, including three independent splits in *Drosophila* species [19,20] and one in the silk moth, *Bombyx mori* [21]. Inversions affecting the transcriptional direction of one or a few genes are also fairly common [18,20]. Insect Hox clusters are larger than those in vertebrates, with longer genes (particularly the introns) and more intervening sequence between genes [18]. Unlike in vertebrates, temporal colinearity and global control elements do not seem to be a factor. Studies have suggested that the main constraint keeping insect Hox genes clustered is the presence of large regulatory regions between genes which limits

the number of places the cluster can be broken without affecting gene function [20].

There are also four microRNA (miRNA) genes conserved within arthropod Hox clusters [18]. At least some of these miRNAs target and repress translation of Hox mRNAs, particularly in the ventral nerve cord [22].

Another group of genes whose functions are closely tied to the Hox genes are the members of the PBX and MEINOX families of TALE class homeodomain proteins [23]. These proteins are best known as cofactors to Hox proteins, although they have Hox-independent functions as well [24–26]. The PBX and MEINOX proteins are very ancient. They are conserved throughout the animal kingdom and a related protein is thought to have been present in the common ancestor of plants and animals [27]. Most insects have one member of the PBX family and two members of the MEINOX family [28,29]. The MEINOX genes consist of one member of the MEIS family and one member of the PREP/PKNOX family. In *Drosophila*, where the PBX and MEINOX proteins are best studied, the PREP/PKNOX ortholog has been lost, leaving only the PBX gene *extradenticle (exd)* [30] and the MEIS gene *homothorax (hth)* [31].

Context

As part of a community gene annotation project, we are curating genes implicated in development, immune function and metabolism from the genome of the Asian citrus psyllid, *Diaphorina citri* (NCBI:txid121845) [32]. This hemipteran insect pest carries the bacterium *Candidatus Liberibacter asiaticus* (CLas), which causes Huanglongbing (citrus

greening disease) and is a serious threat to citrus production worldwide. The Hox genes and their cofactors are potential targets for RNAi-based pest control methods because of their essential role in development. Hox gene sequences have been identified previously in several hemipterans including the milkweed bug *Oncopeltus fasciatus* and the pea aphid *Acyrtosiphon pisum*. These insects seem to have the full complement of Hox genes [29,33–38]. Based on what is known in other insects, hemipteran Hox genes are likely to be clustered, but so far only a few examples of gene linkage have been reported due to the fragmented condition of the available genome assemblies. To date, linkage has only been demonstrated for a few pairs of genes: *zen* - *pb* and *ftz* - *Scr* in the pea aphid [29], and *zen* - *Dfd* in *Oncopeltus* [38].

Hemipteran Hox gene function has been best studied in the milkweed bug, *Oncopeltus fasciatus*. RNAi experiments indicate that *Oncopeltus* Hox genes have functions broadly similar to their orthologs in holometabolous insects such as *Drosophila* and *Tribolium*, although there are many small differences in regulation and function [33,34]. The *Oncopeltus* studies show that RNAi with most Hox genes causes high levels of lethality [33,34]. One exception is RNAi with *pb*, which produces viable nymphs that are unable to feed due to abnormal mouthpart morphology [33]. RNAi with *zen* in the brown planthopper, *Nilaparvata lugens*, caused high levels of embryonic death and prevented normal development [39]. These results suggest that Hox genes could be targets for development of RNAi-based pest management products in hemipteran pests such as *D. citri*. While specific regions of Hox proteins have a high level of conservation throughout

the animal kingdom, their nucleotide sequences differ enough to provide many regions for designing highly specific RNAi-based products.

Figure 1. Protocol for psyllid genome community curation [14].

Methods

D. citri genes were identified by BLAST (NCBI BLAST, RRID:SCR_004870) analysis of *D. citri* sequences [40] with orthologs from *D. melanogaster* and *Tribolium castaneum*. The only exception was *zen*, which was located by using the Apollo BLAT function to search the expected region of the Hox cluster for additional homeodomain-encoding sequences. Reciprocal BLAST was used to confirm orthologs for all *D. citri* genes. Genes were manually annotated in Apollo 2.1.0 (Apollo, RRID:SCR_001936) [41] using RNA-Seq reads, IsoSeq transcripts and *de novo*-assembled transcripts as evidence. A more detailed description of the annotation workflow is available via protocols.io (Figure 1) [42]. Multiple alignments and phylogenetic analysis were performed with MEGA X

(MEGA Software, RRID:SCR_000667) [43]. Orthologs used in phylogenetic analysis are listed in Table 1.

Expression counts for annotated genes were downloaded from the Citrus Greening Expression Network [40]. Heatmaps were produced with the pheatmap package (pheatmap, RRID:SCR_016418) in R (R Project for Statistical Computing, RRID:SCR_001905). Bar charts were created using Microsoft Excel (Microsoft Excel, RRID:SCR_016137).

To search for orthologs of *exd-r* in hemipteran genomes, we used the *D. citri* Exd-r protein as a query to BLAST the assemblies of *Pachypsylla venusta* (GCA_012654025.1), *Bemisia tabaci* (GCA_004919745.1 and GCA_003994315.1), *Trialeurodes vaporariorum* (GCA_011764245.1), and *A. pisum* (GCA_005508785.1).

Table 1. Orthologs used in Phylogenetic Analysis and Multiple Sequence Alignments

Species	Accession	Name in NCBI	Name in Tree
<i>Acyrtosiphon pisum</i>	XP_008182670.1	homeobox protein extradenticle isoform X1	Ap Exd
<i>Nilaparvata lugens</i>	XP_039288084.1	homeobox protein extradenticle	NI Exd
<i>Bemisia tabaci</i>	XP_018913868.1	homeobox protein extradenticle isoform X2	Bt Exd
<i>Apis mellifera</i>	XP_006558038.2	homeobox protein	Am Exd

		extradenticle isoform X4	
<i>Tribolium castaneum</i>	NP_001034501.1	extradenticle	Tc Exd
<i>Drosophila melanogaster</i>	NP_001259592.1	extradenticle, isoform D	Dm Exd
<i>Bombyx mori</i>	XP_004928674.1	homeobox protein extradenticle isoform X4	Bm Exd
<i>Pachypsylla venusta</i>	ScZCZ4B_2870.g10167.t1		Pv Exd
<i>Diaphorina citri</i>	Dcitr08g10400.1.1		Dc Exd
<i>Pachypsylla venusta</i>	ScZCZ4B_298.g6815.t1		Pv Exd-r
<i>Diaphorina citri</i>	Dcitr01g03700.1.1		Dc Exd-r
<i>Homo sapiens</i>	XP_016856884.1	pre-B-cell leukemia transcription factor 1 isoform X4	Hs PBC1
<i>Nilaparvata lugens</i>	XP_039298141.1	homeobox protein Meis1	NI PKNOX
<i>Acyrtosiphon pisum</i>	XP_008181007.1	LOW QUALITY PROTEIN: homeobox protein PKNOX2	Ap PKNOX
<i>Tribolium castaneum</i>	XP_970138.1	PREDICTED: homeobox protein PKNOX2 isoform X2	Tc PKNOX
<i>Apis mellifera</i>	XP_001120618.1	homeobox protein PKNOX2 isoform X1	Am PKNOX
<i>Bombyx mori</i>	XP_037875077.1	homeobox protein PKNOX2 isoform X1	Bm PKNOX
<i>Homo sapiens</i>	BAB83665.1	PKNOX2	Hs PKNOX2
<i>Diaphorina citri</i>	MCOT13578.2.CT		Dc PKNOX
<i>Nilaparvata lugens</i>	XP_039282637.1	homeobox protein homothorax isoform X12	NI Hth
<i>Tribolium castaneum</i>	NP_001034489.1	homothorax	Tc Hth
<i>Apis mellifera</i>	XP_026296417.1	homeobox protein homothorax isoform X2	Am Hth
<i>Drosophila melanogaster</i>	NP_476578.3	homothorax, isoform A	Dm Hth

<i>Bombyx mori</i>	NP_001296493.1	homeobox protein homothorax-like	Bm Hth
<i>Acyrtosiphon pisum</i>	XP_001951115.2	homeobox protein homothorax isoform X2	Ap Hth
<i>Homo sapiens</i>	NP_002390.1	homeobox protein Meis2 isoform f	Hs Meis2
<i>Diaphorina citri</i>	Dcitr03g04000.1.1		Dc Hth

147

148 All orthologs used in phylogenetic analysis and multiple sequence alignment are listed
149 with species name, accession number, full name from NCBI and abbreviated name used
150 in trees. Dcitr and MCOT numbers refer to *D. citri* sequences that can be retrieved from
151 the Citrus Greening web site [44].

152

153 Data Validation and Quality Control

154 Hox genes

155 We identified single orthologs of each of the Hox genes in *D. citri* (Table 2, 3). Notably,
156 *lab* is not found in the *D. citri* v3.0 genome but is found in the Maker, Cufflinks, Oases,
157 Trinity (MCOT) transcriptome and was present in the *D. citri* v1.1 and v2.0 genome
158 assemblies [32,40,45] in its expected position adjacent to *pb*. Thus, its absence in *D.*
159 *citri* genome v3.0 seems to be caused by a local genome misassembly. We also observed
160 additional local misassemblies resulting in false duplications of several Hox genes,
161 including *zen*, *ftz*, *Scr* and *Antp* in both the *D. citri* v2.0 and v3.0 assemblies. This type of
162 duplication, in which almost identical copies of a genome segment are assembled in
163 tandem within a scaffold, is usually caused by the presence of allelic variants that are
164 mistakenly assembled as separate genes [46] and is fairly common in all versions of the
165 *D. citri* genome because of the heterogeneity of the sequenced psyllids [32].

166

167 Table 2. Gene counts in selected insect species

Hox genes	<i>Dm</i>	<i>Am</i>	<i>Tc</i>	<i>Ap</i>	<i>Dc</i>
<i>labial</i>	1	1	1	1	1
<i>proboscipedia</i>	1	1	1	1	1
<i>zerknüllt (Hox 3)</i>	3	1	2	1	1
<i>Deformed</i>	1	1	1	1	1
<i>Sex combs reduced</i>	1	1	1	1	1
<i>fushi tarazu</i>	1	1	1	1	1
<i>Antennapedia</i>	1	1	1	1	1
<i>Ultrabithorax</i>	1	1	1	1	1
<i>abdominal-A</i>	1	1	1	1	1
<i>Abdominal-B</i>	1	1	1	2	1
Hox cluster miRNAs	<i>Dm</i>	<i>Am</i>	<i>Tc</i>	<i>Ap</i>	<i>Dc</i>
<i>miR-10</i>	1	1	1	1	2*
<i>miR-993</i>	1	1	1	1	2*
<i>miR-iab-4</i>	1	1	1	1	3*
<i>miR-iab-8</i>	1	1	1	1	3*
Hox cofactors	<i>Dm</i>	<i>Am</i>	<i>Tc</i>	<i>Ap</i>	<i>Dc</i>
<i>extradenticle</i>	1	1	1	1	2
<i>homothorax</i>	1	1	1	1	1
<i>PKNOX</i>	0	1	1	1	1

168
169 Comparison of gene copy number for Hox genes and Hox cluster miRNAs in *Drosophila*
170 *melanogaster* (*Dm*), *Apis mellifera* (*Am*), *Tribolium castaneum* (*Tc*), *Acyrtosiphon pisum*
171 (*Ap*) and *Diaphorina citri* (*Dc*) [18,29,47–52]. * indicates miRNA genes that appear as
172 multiple copies in *D. citri* genome v3, but are likely to be single copy genes with false
173 duplications due to local misassembly.
174

175 Table 3. Annotated *D. citri* orthologs of Hox cluster and Hox cofactor genes.

Gene	D. citri OGS3 identifier	Complete Gene model	Evidence supporting annotation			
			MCOT	IsoSeq	RNASeq	Ortholog
<i>labial</i> [†]		in v2	X			X
<i>proboscipedia</i>	Dcitr07g11290.1.1 Dcitr07g11290.1.2	X	MCOT18081.0.CO	X	X	X
<i>zerknüllt</i>	Dcitr07g11300.1.1	X	MCOT17344.0.CT		X	X
<i>Deformed</i>	Dcitr07g11310.1.1	X	MCOT13111.0.CT	X	X	
<i>Sex combs</i>	Dcitr07g04890.1.1	X	MCOT12386.0.CT			X

<i>reduced</i>						
<i>fushi tarazu</i>	Dcitr07g04910.1.1	X	MCOT02202.0.CC			
<i>Antennapedia</i>	Dcitr07g04960.1.1	X	MCOT16702.0.CC	X		
<i>Ultrabithorax</i>	Dcitr07g05020.1.1	X	MCOT21836.0.CT	X		X
<i>abdominal-A</i>	Dcitr07g05110.1.1	X			X	X
<i>Abdominal-B</i>	Dcitr07g05150.1.1	X	MCOT02185.0.CT	X		
<i>miR-10</i>		X				X
<i>miR-993</i>		X				X
<i>miR-iab-4</i>		X				X
<i>miR-iab-8</i>		X				X
<i>extradenticle</i>	Dcitr08g10400.1.1	X	MCOT05038.1.CT		X	
<i>extradenticle-retro</i>	Dcitr01g03720.1.1	X				X
<i>homothorax</i>	Dcitr03g04000.1.1	X	MCOT17109.0.CT	X	X	X
	Dcitr03g04000.1.2					
<i>PKNOX[†]</i>	Dcitr00g14970.1.1		MCOT13578.2.CT MCOT13578.1.CT	X	X	X

176

177 Manually annotated genes from *D. citri* have been assigned an OGS3 ID. For protein-
178 coding genes, completeness refers to the presence of the full open reading frame. For
179 miRNA genes, a complete gene model indicates that the entire hairpin-forming portion
180 of the gene (the pre-miRNA) has been annotated. [†]denotes genes that were present in
181 *D. citri* genome v2.0, but are absent or less complete in *D. citri* genome v3.0 [32].
182 Evidence types include the MCOT transcriptome (MCOT), high-quality IsoSeq transcripts
183 (IsoSeq), RNASeq reads (RNASeq), and ortholog sequences (Ortholog). The use of these
184 evidence types has been previously described [42]
185

186 In *D. citri* genome v3.0, the Hox genes (except for *lab* as discussed above) are located in
187 two clusters on chromosome 7, separated by about 6 Mb. *pb*, *zen* and *Dfd* are found in
188 one cluster (*lab* was also part of this cluster in a previous assembly), while *Scr*, *ftz*, *Antp*,
189 *Ubx*, *abd-A* and *Abd-B* are located in a separate cluster (Figure 2). This indicates that a
190 split in the Hox cluster has occurred at some point in the lineage leading to *D. citri*. The
191 break, which is between *Dfd* and *Scr*, is in a different position than any of the previously
192 described splits in insect Hox clusters. In *D. melanogaster*, the Hox cluster is split

193 between *Antp* and *Ubx*, while breaks between *Ubx* and *abd-A* and between *lab* and *pb*
 194 have been reported in other *Drosophila* species [20]. To date, the only other break
 195 described in a non-Drosophilid insect Hox cluster is the separation of *lab* from the rest
 196 of the cluster in the silk moth, *B. mori* [21].
 197

198
 199 Figure 2. Comparison of Hox gene clusters in *D. citri* with the putative ancestral
 200 arthropod Hox cluster. In the ancestral cluster, all ten Hox genes were oriented in the
 201 same direction and at least four miRNAs (black triangles) were located within the cluster
 202 [18]. In *D. citri*, the Hox genes are split between two clusters on the same chromosome.
 203 *lab* is missing from genome v3.0 due to apparent misassembly, but was present in
 204 previous genome versions adjacent to *pb*. *D. citri* *pb* and *Dfd* appear to be oriented in
 205 the opposite direction to the ancestral orientation, but this could be the result of local
 206 misassembly. *D. citri* has all four of the conserved Hox cluster miRNAs in the expected
 207 positions. The additional copies of the miRNAs described in the text are not shown,
 208 since they are likely to be false duplications.

In the current *D. citri* genome assembly, we find that *pb* and *Dfd* are oriented in the opposite direction to what is expected (Figure 2). It is possible that these inversions represent real rearrangements of the Hox cluster, but given the number of local misassemblies affecting the Hox cluster region, and the *D. citri* genome in general, we cannot be certain.

Figure 3. Expression of *D. citri* Hox genes. Heatmap of relative expression levels of Hox genes in RNA-Seq samples of various developmental stages, tissues, host plants and CLas infection states [53,70–73 and NCBI BioProjects PRJNA448935 and PRJNA609978]. Host plants are *Citrus macrophylla* (Cmac), *Citrus reticulata* (Cret), *Citrus* spp., *Citrus medica* (Cmed), and *Citrus sinensis* (Csin). The heat map was scaled by row to allow comparison of expression in different samples for each gene. All ten genes showed highest expression in eggs.

Intact insect Hox clusters range in size from 0.71 Mb in *T. castaneum* to 1.22 Mb in *Anopheles gambiae* [18]. As a comparison, we calculated the size of the two *D. citri* clusters. Since *lab* is missing in the *D. citri* v3.0 genome, we used a previous genome assembly (v2) [32] to determine that the portion of the cluster containing *lab*, *pb*, *zen*, *Dfd* and *miR-10* measures 231 kb. The part of the cluster containing the remaining Hox genes spans 2.17 Mb in the *D. citri* v3.0 genome, but this length is almost certainly inflated by false duplications as described above.

We used publicly available RNA-Seq data in the Citrus Greening Expression Network (CGEN) [40] to assess expression of *D. citri* Hox genes in a variety of stages, tissues and CLas infection states. Expression of Hox genes is generally low in all samples (Table 4), with the highest expression in eggs (Figure 3). We did not observe any effect of CLas infection on Hox gene expression (Figure 3).

240

Table 4. TPM expression values for *D. citri* Hox genes

	Dcitr07	Dcitr07	Dcitr07	Dcitr07	Dcitr07	Dcitr07	Dcitr07	Dcitr07	Dcitr07	Dcitr07
	g11290	g11290	g11300	g11310	g04890	g04910	g04960	g05020	g05110	g05150
Gene ID	.1.1	.1.2	.1.1	.1.1	.1.1	.1.1	.1.1	.1.1	.1.1	.1.1
Gene/Transcript Symbol	<i>pb-RA</i>	<i>pb-RB</i>	<i>zen</i>	<i>Dfd</i>	<i>Scr</i>	<i>ftz</i>	<i>Antp</i>	<i>Ubx</i>	<i>abd-A</i>	<i>Abd-B</i>
Egg C. <i>macrophylla</i>	2.73	16.66	3.29	19.65	21.64	25.35	88.88	11.38	3.14	18.53

Clas- Whole body										
Nymph C. <i>medica</i> Clas+ Low infection Whole body	0.35	0	0	0.31	2.13	0	9.7	3.36	1.29	4.47
Nymph C. <i>sinensis</i> Clas+ High infection Whole body	0.47	0.01	0	0.35	2.41	0.02	8.85	2.09	0.6	3.55
Nymph C. <i>sinensis</i> Clas- Whole body	0.56	0.01	0	0.55	2.25	0	13.56	3.36	1.5	6.03
Nymph C. <i>macrophylla</i> Clas- Whole body	0.66	0	0	0.14	3.3	0	14.55	3.64	0.92	5.62
Nymph	0.44	0	0	0	1.93	0	14.55	2.76	0	1.44

<i>Citrus</i> spp. CLas- Whole body										
Nymph <i>Citrus</i> spp. CLas+ Whole body	0.22	0	0	0	1.12	0	6.6	1.38	0	0.13
Adult <i>C. medica</i> CLas- Gut	0.06	0	0	0	0	0	0.47	0.16	0.01	0
Adult <i>C. medica</i> CLas+ Gut	0.02	0	0	0	0.04	0	0.39	0.07	0	0.02
Adult <i>C. medica</i> CLas+ High infection Whole body	0.16	0.22	0	0.08	2.9	0	8.87	2.58	0.55	5.28
Adult <i>C. medica</i> CLas+ Low	0.15	0.22	0	0.32	2.35	0.06	10.34	2.66	0.42	5.56

infection										
Whole body										
Adult <i>C. medica</i> Clas-										
Whole body	0.54	0.14	0	0.18	2.45	0	12.03	2.28	0.66	5.63
Adult <i>C. macrophylla</i> Clas- Whole										
body	0.46	0	0	0	1.92	0	6.2	1.85	0.47	1.19
Adult <i>Citrus</i> spp. Clas-										
Whole body	1.08	0.34	0	0	3.86	0	17.89	6.25	0	3.32
Adult <i>Citrus</i> spp. Clas+										
Whole body	1.2	0	0	0.56	2.58	0	10.99	5.02	0.41	0.82
Adult <i>Citrus</i> spp. Clas- midgut										
	0	0	0	0	0	0	1.4	0.13	0	0
Adult <i>Citrus</i> spp. Clas+ midgut										
	0	0	0	0	0.29	0	0.42	0.29	0	0.03

Adult <i>C. reticulata</i> Clas- Female abdomen	0.07	0	0	0	0	0	11.06	5.97	0	1.47
Adult <i>C. reticulata</i> Clas- Female antennae	0.2	0	0	0.7	0	0	1.85	0.14	0	0
Adult <i>C. reticulata</i> Clas- Female head	0	1.23	0	3.11	1.05	0	4.65	0.13	0	0
Adult <i>C. reticulata</i> Clas- Female leg	0.83	0.29	0	0.16	0	0	9.84	0.86	0	0
Adult <i>C. reticulata</i> Clas- Female terminal abdomen	0	0	0	0	0	0	0.25	0.87	0	1.37

Adult C. <i>reticulata</i> Clas- Female thorax	1.67	0.37	0	1.45	1.57	0	29.25	3.34	0.22	0
Adult C. <i>reticulata</i> Clas- Male abdomen	0.03	0	0	0	1.17	0	13.81	5.07	0.7	7.25
Adult C. <i>reticulata</i> Clas- Male antennae	0	0.39	0	0.22	0.71	0	2.07	0	0.05	1.06
Adult C. <i>reticulata</i> Clas- Male head	1.18	0.13	0	1.94	1.53	0	6.83	0	0	0.07
Adult C. <i>reticulata</i> Clas- Male leg	0.72	0.74	0	1.21	1.78	0	17.81	1.92	0	0.07
Adult C.	0.06	0	0	0	0.79	0	0.61	1.04	0.16	7.97

<i>reticulata</i> Clas- Male terminal abdomen										
Adult C. <i>reticulata</i> Clas- Male thorax	1.68	0.41	0	1.11	1.32	0	19	2.54	0	0
Adult C. <i>reticulata</i> Clas- Female antennae [53]	0.73	0.48	0	1.07	4.04	0	7.08	0.9	0	0.22
Adult C. <i>reticulata</i> Clas- Female terminal abdomen [53]	0.15	0	0	0	0.38	0	7.98	5.19	0.26	1.92
Adult C. <i>reticulata</i> Clas- Male	1.72	0	0	0.67	3.11	0	7.39	0.06	0	0.75

antennae										
[53]										
Adult C.										
<i>reticulata</i>										
CLas- Male										
terminal										
abdomen										
[53]	0	0	0	0	1.29	0	3.41	3.74	0.58	18.85

242
243 Expression levels of Hox genes used in Figure 3. Counts were obtained from publicly
244 available datasets in CGEN [40] and are reported in transcripts per million
245 (TPM). Developmental stage, citrus host, CLas infection status and source tissue for
246 each sample are noted in the first column.

247

248 Hox cluster miRNAs

249 There are several conserved miRNAs located within insect Hox clusters. *miR-10* is
250 conserved in most animals and is located between the *Dfd* and *Scr* orthologs [54]. Three
251 other miRNAs are widely conserved in arthropods [18]. *miR-993* is found between *Dfd*
252 and *zen*, and the bidirectionally transcribed locus producing the *iab-4* and *iab-8* miRNAs
253 is located between *abd-A* and *Abd-B* [55,56]. We found multiple copies of each of the
254 Hox cluster miRNAs in the *D. citri* v3.0 genome (Table 2). These additional copies are
255 probably false duplications due to local misassembly, particularly since the number
256 often varies between different versions of the assembly. In all but one case, the
257 duplicate copies are adjacent to one another in the genome. The exception is *miR-993*,
258 for which there is one copy on each side of *Dfd* in *D. citri* genome v3.0 (although there

259 was only one copy in the expected position between *Dfd* and *zen* in genome v2.0). The
 260 apparent split in the *D. citri* Hox cluster is between *Scr* and *miR-10*, leaving *miR-10*
 261 adjacent to *Dfd*.
 262

263 Figure 4. Phylogenetic tree of MEINOX family proteins. Full-length proteins were
 264 aligned with ClustalW in MEGAX. The tree was constructed using the neighbor-joining
 265 method with 500 bootstrap replicates. Blue branches are part of the Hth/MEIS clade and
 266 green branches indicate the PKNOX/REP clade. *Apis mellifera* (Am), *Bombyx mori* (Bm),
 267 *Tribolium castaneum* (Tc), *Drosophila melanogaster* (Dm), *Acyrtosiphon pisum* (Ap),
 268 *Nilaparvata lugens* (NI), *Diaphorina citri* (Dc), *Homo sapiens* (Hs). *D. citri* proteins are
 269 denoted by a black circle.
 270

MEINOX genes

MEINOX genes are broadly conserved cofactors of Hox genes. Like most insects, *D. citri* has two MEINOX genes (Table 2, 3). One is orthologous to *hth* and the vertebrate Meis genes, while the other is orthologous to the PREP/PKNOX proteins of insects and vertebrates (Figure 4), so we have named them *hth* and *PKNOX*. The *PKNOX* locus is incomplete in *D. citri* genome v3.0, but the complete protein sequence can be deduced from *de novo*-assembled transcripts (Table 3).

Figure 5. Expression of *exd*, *hth* and *PKNOX* genes in *D. citri* RNA-seq samples. Heatmap of relative expression levels of genes in RNA-Seq samples from various developmental stages, tissues, host plants and CLas infection states [53,70–73 and NCBI BioProjects PRJNA448935 and PRJNA609978]. Host plants are *Citrus macrophylla* (Cmac), *Citrus reticulata* (Cret), *Citrus* spp., *Citrus medica* (Cmed), and *Citrus sinensis* (Csin). The heatmap was scaled by column to allow comparison of expression different genes within samples. *PKNOX* is expressed at higher levels than *hth*-RA in the adult samples, while *hth*-RA is expressed at higher levels than *PKNOX* in many of the nymph samples. Both *exd* and *exd-r* are expressed, raising the possibility that they could both be functional.

292 Comparison of expression levels of *hth* and *PKNOX* in available *D. citri* RNA-Seq datasets
 293 [40] shows a potentially interesting difference in expression between the two genes.
 294 *PKNOX* is expressed at higher levels in most adult samples than is *hth* isoform A (*hth-A*).
 295 Conversely, *hth-A* is expressed at higher levels than *PKNOX* in some, but not all, nymph
 296 samples. (Figure 5, Table 5). Studies in vertebrates have found that the PKNOX/Prep and
 297 *hth*/Meis family proteins not only have different DNA binding sites, but also
 298 preferentially bind to the regulatory regions of different types of genes. Meis tends to
 299 bind to genes involved in embryonic development, while Prep1 is more likely to bind to
 300 genes involved in housekeeping functions [57]. Thus, it is interesting that the *D. citri*
 301 *hth*/Meis ortholog seems to be up-regulated in nymphs, where developmental genes
 302 are expected to be more important than in adults.
 303
 304 *de novo* transcript evidence indicates that *D. citri* *hth* and *PKNOX* each have two
 305 isoforms (Table 3). In both cases, one isoform produces a protein with a MEIS domain
 306 [27] and the homeodomain (Dcitr03g04000.1.1 and MCOT13578.2.CT), while the other
 307 isoform encodes a protein with a MEIS domain but no homeodomain
 308 (Dcitr03g04000.1.2 and MCOT13578.1.CT). Homeodomain-less isoforms of Hth [58] and
 309 PKNOX2 [59] have also been reported in *Drosophila* and vertebrates, respectively. In
 310 *Drosophila*, the homeodomain-less forms of Hth can perform most of the protein's usual
 311 functions, but cannot specify antennal development [58].
 312
 313 Table 5. TPM Expression Values for Hox cofactor genes

Gene ID	Dcitr08g104 00.1.1	Dcitr01g037 20.1.1	Dcitr03g040 00.1.1	Dcitr03g040 00.1.2	Dcitr00g149 70.1.1
Gene/Transcript Symbol					<i>PKNOX</i> partial
Egg C. <i>macrophylla</i> a Clas- Whole body	96.9	33.25	32.58	16.98	23.56
Nymph C. <i>medica</i> Clas+ Low infection Whole body	9.44	7.38	8.61	0.09	0
Nymph C. <i>sinensis</i> Clas+ High infection Whole body	8.52	6.25	5.56	0.4	0

Nymph <i>C. sinensis</i> CLas- Whole body	11.06	9.25	10.7	0.44	0
Nymph <i>C. macrophylla</i> <i>a</i> CLas- Whole body	37.17	10.95	10.35	0.64	7.51
Nymph <i>Citrus</i> spp. CLas- Whole body	21.9	8.24	16.46	0	5.43
Nymph <i>Citrus</i> spp. CLas+ Whole body	18.02	6.1	2.92	0.35	9.76
Adult <i>C. medica</i>	9.97	2.13	2.53	0.29	5.14

Clas- Gut					
Adult C. <i>medica</i> Clas+ Gut	5	2.26	2.88	0.21	5.72
Adult C. <i>medica</i> Clas+ High infection Whole body	12.6	9.73	3.99	0.12	0.04
Adult C. <i>medica</i> Clas+ Low infection Whole body	12.57	7.88	4.94	0.05	0
Adult C. <i>medica</i> Clas- Whole body	14.24	11.24	4.45	0	0

Adult <i>C. macrophylla</i> a Clas- Whole body	12.1	4.08	4.31	1.49	13.61
Adult <i>Citrus</i> spp. Clas- Whole body	6.09	11.49	3.76	0.85	8.69
Adult <i>Citrus</i> spp. Clas+ Whole body	13.16	11.18	2.65	1.28	12.97
Adult <i>Citrus</i> spp. Clas- midgut	26.85	3.38	5.31	0	10.85
Adult <i>Citrus</i> spp. Clas+ midgut	17.99	2.12	2.41	0.2	12.13
Adult <i>C. reticulata</i>	28.77	6.47	2.73	0	23.42

Clas- Female abdomen					
Adult C. <i>reticulata</i> Clas- Female antennae	18.43	6.89	1.57	0.98	16.45
Adult C. <i>reticulata</i> Clas- Female head	19.89	10.41	1.3	0.26	18.13
Adult C. <i>reticulata</i> Clas- Female leg	38.67	10.51	1.76	0.91	20.14
Adult C. <i>reticulata</i> Clas- Female terminal	15.75	1.89	0.35	0	22.38

abdomen					
Adult C. <i>reticulata</i> Clas- Female thorax	32.42	27.35	4.7	2.56	31.49
Adult C. <i>reticulata</i> Clas- Male abdomen	19.28	4.77	2.27	0.27	16.38
Adult C. <i>reticulata</i> Clas- Male antennae	17.31	7.91	2.32	1.63	20.16
Adult C. <i>reticulata</i> Clas- Male head	19.26	12.43	2.41	0.6	22.9
Adult C. <i>reticulata</i> Clas- Male	27.96	21.68	3.79	0.47	17.48

leg					
Adult <i>C. reticulata</i> Clas- Male terminal abdomen	15.32	2.28	0.51	0.18	13.02
Adult <i>C. reticulata</i> Clas- Male thorax	33.75	23.23	5.12	1.17	27.09
Adult <i>C. reticulata</i> Clas- Female antennae [56]	22.51	27.08	5.95	0.41	17.26
Adult <i>C. reticulata</i> Clas- Female terminal abdomen	20.19	5.46	3.47	0.52	17.48

[56]					
Adult <i>C. reticulata</i> CLas- Male antennae					
[56]	27.65	42.67	7.37	0.85	24.39
Adult <i>C. reticulata</i> CLas- Male terminal abdomen					
[56]	16.47	5.2	2.34	0	10.66

314

315 Expression levels of Hox cofactor genes used to create the heatmap in Figure 5. Counts
316 were obtained from publicly available datasets in CGEN [40] and are reported in
317 transcripts per million (TPM). Developmental stage, citrus host, CLas infection status
318 and source tissue for each sample are noted in the first column.

319

320 A

321

322

323

324 B

325

326 Figure 6. Comparison of the two *extradenticle* (*exd*) genes in *D. citri*. A) Structural
327 comparison of the *D. citri* *exd* and *extradenticle-retro* (*exd-r*) transcripts. *exd* has 7
328 exons and six introns, while *exd-r* has a single long exon with no introns. B) Multiple
329 Sequence Alignment of Exd proteins from three hemipteran species: *D. citri* Exd
330 (Dc_Ext), Exd-r (Dc_Ext-r), *Bemisia tabaci* Exd (Bt_Ext) and *Acyrthosiphon pisum* Exd
331 (Ap_Ext). Residues identical in all four proteins are shaded red, while those identical in
332 three out of four are shaded yellow. At yellow-shaded residues, *D. citri* Exd-r often
333 matches the orthologs in other species rather than its *D. citri* paralog Exd.

334

335 PBC genes

336 Exd is the insect member of the PBC class of Hox cofactors. The *D. citri* genome contains
337 two apparent *exd* orthologs rather than one as found in most other insects (Table 2-3).

338 While one of the genes has a typical eukaryotic gene structure with seven exons and six
339 introns, the other gene is unusual in that it contains no introns (Figure 6A). This gene

340 structure suggests that the second ortholog is a retrogene, which forms by

341 retrotransposition of a spliced mRNA [60]. The two *exd* genes encode proteins with 84 %

identity to one other. The scattered differences seen between the two proteins are consistent with gene duplication and divergence (Figure 6B). Moreover, the two genes map to separate chromosomes (*exd* on the X chromosome and *exd-r* on chromosome 1) and are flanked by different genes, strengthening the case for legitimate gene duplication. We have therefore named the intronless gene *extradenticle-retrogene* (*exd-r*) and the more typical gene *exd*. There is a second copy of *exd-r* adjacent to the annotated locus, but it appears to be a false duplication resulting from incomplete collapse of haplotypes during genome assembly.

exd-r does not seem to have retained the 3' poly(A) region (the vestige of a poly(A) tail) that is still present in some retrogenes [61]. This is not extremely surprising, as a study of more than 20 young retrogenes in the *D. melanogaster* genome found that almost all of them had already lost the poly(A) tail [62]. At the 5' end of *exd-r*, identity to *exd* at the nucleotide level starts at the translation start site. Such truncation is also common in retrogenes, possibly because of incomplete reverse transcription during the retrotransposition process [63].

Retrogenes do not usually include the regulatory elements of the parent gene, so they are not likely to retain their original expression pattern. This means that a retrogene will only be expressed if it is activated by regulatory elements close to its new position. If a retrogene is not expressed and therefore not subject to selection, it will often accumulate mutations and become a pseudogene. However, the *exd-r* ORF seems to be

complete, with no premature stop codons. Moreover, RNA-Seq evidence from CGEN indicates that the *exd-r* locus is expressed at a low to moderate level in most tissues and developmental stages (Figure 5, Table 5). To be sure that the expression seen in CGEN was not the result of multi-mapped or mismapped reads actually produced by the original *exd* locus, we examined the reads mapping to *exd* or *exd-r* in Apollo. The mapped reads show very few mismatches with the genomic sequence at each locus, indicating that transcripts from both genes are present in the RNA-Seq samples. Taken together, these observations suggest that *exd-r* has not become a pseudogene and could be a functional gene.

To determine if *exd-r* is evolutionarily conserved, we searched other hemipteran genomes for *exd* homologs. The closest relative of *D. citri* that has been sequenced is the hackberry petiole gall psyllid, *P. venusta* [64]. We found orthologs of both *Exd-r* (98% identity) and *Exd* (98.5% identity) encoded in the *P. venusta* genome. However, we found no evidence of an *exd* retrogene in any other genomes, including those of three other sequenced members of the suborder Sternorrhyncha: the sweet potato whitefly, *B. tabaci* [65], the greenhouse whitefly, *T. vaporariorum* [66], and the pea aphid, *A. pisum* [67]. Thus, we can place the origin of the retrogene after the separation of the psyllid lineage from the other sternorrhynchans, but before the divergence of *D. citri* and *P. venusta*.

Figure 7. Phylogenetic tree of Exd/PBC family proteins. Full-length proteins were aligned using ClustalW within MEGAX. A neighbor-joining tree was constructed with 500 bootstrap replicates. *Apis mellifera* (Am), *Bombyx mori* (Bm), *Tribolium castaneum* (Tc), *Drosophila melanogaster* (Dm), *Acyrtosiphon pisum* (Ap), *Nilaparvata lugens* (Nl), *Bemisia tabaci* (Bt), *Pachypsylla venusta* (Pv), *Diaphorina citri* (Dc), *Homo sapiens* (Hs). *D. citri* proteins are marked by a black circle.

Surprisingly, when the two *D. citri* Exd proteins are compared to other hemipteran Exd proteins, Exd-r has a higher percent identity to the orthologs (>90%) than does the “typical” *D. citri* Exd protein (80-83%). In fact, Exd-r has higher overall identity to its hemipteran orthologs than it does to the other Exd protein from *D. citri* (82.5%). Only at the extreme N-terminus does *D. citri* Exd-r appear more similar to *D. citri* Exd than to

398 Exd proteins from other hemipterans. (Figure 6B). Phylogenetic analysis is consistent
399 with these observations. The *D. citri* and *P. venusta* Exd-r proteins cluster with other
400 hemipteran Exd proteins, while their Exd proteins cluster as an outgroup to the insect
401 Exd cluster (Figure 7).

402

403 The high level of conservation between Exd-r and the Exd proteins of other species
404 supports the possibility that Exd-r still has a function. Functional retrogenes are not
405 unheard of; the *Drosophila* genome has almost 100 retrogenes that appear to be
406 functional [68]. It is more difficult to explain the observation that Exd-r more closely
407 resembles its hemipteran orthologs than it does its paralog Exd, since the paralogs
408 diverged more recently and would be expected to be more closely related. It appears
409 that Exd, rather than Exd-r as might have been expected, has diverged significantly since
410 the duplication event, resulting in decreased identity to both Exd-r and their orthologs.

411 Since the *D. citri* and *P. venusta* Exd sequences are almost identical, most of the
412 divergence must have occurred before separation of these psyllid lineages. One
413 explanation for this scenario is that *exd-r* inserted in a location that allowed it to
414 maintain its original expression pattern and function, freeing *exd* to diverge. While it is
415 usually the retrogene that acquires a new function after duplication, there is precedent
416 for the parental gene diverging in expression pattern and function. For example, the
417 *Drosophila* retrogene *e(y)2* has replaced the function of its parental gene *e(y)2b* [69].

418 The lack of additional divergence between the *D. citri* and *P. venusta* Exd proteins is
419 intriguing and suggests that Exd may have acquired a new function in the ancestral

psyllid that constrained further sequence change. Functional studies of both Exd and Exd-r will be needed to test this hypothesis.

Re-use Potential

D. citri is the focus of a major pest control effort because of its economic impact on the citrus industry as the vector of citrus greening disease. Gene-based control methods are desirable, since pesticides are not a sustainable solution. Our improved annotations will allow researchers to design more accurate, detailed experiments on genes of interest. Our revised gene models will be part of a new official gene set that will be available for BLAST analysis and expression profiling on the Citrus Greening website and the CGEN [44].

Data Availability

The *D. citri* genome assembly, gene sets, and transcriptomic data can be accessed through the Citrus Greening website [44]. Accession numbers for all genes used in phylogenetic analysis are provided within this report, and all additional data are available via the *GigaScience* GigaDB repository.

Declarations

List of Abbreviations

abd-A: abdominal-A; *Abd-B*: Abdominal-B; *Am*: *Apis mellifera*; *Antp*: *Antennapedia*; *Ap*: *Acyrtosiphon pisum*; *Bm*: *Bombyx mori*; *Bt*: *Bemisia tabaci*; CGEN: Citrus Greening Expression Network; CLas: *Candidatus Liberibacter asiaticus*; *Dc*: *Diaphorina citri*; *Dfd*: *Deformed*; *Dm*: *Drosophila melanogaster*; *exd*: *extradenticle*; *exd-r*: *extradenticle-*

442 *retrogene; ftz: fushi tarazu; Hs: Homo sapiens; hth: homothorax; lab: labial; miRNA:*
 443 *microRNA; MCOT: Maker, Cufflinks, Oases, Trinity; NCBI: National Center for*
 444 *Biotechnology Information; Nl: Nilaparvata lugens; pb: proboscipedia; Pv: Pachypsylla*
 445 *venusta; Scr: Sex combs reduced; Tc: Tribolium castaneum; TPM: transcripts per million;*
 446 *Ubx: Ultrabithorax; zen: zerknüllt*

447 **Ethical Approval**

448 Not applicable

449 **Consent for Publication**

450 Not applicable

451 **Competing Interests**

452 The authors declare that they have no competing interests.

453 **Funding**

454 This research was funded by USDA-NIFA grant 2015-70016-23028, HSI 2020-70029-
 455 33199, 2020-70029-33199 and an Institutional Development Award (IDeA) from the
 456 National Institute of General Medical Sciences of the National Institutes of Health under
 457 grant number P20GM103418.

458 **Author Contributions**

459 WBH, SJB, TD and LAM conceptualized the study; TD, SS, TDS and SJB supervised the
 460 study; SJB, TD, SS, and LAM contributed to project administration; TDS conducted the
 461 investigation; PH, MF-G, and SS contributed to software development; SS, TDS, PH, and
 462 MF-G developed methodology; SJB, TD, WBH, and LAM acquired funding; TDS prepared
 463 and wrote the original draft; SS, WBH, TD and SJB reviewed and edited the draft.

464

465 **Acknowledgements**

466 We thank Dr. Josh Benoit for assistance with using pheatmap to visualize expression
467 data. This work was supported by USDA-NIFA grant 2015- 70016-23028 and an
468 Institutional Development Award (IDeA) from the National Institute of General Medical
469 Sciences of the National Institutes of Health under grant number P20GM103418.

470

471 **References**

- 472 1. Lemons D, McGinnis W. Genomic evolution of Hox gene clusters. *Science*,
473 2006;313:1918–22.
- 474 2. Lewis EB. A gene complex controlling segmentation in *Drosophila*. *Nature*,
475 1978;276:565–70.
- 476 3. Lewis RA, Kaufman TC, Denell RE, Tollerico P. Genetic Analysis of the Antennapedia
477 Gene Complex (Ant-C) and Adjacent Chromosomal Regions of *DROSOPHILA*
478 *MELANOGASTER*. I. Polytene Chromosome Segments 84b-D. *Genetics*, 1980;95:367–81.
- 479 4. Lewis RA, Wakimoto BT, Denell RE, Kaufman TC. Genetic Analysis of the Antennapedia
480 Gene Complex (Ant-C) and Adjacent Chromosomal Regions of *DROSOPHILA*
481 *MELANOGASTER*. II. Polytene Chromosome Segments 84A-84B1,2. *Genetics*,
482 1980;95:383–97.
- 483 5. Lewis EB. Regulation of the Genes of the Bithorax Complex in *Drosophila*. *Cold Spring*
484 *Harb Symp Quant Biol.*, 1985;50:155–64.
- 485 6. Duboule D. The rise and fall of Hox gene clusters. *Development*, 2007;134:2549–60.

- 486 7. Duboule D. Temporal colinearity and the phylotypic progression: A basis for the
- 487 stability of a vertebrate Bauplan and the evolution of morphologies through
- 488 heterochrony. *Development. Dev Suppl*, 1994;120:135–42.
- 489 8. Chambeyron S, Bickmore WA. Chromatin decondensation and nuclear reorganization
- 490 of the *HoxB* locus upon induction of transcription. *Genes Dev.*, 2004;18:1119–30.
- 491 9. Monteiro AS, Ferrier DEK. Hox genes are not always Colinear. *Int. J. Biol. Sci.*,
- 492 2006;2:95–103.
- 493 10. Gould A, Morrison A, Sproat G, White RAH, Krumlauf R. Positive cross-regulation and
- 494 enhancer sharing: Two mechanisms for specifying overlapping Hox expression patterns.
- 495 *Genes Dev.*, 1997;11:900–13.
- 496 11. Zákány J, Gérard M, Favier B, Duboule D. Deletion of a *HoxD* enhancer induces
- 497 transcriptional heterochrony leading to transposition of the sacrum. *EMBO J.*,
- 498 1997;16:4393–402.
- 499 12. Sharpe J, Nonchev S, Gould A, Whiting J, Krumlauf R. Selectivity, sharing and
- 500 competitive interactions in the regulation of *Hoxb* genes. *EMBO J.*, 1998;17:1788–98.
- 501 13. Spitz F, Gonzalez F, Duboule D. A global control region defines a chromosomal
- 502 regulatory landscape containing the *HoxD* cluster. *Cell*, 2003;113:405–17.
- 503 14. Spitz F, Herkenne C, Morris MA, Duboule D. Inversion-induced disruption of the
- 504 *Hoxd* cluster leads to the partition of regulatory landscapes. *Nat Genet.*, 2005;37:889–
- 505 93.
- 506 15. Gaunt SJ. The significance of Hox gene collinearity. *Int J Dev Biol.*, 2015;59:159–70.
- 507 16. Beeman RW, Stuart JJ, Haas MS, Denell RE. Genetic analysis of the homeotic gene

508 complex (HOM-C) in the beetle *Tribolium castaneum*. Dev Biol., 1989;133:196–209.

509 17. Garcia-Fernàndez J, Holland PWH. Archetypal organization of the amphioxus Hox
510 gene cluster. Nature, 1994;370:563–6.

511 18. Pace RM, Grbić M, Nagy LM. Composition and genomic organization of arthropod
512 Hox clusters. Evodevo, 2016;https://doi.org/10.1186/s13227-016-0048-4.

513 19. Von Allmen G, Hogga I, Spierer A, Karch F, Bender W, Gyurkovics H, et al. Splits in
514 fruitfly Hox gene complexes. Nature, 1996;380:116.

515 20. Negre B, Ruiz A. HOM-C evolution in Drosophila: is there a need for Hox gene
516 clustering? Trends Genet., 2007;23:55–9.

517 21. Yasukochi Y, Ashakumary LA, Wu C, Yoshido A, Nohata J, Mita K, et al. Organization
518 of the Hox gene cluster of the silkworm, *Bombyx mori*: A split of the Hox cluster in a
519 non-Drosophila insect. Dev Genes Evol., 2004;214:606–14.

520 22. Garaulet DL, Lai EC. Hox miRNA regulation within the Drosophila Bithorax complex:
521 Patterning behavior. Mech Dev., 2015;138:151–9.

522 23. Moens CB, Selleri L. Hox cofactors in vertebrate development. Dev Biol.,
523 2006;291:193–206.

524 24. Casares F, Mann RS. Control of antennal versus leg development in Drosophila.
525 Nature, 1998;392:723–6.

526 25. Aldaz S, Morata G, Azpiazu N. Patterning function of *homothorax/extradenticle* in
527 the thorax of Drosophila. Development, 2005;132:439–46.

528 26. Laurent A, Bihan R, Omilli F, Deschamps S, Pellerin I. PBX proteins: Much more than
529 Hox cofactors. Int. J. Dev. Biol., 2008;52:9–20.

530 27. Brglin TR. Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois,
531 TGIF) reveals a novel domain conserved between plants and animals. Nucleic Acids Res.,
532 1997;25:4173–80.

533 28. Mukherjee K, Brglin TR. Comprehensive analysis of animal TALE homeobox genes:
534 New conserved motifs and cases of accelerated evolution. J Mol Evol., 2007;65:137–53.

535 29. Shigenobu S, Bickel RD, Brisson JA, Butts T, Chang C, Christiaens O, et al.
536 Comprehensive survey of developmental genes in the pea aphid, *Acyrtosiphon pisum*:
537 frequent lineage-specific duplications and losses of developmental genes. Insect Mol
538 Biol., 2010;19:47–62.

539 30. Rauskolb C, Peifer M, Wieschaus E. extradenticle, a regulator of homeotic gene
540 activity, is a homolog of the homeobox-containing human proto-oncogene *pbx1*. Cell,
541 1993;74:1101–12.

542 31. Kurant E, Pai CY, Sharf R, Halachmi N, Sun YH, Salzberg A. *dorsotonals/homothorax*,
543 the Drosophila homologue of *meis1* interacts with *extradenticle* in patterning of the
544 embryonic PNS. Development, 1998;125:1037–48.

545 32. Hosmani P, Flores-Gonzalez M, Shippy T, Vosburg C, Massimino C, Tank W, et al.
546 Chromosomal length reference assembly for *Diaphorina citri* using single-molecule
547 sequencing and Hi-C proximity ligation with manually curated genes in developmental,
548 structural and immune pathways. bioRxiv. Cold Spring Harbor Laboratory; 2019;869685.

549 33. Hughes CL, Kaufman TC. RNAi analysis of *Deformed*, *proboscipedia* and *Sex combs*
550 *reduced* in the milkweed bug *Oncopeltus fasciatus*: novel roles for Hox genes in the
551 hemipteran head. Development, 2000;127:3683–94.

552 34. Angelini DR, Liu PZ, Hughes CL, Kaufman TC. Hox gene function and interaction in the
553 milkweed bug *Oncopeltus fasciatus* (Hemiptera). Dev Biol., 2005;287:440–55.

554 35. International Aphid Genomics Consortium. Genome Sequence of the Pea Aphid
555 *Acyrtosiphon pisum*. PLoS Biol, 2010;8:e1000313.

556 36. Panfilio KA, Liu PZ, Akam M, Kaufman TC. *Oncopeltus fasciatus zen* is essential for
557 serosal tissue function in katatrepsis. Dev Biol., 2006;292:226-43;
558 <https://doi.org/10.1016/j.ydbio.2005.12.028>.

559 37. Pick L. Hox genes, evo-devo, and the case of the *ftz* gene. Chromosoma,
560 2016;125:535–51.

561 38. Panfilio KA, Vargas Jentsch IM, Benoit JB, Erezylmaz D, Suzuki Y, Colella S, et al.
562 Molecular evolutionary trends and feeding ecology diversification in the Hemiptera,
563 anchored by the milkweed bug genome. Genome Biol.,
564 2019;<https://doi.org/10.1186/s13059-019-1660-0>

565 39. Ren Z-W, Zhuo J-C, Zhang C-X, Wang D. Characterization of *NIHox3*, an essential gene
566 for embryonic development in *Nilaparvata lugens*. Arch Insect Biochem Physiol.,
567 2018;98:e21448.

568 40. Flores-Gonzalez M, Hosmani P, Fernandez-Pozo N, Mann M, Humann J, Main D, et
569 al. Citrusgreening.org: An open access and integrated systems biology portal for the
570 Huanglongbing (HLB) disease complex. bioRxiv. Cold Spring Harbor Laboratory;
571 2019;868364.

572 41. Dunn NA, Unni DR, Diesh C, Munoz-Torres M, Harris NL, Yao E, et al. Apollo:
573 Democratizing genome annotation. PLoS Comput Biol., 2019;15:e1006790.

574 42. Shippy, TD; Miller, S; Massimino, C; Vosburg, C; Hosmani, PS; Flores-Gonzalez, M;
575 Mueller, LA; Hunter, WB; Benoit, JB; Brown, SJ; D’elia, T; Saha S. Annotating genes in
576 *Diaphorina citri* genome version 3. protocols.io. 2020;
577 <https://dx.doi.org/10.17504/protocols.io.bniimcce>.

578 43. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary
579 genetics analysis across computing platforms. Mol Biol Evol., 2018;35:1547–9.

580 44. Citrus Greening Solutions. <https://www.citrusgreening.org>
581 (2019). Accessed 31 Aug 2021.

582 45. Saha S, Hosmani PS, Villalobos-Ayala K, Miller S, Shippy T, Flores M, et al. Improved
583 annotation of the insect vector of citrus greening disease: biocuration by a diverse
584 genomics community. Database, 2017;<https://doi.org/10.1093/database/bax032>.

585 46. Denton JF, Lugo-Martinez J, Tucker AE, Schrider DR, Warren WC, Hahn MW.
586 Extensive Error in the Number of Genes Inferred from Draft Genome Assemblies. PLoS
587 Comput Biol., 2014;<https://doi.org/10.1371/journal.pcbi.1003998>.

588 47. Flybase. <http://flybase.org> (2021). Accessed 1 Sep 2021.

589 48. Dearden PK, Wilson MJ, Sablan L, Osborne PW, Havler M, McNaughton E, et al.
590 Patterns of conservation and change in honey bee developmental genes. Genome Res.,
591 2006;16:1376–84.

592 49. Shippy TD, Ronshaugen M, Cande J, He JP, Beeman RW, Levine M, et al. Analysis of
593 the *Tribolium* homeotic complex: Insights into mechanisms constraining insect Hox
594 clusters. Dev Genes Evol., 2008;218:127-39.

595 50. O’Neill MP. A sub-set of the Hox genes in the pea aphid, *Acyrtosiphon pisum*.

596 University of Otago; 2012.

597 51. Legeai F, Rizk G, Walsh T, Edwards O, Gordon K, Lavenier D, et al. Bioinformatic
598 prediction, deep sequencing of microRNAs and expression analysis during phenotypic
599 plasticity in the pea aphid, *Acyrtosiphon pisum*. BMC Genomic, 2010;[https://doi.org/](https://doi.org/10.1186/1471-2164-11-281)
600 10.1186/1471-2164-11-281.

601 52. Miura S, Nozawa M, Nei M. Evolutionary changes of the target sites of two
602 MicroRNAs encoded in the Hox gene cluster of Drosophila and other insect species.
603 Genome Biol Evol., 2011;3:129–39.

604 53. Wu Z, Zhang H, Bin S, Chen L, Han Q, Lin J. Antennal and abdominal transcriptomes
605 reveal chemosensory genes in the Asian Citrus Psyllid, *Diaphorina citri*. PLoS One,
606 2016;<https://doi.org/10.1371/journal.pone.0159372>.

607 54. Tehler D, Høyland-Kroghsbo NM, Lund AH. The miR-10 microRNA precursor family.
608 RNA Biol., 2011;<https://doi.org/10.4161/rna.8.5.16324>.

609 55. Stark A, Bushati N, Jan CH, Kheradpour P, Hodges E, Brennecke J, et al. A single Hox
610 locus in Drosophila produces functional microRNAs from opposite DNA strands. Genes
611 Dev., 2008;22:8–13.

612 56. Hui JHL, Marco A, Hunt S, Melling J, Griffiths-Jones S, Ronshaugen M. Structure,
613 evolution and function of the bi-directionally transcribed iab-4/iab-8 microRNA locus in
614 arthropods. Nucleic Acids Res., 2013;41:3352–61.

615 57. Penkov D, San Martín DM, Fernandez-Díaz LC, Rosselló CA, Torroja C, Sánchez-Cabo
616 F, et al. Analysis of the DNA-Binding Profile and Function of TALE Homeoproteins
617 Reveals Their Specialization and Specific Interactions with Hox Genes/Proteins. Cell Rep.,

618 2013;3:1321–33.

619 58. Noro B, Culi J, McKay DJ, Zhang W, Mann RS. Distinct functions of homeodomain-
620 containing and homeodomain-less isoforms encoded by *homothorax*. Genes Dev.,
621 2006;20:1636–50.

622 59. Longobardi E, Penkov D, Mateos D, De Florian G, Torres M, Blasi F. Biochemistry of
623 the tale transcription factors PREP, MEIS, and PBX in vertebrates. Dev. Dyn., 2014;8:59–
624 75.

625 60. Brosius J. Retroposons--seeds of evolution. Science, 1991;251:753.

626 61. Vanin EF. Processed Pseudogenes: Characteristics and Evolution. Annu Rev Genet.,
627 1985;19:253–72.

628 62. Betrán E, Thornton K, Long M. Retroposed new genes out of the X in Drosophila.
629 Genome Res., 2002;12:1854–9.

630 63. Pavlíček A, Pačes J, Zíka R, Hejnar J. Length distribution of long interspersed nuclear
631 elements (LINEs) and processed pseudogenes of human endogenous retroviruses:
632 Implications for retrotransposition and pseudogene detection. Gene, 2002;300:189–94.

633 64. Li Y, Zhang B, Moran NA. The Aphid X Chromosome is a Dangerous Place for
634 Functionally Important Genes: Diverse Evolution of Hemipteran Genomes Based on
635 Chromosome-level Assemblies. Mol Biol Evol., 2020;37:2357-2368.

636 65. Chen W, Wosula EN, Hasegawa DK, Casinga C, Shirima RR, Fiaboe KKM, et al.
637 Genome of the African cassava whitefly *Bemisia tabaci* and distribution and genetic
638 diversity of cassava-colonizing whiteflies in Africa. Insect Biochem Mol Biol.,
639 2019;110:112–20.

640 66. Xie W, He C, Fei Z, Zhang Y. Chromosome-level genome assembly of the greenhouse
641 whitefly (*Trialeurodes vaporariorum* Westwood). *Mol Ecol Resour.*, 2020;20:995–1006.

642 67. Li Y, Park H, Smith TE, Moran NA. Gene Family Evolution in the Pea Aphid Based on
643 Chromosome-Level Genome Assembly. *Mol Biol Evol.*, 2019;36:2143–56.

644 68. Bai Y, Casola C, Feschotte C, Betrán E. Comparative genomics reveals a constant rate
645 of origination and convergent acquisition of functional retrogenes in *Drosophila*.
646 *Genome Biol.*, 2007;<https://doi.org/10.1186/gb-2007-8-1-r11>.

647 69. Krasnov AN, Kurshakova MM, Ramensky VE, Mardanov P V., Nabirochkina EN,
648 Georgieva SG. A retrocopy of a gene can functionally displace the source gene in
649 evolution. *Nucleic Acids Res.*, 2005;<https://doi.org/10.1093/NAR/GKI969>.

650 70. Kruse A, Fattah-Hosseini S, Saha S, Johnson R, Warwick ER, Sturgeon K, et al.
651 Combining omics and microscopy to visualize interactions between the Asian citrus
652 psyllid vector and the Huanglongbing pathogen *Candidatus Liberibacter asiaticus* in the
653 insect gut. *PLoS One*, 2017;<https://doi.org/10.1371/journal.pone.0179531>.

654 71. Reese J, Christenson MK, Leng N, Saha S, Cantarel B, Lindeberg M, et al.
655 Characterization of the Asian Citrus Psyllid Transcriptome. *J Genomics*, 2013;2:54–8.

656 72. Vyas M, Fisher TW, He R, Nelson W, Yin G, Cicero JM, et al. Asian citrus psyllid
657 expression profiles suggest *Candidatus Liberibacter asiaticus*-mediated alteration of
658 adult nutrition and metabolism, and of nymphal development and immunity. *PLoS One*,
659 2015;<https://doi.org/10.1371/journal.pone.0130328>.

660 73. Yu HZ, Li NY, Zeng XD, Song JC, Yu XD, Su HN, et al. Transcriptome analyses of
661 *Diaphorina citri* midgut responses to *Candidatus Liberibacter asiaticus* infection. *Insects*,

662 2020;<https://doi.org/10.3390/insects11030171>.

663

664