



HAL
open science

**Annotation of transcription factors,
chromatin-associated factors, and basal transcription
machinery in the pea aphid, *Acyrtosiphon pisum*, and
development of the ATFdb database, a resource for
studies of transcriptional regulation**

Nicolas Parisot, Mélanie Ribeiro Lopes, Sergio Peignier, Patrice
Baa-Puyoulet, Hubert Charles, Federica Calevro, Patrick Callaerts

► **To cite this version:**

Nicolas Parisot, Mélanie Ribeiro Lopes, Sergio Peignier, Patrice Baa-Puyoulet, Hubert Charles, et al.. Annotation of transcription factors, chromatin-associated factors, and basal transcription machinery in the pea aphid, *Acyrtosiphon pisum*, and development of the ATFdb database, a resource for studies of transcriptional regulation. *Insect Biochemistry and Molecular Biology*, 2024, pp.104217. 10.1016/j.ibmb.2024.104217 . hal-04832131

HAL Id: hal-04832131

<https://hal.science/hal-04832131v1>

Submitted on 12 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Annotation of transcription factors, chromatin-associated factors, and basal transcription machinery in the pea aphid, *Acyrtosiphon pisum*, and development of the ATFdb database, a resource for studies of transcriptional regulation

Nicolas Parisot^{a,*}, Mélanie Ribeiro Lopes^a, Sergio Peignier^a, Patrice Baa-Puyoulet^a, Hubert Charles^a, Federica Calevro^{a,**}, Patrick Callaerts^{b,***}

^a INSA Lyon, INRAE, BF2I, UMR0203, F-69621, Villeurbanne, France

^b KU Leuven, University of Leuven, Department of Human Genetics, Laboratory of Behavioral and Developmental Genetics, B-3000, Leuven, Belgium

ARTICLE INFO

Keywords:

Hemiptera
Pea aphid
Transcription factors
Chromatin
Basal transcription machinery
TF database

ABSTRACT

The pea aphid, *Acyrtosiphon pisum*, is an emerging model system in functional and comparative genomics, in part due to the availability of new genomic approaches and the different sequencing and annotation efforts that the community has dedicated to this important crop pest insect. The pea aphid is also used as a model to study fascinating biological traits of aphids, such as their extensive polyphenisms, their bacteriocyte-confined nutritional symbiosis, or their adaptation to the highly unbalanced diet represented by phloem sap. To get insights into the molecular basis of all these processes, it is important to have an appropriate annotation of transcription factors (TFs), which would enable the reconstruction/inference of gene regulatory networks in aphids. Using the latest version of the *A. pisum* genome assembly and annotation, which represents the first chromosome-level pea aphid genome, we annotated the complete repertoire of *A. pisum* TFs and complemented this information by annotating genes encoding chromatin-associated and basal transcription machinery proteins. These annotations were done combining information from the model *Drosophila melanogaster*, for which we also provide a revisited list of these proteins, and *de novo* prediction. The comparison between the two model systems allowed the identification of major losses or expansions in each genome, while a deeper analysis was made of ZNF TFs (with certain families expanded in the pea aphid), and the Hox gene cluster (showing reorganization in gene position in the pea aphid compared to *D. melanogaster*). All annotations are available to the community through the Aphid Transcription Factors database (ATFdb), consolidating the various annotations we generated. ATFdb serves as a valuable resource for gene regulation studies in aphids.

1. Introduction

Changes in gene expression arise in response to internal and external signals and are essential for translating genotypes into phenotypes at cellular, tissue, and organismal levels. Gene expression alterations are observed, for example, during development (Martín et al., 2016; Spitz and Furlong, 2012), upon cellular differentiation (Peñalosa-Ruiz et al., 2019), in metabolic homeostasis and physiological regulations (Desvergne et al., 2006), and in response to environmental stimuli (Han

and Kaufman, 2017; Song et al., 2016). Gene expression is controlled to an important extent by Transcription Factors (TFs), which bind accessible cis-regulatory DNA elements to positively or negatively regulate the expression of target genes (Hammonds et al., 2013; Hens et al., 2011; Kim and Wysocka, 2023; Lambert et al., 2018). The accessibility of DNA depends on chromatin state and can be enhanced by TFs (Zaret, 2020). Sets of TFs and their target genes are organized in so-called Gene Regulatory Networks (GRNs) that govern cellular gene expression (Aerts, 2012; Davidson and Levine, 2008; Singh et al., 2018). Understanding

* Corresponding author.

** Corresponding author.

*** Corresponding author.

E-mail addresses: nicolas.parisot@insa-lyon.fr (N. Parisot), federica.calevro@insa-lyon.fr (F. Calevro), patrick.callaerts@kuleuven.be (P. Callaerts).

<https://doi.org/10.1016/j.ibmb.2024.104217>

Received 29 March 2024; Received in revised form 15 October 2024; Accepted 19 November 2024

Available online 22 November 2024

0965-1748/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

GRNs is essential for deciphering the complex regulation of gene expression, which often involves multiple players acting synergistically or antagonistically. Altered specificity or activity of TFs or changes in cis-regulatory sequences impact the structure or activity of GRNs and represent an important source of phenotypic diversity and evolutionary adaptation (Carroll, 2008; Schember and Halfon, 2022).

TFs are a large group of evolutionary conserved regulatory proteins. Their modular structure determines their ability to interact with DNA sequences through DNA binding domains (DBDs), to exert their function as positive or negative regulators of target gene expression through transcriptional regulatory domains, or even to associate with each other or with co-activating/repressing proteins through oligomerization (Gonzalez, 2016; Lemon and Tjian, 2000; Näär et al., 2001). This modular nature is one of the key drivers of TF evolution and functional diversification, with domains that can individually undergo mutation and be lost or gained, generating new combinations over time (Cheatle Jarvela and Hinman, 2015). TFs are classified into several superfamilies, families and sub-families, based on structural similarities, primarily of their DBDs (Luscombe et al., 2000). Three major superfamilies are the C₂H₂ zinc finger, homeodomain and helix-loop-helix, which together account for more than 80% of all predicted human TFs (Vaquerizas et al., 2009) and around 70% of TFs in the fruit fly *Drosophila melanogaster* (Gramates et al., 2022).

For over a century, *D. melanogaster* has been used as a model system to dissect and understand genetics, developmental biology, organ specification and function, physiology and metabolism (Bellen and Yamamoto, 2015; Dow et al., 2022; Fischer et al., 2023; Helfand and Rogina, 2003; Mohr and Perrimon, 2019). Many sophisticated biochemical, molecular, and cellular approaches have been developed in *D. melanogaster* and were used to functionally validate TFs (Shokri et al., 2019). Consequently, the *Drosophila* genome is one of the best characterized metazoan genomes in terms of functionally annotated genes and regulatory elements (Celniker and Rubin, 2003; Chen et al., 2014; Yang et al., 2018). In recent years, the increasing number of sequenced genomes has accelerated TF predictions and annotations in diverse insects beyond *Drosophila* (i5K Consortium, 2013; Sproul et al., 2023). The pea aphid *Acyrtosiphon pisum* has emerged as an important model system in functional and comparative genomics (The International Aphid Genomics Consortium, 2010), in part due to the availability of new genomic approaches (Calevro et al., 2019; Le Trionnaire et al., 2019; Sapountzis et al., 2014; Tagu et al., 2016) and the high annotation quality of its genome (Li et al., 2019). Aphids are a great model to study extensive polyphenisms, including parthenogenetic and sexual reproduction and the presence of winged and non-winged morphs (Davis et al., 2021; Ogawa and Miura, 2014). They are one of the best studied models of nutritional symbioses and have bacteriocytes, a specialized cell type that houses the obligate endosymbiont *Buchnera aphidicola* (Baumann, 2005; Calevro et al., 2023; Simonet et al., 2018). Not only the developmental origin of this novel cell type remains unknown, but recent work has shown extensive plasticity and transcriptional reprogramming of these cells in response to environmental challenges (Colella et al., 2018; Ribeiro Lopes et al., 2022). Finally, aphids are phloem sap-sucking insects that are highly adapted to this unbalanced diet (Douglas, 2015). We propose that all these processes are supported by the interaction of TFs with their target genes and by GRNs that are either modified from preexisting GRNs or are novel and possibly aphid specific.

To facilitate studies on gene regulation and GRN inference in aphids, we annotated the TFs, chromatin-associated proteins, and basal transcription machinery in the pea aphid, at both the family and gene levels. A first effort by Shigenobu et al. (2010) made use of the first pea aphid genome assembly but only precisely annotated developmentally important TFs, limiting the rest of the annotation to a DBD-based classification. Other analyses, focusing on the basic helix-loop-helix TF family in *A. pisum* (Dang et al., 2011) or on the annotation of TFs and chromatin-related proteins in different hemipteran species including *A. pisum* (Vidal et al., 2016), have also been made on the first pea aphid

genome assembly. Since then, the availability of the first chromosome-scale assembly for the pea aphid (Li et al., 2019) resulted in a significant adjustment of predicted protein-coding gene numbers, making a new and comprehensive annotation a necessity. This new assembly represents a significant improvement over previous versions, which were highly fragmented and known to contain many mis-assembled scaffolds. In addition to enabling a more reliable prediction of the number of genes making up each TF family, by substantially reducing redundancy of the gene set, it allows to study the position of genes relative to each other on chromosomes, which is particularly useful in the context of conserved gene clusters such as the Hox gene cluster in our study. Taking advantage of the knowledge accumulated in *Drosophila*, and integrating it with a *de novo* annotation approach using TFs databases including information from other organisms, we predicted 854 TF-coding genes in *A. pisum*. In addition, we annotated 230 chromatin-associated genes and 67 genes belonging to the basal transcription machinery. Expression of selected genes in different tissues was analyzed using publicly available RNA-seq libraries, confirming that their expression in the different aphid tissues is consistent with the biological functions that could be predicted on the basis of our annotation. All annotations are compiled in a dedicated database, the Aphid Transcription Factors database (ATFdb) (<http://atf.cycadsys.org>). This complete annotation is expected to be of use to the community studying aphids and other hemipterans, prominent crop pest insects, as well as to researchers studying insect development and evolution.

2. Materials and methods

2.1. Definition of a list of bona fide *D. melanogaster* transcription factors (TFs)

D. melanogaster TF annotations were retrieved from Hens et al. (2011) and Hammonds et al. (2013). These lists were merged and then filtered to remove splicing variants that had been annotated as independent genes, or genes that have since been (re)annotated as not being TFs or as being a component of the basal transcription machinery or associated with chromatin. The remaining TFs were compared to the “Transcription factors” gene group (FBgg0000745) available in FlyBase release 6.55 (Gramates et al., 2022), to the precompiled *Drosophila* TF predictions from the DBD database v2.0 (Kummerfeld and Teichmann, 2006), and to the TF lists extracted from the relevant databases FlyMine v48 (Lyne et al., 2007) and FlyNet from release 6.54 of FlyBase (Tian et al., 2009). *De novo* predictions were made based on the *Drosophila melanogaster* genome release 6.55 using the transcription factor prediction tools from PlantTFDB (Jin et al., 2017) and Pfam using TF domains of the DBD database (Mistry et al., 2021; Wenger, 2018; Wilson et al., 2008). The refinement strategy is summarized in the upper part of Fig. 1.

2.2. Prediction of *Acyrtosiphon pisum* transcription factors

All predicted proteins of *A. pisum* were downloaded from the NCBI datasets using the latest RefSeq Annotation available (Annotation release 103, dated May 24, 2021) and the reference pea aphid genome assembly (assembly ID: GCF_005508785.2), also published in Li et al. (2019). A first BLASTP (Altschul et al., 1990) analysis was performed against the *A. pisum* proteins using our new list of bona fide *D. melanogaster* TF protein sequences as a query with an e-value cutoff of 1e-5. BLASTP results were manually curated. To confirm the homology relationships and to remove proteins that are not homologs, reverse BLASTP analyses (*i.e.* using the *A. pisum* candidate protein as query against *D. melanogaster*) were performed using as query the *A. pisum* proteins corresponding to the top ten first BLASTP hits against all *D. melanogaster* proteins. When both BLASTP analyses yielded the same *A. pisum/D. melanogaster* homologous protein pairs (one to one relationships) the *A. pisum* orthologous protein was annotated as such according to *D. melanogaster* nomenclature. When multiple *A. pisum* TFs

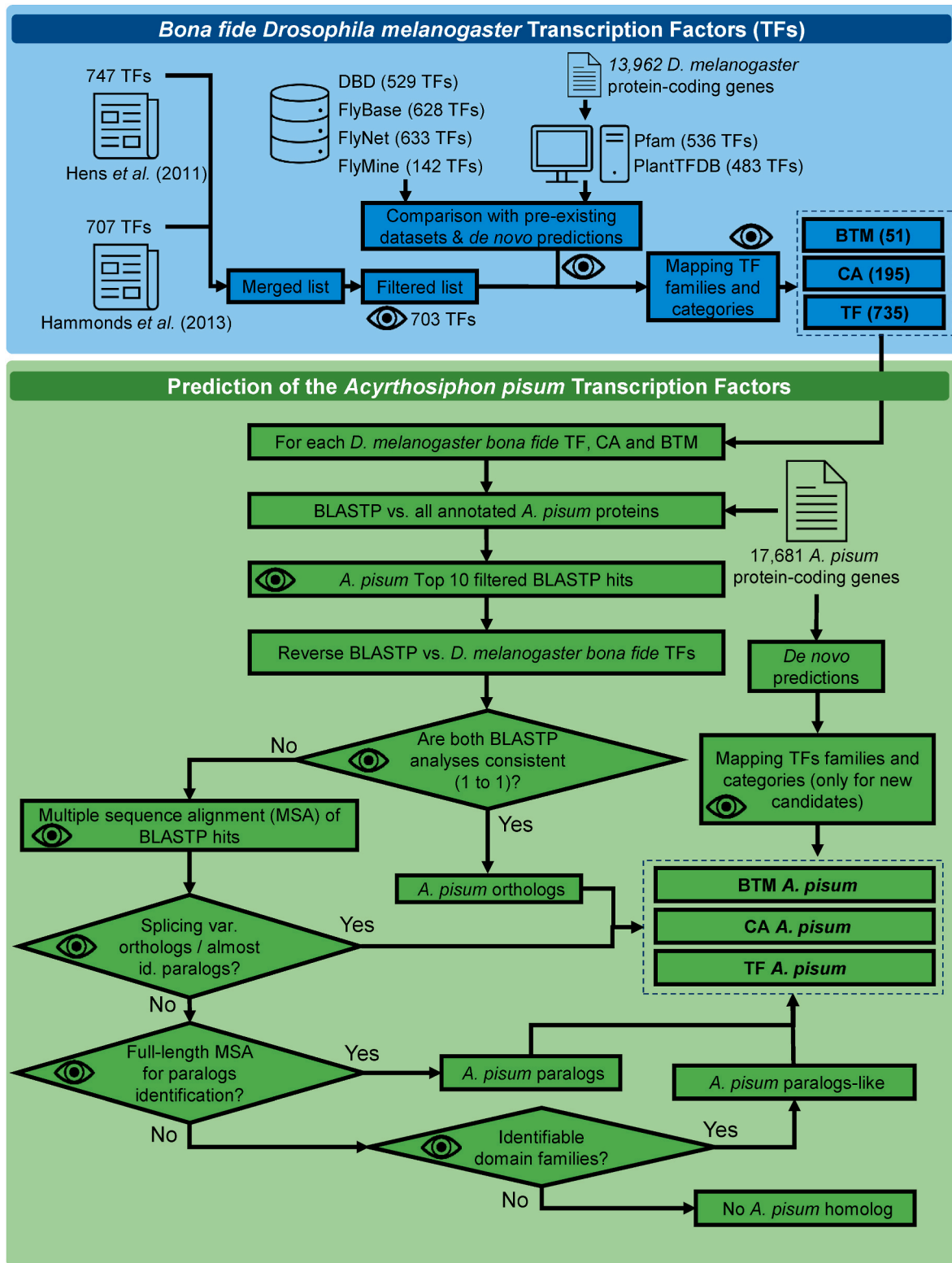


Fig. 1. Transcription factor annotation pipeline in *Drosophila melanogaster* (top blue) and *Acyrthosiphon pisum* (bottom green). BTM: Basal Transcription Machinery, CA: Chromatin-Associated, TF: Transcription Factor. The eye symbol indicates when manual annotation was required.

gave the same *Drosophila* TF as the top hit, then the *A. pisum* TF sequences were aligned pairwise using BLAST Global Alignment to identify identical protein-coding sequences, i.e., putative splicing variants in *A. pisum* or almost identical copies homologous to one *Drosophila* TF. Nearly identical sequences in *A. pisum* were annotated based on *Drosophila* nomenclature, with chromosomal localization used to

discriminate between splicing variants (one localization) and paralogs (several localizations). In cases where the divergence between *A. pisum* and *D. melanogaster* hits did not allow an orthologous relationship to be defined, the full-length amino acid sequences were aligned and a manual multi-criteria expert judgement was used to distinguish the paralogs, which were then numbered and named according to the *Drosophila*

nomenclature. In most cases, sequence bordering the DNA-binding domain contains unique residues that allow unambiguous identification. If the divergence from the *D. melanogaster* sequence was too strong, but the DNA-binding domain identified it as belonging to the same TF family, the suffix "-like" was added to the end of the gene name symbol. Finally, any remaining sequences that were too short or too divergent and lacked an identifiable DNA-binding domain were removed. Since the approach described above would not identify TFs that are not homologous to *Drosophila*, a *de novo* prediction was done in parallel on all 17,681 *A. pisum* protein-coding genes using the same prediction tools as for *Drosophila* (i.e., PlantTFDB and Pfam).

The global strategy we used to annotate the *A. pisum* TFs is illustrated in the lower part of Fig. 1.

2.3. Prediction of chromatin associated proteins and basal transcription machinery

Analysis of the gene lists provided by Hens et al. (2011) and Hammonds et al. (2013) revealed some overlap between TFs and chromatin-associated proteins, or proteins belonging to the basal transcription machinery. This stimulated us to produce a complete list of the latter, for which an exhaustive single repository was not available, using information derived from FlyBase, and expert literature. A master list of *D. melanogaster* chromatin-associated proteins was thus compiled and used to search for homologs in the *A. pisum* genome, as detailed above (see § 2.2 and Fig. 1). For chromatin-associated proteins also functioning as TFs, they were double-listed (as TFs and as chromatin-associated proteins). Lastly, the basal transcription machinery of *A. pisum* was annotated following the same procedure. All identified TFs, chromatin-associated genes and genes involved in the transcription machinery were classified into families and subfamilies based on the available information listed in § 2.1.

2.4. Analysis of *A. pisum* TF gene expression

A total of 123 public RNA-seq libraries from 7 different *A. pisum* tissues (Suppl. File 1) were automatically retrieved from the NCBI Short Read Archive using fasterq-dump v3.0.1 (<https://github.com/ncbi/sra-tools>). Raw reads were processed using Trimmomatic v0.39 (Bolger et al., 2014) to remove adapters and filter out reads shorter than 35 bp and reads that had a mean quality value lower or equal to 25. Clean reads were then mapped against the *A. pisum* genome assembly (GCF_005508785.1) using the subread-align method from the Subread package v2.0.1 (Liao et al., 2019). Strand-specific gene counts from each library were obtained with featureCounts (Liao et al., 2019) and concatenated in a single count matrix, which was normalized using standard functions from the Scanpy library (Wolf et al., 2018). First we applied the `pp.normalize_total` function to normalize each library by its total number of counts over all genes, in order to avoid differences in total number of counts for the different libraries. Here the final sum of counts per library after normalization was set to 1 million reads. Then, we applied a logarithmic transformation to the counts, using the `pp.log1p` function with a pseudocount equal to one read. Next, we applied a first z-score along each library vector of counts, to make the level of expression per library comparable, centered around zero and with unitary standard deviation. Finally, a second gene-wise z-score was applied to ensure that genes had comparable distributions in terms of expression levels. To identify marker genes that characterize a given tissue, we used the `tl.rank_genes_groups` function from the scanpy library (Wolf et al., 2018) with default parameters, to detect the five most highly differential genes in the tissue of interest.

2.5. Annotation of HNF4 homologs in Hemiptera

Following the identification of a HNF4 homolog in *A. pisum*, the search was extended to 14 additional species: six aphid species (*Aphis*

gossypii, *Diuraphis noxia*, *Melanaphis sacchari*, *Myzus persicae*, *Rhopalosiphum maidis* and *Sipha flava*, all members of the Aphididae family), two aphid-related species (*Adelges cooleyi*, member of the Adelgidae family and *Daktulosphaira vitifoliae*, member of the Phylloxeridae family) and six other hemipterans (*Bemisia tabaci*, *Cimex lectularius*, *Diaphorina citri*, *Halyomorpha halys*, *Homalodisca vitripennis* and *Nilavarvata lugens*) (Table 1). We specifically selected hemipteran genomes with NCBI RefSeq annotations for analysis due to their quality reference for genome annotation and gene identification. The protein sequences of *A. pisum* HNF4 (identified in this study) and *D. melanogaster* HNF4 were used as query to perform BLASTP searches against the RefSeq proteome of the 14 insect species listed above. For each species, reverse BLASTP searches (i.e. using the candidate protein as query against the *D. melanogaster* and *A. pisum* protein sets) were performed on the top ten BLASTP hits. Additionally, candidate sequences were analyzed with InterProScan v5.59–91.0 to identify functional domains (Paysan-Lafosse et al., 2023). Candidate proteins were considered *bona fide* HNF4 homologs only if the reverse BLASTP analyses yielded HNF4 in the top ten hits and if the InterProScan analyses revealed the presence of a DBD (InterPro signature IPR049636) and/or ligand binding domain (InterPro signature IPR049635) that are typical of HNF4 TFs. Alignments were performed using the Kalign multiple sequence alignment web service from EMBL-EBI (Madeira et al., 2022) with default parameters. Graphical representations of the alignments were performed using the ESPript v3.0 web service (Robert and Gouet, 2014).

2.6. Hox cluster gene annotation in Hemiptera

To better understand the evolution of the Hox gene cluster in hemipterans, genes belonging to this cluster were annotated in the 14 additional species listed in Table 1 (see § 2.4). For four of them (*A. gossypii*, *R. maidis*, *H. vitripennis* and *N. lugens*) chromosome-level assemblies were available, and for one (*D. vitifoliae*) genes belonging to the Hox gene cluster were located on the same genomic scaffold. This permitted cross-species comparison of the chromosomal organization of the Hox gene cluster. Protein sequences encoded by the *A. pisum* Hox cluster genes identified in this study were used as query to perform BLASTP searches against the NCBI RefSeq proteomes of the 14 selected insect species. For each species, the BLASTP hits with the 10% highest identity scores were subsequently blasted back against the *A. pisum* reference proteins. When both BLASTP analyses yielded the same homologous protein pairs as top hits, these were annotated as such. When there were discrepancies between the two BLASTP results, semi-manual curation involved (i) repeating the analysis with the *D. melanogaster* Hox cluster protein sequences as query, (ii) looking at the relative position of each Hox cluster gene relative to the others on the genome and (iii) looking for conserved residues unique to certain Hox orthology groups, which are shared across bilaterians (de Rosa et al., 1999). The annotation was further confirmed through inference of the phylogenetic relationships between the annotated Hox genes. Phylogenetic analyses were performed using amino acid sequences corresponding to the homeodomain and the ten amino acids flanking the N- and C-terminal ends, in a strategy adapted from previous studies (Cook et al., 2001; Fröblius and Funch, 2017). Sequences were aligned with Kalign (Lassmann and Sonnhammer, 2005) using default parameters and a bayesian tree was computed with MrBayes v3.2.7 (Ronquist et al., 2012) with a mixed model of amino acid substitution. MrBayes was run with one chain for 21 million generations and trees were sampled every 500 generations. Phylogenetic trees were visualized and annotated using Itol v6 (<https://itol.embl.de/itol.cgi>) (Letunic and Bork, 2019). Systematic BLASTP searches were performed to compare the protein sequences of each predicted Hox gene with the complete set of homologs identified in hemipterans or model species, enabling the level of inter- and intra-specific divergence within the Hox cluster to be assessed (see § 3, Results and Discussion).

This annotation strategy was also used to reannotate the canonical

Table 1

Genomic information (assembly and annotation numbers) used for the annotation of HNF4 and the Hox cluster genes in hemipteran insects and model species representative of Diptera, Coleoptera and Lepidoptera.

Species	Assembly accession number	Assembly size (Mbp)	Assembly level	Annotation version	Gene number	Gene mean size (bp)	Hox cluster size (Mbp)
Model species							
<i>Drosophila melanogaster</i>	GCF_000001215.4	143.7	Chromosome	FlyBase Release 6.55	17,873	5,836	10.3
<i>Tribolium castaneum</i>	GCF_000002335.3	165.9	Chromosome	NCBI Annotation Release 103	14,322	8,032	0.76
<i>Bombyx mori</i>	GCF_014905235.1	460.3	Chromosome	NCBI Annotation Release 103	17,047	18,362	10.7
Aphididae							
<i>Acyrtosiphon pisum</i>	GCF_005508785.2	533.6	Chromosome	NCBI Annotation Release 103	20,307	11,619	40.1
<i>Aphis gossypii</i>	GCF_020184175.1	334.9	Chromosome	NCBI Annotation Release 101	18,811	11,844	19.9
<i>Diuraphis noxia</i> ^a	GCF_001186385.1	395.1	Scaffold	NCBI Annotation Release 100	13,004	11,582	N/A
<i>Melanaphis sacchari</i> ^a	GCF_002803265.2	300.3	Scaffold	NCBI Annotation Release 100	13,619	12,460	N/A
<i>Myzus persicae</i> ^a	GCF_001856785.1	347.3	Scaffold	NCBI Annotation Release 100	17,052	11,027	N/A
<i>Rhopalosiphum maidis</i>	GCF_003676215.2	326.0	Chromosome	NCBI Annotation Release 101	13,624	13,342	29.1
<i>Sipha flava</i> ^a	GCF_003268045.1	353.2	Scaffold	NCBI Annotation Release 100	15,476	13,001	N/A
Adelgidae							
<i>Adelges cooleyi</i> ^a	GCF_023614345.1	270.2	Scaffold	NCBI Annotation Release 100	15,325	10,785	N/A
Phylloxeroidea							
<i>Daktulosphaira vitifoliae</i> ^a	GCF_025091365.1	282.6	Scaffold	NCBI Annotation Release 100	17,104	11,405	16.8
Other hemipterans							
<i>Bemisia tabaci</i> ^a	GCF_001854935.1	615.0	Scaffold	NCBI Annotation Release 100	15,283	28,281	N/A
<i>Cimex lectularius</i> ^a	GCF_000648675.2	510.8	Scaffold	NCBI Annotation Release 101	14,647	22,372	N/A
<i>Diaphorina citri</i>	GCF_000475195.1	485.7	Scaffold	NCBI Annotation Release 102	26,710	7,714	N/A
<i>Halyomorpha halys</i> ^a	GCF_000696795.2	998.2	Scaffold	NCBI Annotation Release 101	16,831	26,469	N/A
<i>Homalodisca vitripennis</i>	GCF_021130785.1	2 305.0	Chromosome	NCBI Annotation Release 100	22,591	31,968	3.8
<i>Nilaparvata lugens</i>	GCF_014356525.1	1 087.8	Chromosome	NCBI Annotation Release 101	21,385	32,416	7.8

^a Species for which Hox genes were found on different scaffolds, preventing the estimation of Hox cluster size. Hox cluster sizes were calculated by considering both canonical and non-canonical Hox genes.

and non-canonical Hox genes of the red flour beetle *Tribolium castaneum* and the domestic silk moth *Bombyx mori*, insect models representative of the orders Coleoptera and Lepidoptera, respectively, using the latest NCBI RefSeq genome sequences available (GCF_000002335.3 and GCF_014905235.1). The results we obtained were consistent with the annotation of the *T. castaneum* and *B. mori* Hox gene cluster annotation available in previous studies (we identified the same genes and confirmed their relative position in the Hox cluster), validating our approach (Chai et al., 2008; Mulhair and Holland, 2022; Pace et al., 2016).

2.7. Database construction and web interface

All predicted TF sequences and annotations were stored in a MySQL relational database on a Linux server. This includes sequences from *A. pisum* TFs, transcription basal machinery and chromatin-related genes, as well as sequences of HNF4 and Hox genes in 14 additional hemipteran genomes. Queries to the database were implemented in PHP scripts running in an Apache/PHP environment. ATFdb is accessible

online (<https://atf.cycadsys.org/>) and allows users to browse by TF family. Users can search ATFdb by gene ID, gene name, TF family or by keywords. Users can also run BLAST searches against all the sequences available in the database. The lists of TFs as well as their nucleotide and protein sequences are available for download. For each aphid TF, if there is an ortholog in *D. melanogaster*, the ATFdb gene page displays a clickable link to the *D. melanogaster* gene page on FlyBase (Suppl. Fig. 1). Similarly, for each aphid TF, the "view full information in external database" link available on the ATFdb gene page redirects to the corresponding gene page on ArthropodaCyc (Baa-Puyoulet et al., 2016), which aggregates various functional annotations (e.g. conserved protein domains, Gene Ontology terms) and provides cross-references to other databases such as AphidBase (Legeai et al., 2010), NCBI, KEGG (Kanehisa and Goto, 2000) and Interpro (Paysan-Lafosse et al., 2023).

3. Results and Discussion

3.1. A new annotation of *D. melanogaster* TFs, chromatin-associated and basal transcription machinery-associated genes

The analysis of the 755 putative TFs of Hens et al. (2011) revealed that they correspond to 747 independent genes. The difference is due to the independent listing of splicing variants and the duplicate listing of a few TFs. The overlap between these 747 TFs and the 707 TFs from Hammonds et al. (2013) includes 643 TFs, with 104 putative TFs that were listed only in Hens et al. (2011) and 64 only in Hammonds et al. (2013). By removing genes that had since been (re)annotated as not being TFs or as being a component of the basal transcription machinery or associated with chromatin, we obtained a final list of 703 putative *D. melanogaster* TFs. This list was then complemented with annotation datasets from FlyBase (FBgg0000745, 628 candidates) and retrieving information from FlyMine (142), FlyNet (633) and TranscriptionFactor.org (529). *De novo* predictions were also performed on the 13,962 protein-coding genes of the *D. melanogaster* genome using the Pfam domain (536) and the PlantTFDB (483) prediction tools. These analyses yielded an additional 32 candidate TFs (upper part of Fig. 1). Thus, the complete list of *bona fide* *D. melanogaster* TFs that was used to annotate the *A. pisum* genome comprised 735 unique TF-coding genes (Suppl. File 2).

We also annotated 195 unique *D. melanogaster* chromatin-associated genes (Suppl. File 3). Thirty-two of those genes function as TFs and are also listed in our *D. melanogaster* TF list (Suppl. File 2). Lastly, we report in Suppl. File 4 the annotation of the 51 *D. melanogaster* genes involved in the basal transcription machinery. Two of them are double-listed in Suppl. File 3, as they also belong to the chromatin-associated machinery.

3.2. The *A. pisum* repertoire of TFs, chromatin-associated and basal transcription machinery-associated genes

Using the curated TF list from *D. melanogaster* as query in combination with our *de novo* prediction approach, we identified 854 putative TFs in the pea aphid genome (Suppl. File 2), including 16 chromatin-associated genes. Those were double-listed in Suppl. File 3, which contains the 230 unique *A. pisum* chromatin-associated genes. We also annotated 67 basal transcription machinery associated genes (Suppl. File 4), two of which also belong to the class of chromatin associated genes. All this information is available in ATFDdb. Out of the 1,044 genes that were annotated in this study, 97.8% were expressed, as demonstrated by their presence in at least one publicly available RNA-seq dataset, providing support for the structural annotation used in our work. (Fig. 2A). The basal transcription machinery did not display important differences between *D. melanogaster* and *A. pisum*. The TFs and the chromatin-associated genes did, however, show remarkable differences. While some families are well conserved relative to the ones present in the *D. melanogaster* genome, others show independent expansions and losses (Table 2) and are discussed below. TFs play critical roles in body plan determination, cell fate specification and differentiation, and physiological regulation. For this reason, gain and loss of TFs are major contributors to phenotypic diversity (Romero et al., 2012) and the differences in terms of TF numbers between the pea aphid and the fruit fly could reflect the observed morphological, developmental and physiological diversity between the two species. While the holo-vs hemimetabolous nature of their metamorphosis could account for part of the divergence in TF repertoire between the two species, Vidal and colleagues (Vidal et al., 2016) previously showed that the *A. pisum* TF complement also differs from the one of other hemimetabolous insects suggesting that it is related to regulatory novelty unique to this species, or to aphids in general.

In line with this, among the 854 unique *A. pisum* TF genes, 215 were exclusively found in *A. pisum* and did not have homologs in *D. melanogaster*. None of these aphid-specific genes fall into a TF class

not already present in *Drosophila*.

Importantly, a systematic comparison with previous annotations of TFs in the *A. pisum* genome was not possible as we worked on the latest pea aphid reference assembly (Li et al., 2019), in which the total gene number is reduced compared to the versions previously used (20,307 predicted genes in the latest reference assembly compared to 34,604 in the first assembly used by Shigenobu et al. (2010) and Dang et al. (2011), or to 36,939 of the assembly used by Vidal et al. (2016)) and contains a significantly lower number of duplications. It is also important to note that the *A. pisum* genome assembly used in our study is the first one to have been resolved at the chromosomal level, which permitted a deeper comparative genomic analysis with the annotation available in *D. melanogaster* and other insects (see for example our analysis of the Hox cluster genes).

3.3. The expansion of distinct groups of chromatin-associated genes suggests an important role for epigenetic regulation in the pea aphid

Annotation revealed 230 chromatin-associated genes in the *A. pisum* genome, compared to 195 in the *D. melanogaster* genome. The expansions in *A. pisum* include multiple homologs of the Polycomb-repressive Complex 1 (PRC1) associated genes *Kdm2* (histone demethylase specific for H3 lysine 36; Lagarou et al., 2008), *Scp* (*Sex combs extra*, E3 ubiquitin ligase that monoubiquitinates H2A; Gutiérrez et al., 2012) and *Scm* (*Sex combs on midleg*, enables PRC1 binding; Bornemann et al., 1996), as well as the SET-domain lysine methyltransferase *egg* (*eggless*, histone H3-lysine(9) N-trimethyltransferase; Wang et al., 2011). In addition, we found expansions of histone lysine acetyltransferase 6 complex genes (*enok - enoki mushroom* (Scott et al., 2001), *Ing5 - Inhibitor of growth family member 5* (Huang et al., 2016)) and histone deacetylase genes (*HDAC1*; Mannervik and Levine, 1999). Lastly, the heterochromatin protein 1 (HP1) family, which binds histone H3 tails, is also larger in *A. pisum* (Mendez et al., 2011). Combined, these results suggest that epigenetic mechanisms may play a more prominent role in gene and transcriptional regulation in the pea aphid than in *Drosophila*, a hypothesis already advanced by Vidal et al. (2016) and further supported by studies suggesting that gene expression in *A. pisum* is regulated by multiple histone acetyltransferase and histone deacetylase genes (Kirfel et al., 2020). Aphids are known for their ability to express multiple phenotypes from the same genetic repertoire, a phenomenon also known as polyphenisms. They are in particular able to switch between sexual and asexual parthenogenetic reproduction or between winged and apterous morphs in response to various environmental cues (e.g. population density, change in photoperiod or temperature, presence of predators and parasites) (Ogawa and Miura, 2014). Previous studies have shown that epigenetic changes, in particular changes in DNA methylation, are involved in some of these processes (Mathers et al., 2019). Thus, the phenotypic plasticity that characterizes aphids may well find its basis in the expansion of chromatin-associated genes, a hypothesis also proposed by Rider Jr et al. (2010) and Vidal et al. (2016). This is further supported by the fact that chromatin associated protein expansion was detected in *Daphnia pulex*, another arthropod able to reproduce through cyclical parthenogenesis and displaying phenotypic plasticity in presence of predators (Schurko et al., 2009; Pappaluskas et al., 2024), but not in *Rhodnius prolixus*, a hemipteran with a classic hemimetabolous development without extensive morphological or physiological variation (Vidal et al., 2016).

3.4. TF family-specific differences in *A. pisum* and *D. melanogaster*

An overall observation was that the genomes of *A. pisum* and *D. melanogaster* show differences in the numbers of TFs in distinct families, suggesting lineage-specific expansions and losses (Table 2). We here highlight some of the most notable differences.

The most prominent differences were found in the zinc finger (ZNF) TFs. Using a previous version of the *A. pisum* genome assembly,

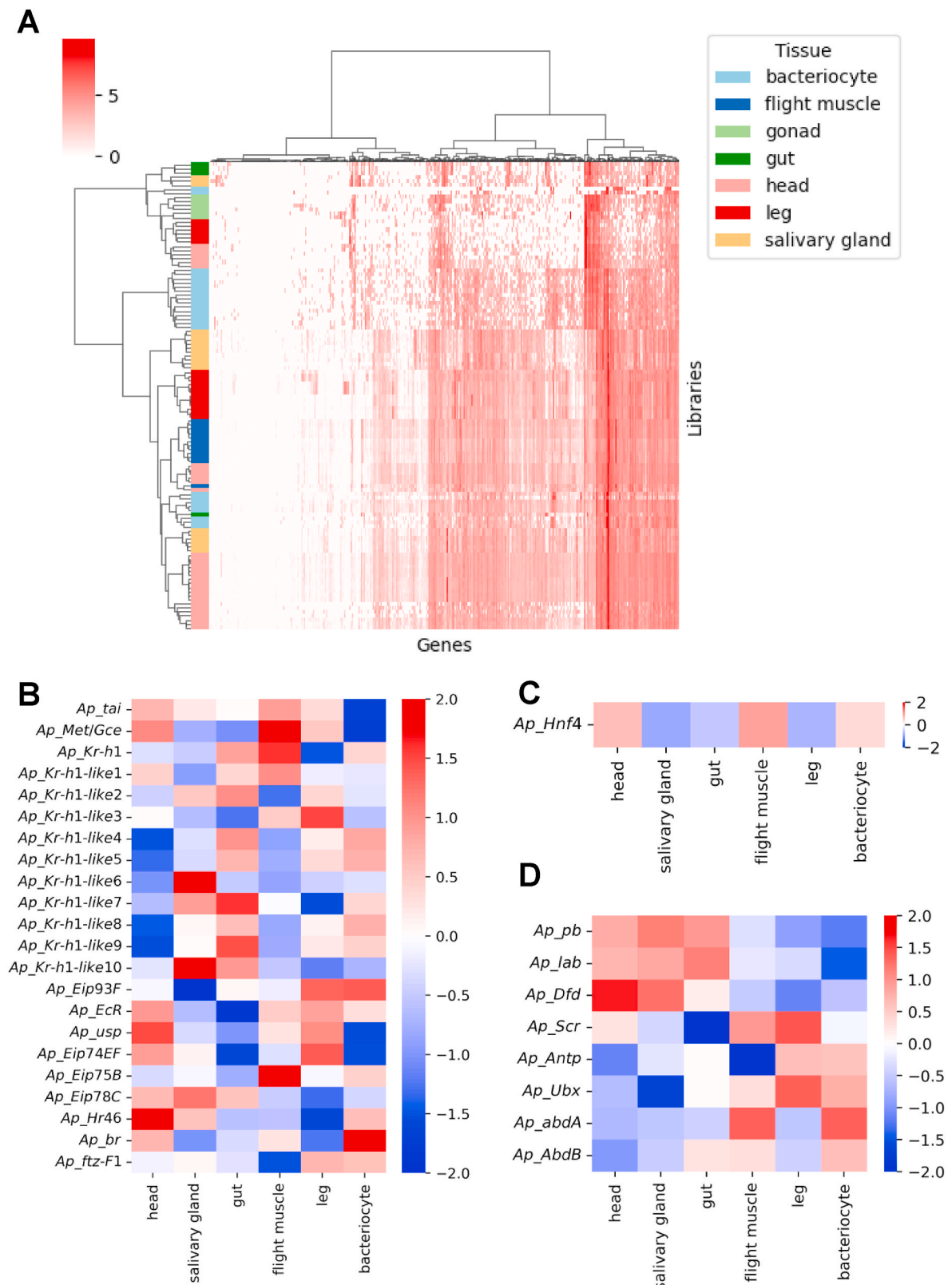


Fig. 2. Expression analysis of *Acyrthosiphon pisum* TFs and chromatin-associated factors. (A) ClusterMap representing the normalized level of expression in different tissues, of the genes annotated in this study. Expression levels are represented in a logarithmic scale, following a gradient from least (white) to most (red) expressed. The left color track represents the tissue of origin of the RNA-seq libraries. (B–C) Heatmap representing the average expression level per tissue of pea aphid TF involved in the hormonal response to Juvenile Hormone and/or Ecdysone (B) and *hnf4* homolog (C). The levels of expression are represented following a gradient from least (blue) to most (red) expressed. (D) Heatmap representing the average level of expression per tissue, for the nine canonical *Hox* genes of the pea aphid. Z-score normalization has been applied column-wise to ensure each column vector has zero mean and unitary standard deviation. The levels of expression are represented following a gradient from least (blue) to most (red) expressed.

Table 2
Transcription factors in *Acyrtosiphon pisum* and *Drosophila melanogaster* classified by family. The table is nonredundant: genes are counted only once, regardless of whether they have splicing variants.

Transcription factor families	Predicted number of genes per family	
	<i>A. pisum</i> GCF.005508785.2 NCBI Annotation Release 103	<i>D. melanogaster</i> FlyBase Release 6.55
ARID	9	5
bHLH	49	60
bHLH-PAS	2	3
bZIP	22	22
DM domain	3	4
E2F/DP	6	5
ETS	9	8
FKH (winged helix)	16	20
GCM	1	2
HMG	19	22
Homeobox CUT	3	3
Homeobox HOX-like	15	17
Homeobox LIM	8	7
Homeobox NK-like	25	30
Homeobox POU	4	5
Homeobox PRD-like	18	19
Homeobox PROS	1	1
Homeobox SIX	3	3
Homeobox TALE	5	8
Homeobox ZF	2	2
Homeobox Other	2	1
HTH-Psq	7	15
MAD homology domain	10	5
MADF-BESS	57	46
MADS	5	2
NF-KB	2	3
p53/RUNT	8	7
PAX	9	10
SANT-MYB	20	20
T-box	5	8
ZNF-Classical/C2H2	269	149
ZNF-AD	29	83
ZNF-AN1	2	2
ZNF-BED	29	6
ZNF-BTB/POZ	20	14
ZNF-LIM	3	1
ZNF-NHR/GATA	27	27
ZNF-RING/FYBVE/PHD	38	14
ZNF-SP1/KLF	19	13
ZNF-THAP	17	7
Other DNA-BD and cofactors	56	56

Shigenobu et al. (2010) described significantly more zinc finger containing TFs of the C₂H₂ (also called classical ZNF), GATA, BED, and MIZ families. We have now systematically reannotated all these families and assigned each ZNF TF in *A. pisum* to its precise family and observed species-specific expansions of ZNF TFs. The most striking expansions concern the following families: “ZNF-classical transcription factors” (with 269 genes in *A. pisum* and 149 genes in *Drosophila*), ZNF-BED (29 genes in *A. pisum* vs. six in *D. melanogaster*), ZNF-THAP (17 genes in *A. pisum* vs. seven in *D. melanogaster*), ZNF-SP1/KLF (19 genes in *A. pisum* vs. 13 in *D. melanogaster*) and ZNF-RING/FYBVE/PHD (also known as ZNF-MIZ, 38 genes in *A. pisum* vs. 14 in *D. melanogaster*). By contrast, the family ZNF-AD displays an expansion in *D. melanogaster* (83 TFs). Several of the ZNF-AD TFs are lacking in the *A. pisum* genome (29 TFs), yet we noted seven *A. pisum* homologs for the *D. melanogaster* Meics gene. There is no notable difference in the ZNF-GATA family. Concerning the ZNF-classical family, we were able to subdivide this expansion in *A. pisum* based on homology with known *D. melanogaster* genes. We identified 18 homologs for CG12299, 12 for *clamp* (*clp*), 135 for *crooked legs* (*cro*), and seven for *Phaser*. Overall, these results suggest that an independent evolution of ZNF TFs occurred in the two species. Panfilio et al. (2019) described a different expansion of the C₂H₂ ZNF TFs in the milkweed bug *Oncopeltus fasciatus* therefore suggesting that

this family of TFs is particularly prone to expansions. Whether and how this is reflected functionally remains to be discovered. However, similar expansions of ZNF TFs have been observed in vertebrates suggesting that the modularity of the zinc fingers and the possibility of changing amino acids that interact with DNA are two factors that contribute to the fact that they are the largest group of TFs (Nardelli et al., 1991; Vaquerizas et al., 2009). The different classes of ZNF TFs in the *A. pisum* genome are listed in Table 2.

Other less prominent differences were found. In *A. pisum* we annotated larger numbers of TFs of the following families: ARID, MAD homology domain, and MADS. On the other hand, in the *D. melanogaster* genome, we found an overall higher number of TFs belonging to the following TF families: bHLH, Homeobox-TALE, HTH-Psq and T-box.

The families that are not explicitly listed here comprised the same or nearly the same numbers of TFs in *A. pisum* and *D. melanogaster*. However, in some instances, and even though the overall numbers were the same in a TF family, expansions of some subfamilies were seen in either species, e.g. for MADF-BESS. In addition, there were numerous cases where a single homolog was present in *A. pisum* and two homologs in *D. melanogaster* (Suppl. File 2). This was the case for *Met/gce* (*Methoprene tolerant/germ cell-expressed bHLH-PAS*; Juvenile Hormone signaling; Baumann et al., 2010; Miura et al., 2005), *Rbf/Rbf2* (retinoblastoma family protein; cell cycle regulation; Stevaux et al., 2002), *slp1/slp2* (*sloppy paired 1 and 2*; segmentation; Grossniklaus et al., 1992), *gcm/gcm2* (*glial cell missing 1 and 2*; glia and plasmatocyte development; Chotard et al., 2005), *B-H1/B-H2* (Bar; eye and leg development; Higashijima et al., 1992), *en/inv* (*engrailed/invented*; segmentation; Cheng et al., 2014), *lbe/lbl* (*ladybird early/late*; heart development; Jagla et al., 1997), *Vsx1/Vsx2* (*visual system homeobox 1 and 2*; Erclik et al., 2008), *bab1/bab2* (*bric-a-brac*; Lours et al., 2003), *tsh/tio* (*teashirt/tiptop*; head and trunk development; Datta et al., 2011), and *eyg/toe* (*eyegone/twin of eyegone*; eye development; Yao et al., 2008). The presence of a single *engrailed/invented* homolog is likely due to secondary loss since it was previously shown that they arose through tandem duplication prior to the radiation of hexapods (Peel et al., 2006).

3.5. Differences and similarities in TFs regulating early development and organ development of *D. melanogaster* and *A. pisum*

We confirm the absence in *A. pisum* of homologs for the early developmental genes *bicoid* (*bcd*), *huckebein* (*hkb*), *buttonhead* (*btd*) and *giant* (*gt*), as originally reported by Shigenobu et al. (2010). *Bicoid* is only found in higher Diptera (Gregor et al., 2008) and its absence in the *A. pisum* genome was therefore expected. By contrast, the absence of homologs for *huckebein*, *buttonhead* and *giant* was less expected. *Giant* is conserved in many insects and has been shown to act as a gap gene in the hemipteran *O. fasciatus* (Ben-David and Chipman, 2010). While we did not find a *bona fide* homolog for *buttonhead* in the *A. pisum* genome, we did identify an *Sp1* gene homolog. Given the close relatedness of *buttonhead* and *Sp1* and similar functions in appendage development in *Drosophila*, *T. castaneum* and *O. fasciatus* (Schaeper et al., 2009), we propose that the presence of *Sp1* in *A. pisum* likely reflects the fact that it retained this role as a regulator of appendage development. *Huckebein* is present in *O. fasciatus* where its expression pattern suggests a role different from what was described in *Drosophila* where it is a terminal gap gene (Weisbrod et al., 2013).

Key TFs involved in the development of various organs are conserved in *A. pisum*. These include development of (i) wings: *apterous* (*ap*; Cohen et al., 1992) and *vestigial* (*vg*; Williams et al., 1991); (ii) salivary gland: *forkhead* (*fkh*; Mach et al., 1996); (iii) muscle: *Mef2* (Kelly et al., 2002); (iv) heart: *tinman* (*tin*; Bodmer, 1993). For eye (and head) development we found that many of the genes that constitute the retinal determination pathway in *Drosophila* (Domínguez and Casares, 2005; Kumar, 2010), are conserved in *A. pisum*, namely *so*, *lz*, *ey*, *toy*, *eyg*, *eya*, *hth*, *exd*, *oc*. At the same time, the absence of *dan*, *danr*, *toe* and the presence of a single *ara/caup/mirror* homolog indicates that the pathway has likely

undergone modifications. A final interesting observation was the complete absence of homologs for *Doc1*, *Doc2* and *Doc3*, which are involved in the amnioserosa development in *Drosophila* (Hamaguchi et al., 2004). Hemimetabolous insects, including the pea aphid, have two extraembryonic membranes (amnion and serosa) that are involved in blastokinesis (comprising anatrepsis and katatrepsis) (Panfilio, 2008; Schmidt-Ott, 2000). It contrasts with *Drosophila* where extraembryonic development is extremely reduced, thus making it likely that the underlying genetic mechanisms will also show considerable differences.

3.6. TFs regulating developmental transitions are conserved

Ecdysone and juvenile hormone (JH) are essential hormones regulating developmental transitions. All TFs implicated therein in *Drosophila* are conserved in *A. pisum*. *Ecdysone Receptor (EcR)* and *Ultraspiracle (Usp)* for ecdysone signaling (Yamanaka et al., 2013), and *Met/Gce*, *Taiman*, *Kr-h1* and *Eip93F* for JH signaling (Jindra et al., 2015). Interestingly, *Kr-h1*, the primary JH-response gene which reflects the level of JH signaling (Li et al., 2018), is expanded in the pea aphid compared to *D. melanogaster* (11 homologs vs. 1). Moreover, the pea aphid *Kr-h1* homologs show different tissue expression patterns suggesting functional diversification: while most genes are highly expressed in the gonad and gut compared to other tissues, *Ap_Kr-h1-like1*, 3 and 6 show high expression in the flight muscles, legs and salivary glands, respectively (Fig. 2B and Suppl. File 5). It was recently found that JH signaling is involved in flight muscle degeneration in the pea aphid through upregulation of *Met* and *Kr-h1* following feeding (Bai et al., 2022). Consistent with this, in addition to *Ap_Kr-h1-like1*, several predicted members of the JH signaling pathway are highly expressed in flight muscles (*Ap_Eip75B*, *Ap_tai*, *Ap_Met/Gce*) (Fig. 2B and Suppl. File 4).

3.7. Key TFs of signaling pathways are conserved, but TCF and MAD TFs show expansion in *A. pisum*

The activity of signaling pathways ultimately leads to transcriptional changes mediated by specific TFs. As expected, we identified homologs of all these TFs in the *A. pisum* genome. *Yan/Aop* (anterior open; Rebay and Rubin, 1995) and *Pointed* (O'Neill et al., 1994) are central negative and positive transcriptional regulators downstream of receptor tyrosine kinases (EGFR, FGFR, PVR). Others include *Foxo* (insulin signaling; Puig et al., 2003), *cubitus interruptus (ci)* (hedgehog signaling; Alexandre et al., 1996), *scalloped* (Wu et al., 2008) and *yorkie* (Hippo signaling; Huang et al., 2005), *Su(H)* (Notch signaling; Schweisguth and Posakony, 1992), and *Jra/kay = JUN/cFos* (JNK signaling; Perkins et al., 1990). Interestingly, we observed expansions of TFs associated with Wnt signaling (pangolin and TCF TFs; Brunner et al., 1997), JAK-STAT signaling (three homologs of *Stat92E*; Hou et al., 1996) and TGF-beta signaling (SMAD TFs; Sekelsky et al., 1995). The expansion of TCF TFs is intriguing in light of previous observations of duplications of the Wnt pathway component armadillo/b-catenin in *O. fasciatus* and other hemimetabolous insects (Bao et al., 2012; Panfilio et al., 2019). Previously, Panfilio et al. (2019) also reported an expansion of SMAD TFs in *O. fasciatus*, and suggested that it was specific to *Oncopeltus*. The fact that we also found an expansion in the pea aphid seems to suggest that the amplification may be more ancient in insect evolution. bHLH TFs of the E(Spl) complex acting in Notch signaling show a strongly reduced number as well as an altered organization of the complex. Dearden (2015) described and discussed these differences for the E(Spl) complex in arthropods.

3.8. Identification of an *A. pisum* homolog of HNF4, a key TF regulating metabolic homeostasis

Previously it was reported that there was no HNF4 homolog in *A. pisum* (Shigenobu et al., 2010). Given its central importance in

regulating metabolic homeostasis (Palanker et al., 2009), this was unexpected. We here report the identification of a *bona fide* HNF4 homolog in *A. pisum* (corresponding to different isoforms, listed in Table 3). This annotation was supported by the presence of two domains characteristic of HNF4 TFs: (i) a N-terminal ZNF-GATA-type DNA binding domain (DBD) (identified by InterPro with the signature IPR049636 corresponding to a "Hepatocyte Nuclear Factor 4-like, DNA binding domain") and (ii) a C-terminal nuclear receptor-type ligand binding domain (LBD) (identified by InterPro with the signature IPR049635 corresponding to a "Hepatocyte nuclear factor 4, ligand-binding domain"). The isoforms encoded by the *D. melanogaster* HNF4 gene are longer than those of the pea aphid (between 666 and 732 amino acids vs. between 375 and 420 amino acids, respectively). Nonetheless, *D. melanogaster* and *A. pisum* HNF4 proteins are highly similar, with 65% identity over the whole sequence and 72% when only functional domains are considered. We next searched the sequence of HNF4 in 14 additional hemipteran species (see Table 1 and § 2.4). At least one copy of HNF4 was found in each of these species (Table 3), with very high conservation of functional domains, especially the DNA binding domain (Fig. 3). Three species (*B. tabaci*, *D. citri* and *D. noxia*) appeared to have a duplication of HNF4. In *B. tabaci*, we identified three genes that have a size consistent with HNF4 and have DBD and LBD that were successfully identified by InterProScan. Importantly, the sequences of the *B. tabaci* proteins encoded by the LOC109043749 and LOC109038961 genes differ significantly from those of the other HNF4 homologs. On the one hand, the position of LOC109043749 in the *B. tabaci* genome of *B. tabaci*, in tandem with LOC109043751, suggests that this gene could have appeared through tandem duplication, and diverged afterwards. On the other hand, LOC109038961 probably appeared through duplication of LOC109043749. Two HNF4 homologs were also identified for *D. citri* and *D. noxia*, but the presence of two genes seems to be the result of assembly errors. In the case of *D. citri*, one of the genes codes for a very short protein of 90 amino acids (LOC103508991), not compatible with a functional HNF4. In the case of *D. noxia*, one sequence encodes a protein with only a LBD (LOC107163700), and the other sequence corresponds to a single gene on a very short contig, encoding a protein with only a DBD (LOC107172291). In the latter case, it is possible that the two genes identified as homologs actually correspond to a single gene that was not reconstructed during assembly.

The conservation of HNF4 in all hemipteran genomes we analyzed supports the hypothesis, emerging from recent studies (Cheng et al., 2020; Wang et al., 2023), that HNF4 has important roles in hemipteran physiology as well. Moreover, the observation that the length of HNF4 proteins encoded in all these hemipteran genomes is closer to the size of the isoforms we found in *A. pisum* and shorter than the isoforms found in *D. melanogaster*, suggests the presence of potential lineage-specific features. Consistent with this, the pea aphid HNF4 ortholog expression pattern differs from what has been described in *D. melanogaster*, showing high expression in the flight muscles, head and bacteriocytes and a lower expression in the gut (Fig. 2C and Suppl. File 5) (Palanker et al., 2009). Differences between the two species can be linked to differences in their morphology as aphids do not have Malpighian tubules, one of the tissues where HNF4 is highly expressed in the fly. On the other hand, bacteriocytes are absent in *Drosophila* but are known to play an important, if not primordial, role in metabolic homeostasis in the pea aphid. Further studies are needed to better define the spatiotemporal distribution of HNF4 transcripts in the pea aphid, in particular to determine in which part of the digestive tract HNF4 is expressed and whether expression in the head is localized in the brain, salivary glands and/or surrounding tissues.

3.9. TFs regulating essential cellular and organismal processes

We also analyzed TFs with central roles in important cellular and organismal processes in *D. melanogaster*. For the majority of these we identified homologs in *A. pisum*. Specifically, we found homologs for

Table 3

Annotation of HNF4 transcription factor homologs in hemipteran insects and model species representative of Diptera, Coleoptera and Lepidoptera. For each NCBI gene ID, all corresponding protein entries present in the database are listed. Each protein identifier corresponds to a protein isoform encoded by a unique predicted messenger RNA. Some isoforms are identical but are encoded by mRNAs that differ in their UTR regions, suggesting differences in the regulation of their expression.

Species	NCBI gene ID	NCBI protein ID	Length (aa)	Species	NCBI gene identifier	NCBI protein identifier	Length (aa)	
Model species				Adelgidae				
<i>Drosophila melanogaster</i>	Hnf4	NP_723413.1	732	<i>Adelges cooleyi</i>	LOC126835753	XP_050424501.1	418	
		NP_001285758.1	666			XP_050424502.1	417	
		NP_001097126.1	708			XP_050424503.1	381	
		NP_476887.2	704			XP_050424504.1	373	
		NP_723414.2	704					
Phylloxeroidea				Other hemipterans				
<i>Tribolium castaneum</i>	LOC657032	XP_008196266.1	507	<i>Daktulosphaira vitifoliae</i>	LOC126902027	XP_050535010.1	413	
		XP_008196267.1	507					
<i>Bombyx mori</i>	Hnf-4	XP_008196268.1	506	<i>Bemisia tabaci</i>	LOC109043751	XP_018916593.1	443	
		XP_008196269.1	503			XP_018916594.1	442	
		XP_008196270.1	502			XP_018916595.1	421	
		XP_008196271.1	491			XP_018916597.1	421	
		XP_008196272.1	490			XP_018916598.1	421	
		XP_008196273.1	471			XP_018916599.1	403	
		XP_008196274.1	470			XP_018916600.1	395	
		XP_008196275.1	462			XP_018909785.1	394	
		XP_008196276.1	498			XP_018916591.1	434	
		XP_037873397.1	466			XP_018916592.1	434	
		NP_001037474.1	436			XP_014243592.1	485	
		NP_001166834.1	410			XP_014243593.1	469	
		XP_037873398.1	462			XP_014243594.1	449	
Aphididae				<i>Cimex lectularius</i>	LOC106663339	XP_014243595.1	443	
<i>Acyrtosiphon pisum</i>	LOC100168732	XP_001946928.2	383			XP_014243596.1	437	
		XP_003246275.1	420			XP_014243598.1	433	
		XP_008185671.1	419			XP_014243599.1	427	
		XP_008185672.1	375			XP_014243600.1	417	
<i>Aphis gossypii</i>	LOC114123835	XP_027842742.2	420	<i>Diaphorina citri</i>	LOC103508978	XP_026679248.1	501	
		XP_027842743.2	419		LOC103508991	XP_008471805.1	90	
		XP_027842744.2	384	<i>Halyomorpha halys</i>	LOC106684564	XP_014282200.1	432	
<i>Diuraphis noxia</i>	LOC107163700	XP_027842745.2	376			XP_014282201.1	428	
		XP_015366730.1	242			XP_014282202.1	427	
<i>Melanaphis sacchari</i>	LOC107172291	XP_015378068.1	147	<i>Homalodisca vitripennis</i>	LOC124355498	XP_014282203.1	419	
		LOC112603138	XP_025207351.1			419	XP_046662611.1	503
		XP_025207353.1	418			XP_046662612.1	491	
		XP_025207354.1	383			XP_046662613.1	491	
		XP_025207355.1	375			XP_046662614.1	487	
<i>Myzus persicae</i>	LOC111038891	XP_025207356.1	375			XP_046662615.1	484	
		XP_022177849.1	419	<i>Nilaparvata lugens</i>	LOC111045151	XP_046662616.1	480	
		XP_022177850.1	418			XP_022186160.2	479	
		XP_022177851.1	383			XP_022186161.2	469	
XP_022177852.1	382	XP_022186162.2	468					
<i>Rhopalosiphum maidis</i>	LOC113556845	XP_026817824.1	419					
		XP_026817825.1	418					
		XP_026817826.1	396					
		XP_026817827.1	383					
		XP_026817828.1	375					
		XP_026817829.1	375					
<i>Sipha flava</i>	LOC112694256	XP_025425456.1	423					
		XP_025425457.1	422					
		XP_025425458.1	377					
		XP_025425459.1	377					
		XP_025425460.1	377					
		XP_025425461.1	377					

SREBP (HLH106 - lipid homeostasis; Theopold et al., 1996), *Atf6* (endoplasmic reticulum stress response; Allen and Seo, 2018), *cnc* (oxidative stress response; Sykiotis and Bohmann, 2008), *Xbp1* (UPR pathway; Huang et al., 2017), *REPTOR/REPTOR-bp* (TOR mediator complex; Tiebe et al., 2015). Intriguingly, we were not able to identify homologs for *Myc* (cell growth, cell competition, proliferation; Gallant et al., 1996) and *Mitf* (regulator of V-ATPase and lysosomal – autophagic pathway; Zhang et al., 2015). We did find a distant relative of *Myc*, namely *Mnt* (Loo et al., 2005) to be conserved. Since both are dimerization partners for *Max*, it is possible that one or both functionally replace *Myc*.

As an example of an organismal process, we investigated the central circadian clock and found homologs for all relevant TFs, *Clk*, *cwo*, *cycle*, *vri* and *Pdp1* suggesting that this process is fundamentally conserved (Hardin, 2011; Tataroglu and Emery, 2014).

3.10. *A. pisum* has no homolog for *doublesex*, a key TF in the *Drosophila* sex determination pathway

The genome of *D. melanogaster* contains four related TFs (*dmrt11E*, *dmrt93B*, *dmrt99B* and *dsx*) of the DM domain family. *doublesex* (*dsx*) encodes a key TF in the *Drosophila* sex determination pathway

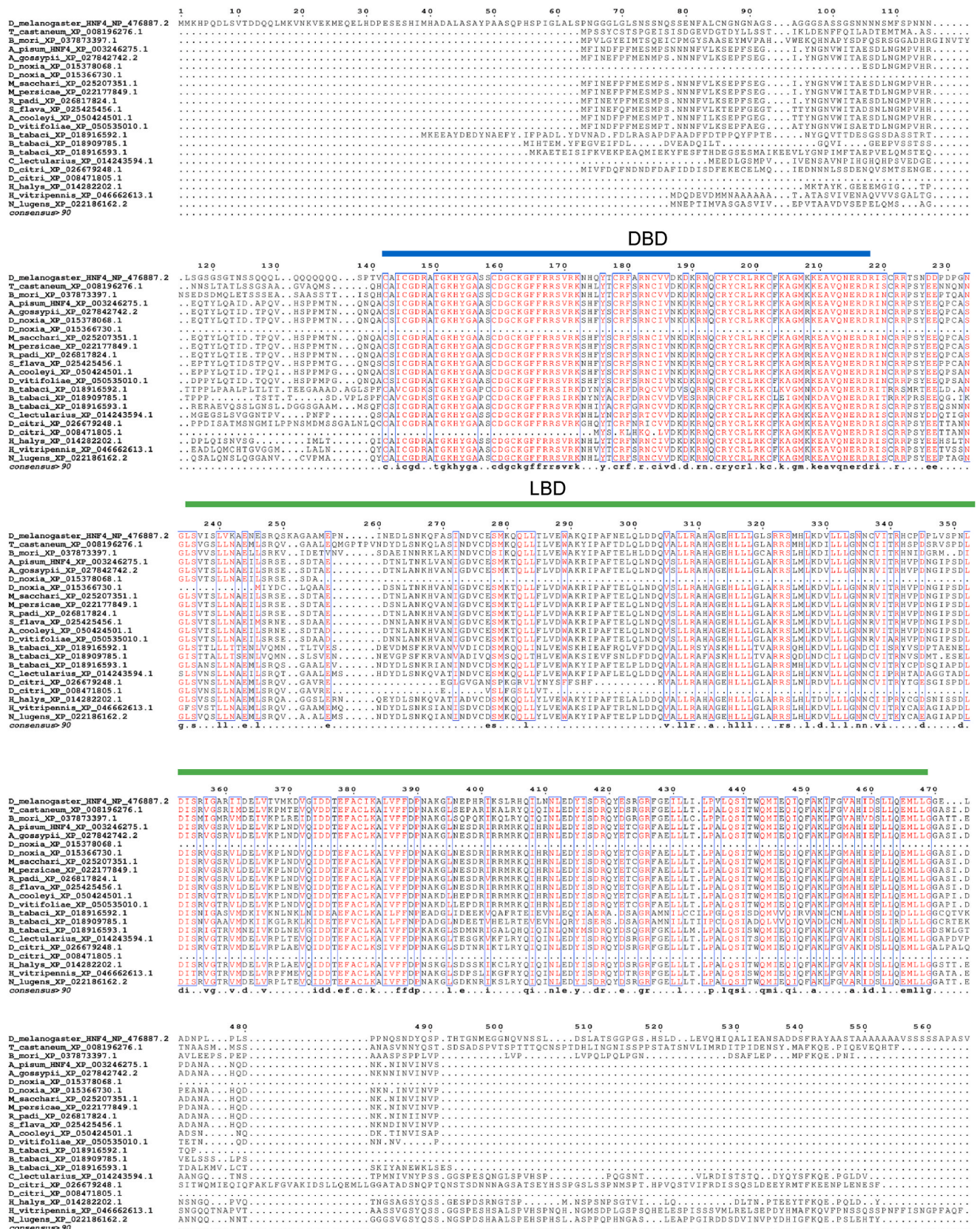


Fig. 3. Alignment of HNF4 transcription factor homologs in hemipteran insects and model species representative of Diptera, Coleoptera and Lepidoptera. Amino acids conserved in at least 90% of all sequences are indicated below the alignment (consensus >90). The regions corresponding to the DNA binding domain (DBD) and ligand binding domain (LBD) are indicated above the alignment by blue and green bars, respectively.

(Coschigano and Wensink, 1993), while *dmrt11E* is required in testis somatic cells for male fertility (Yu et al., 2015). The two other genes have not been functionally characterized. We did not find direct homologs for *dsx* and *dmrt11E*, but homologs for *Drosophila dmrt93b* and *dmrt99b* are present in *A. pisum*. The absence of a *dsx* homolog may reflect the differences in sex determination between *A. pisum* and *D. melanogaster*. Whether the *A. pisum* homologs *dmrt93b* and *dmrt99b* have a role in spermatogenesis remains to be investigated. Consistent with a different organization of sex determination is the observation that the *A. pisum* genome does not encode a *fruitless* homolog, a BTB transcription factor that regulates male sexual behavior in *Drosophila* (Just et al., 2023; Laslo et al., 2023).

3.11. The *A. pisum* Hox cluster is reorganized and lacks a bona fide *fushi tarazu* homolog

The insect Hox cluster is thought to have been composed, in a bilaterian ancestor, of 10 genes: *labial* (*lab*), *proboscipedia* (*pb*), *zerknüllt* (*zen*), *Deformed* (*Dfd*), *Sex combs reduced* (*Scr*), *fushi tarazu* (*ftz*), *Antennapedia* (*Antp*), *Ultrabithorax* (*Ubx*), *abdominal-A* (*abdA*), and *Abdominal-B* (*AbdB*), (listed here in the order they occupy in the cluster in a metazoan hypothetical ancestor). Unlike other arthropods (e.g. Crustacea), that can lack some of them, insects appear to have conserved all eight canonical Hox genes (*lab*, *pb*, *Dfd*, *Scr*, *Antp*, *Ubx*, *abdA*, *AbdB*), essential in determining positional identity along the body axis (Mulhair and Holland, 2022). The other two genes of the cluster, *zen* and *ftz*, which had ancestral homeotic functions, evolved novel roles in insects (extraembryonic membrane patterning for *zen* and segmentation for *ftz*), and are often referred to as the non-canonical Hox genes. We focused on the Hox cluster since it is one of only few gene clusters in arthropod genomes and, given the availability of the chromosomal assignment of the *A. pisum* genome, we analyzed the structure of the cluster and determined whether the linear arrangement is conserved in *A. pisum* or whether it is rearranged.

The *A. pisum* Hox gene cluster is localized on chromosome A1 and contains the eight canonical Hox genes. The fact that all insects have retained these eight could have functional implications: segment number and tagmatization are constant in insects, despite their high species radiation, leaving no place for redundancy or loss for these genes essential for the establishment of the insect body plan (Mulhair and Holland, 2022). In Fig. 4 we compare the *A. pisum* Hox cluster organization to the ones present in three model species, *D. melanogaster*,

T. castaneum and *B. mori*. This shows several differences in the *A. pisum* Hox cluster when compared to other insects: (i) the genomic size of the cluster is four times bigger than in *D. melanogaster* or *T. castaneum*, with longer average gene length and intergenic distances, (ii) the split between the anterior and the posterior parts of the cluster observed in *D. melanogaster* is not present, and a shorter split separates the first three genes from the others, (iii) while the respective positions of *Dfd*, *Scr*, *Antp*, *Ubx*, *abd-A* and *Abd-B* are the same as in the three other species, the first two genes of the cluster are reorganized, with inversion of the respective positions of the *lab* and *pb* genes. These observations are in agreement with the work of Mulhair and Holland (2022), who analyzed the Hox cluster arrangements in 243 insect genomes and showed that the most anterior genes of the cluster have undergone several rearrangements and that intergenic distances can vary greatly.

Importantly, the study of Hox gene expression in adult pea aphid tissues reveals a conservation of spatial collinearity common to all bilaterians, with the most anterior genes on the chromosome expressed in the anterior parts of the animal and vice versa (Fig. 2D and Suppl. File 5). Additionally, our results are consistent with previous studies that showed expression of both *Ubx* and *abd-A* in the pea aphid bacteriocytes (Braendle et al., 2003). While the evolutionary origin of bacteriocytes remains elusive in the pea aphid, it is also interesting to note that these cells show a pattern of TF expression that differs significantly from what is observed in other aphid tissues, including the gut (Fig. 2D and Suppl. File 5). While this was not unexpected, given the differences in morphology and function of bacteriocytes and gut, it also suggests that these two tissues have different developmental origins.

Concerning the non-canonical Hox, we have identified a unique *HoxB4-like* gene in *A. pisum* instead of the two *zen* genes (*zen* and *zen-2*) present in Diptera and Coleoptera (Fig. 4). The *A. pisum* genome also lacks another Hox-related gene, *bicoid* (see also § 3.5), which is considered a Diptera-specific duplication of *zen* (Stauber et al., 1999). Moreover, it was difficult to identify a *ftz* homolog. A homeobox-containing gene is present at the usual position for this gene, located between *Scr* and *Antp*, but it encodes a shorter protein (237 predicted amino acids vs. 410 in *D. melanogaster*) (Fig. 4 and Table 4B). Moreover, blasting it back to the *Drosophila* genome gave *Scr* as best hit, and not *ftz*. This difficulty in identifying a clear homolog of *ftz* was already reported in the first annotation of TFs in the pea aphid genome (Shigenobu et al., 2010), and was attributed to the lack of a chromosome-level assembly for the *A. pisum* genome. Our analysis has been made on the first chromosome-level genome assembly for this

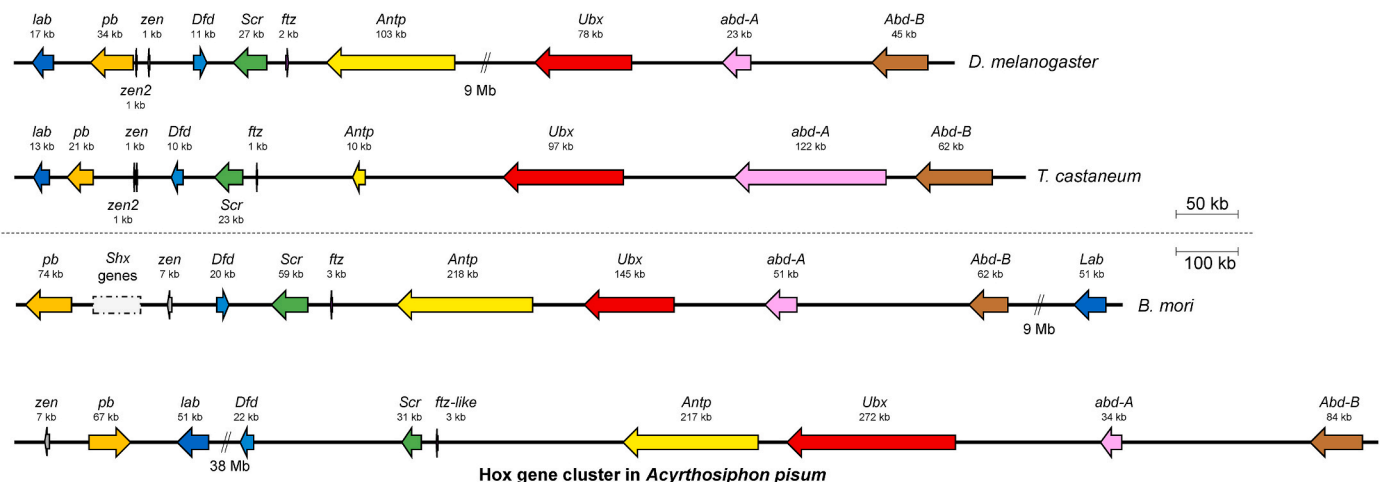


Fig. 4. Comparison of the Hox gene cluster in *Acyrthosiphon pisum* with the Hox clusters of model species representative of Diptera, Coleoptera and Lepidoptera. The orientation of each arrow indicates the transcriptional orientation of each canonical and non-canonical Hox gene. Genes are represented with their actual lengths and genomic distances. In *B. mori*, the « *Shx* genes » box refers to 12 genes obtained through extensive tandem gene duplication of *zen* (Chai et al., 2008). In *T. castaneum*, orthologs of *Antp*, *Scr* and *pb* are often referred to as *ptl* (*prothoraxless*, Brown et al., 2002), *Cx* (*cephalothorax*, Curtis et al., 2001), and *mxxp* (*maxillopedia*, Shippy et al., 2000), respectively.

Table 4
 Predicted Hox gene length (in base pair) (A) and Hox protein length (in amino acid) (B) from different hemipteran genomes and model species representative of Diptera, Coleoptera and Lepidoptera.

A																		
Gene	Species																	
	<i>Dmel</i>	<i>Tcas</i>	<i>Bmor</i>	<i>Apis</i>	<i>Acool</i>	<i>Agos</i>	<i>Btab</i>	<i>Clec</i>	<i>Dvit</i>	<i>Dcit</i>	<i>Dnox</i>	<i>Hhal</i>	<i>Hvit</i>	<i>Msc</i>	<i>Mper</i>	<i>Nlug</i>	<i>Rmai</i>	<i>Sfla</i>
<i>lab</i>	17,163	12,802	51,400	50,984	13,202	38,805	36,743	64,684	25,771	27,300	40,305	31,056	101,795	42,518	40,559	84,477	41,532	39,986
<i>pb</i>	34,263	20,972	73,989	67,078	50,728	60,959	100,766	108,990	39,940	63,833	61,501	161,831	103,957	60,664	61,726	89,565	61,729	63,599
<i>zen^a</i>	1330	941	7027	7142	4910	7777	8108	14,007	3613	27,435	6179	N/A	N/A	7275	6197	19,341	6926	8083
<i>Dfd</i>	10,592	9810	19,605	22,234	12,323	20,561	36,241	40,364	10,160	9369	20,993	70,739	61,213	19,862	20,937	39,098	19,432	17,831
<i>Scr</i>	26,861	22,556	58,606	31,371	21,749	30,078	68,258	67,342	18,044	29,453	2865	92,514	72,445	29,552	30,861	63,652	31,206	31,904
<i>ftz^a</i>	1920	1082	2562	2688	709	2436	1641	1232	6427	5416	2627	6373	11,265	2218	2668	8982	2506	2105
<i>Antp</i>	102,975	10,347	218,095	217,154	133,110	183,470	283,087	294,612	107,130	23,014	176,528	413,287	273,220	158,063	173,894	388,425	153,704	15,482
<i>Ubx</i>	77,803	96,642	144,972	271,572	86,845	235,207	248,978	348,362	77,617	1338	57,593	235,466	306,868	65,567	166,471	481,855	239,015	N/A
<i>abdA</i>	22,835	122,123	51,244	33,390	26,754	35,439	73,317	78,173	25,085	2628	33,913	97,192	83,412	31,919	33,487	132,293	32,160	33,586
<i>AbdB</i>	45,024	61,885	62,456	84,274	52,297	109,087	65,241	201,920	60,925	836	6984	264,994	231,456	75,137	40,674	162,794	67,972	106,033
B																		
Protein	Species																	
	<i>Dmel</i>	<i>Tcas</i>	<i>Bmor</i>	<i>Apis</i>	<i>Acool</i>	<i>Agos</i>	<i>Btab</i>	<i>Clec</i>	<i>Dvit</i>	<i>Dcit</i>	<i>Dnox</i>	<i>Hhal</i>	<i>Hvit</i>	<i>Msc</i>	<i>Mper</i>	<i>Nlug</i>	<i>Rmai</i>	<i>Sfla</i>
<i>lab</i>	213–629	353	336	603	419	620	466	279	436–450	247	597	202	388	623	607	414	617	642
<i>pb</i>	772–782	654	675	893–905	681–807	907–919	744–747	601	779–790	300	891–903	586	688	900–912	889–901	719	906–918	919–930
<i>zen^a</i>	353	246	549	408	294	534	339	343	267	118	407	N/A	N/A	424	405	567	416	424
<i>Dfd</i>	586	412–423	392	531	471	532	409	337–356	463	235	519	307–332	382	530	537	383–392	422–542	523
<i>Scr</i>	417–564	312–320	356	392	346	391	338–346	323	347	327	102	299	333	391	390	316	391	399
<i>ftz^a</i>	410	290	447	237	184	226	163	235	137–271	100	239	290	370–371	224	232	390	228	243
<i>Antp</i>	297–378	325	259–303	389–393	337–341	318–393	326–327	302	344	317	387	291–293	352	391	338–342	329–330	384	386–390
<i>Ubx</i>	346–389	309–314	254	339–343	332–336	262–342	296	249–303	319–323	94	57	166	318	106	388–392	259–314	337–341	N/A
<i>abdA</i>	330–590	343–347	343–352	369	351	372	335–365	345	348	149	371	339	284–359	370	372	337–351	372	376
<i>AbdB</i>	270–493	351–364	306–353	489	443	495	551	343–355	475	174	195	314	371	492	492	370	493	490

Abbreviations: *Apis*, *Acyrtosiphon pisum*; *Acool*, *Adelges cooleyi*; *Agos*, *Aphis gossypii*; *Btab*, *Bemisia tabaci*; *Bmor*, *Bombyx mori*; *Clec*, *Cimex lectularius*; *Dvit*, *Daktulosphaira vitifoliae*; *Dcit*, *Diaphorina citri*; *Dmel*, *Drosophila melanogaster*; *Dnox*, *Diuraphis noxia*; *Hhal*, *Halyomorpha halys*; *Hvit*, *Homalodisca vitripennis*; *Msc*, *Melanaphis sacchari*; *Mper*, *Myzus persicae*; *Nlug*, *Nilaparvata lugens*; *Rmai*, *Rhopalosiphum maidis*; *Sfla*, *Sipha flava*; *Tcas*, *Tribolium, castaneum*.

^a Non-canonical Hox genes.

was not resolved (Fig. 5). In the case of *ftz*-like proteins in particular, this polyphyly suggests that some of these genes might have different evolutionary origins. In line with this, sequences from aphid species formed a monophyletic group more closely related to members of the Abd-B orthology group than to other *ftz*-like proteins. Furthermore, according to our phylogeny, the predicted *ftz*-like TF of *Diaphorina citri* is part of the Scr orthology group, which suggests that this protein may in fact be the result of a more recent partial duplication of the *scr* gene. No other *ftz*-like gene was predicted in this species.

To gain a better understanding of the levels of intra- and inter-specific divergence within the Hox cluster in Hemiptera, we compared the protein sequences of the genes that make up this cluster in the species listed in Table 1 through systematic BLASTP. Results are presented in Suppl. File 6 and include the percentages of coverage and identity for each pairwise comparison. This latter analysis shows that some genes are more conserved than others. For example, when we compare the Hox genes of *D. melanogaster* with those of the pea aphid, we find coverage and identity percentages of 56% and 61%, respectively, for *abd-A* versus 30% and 56% for *ftz*. When considering the average values of coverage and identity for each group of homologs, we find low coverage rates, averaging 27% (Suppl. File 7). Significant identity is primarily observed in the homeodomain. Identity percentages vary between 56% (using *pb* proteins as queries) and 66% (using *Antp* proteins as queries), on average. These values reflect the level of divergence between proteins encoded by Hox cluster genes, despite the presence of the conserved homeodomain. When only orthologs are compared, coverage percentages are higher, with a minimum of 48% (57% when only canonical Hox proteins are considered) and a maximum of 88%. The *pb* and *lab* proteins are the most divergent, with coverage and identity percentages averaging no more than 60% and 63% respectively. Conversely, the *Abd-B*, *abd-A*, *Ubx*, *Antp* and *Scr* proteins show higher conservation, with coverage and identity percentages reaching up to 88% and 81%, respectively. These results are intriguing in light of the fact that most hemipteran Hox clusters present a split between the posterior region of the cluster (from *Abd-B* to *Scr*) and the anterior region of the cluster (comprising *lab* and *pb*), suggesting that the evolutionary constraints are higher on the former while *lab* and *pb* diverge more rapidly. *Dfd*, which can be associated to either the posterior or anterior region depending on the species, has an intermediate status, with coverage and identity averaging 65% and 66% respectively. Consistent with the results of the phylogenetic analysis, *ftz* and *zen* homologs appear to be highly divergent compared to canonical Hox gene

products.

A comparison of genomic organization and genomic distances between genes of the Hox clusters was possible for a small group of Hemiptera, for which all the Hox genes were on the same chromosome/scaffold (Fig. 6). This analysis shows how, in the Hemiptera group, the general organization of the Hox cluster, in terms of relative position of the genes and the splits, can vary greatly from each other. General observations made by Mulhair and Holland (2022) in other insects, for instance that intergenic distances can be large in the anterior part of the cluster, are reduced in the middle and become even larger in the posterior part of the Hox cluster do not apply to the hemipteran genomes we have analyzed here (Fig. 6). Moreover, contrary to what has been found in other insect genomes, *Ubx*, *abd-A* and *Abd-B* are not always present in the cluster in this specific order in hemipteran genomes.

Concerning the function of TFs in *A. pisum*, very little information is available. A few studies have focused on the Hox genes and have shown that they evolved new roles in aphids, related to their symbiotic status or certain polyphenisms. Immunohistochemical studies have demonstrated that *Ubx* and *AbdA/B* are localized in the bacteriocytes of aphid embryos and nymphs, which supports the hypothesis that they are associated with the early development and the differentiation of these symbiont-containing cells (Braendle et al., 2003). The important role of *Ubx* in the development of insect bacteriocytes has subsequently been confirmed in the hemipteran *Nysius plebeius* (Matsuura et al., 2015), but no study has since been performed in *A. pisum*. Other roles for the Hox TFs seem to be related to the wing polyphenisms, as *Scr*, *Antp* and *Ubx* are up-regulated in apterae vs. alate aphids and are differentially expressed between the two morphs during development (Zhang et al., 2019). Future availability of high-quality insect genomes with chromosomal annotation will permit further comparative studies of the Hox cluster, not only to infer its evolutionary origin and dynamics, but also to determine what - if any - the developmental and functional consequences are of the rearrangements.

4. Conclusion

In the present work, we have (re)annotated all TFs, chromatin-associated genes and genes associated with the basal transcriptional machinery in the model aphid *A. pisum* using the latest genomic data. We describe the procedure we used for the annotation, combining homology-based and manual curation of *D. melanogaster* TFs and *de novo* predictions. The results of these annotations are made available to the

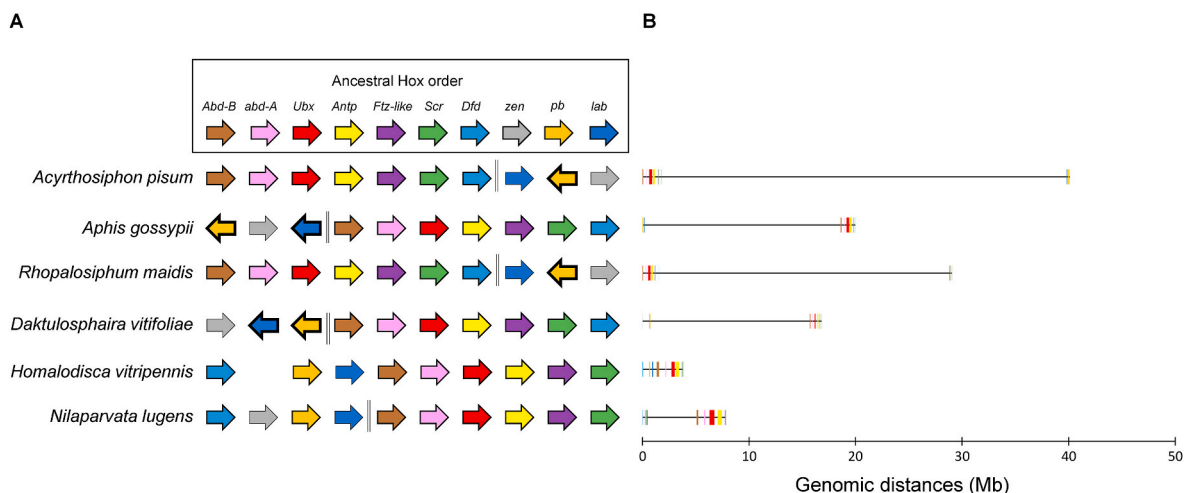


Fig. 6. Genomic organization and gene orientation across hemipteran Hox clusters. (A) Order and transcriptional orientation of canonical and non-canonical Hox genes in each species. Splits within the Hox cluster are denoted by double black lines and inversions with respect to the predicted ancestral Hox cluster are annotated with a black border around the gene. (B) Organization of the Hox cluster per species, shown using actual genomic distances. Each line represents a Hox gene as it occurs in the genome. Genomic distances are shown in Megabases (Mb).

scientific community through the ATF database. This approach may be of interest to researchers who want to annotate other aphids to expand the ATFdb database. As this database already contains the annotations of the *hnf4* and *Hox* genes in the hemipteran genomes with a RefSeq annotation, ATFdb could also be the starting point to develop a larger database, including the annotation of the full TF complement in hemipterans. To achieve this, further efforts are needed to increase the number of high-quality hemipteran genomes available, as only five of those genomes meet the criteria required for making a relevant comparative genomic analysis: the availability of a RefSeq annotation and of a genome resolved at the chromosome level.

Despite the evolutionary divergence between the two species, homologs for most *D. melanogaster* TF families have been found in *A. pisum*. This includes the identification of an *A. pisum* homolog of HNF4, a key TF regulating metabolic homeostasis and that had not been previously described in aphids. We propose that the major differences in development, physiology and reproduction between the two species (e.g. holometabolous vs. hemimetabolous development, strictly sexual vs. reproductive polyphenisms including alternation of sexual and asexual parthenogenetic reproduction, strictly winged adults vs. wing polyphenism with alate and apterous individuals, polyphagous vs. specialized, phloem-restricted) can be explained by the numerous differences they display in their TF repertoires. In particular, the expansion of distinct sets of chromatin-associated genes suggests an important role for epigenetic regulation in the pea aphid. Finally, we performed and discussed in-depth analyses of the ZNF TFs with certain families expanded in the pea aphid, and of the *Hox* gene cluster, which shows a reorganization of gene position in the pea aphid compared to *Drosophila* and other insect models.

CRedit authorship contribution statement

Nicolas Parisot: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Mélanie Ribeiro Lopes:** Writing – original draft, Methodology, Investigation, Formal analysis. **Sergio Peignier:** Writing – review & editing, Validation, Resources, Data curation. **Patrice Baa-Puyoulet:** Writing – review & editing, Software, Data curation. **Hubert Charles:** Writing – review & editing, Validation, Formal analysis. **Federica Calevro:** Writing – review & editing, Writing – original draft, Supervision, Investigation, Funding acquisition, Conceptualization. **Patrick Callaerts:** Writing – review & editing, Writing – original draft, Supervision, Investigation, Funding acquisition, Conceptualization.

Funding

This work was supported by INRAE (Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement), INSA Lyon (Institut National des Sciences Appliquées de Lyon), KU Leuven Grants C14/17/099 and C14/22/126, and FWO (Flemish Fund for Scientific Research) grants G050822N, G065408.N10 and G078914N.

Acknowledgements

The authors would like to acknowledge Karen Gaget for her help at the beginning of the project and Aurélie Herbomez for secretarial assistance.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ibmb.2024.104217>.

References

- Aerts, S., 2012. Chapter five - computational strategies for the genome-wide identification of cis-regulatory elements and transcriptional targets. In: Plaza, S., Payre, F. (Eds.), *Current Topics in Developmental Biology, Transcriptional Switches during Development*. Academic Press, pp. 121–145. <https://doi.org/10.1016/B978-0-12-386499-4.00005-7>.
- Alexandre, C., Jacinto, A., Ingham, P.W., 1996. Transcriptional activation of hedgehog target genes in *Drosophila* is mediated directly by the cubitus interruptus protein, a member of the GLI family of zinc finger DNA-binding proteins. *Gene Dev.* 10, 2003–2013. <https://doi.org/10.1101/gad.10.16.2003>.
- Allen, D., Seo, J., 2018. ER stress activates the TOR pathway through Atf6. *J. Mol. Signal.* 13, 1. <https://doi.org/10.5334/1750-2187-13-1>.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- Baa-Puyoulet, P., Parisot, N., Febvay, G., Huerta-Cepas, J., Vellozo, A.F., Gabaldón, T., Calevro, F., Charles, H., Colella, S., 2016. ArthropodaCyc: a CycADS powered collection of BioCyc databases to analyse and compare metabolism of arthropods. Database. <https://doi.org/10.1093/database/baw081> (Oxford) 2016:baw081.
- Bao, R., Fischer, T., Bolognesi, R., Brown, S.J., Friedrich, M., 2012. Parallel duplication and partial subfunctionalization of β -Catenin/Armadillo during insect evolution. *Mol. Biol. Evol.* 29, 647–662. <https://doi.org/10.1093/molbev/msr219>.
- Baumann, A., Fujiwara, Y., Wilson, T.G., 2010. Evolutionary divergence of the paralogs Methoprene tolerant (Met) and germ cell expressed (gce) within the genus *Drosophila*. *J. Insect Physiol.* 56, 1445–1455. <https://doi.org/10.1016/j.jinsphys.2010.05.001>. *Insect Molecular Physiology - Basic Science to Application, A Special Issue in Honour of Dr. Judith H. Willis*.
- Baumann, P., 2005. Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. *Annu. Rev. Microbiol.* 59, 155–189. <https://doi.org/10.1146/annurev.micro.59.030804.121041>.
- Bellen, H.J., Yamamoto, S., 2015. Morgan's legacy: fruit flies and the functional annotation of conserved genes. *Cell* 163, 12–14. <https://doi.org/10.1016/j.cell.2015.09.009>.
- Ben-David, J., Chipman, A.D., 2010. Mutual regulatory interactions of the trunk gap genes during blastoderm patterning in the hemipteran *Oncopeltus fasciatus*. *Dev. Biol.* 346, 140–149. <https://doi.org/10.1016/j.ydbio.2010.07.010>.
- Bodmer, R., 1993. The gene tinman is required for specification of the heart and visceral muscles in *Drosophila*. *Development* 118, 719–729. <https://doi.org/10.1242/dev.118.3.719>.
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- Bornemann, D., Miller, E., Simon, J., 1996. The *Drosophila* Polycomb group gene Sex comb on midleg (Scm) encodes a zinc finger protein with similarity to polyhomeotic protein. *Development* 122, 1621–1630. <https://doi.org/10.1242/dev.122.5.1621>.
- Braendle, C., Miura, T., Bickel, R., Shingleton, A.W., Kambhampati, S., Stern, D.L., 2003. Developmental origin and evolution of bacteriocytes in the aphid-*Buchnera* symbiosis. *PLoS Biol.* 1, e21. <https://doi.org/10.1371/journal.pbio.0000021>.
- Brown, S.J., Fellers, J.P., Shippy, T.D., Richardson, E.A., Maxwell, M., Stuart, J.J., Denell, R.E., 2002. Sequence of the *Tribolium castaneum* homeotic complex: the region corresponding to the *Drosophila melanogaster* Antennapedia complex. *Genetics* 160, 1067–1074. <https://doi.org/10.1093/genetics/160.3.1067>.
- Brunner, E., Peter, O., Schweizer, L., Basler, K., 1997. Pangolin encodes a Lef-1 homologue that acts downstream of Armadillo to transduce the Wingless signal in *Drosophila*. *Nature* 385, 829–833. <https://doi.org/10.1038/385829a0>.
- Calevro, F., Callaerts, P., Matsuura, Y., Michalik, A., 2023. Editorial: symbiotic organs in insects: development, metabolism, and physiological regulation. *Front. Physiol.* 14. <https://doi.org/10.3389/fphys.2023.124865>.
- Calevro, F., Tagu, D., Callaerts, P., 2019. Genome of the Month: *Acyrtosiphon pisum*. *Trends Genet.* 35, 781–782. <https://doi.org/10.1016/j.tig.2019.07.003>.
- Carroll, S.B., 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134, 25–36. <https://doi.org/10.1016/j.cell.2008.06.030>.
- Celniker, S.E., Rubin, G.M., 2003. The *Drosophila melanogaster* genome. *Annu. Rev. Genom. Hum. Genet.* 4, 89–117. <https://doi.org/10.1146/annurev.genom.4.070802.110323>.
- Chai, C.-L., Zhang, Z., Huang, F.-F., Wang, X.-Y., Yu, Q.-Y., Liu, B.-B., Tian, T., Xia, Q.-Y., Lu, C., Xiang, Z.-H., 2008. A genomewide survey of homeobox genes and identification of novel structure of the Hox cluster in the silkworm, *Bombyx mori*. *Insect Biochem. Mol. Biol.* 38, 1111–1120. <https://doi.org/10.1016/j.ibmb.2008.06.008>. *Special Issue on the Silkworm Genome*.
- Cheatle Jarvela, A.M., Hinman, V.F., 2015. Evolution of transcription factor function as a mechanism for changing metazoan developmental gene regulatory networks. *EvoDevo* 6, 3. <https://doi.org/10.1186/2041-9139-6-3>.
- Chen, Z.-X., Sturgill, D., Qu, J., Jiang, H., Park, S., Boley, N., Suzuki, A.M., Fletcher, A.R., Plachetzki, D.C., FitzGerald, P.C., Artieri, C.G., Atallah, J., Barmina, O., Brown, J.B., Blankenbush, K.P., Clough, E., Dasgupta, A., Gubbala, S., Han, Y., Jayaseelan, J.C., Kalra, D., Kim, Y.-A., Kovar, C.L., Lee, S.L., Li, M., Malley, J.D., Malone, J.H., Mathew, T., Mattiuzzo, N.R., Munidasa, M., Muzny, D.M., Ongeri, F., Perales, L., Przytycka, T.M., Pu, L.-L., Robinson, G., Thornton, R.L., Saada, N., Scherer, S.E., Smith, H.E., Vinson, C., Warner, C.B., Worley, K.C., Wu, Y.-Q., Zou, X., Cherbas, P., Kellis, M., Eisen, M.B., Piano, F., Kionte, K., Fitch, D.H., Sternberg, P.W., Cutter, A. D., Duff, M.O., Hoskins, R.A., Graveley, B.R., Gibbs, R.A., Bickel, P.J., Kopp, A., Carninci, P., Celniker, S.E., Oliver, B., Richards, S., 2014. Comparative validation of

- the *D. melanogaster* modENCODE transcriptome annotation. *Genome Res.* 24, 1209–1223. <https://doi.org/10.1101/gr.159384.113>.
- Cheng, Y., Brunner, A.L., Kremer, S., DeVido, S.K., Stefaniuk, C.M., Kassis, J.A., 2014. Co-regulation of inverted and engrailed by a complex array of regulatory sequences in *Drosophila*. *Dev. Biol.* 395, 131–143. <https://doi.org/10.1016/j.ydbio.2014.08.021>.
- Cheng, Y., Li, Y., Li, W., Song, Y., Zeng, R., Lu, K., 2020. Effect of hepatocyte nuclear factor 4 on the fecundity of *Nilaparvata lugens*: insights from RNA interference combined with transcriptomic analysis. *Genomics* 112, 4585–4594. <https://doi.org/10.1016/j.ygeno.2020.08.002>.
- Chotard, C., Leung, W., Salecker, I., 2005. Glial cells missing and *gcm2* cell autonomously regulate both glial and neuronal development in the visual system of *Drosophila*. *Neuron* 48, 237–251. <https://doi.org/10.1016/j.neuron.2005.09.019>.
- Cohen, B., McGuffin, M.E., Pfeifle, C., Segal, D., Cohen, S.M., 1992. *apterous*, a gene required for imaginal disc development in *Drosophila* encodes a member of the LIM family of developmental regulatory proteins. *Genes Dev.* 6, 715–729. <https://doi.org/10.1101/gad.6.5.715>.
- Colella, S., Parisot, N., Simonet, P., Galet, K., Duport, G., Baa-Puyoulet, P., Rahbé, Y., Charles, H., Febvay, G., Callaerts, P., Calevro, F., 2018. Bacteriocyte reprogramming to cope with nutritional stress in a phlebotomus feeding hemipteran, the pea aphid *Acyrtosiphon pisum*. *Front. Physiol.* 25 (9), 1498. <https://doi.org/10.3389/fphys.2018.01498>.
- Cook, C.E., Smith, M.L., Telford, M.J., Bastianello, A., Akam, M., 2001. *Hox* genes and the phylogeny of the arthropods. *Curr. Biol.* 11, 759–763. [https://doi.org/10.1016/S0960-9822\(01\)00222-6](https://doi.org/10.1016/S0960-9822(01)00222-6).
- Coschigano, K.T., Wensink, P.C., 1993. Sex-specific transcriptional regulation by the male and female doublesex proteins of *Drosophila*. *Gene Dev.* 7, 42–54. <https://doi.org/10.1101/gad.7.1.42>.
- Curtis, C.D., Brisson, J.A., DeCamillis, M.A., Shipley, T.D., Brown, S.J., Denell, R.E., 2001. Molecular characterization of Cephalothorax, the *Tribolium* ortholog of Sex combs reduced. *Genesis* 30, 12–20. <https://doi.org/10.1002/gene.1027>.
- Dang, C.-W., Wang, Y., Chen, K.-P., Yao, Q., Zhang, D., Guo, M., 2011. The basic helix-loop-helix transcription factor family in the pea aphid, *Acyrtosiphon pisum*. *J. Insect Sci.* 11, 84. <https://doi.org/10.1673/031.011.8401>.
- Datta, R.R., Weasner, B.P., Kumar, J.P., 2011. A dissection of the teashirt and tiptop genes reveals a novel mechanism for regulating transcription factor activity. *Dev. Biol.* 360, 391–402. <https://doi.org/10.1016/j.ydbio.2011.09.030>.
- E.H. Davidson, M.S. Levine, Properties of developmental gene regulatory networks, *Proc. Natl. Acad. Sci. U.S.A.* 105 (51) 20063–20066, <https://doi.org/10.1073/pnas.0806007105> (2008).
- Davis, G.K., Brisson, J.A., Bickel, R.D., 2021. Evo-devo lessons learned from aphids. In: Nuño de la Rosa, L., Müller, G.B. (Eds.), *Evolutionary Developmental Biology: A Reference Guide*. Springer International Publishing, Cham, pp. 817–829. https://doi.org/10.1007/978-3-319-32979-6_182.
- Dearden, P.K., 2015. Origin and evolution of the enhancer of split complex. *BMC Genom.* 16, 712. <https://doi.org/10.1186/s12864-015-1926-1>.
- De Rosa, R., Grenier, J., Andreeva, T., Cook, E.C., Adoutte, A., Akam, M., Carroll, S.B., Balavoine, G., 1999. *Hox* genes in brachiopods and priapulids and protostome evolution. *Nature* 399, 772–776. <https://doi.org/10.1038/21631>.
- Desvergne, B., Michalik, L., Wahli, W., 2006. Transcriptional regulation of metabolism. *Physiol. Rev.* 86, 465–514. <https://doi.org/10.1152/physrev.00025.2005>.
- Domínguez, M., Casares, F., 2005. Organ specification-growth control connection: new insights from the *Drosophila* eye-antennal disc. *Dev. Dynam.* 232, 673–684. <https://doi.org/10.1002/dvdy.20311>.
- Douglas, A.E., 2015. Multiorganismal insects: diversity and function of resident microorganisms. *Annu. Rev. Entomol.* 60, 17–34. <https://doi.org/10.1146/annurev-ento-010814-020822>.
- Dow, J.A.T., Simons, M., Romero, M.F., 2022. *Drosophila melanogaster*: a simple genetic model of kidney structure, function and disease. *Nat. Rev. Nephrol.* 18, 417–434. <https://doi.org/10.1038/s41581-022-00561-4>.
- Erlick, T., Hartenstein, V., Lipshitz, H.D., McInnes, R.R., 2008. Conserved role of the *Vsx* genes supports a monophyletic origin for bilaterian visual systems. *Curr. Biol.* 18, 1278–1287. <https://doi.org/10.1016/j.cub.2008.07.076>.
- Fischer, F.P., Karge, R.A., Weber, Y.G., Koch, H., Wolkung, S., Voigt, A., 2023. *Drosophila melanogaster* as a versatile model organism to study genetic epilepsies: an overview. *Front. Mol. Neurosci.* 16–2023. <https://doi.org/10.3389/fnmol.2023.1116000>.
- Fröblius, A.C., Funch, P., 2017. Rotiferan *Hox* genes give new insights into the evolution of metazoan bodyplans. *Nat. Commun.* 8, 9. <https://doi.org/10.1038/s41467-017-00020-w>.
- Gallant, P., Shiio, Y., Cheng, P.F., Parkhurst, S.M., Eisenman, R.N., 1996. Myc and Max homologs in *Drosophila*. *Science* 274, 1523–1527. <https://doi.org/10.1126/science.274.5292.1523>.
- Gonzalez, D.H., 2016. Chapter 1 - introduction to transcription factor structure and function. In: Gonzalez, D.H. (Ed.), *Plant Transcription Factors*. Academic Press, Boston, pp. 3–11. <https://doi.org/10.1016/B978-0-12-800854-6.00001-4>.
- Gramates, L.S., Agapite, J., Attrill, H., Calvi, B.R., Crosby, M.A., dos Santos, G., Goodman, J.L., Goutte-Gattat, D., Jenkins, V.K., Kaufman, T., Larkin, A., Matthews, B.B., Millburn, G., Strelets, V.B., the FlyBase Consortium, 2022. FlyBase: a guided tour of highlighted features. *Genetics* 220, iyac035. <https://doi.org/10.1093/genetics/iyac035>.
- Gregor, T., McGregor, A.P., Wieschaus, E.F., 2008. Shape and function of the Bicoid morphogen gradient in dipteran species with different sized embryos. *Dev. Biol.* 316, 350–358. <https://doi.org/10.1016/j.ydbio.2008.01.039>.
- Grossniklaus, U., Pearson, R.K., Gehring, W.J., 1992. The *Drosophila* sloppy paired locus encodes two proteins involved in segmentation that show homology to mammalian transcription factors. *Gene Dev.* 6, 1030–1051. <https://doi.org/10.1101/gad.6.6.1030>.
- Gutiérrez, L., Oktaba, K., Scheuermann, J.C., Gambetta, M.C., Ly-Hartig, N., Müller, J., 2012. The role of the histone H2A ubiquitinase Sce in Polycomb repression. *Development* 139, 117–127. <https://doi.org/10.1242/dev.074450>.
- Hamaguchi, T., Yabe, S., Uchiyama, H., Murakami, R., 2004. *Drosophila* Tbx6-related gene, Dorsocross, mediates high levels of Dpp and Scw signal required for the development of amnioserosa and wing disc primordium. *Dev. Biol.* 265, 355–368. <https://doi.org/10.1016/j.ydbio.2003.09.034>.
- Hammonds, A.S., Bristow, C.A., Fisher, W.W., Weiszmann, R., Wu, S., Hartenstein, V., Kellis, M., Yu, B., Frise, E., Celniker, S.E., 2013. Spatial expression of transcription factors in the unfolded protein organ development. *Genome Biol.* 14, R140. <https://doi.org/10.1186/gb-2013-14-12-r140>.
- Han, J., Kaufman, R.J., 2017. Physiological/pathological ramifications of transcription factors in the unfolded protein response. *Gene Dev.* 31, 1417–1438. <https://doi.org/10.1101/gad.297374.117>.
- Hardin, P.E., 2011. Chapter 5 - molecular genetic analysis of circadian timekeeping in *Drosophila*. In: Brody, S. (Ed.), *Advances in Genetics, the Genetics of Circadian Rhythms*. Academic Press, pp. 141–173. <https://doi.org/10.1016/B978-0-12-387690-4.00005-2>.
- Helfand, S.L., Rogina, B., 2003. Genetics of aging in the fruit fly, *Drosophila melanogaster*. *Annu. Rev. Genet.* 37, 329–348. <https://doi.org/10.1146/annurev.genet.37.040103.095211>.
- Hens, K., Feuz, J.-D., Isakova, A., Iagovitina, A., Massouras, A., Bryois, J., Callaerts, P., Celniker, S.E., Deplancke, B., 2011. Automated protein-DNA interaction screening of *Drosophila* regulatory elements. *Nat. Methods* 8, 1065–1070. <https://doi.org/10.1038/nmeth.1763>.
- Higashijima, S., Kojima, T., Michiue, T., Ishimaru, S., Emori, Y., Saigo, K., 1992. Dual Bar homeo box genes of *Drosophila* required in two photoreceptor cells, R1 and R6, and primary pigment cells for normal eye development. *Gene Dev.* 6, 50–60. <https://doi.org/10.1101/gad.6.1.50>.
- Hou, X.S., Melnick, M.B., Perrimon, N., 1996. Marelle acts downstream of the *Drosophila* HOP/JAK kinase and encodes a protein similar to the mammalian STATs. *Cell* 84, 411–419. [https://doi.org/10.1016/S0092-8674\(00\)81286-6](https://doi.org/10.1016/S0092-8674(00)81286-6).
- Huang, F., Saraf, A., Florens, L., Kusch, T., Swanson, S.K., Szerszen, L.T., Li, G., Dutta, A., Washburn, M.P., Abmayr, S.M., Workman, J.L., 2016. The Enok acetyltransferase complex interacts with Elg1 and negatively regulates PCNA unloading to promote the G1/S transition. *Gene Dev.* 30 (10), 1198–1210. <https://doi.org/10.1101/gad.271429.115>.
- Huang, H.-W., Zeng, X., Rhim, T., Ron, D., Ryoo, H.D., 2017. The requirement of IRE1 and XBP1 in resolving physiological stress during *Drosophila* development. *J. Cell Sci.* 130, 3040–3049. <https://doi.org/10.1242/jcs.203612>.
- Huang, J., Wu, S., Barrera, J., Matthews, K., Pan, D., 2005. The Hippo signaling pathway coordinately regulates cell proliferation and apoptosis by inactivating Yorkie, the *Drosophila* homolog of YAP. *Cell* 122, 421–434. <https://doi.org/10.1016/j.cell.2005.06.007>.
- i5K Consortium, 2013. The i5K Initiative: advancing arthropod genomics for knowledge, human health, agriculture, and the environment. *J. Hered.* 104, 595–600. <https://doi.org/10.1093/jhered/est050>.
- Jagla, K., Frasch, M., Jagla, T., Dretzen, G., Bellard, F., Bellard, M., 1997. ladybird, a new component of the cardiogenic pathway in *Drosophila* required for diversification of heart precursors. *Development* 124, 3471–3479. <https://doi.org/10.1242/dev.124.18.3471>.
- Jin, J., Tian, F., Yang, D.-C., Meng, Y.-Q., Kong, L., Luo, J., Gao, G., 2017. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* 45, D1040–D1045. <https://doi.org/10.1093/nar/gkw982>.
- Jindra, M., Bellés, X., Shinoda, T., 2015. Molecular basis of juvenile hormone signaling. *Curr. Opin. Insect Sci.* 11, 39–46. <https://doi.org/10.1016/j.cois.2015.08.004>.
- Global change biology * Molecular physiology.
- Just, J., Laslo, M., Lee, Y.-J., Yarnell, M., Zhang, Z., Angelini, D.R., 2023. Distinct developmental mechanisms influence sexual dimorphisms in the milkweed bug *Oncopeltus fasciatus*. *Proc. Biol. Sci.* 290, 20222083. <https://doi.org/10.1098/rspb.2022.2083>.
- Kanehisa, M., Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. <https://doi.org/10.1093/nar/28.1.27>.
- Kelly, K.K., Meadows, S.M., Cripps, R.M., 2002. *Drosophila* MEF2 is a direct regulator of Actin57B transcription in cardiac, skeletal, and visceral muscle lineages. *Mech. Dev.* 110, 39–50. [https://doi.org/10.1016/S0925-4773\(01\)00586-X](https://doi.org/10.1016/S0925-4773(01)00586-X).
- Kim, S., Wysocka, J., 2023. Deciphering the multi-scale, quantitative cis-regulatory code. *Mol. Cell* 83, 373–392. <https://doi.org/10.1016/j.molcel.2022.12.032>. Reimagining the Central Dogma.
- Kirfel, P., Skaljac, M., Grotmann, J., Kessel, T., Seip, M., Michaelis, K., Vilcinskas, A., 2020. Inhibition of histone acetylation and deacetylase enzymes affects longevity, development, and fecundity in the pea aphid (*Acyrtosiphon pisum*). *Arch. Insect Biochem. Physiol.* 103, e21614. <https://doi.org/10.1002/arch.21614>.
- Kumar, J.P., 2010. Chapter one - retinal Determination: the beginning of eye development. In: Cagan, R.L., Reh, T.A. (Eds.), *Current Topics in Developmental Biology, Invertebrate and Vertebrate Eye Development*. Academic Press, pp. 1–28. <https://doi.org/10.1016/B978-0-12-385044-7.00001-1>.
- Kummerfeld, S.K., Teichmann, S.A., 2006. DBD: a transcription factor prediction database. *Nucleic Acids Res.* 34, D74–D81. <https://doi.org/10.1093/nar/gkj131>.
- Lagarou, A., Mohd-Sarip, A., Moshkin, Y.M., Chalkley, G.E., Bezstarosti, K., Demmers, J. A., Verrijzer, C.P., 2008. dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. *Genes Dev.* 22, 2799–2810. <https://doi.org/10.1101/gad.484208>.
- Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., Weirauch, M.T., 2018. The human transcription factors. *Cell* 172, 650–665. <https://doi.org/10.1016/j.cell.2018.01.029>.

- Laslo, M., Just, J., Angelini, D.R., 2023. Theme and variation in the evolution of insect sex determination. *J. Exp. Zool. B Mol. Dev. Evol.* 340, 162–181. <https://doi.org/10.1002/jez.b.23125>.
- Lassmann, T., Sonnhammer, E.L., 2005. Kalign – an accurate and fast multiple sequence alignment algorithm. *BMC Bioinf.* 6, 298. <https://doi.org/10.1186/1471-2105-6-298>.
- Legeai, F., Shigenobu, S., Gauthier, J.P., Colbourne, J., Rispe, C., Collin, O., Richards, S., Wilson, A.C., Murphy, T., Tagu, D., 2010. AphidBase: a centralized bioinformatic resource for annotation of the pea aphid genome. *Insect Mol. Biol.* 19 (Suppl. 2), 5–12. <https://doi.org/10.1111/j.1365-2583.2009.00930.x>.
- Le Trionnaire, G., Tanguy, S., Hudaverdian, S., Gleonnec, F., Richard, G., Cayrol, B., Monsion, B., Pichon, E., Deshoux, M., Webster, C., Uzeit, M., Herpin, A., Tagu, D., 2019. An integrated protocol for targeted mutagenesis with CRISPR-Cas9 system in the pea aphid. *Insect Biochem. Mol. Biol.* 110, 34–44. <https://doi.org/10.1016/j.ibmb.2019.04.016>.
- Letunic, I., Bork, P., 2019. Interactive Tree of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. <https://doi.org/10.1093/nar/gkz239>.
- Lemon, B., Tjian, R., 2000. Orchestrated response: a symphony of transcription factors for gene control. *Gene Dev.* 14, 2551–2569. <https://doi.org/10.1101/gad.831000>.
- Liao, Y., Smyth, G.K., Shi, W., 2019. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res.* 47, e47, 110.1093/nar/gkz114.
- Li, K., Jia, Q., Li, S., 2018. Juvenile hormone signaling – a mini review. *Insect Sci.* 26, 600–606. <https://doi.org/10.1111/1744-7917.12614>.
- Li, Y., Park, H., Smith, T.E., Moran, N.A., 2019. Gene family evolution in the pea aphid based on chromosome-level genome assembly. *Mol. Biol. Evol.* 36, 2143–2156. <https://doi.org/10.1093/molbev/msz138>.
- Loo, L.W.M., Secombe, J., Little, J.T., Carlos, L.-S., Yost, C., Cheng, P.-F., Flynn, E.M., Edgar, B.A., Eisenman, R.N., 2005. The transcriptional repressor dMnt is a regulator of growth in *Drosophila melanogaster*. *Mol. Cell Biol.* 25, 7078–7091. <https://doi.org/10.1128/MCB.25.16.7078-7091.2005>.
- Lours, C., Bardot, O., Godt, D., Laski, F.A., Couderc, J., 2003. The *Drosophila melanogaster* BTB proteins bric à brac bind DNA through a composite DNA binding domain containing a pipsqueak and an AT-hook motif. *Nucleic Acids Res.* 31, 5389–5398. <https://doi.org/10.1093/nar/gkg724>.
- Luscombe, N.M., Austin, S.E., Berman, H.M., Thornton, J.M., 2000. An overview of the structures of protein-DNA complexes. *Genome Biol.* 1. <https://doi.org/10.1186/gb-2000-1-1-reviews001> reviews001.1.
- Lyne, R., Smith, R., Rutherford, K., Wakeling, M., Varley, A., Guillier, F., Janssens, H., Ji, W., McLaren, P., North, P., Rana, D., Riley, T., Sullivan, J., Watkins, X., Woodbridge, M., Lilley, K., Russell, S., Ashburner, M., Mizuguchi, K., Micklem, G., 2007. FlyMine: an integrated database for *Drosophila* and *Anopheles* genomics. *Genome Biol.* 8, R129. <https://doi.org/10.1186/gb-2007-8-7-r129>.
- Mach, V., Ohno, K., Kokubo, H., Suzuki, Y., 1996. The *Drosophila* fork head factor directly controls larval salivary gland-specific expression of the glue protein gene Sgs3. *Nucleic Acids Res.* 24, 2387–2394. <https://doi.org/10.1093/nar/24.12.2387>.
- Madeira, F., Pearce, M., Tivey, A.R.N., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N., Kolesnikov, A., Lopez, R., 2022. Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* 50, W276–W279. <https://doi.org/10.1093/nar/gkac240>.
- Mannervik, M., Levine, M., 1999. The Rpd3 histone deacetylase is required for segmentation of the *Drosophila* embryo. *Proc. Natl. Acad. Sci. USA* 96, 6797–6801. <https://doi.org/10.1073/pnas.96.12.6797>.
- Martín, M., Organista, M.F., de Celis, J.F., 2016. Structure of developmental gene regulatory networks from the perspective of cell fate-determining genes. *Transcription* 7, 32–37. <https://doi.org/10.1080/21541264.2015.1130118>.
- Mathers, T.C., Mugford, S.T., Percival-Alwyn, L., Chen, Y., Kaithakotti, G., Swarbreck, D., Hogenhout, S.A., van Oosterhout, C., 2019. Sex-specific changes in the aphid DNA methylation landscape. *Mol. Ecol.* 28, 4228–4241. <https://doi.org/10.1111/mec.15216>.
- Matsuura, Y., Kikuchi, Y., Miura, T., Fukatsu, T., 2015. Ultrathorax is essential for bacteriocyte development. *Proc. Natl. Acad. Sci. USA* 112, 9376–9381. <https://doi.org/10.1073/pnas.1503371112>.
- Mendez, D.L., Kim, D., Chruszcz, M., Stephens, G.E., Minor, W., Khorasanizadeh, S., Elgin, S.C.R., 2011. The HP1a disordered C terminus and chromo shadow domain cooperate to select target peptide partners. *Chembiochem* 12, 1084–1096. <https://doi.org/10.1002/cbic.201000598>.
- Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., Finn, R.D., Bateman, A., 2021. Pfam: the protein families database in 2021. *Nucleic Acids Res.* 49, D412–D419. <https://doi.org/10.1093/nar/gkaa913>.
- Miura, K., Oda, M., Makita, S., Chinzei, Y., 2005. Characterization of the *Drosophila* Methoprene-tolerant gene product. *FEBS J.* 272, 1169–1178. <https://doi.org/10.1111/j.1742-4658.2005.04552.x>.
- Mohr, S.E., Perrimon, N., 2019. *Drosophila melanogaster*: a simple system for understanding complexity. *Dis. Model. Mech.* 12, dmm041871. <https://doi.org/10.1242/dmm.041871>.
- Mulhair, P.O., Holland, P.W.H., 2022. Evolution of the insect Hox gene cluster: comparative analysis across 243 species. *Semin. Cell Dev. Biol.* 152–153, 4–15. <https://doi.org/10.1016/j.semcdb.2022.11.010>. Hox genes: The Original Body Builders.
- Näär, A.M., Lemon, B.D., Tjian, R., 2001. Transcriptional coactivator complexes. *Annu. Rev. Biochem.* 70, 475–501. <https://doi.org/10.1146/annurev.biochem.70.1.475>.
- Nardelli, J., Gibson, T.J., Vesque, C., Charnay, P., 1991. Base sequence discrimination by zinc-finger DNA-binding domains. *Nature* 349, 175–178. <https://doi.org/10.1038/349175a0>.
- Ogawa, K., Miura, T., 2014. Aphid polyphenisms: trans-generational developmental regulation through viviparity. *Front. Physiol.* 5, 1. <https://doi.org/10.3389/fphys.2014.00001>.
- O'Neill, E.M., Rebay, I., Tjian, R., Rubin, G.M., 1994. The activities of two Ets-related transcription factors required for *Drosophila* eye development are modulated by the Ras/MAPK pathway. *Cell* 78, 137–147. [https://doi.org/10.1016/0092-8674\(94\)90580-0](https://doi.org/10.1016/0092-8674(94)90580-0).
- Pace, R.M., Grbić, M., Nagy, L.M., 2016. Composition and genomic organization of arthropod Hox clusters. *EvoDevo* 7, 11. <https://doi.org/10.1186/s13227-016-0048-4>.
- Palanker, L., Tennesen, J.M., Lam, G., Thummel, C.S., 2009. *Drosophila* HNF4 regulates lipid mobilization and β -oxidation. *Cell Metabol.* 9, 228–239. <https://doi.org/10.1016/j.cmet.2009.01.009>.
- Panfili, K.A., 2008. Extraembryonic development in insects and the acrobatics of blastokinesis. *Dev. Biol.* 313, 471–491. <https://doi.org/10.1016/j.ydbio.2007.11.004>.
- Panfili, K.A., Vargas Jentsch, I.M., Benoit, J.B., Erezylmaz, D., Suzuki, Y., Colella, S., Robertson, H.M., Poelchau, M.F., Waterhouse, R.M., Ioannidis, P., Weirauch, M.T., Hughes, D.S.T., Murali, S.C., Werren, J.H., Jacobs, C.G.C., Duncan, E.J., Armisén, D., Vreede, B.M.L., Baa-Puyoulet, P., Berger, C.S., Chang, C., Chao, H., Chen, M.-J.M., Chen, Y.-T., Childers, C.P., Chipman, A.D., Cridge, A.G., Crumière, A.J.J., Dearden, P.K., Didion, E.M., Dinh, H., Doddapaneni, H.V., Dolan, A., Dugan, S., Extavour, C.G., Febvay, G., Friedrich, M., Ginzburg, N., Han, Y., Heger, P., Holmes, C.J., Horn, T., Hsiao, Y., Jennings, E.C., Johnston, J.S., Jones, T.E., Jones, J. W., Khila, A., Koelzer, S., Kovacova, V., Leask, M., Lee, S.L., Lee, C.-Y., Lovegrove, M. R., Lu, H., Lu, Y., Moore, P.J., Munoz-Torres, M.C., Muzny, D.M., Palli, S.R., Parisot, N., Pick, L., Porter, M.L., Qu, J., Refki, P.N., Richter, R., Rivera-Pomar, R., Rosendale, A.J., Roth, S., Sachs, L., Santos, M.E., Seibert, J., Sghaier, E., Shukla, J.N., Stancliffe, R.J., Tidswell, O., Traverso, L., van der Zee, M., Viala, S., Worley, K.C., Zdobnov, E.M., Gibbs, R.A., Richards, S., 2019. Molecular evolutionary trends and feeding ecology diversification in the Hemiptera, anchored by the milkweed bug genome. *Genome Biol.* 20, 64. <https://doi.org/10.1186/s13059-019-1660-0>.
- Paplauskas, S., Morton, O., Hunt, M., Courage, A., Swanney, S., Dennis, S.R., Becker, D., Auld, S.K.J.R., Beckerman, A.P., 2024. Predator-induced shape plasticity in *Daphnia pulex*. *Ecol. Evol.* 14, 2. <https://doi.org/10.1002/ece3.10913>.
- Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B.L., Salazar, G.A., Bileschi, M.L., Bork, P., Bridge, A., Colwell, L., Gough, J., Haft, D.H., Letunic, I., Marchler-Bauer, A., Mi, H., Natale, D.A., Orengo, C.A., Pandurangan, A.P., Rivoire, C., Sigrist, C.J.A., Sillitoe, I., Thanki, N., Thomas, P.D., Tosatto, S.C.E., Wu, C.H., Bateman, A., 2023. InterPro in 2022. *Nucleic Acids Res.* 51, D418–D427. <https://doi.org/10.1093/nar/gkac993>.
- Peel, A.D., Telford, M.J., Akam, M., 2006. The evolution of hexapod engrailed-family genes: evidence for conservation and concerted evolution. *Proc. Biol. Sci.* 273, 1733–1742. <https://doi.org/10.1098/rspb.2006.3497>.
- Peñalosa-Ruiz, G., Bright, A.R., Mulder, K.W., Veenstra, G.J.C., 2019. The interplay of chromatin and transcription factors during cell fate transitions in development and reprogramming. *Biochim. Biophys. Acta (BBA) - Gene Regul. Mech.* 1862, 194407. <https://doi.org/10.1016/j.bbagg.2019.194407>.
- Perkins, K.K., Admon, A., Patel, N., Tjian, R., 1990. The *Drosophila* Fos-related AP-1 protein is a developmentally regulated transcription factor. *Genes Dev.* 4, 822–834. <https://doi.org/10.1101/gad.4.5.822>.
- Puig, O., Marr, M.T., Ruhf, M.L., Tjian, R., 2003. Control of cell number by *Drosophila* FOXO: downstream and feedback regulation of the insulin receptor pathway. *Genes Dev.* 17, 2006–2020. <https://doi.org/10.1101/gad.1098703>.
- Rebay, I., Rubin, G.M., 1995. Yan functions as a general inhibitor of differentiation and is negatively regulated by activation of the Ras1/MAPK pathway. *Cell* 81, 857–866. [https://doi.org/10.1016/0092-8674\(95\)90006-3](https://doi.org/10.1016/0092-8674(95)90006-3).
- Ribeiro Lopes, M., Gaget, K., Renoz, F., Dupont, G., Balmant, S., Charles, H., Callaerts, P., Calevro, F., 2022. Bacteriocyte plasticity in pea aphids facing amino acid stress or starvation during development. *Front. Physiol.* 13, 982920. <https://doi.org/10.3389/fphys.2022.982920>.
- Rider Jr, S.D., Srinivasan, D.G., Hilgarth, R.S., 2010. Chromatin-remodelling proteins of the pea aphid, *Acyrtosiphon pisum* (Harris). *Insect Mol. Biol.* 19, 201–214. <https://doi.org/10.1111/j.1365-2583.2009.00972.x>.
- Robert, X., Gouet, P., 2014. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* 42, W320–W324. <https://doi.org/10.1093/nar/gku316>.
- Romero, I.G., Ruvinsky, I., Gilad, Y., 2012. Comparative studies of gene expression and the evolution of gene regulation. *Nat. Rev. Genet.* 13, 505–516. <https://doi.org/10.1038/nrg3229>.
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. <https://doi.org/10.1093/sysbio/sys029>.
- Sapountzis, P., Dupont, G., Balmant, S., Gaget, K., Jaubert-Possamai, S., Febvay, G., Charles, H., Rahbé, Y., Colella, S., Calevro, F., 2014. New insight into the RNA interference response against cathepsin-L gene in the pea aphid, *Acyrtosiphon pisum*: molting or gut phenotypes specifically induced by injection or feeding treatments. *Insect Biochem. Mol. Biol.* 51, 20–32. <https://doi.org/10.1016/j.ibmb.2014.05.005>.
- Schaeper, N.D., Prpic, N.-M., Wimmer, E.A., 2009. A conserved function of the zinc finger transcription factor Sp8/9 in allometric appendage growth in the milkweed bug *Oncopeltus fasciatus*. *Dev. Gene. Evol.* 219, 427–435. <https://doi.org/10.1007/s00427-009-0301-0>.

- Schember, I., Halfon, M.S., 2022. Common themes and future challenges in understanding gene regulatory network evolution. *Cells* 11, 510. <https://doi.org/10.3390/cells11030510>.
- Schmidt-Ott, U., 2000. The amnioserosa is an apomorphic character of cyclorrhaphan flies. *Dev. Gene. Evol.* 210, 373–376. <https://doi.org/10.1007/s004270000068>.
- Schweisguth, F., Posakony, J.W., 1992. Suppressor of Hairless, the *Drosophila* homolog of the mouse recombination signal-binding protein gene, controls sensory organ cell fates. *Cell* 69, 1199–1212. [https://doi.org/10.1016/0092-8674\(92\)90641-O](https://doi.org/10.1016/0092-8674(92)90641-O).
- Schurko, A.M., Logsdon, J.M., Eads, B.D., 2009. Meiosis genes in *Daphnia pulex* and the role of parthenogenesis in genome evolution. *BMC Evol. Biol.* 9, 78. <https://doi.org/10.1186/1471-2148-9-78>.
- Scott, E.K., Lee, T., Luo, L., 2001. Enok encodes a *Drosophila* putative histone acetyltransferase required for mushroom body neuroblast proliferation. *Curr. Biol.* 11, 99–104. [https://doi.org/10.1016/S0960-9822\(01\)00020-3](https://doi.org/10.1016/S0960-9822(01)00020-3).
- Sekelsky, J.J., Newfeld, S.J., Raftery, L.A., Chartoff, E.H., Gelbart, W.M., 1995. Genetic characterization and cloning of mothers against dpp, a gene required for decapentaplegic function in *Drosophila melanogaster*. *Genetics* 139, 1347–1358. <https://doi.org/10.1093/genetics/139.3.1347>.
- Shigenobu, S., Bickel, R.D., Brisson, J.A., Butts, T., Chang, C., Christiaens, O., Davis, G.K., Duncan, E.J., Ferrier, D.E.K., Iga, M., Janssen, R., Lin, G.-W., Lu, H.-L., McGregor, A.P., Miura, T., Smagghe, G., Smith, J.M., Van Der Zee, M., Velarde, R.A., Wilson, M.J., Dearden, P.K., Stern, D.L., 2010. Comprehensive survey of developmental genes in the pea aphid, *Acyrtosiphon pisum*: frequent lineage-specific duplications and losses of developmental genes. *Insect Mol. Biol.* 19, 47–62. <https://doi.org/10.1111/j.1365-2583.2009.00944.x>.
- Shippy, T.D., Brown, S.J., Denell, R.E., 2000. Maxillopedia is the *Tribolium* ortholog of proscopia. *Evol. Dev.* 2, 145–151. <https://doi.org/10.1046/j.1525-142x.2000.00055.x>.
- Shokri, L., Inukai, S., Hafner, A., Weinand, K., Hens, K., Vedenko, A., Gisselbrecht, S.S., Dainese, R., Bischof, J., Furger, E., Feuz, J.-D., Basler, K., Deplancke, B., Bulyk, M.L., 2019. A comprehensive *Drosophila melanogaster* transcription factor interactome. *Cell Rep.* 27, 955–970.e7. <https://doi.org/10.1016/j.celrep.2019.03.071>.
- Simonet, P., Gaget, K., Balmard, S., Ribeiro Lopes, M., Parisot, N., Buhler, K., Dupont, G., Vulsteke, V., Febvay, G., Heddi, A., Charles, H., Callaerts, P., Calevro, F., 2018. Bacteriocyte cell death in the pea aphid/*Buchnera* symbiotic system. *Proc. Natl. Acad. Sci. USA* 115, E1819–E1828. <https://doi.org/10.1073/pnas.1720237115>.
- Singh, A.J., Ramsey, S.A., Filtz, T.M., Kioussi, C., 2018. Differential gene regulatory networks in development and disease. *Cell. Mol. Life Sci.* 75, 1013–1025. <https://doi.org/10.1007/s00018-017-2679-6>.
- Song, L., Huang, S.C., Wise, A., Castanon, R., Nery, J.R., Chen, H., Watanabe, M., Thomas, J., Bar-Joseph, Z., Ecker, J.R., 2016. A transcription factor hierarchy defines an environmental stress response network. *Science* 354, aag1550. <https://doi.org/10.1126/science.aag1550>.
- Spitz, F., Furlong, E.E.M., 2012. Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* 13, 613–626. <https://doi.org/10.1038/nrg3207>.
- Sproul, J.S., Hotaling, S., Heckenhauer, J., Powell, A., Marshall, D., Larracuent, A.M., Kelley, J.L., Pauls, S.U., Frandsen, P.B., 2023. Analyses of 600+ insect genomes reveal repetitive element dynamics and highlight biodiversity-scale repeat annotation challenges. *Genome Res.* 33, 1708–1717. <https://doi.org/10.1101/gr.277387.122>.
- Stauber, M., Jäckle, H., Schmidt-Ott, U., 1999. The anterior determinant bicoid of *Drosophila* is a derived Hox class 3 gene. *Proc. Natl. Acad. Sci. USA* 96, 3786–3789. <https://doi.org/10.1073/pnas.96.7.3786>.
- Stevaux, O., Dimova, D., Frolov, M.V., Taylor-Harding, B., Morris, E., Dyson, N., 2002. Distinct mechanisms of E2F regulation by *Drosophila* RBF1 and RBF2. *EMBO J.* 21, 4927–4937. <https://doi.org/10.1093/emboj/cdf501>.
- Sykitiotis, G.P., Bohmann, D., 2008. Keap1/Nrf2 signaling regulates oxidative stress tolerance and lifespan in *Drosophila*. *Dev. Cell* 14, 76–85. <https://doi.org/10.1016/j.devcel.2007.12.002>.
- Tagu, D., Calevro, F., Colella, S., Gabaldon, T., Sugio, A., 2016. Functional and evolutionary genomics in aphids. In: Vilcinskas, A. (Ed.), *Biology and Ecology of Aphids*. CRC Press, Taylor and Francis Group, pp. 52–88, doi: <https://doi.org/10.1201/b19967>.
- Tataroglu, O., Emery, P., 2014. Studying circadian rhythms in *Drosophila melanogaster*. *Methods* 68, 140–150. <https://doi.org/10.1016/j.ymeth.2014.01.001>. *Drosophila* developmental biology methods.
- The International Aphid Genomics Consortium, 2010. Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS Biol.* 8, e1000313. <https://doi.org/10.1371/journal.pbio.1000313>.
- Theopold, U., Ekengren, S., Hultmark, D., 1996. HLH106, a *Drosophila* transcription factor with similarity to the vertebrate sterol responsive element binding protein. *Proc. Natl. Acad. Sci. USA* 93, 1195–1199. <https://doi.org/10.1073/pnas.93.3.1195>.
- Tian, F., Shah, P.K., Liu, X., Negre, N., Chen, J., Karpenko, O., White, K.P., Grossman, R. L., 2009. Flynet: a genomic resource for *Drosophila melanogaster* transcriptional regulatory networks. *Bioinformatics* 25, 3001–3004. <https://doi.org/10.1093/bioinformatics/btp469>.
- Tiebe, M., Lutz, M., De La Garza, A., Buechling, T., Boutros, M., Teleman, A.A., 2015. REPTOR and REPTOR-BP regulate organismal metabolism and transcription downstream of TORC1. *Dev. Cell* 33, 272–284. <https://doi.org/10.1016/j.devcel.2015.03.013>.
- Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., Luscombe, N.M., 2009. A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.* 10, 252–263. <https://doi.org/10.1038/nrg2538>.
- Vidal, N.M., Graziotin, A.L., Iyer, L.M., Aravind, L., Venancio, T.M., 2016. Transcription factors, chromatin proteins and the diversification of Hemiptera. *Insect Biochem. Mol. Biol.* 69, 1–13. <https://doi.org/10.1016/j.ibmb.2015.07.001>. Special issue: *Rhodnius prolixus* genome.
- Wang, K., Yang, Z., Li, X., Liu, S., Wang, L., Zhang, H., Yu, H., 2023. A hepatocyte nuclear factor BtabhNF4 mediates desiccation tolerance and fecundity in whitefly (*Bemisia tabaci*). *Environ. Entomol.* 52, 138–147. <https://doi.org/10.1093/ee/nvac103>.
- Wang, X., Pan, L., Wang, S., Zhou, J., McDowell, W., Park, J., Haug, J., Staehling, K., Tang, H., Xie, T., 2011. Histone H3K9 trimethylase eggless controls germline stem cell maintenance and differentiation. *PLoS Genet.* 7, e1002426. <https://doi.org/10.1371/journal.pgen.1002426>.
- Weisbrod, A., Cohen, M., Chipman, A.D., 2013. Evolution of the insect terminal patterning system—insights from the milkweed bug, *Oncopeltus fasciatus*. *Dev. Biol.* 380, 125–131. <https://doi.org/10.1016/j.ydbio.2013.04.030>.
- Wenger, Y., 2018. Pfam 18 IDs from. <http://www.transcriptionfactor.org>. <https://doi.org/10.6084/m9.figshare.6610229.v1>.
- Williams, J.A., Bell, J.B., Carroll, S.B., 1991. Control of *Drosophila* wing and haltere development by the nuclear vestigial gene product. *Genes Dev.* 5, 2481–2495. <https://doi.org/10.1101/gad.5.12b.2481>.
- Wilson, D., Charoensawan, V., Kummerfeld, S.K., Teichmann, S.A., 2008. DBD—taxonomically broad transcription factor predictions: new content and functionality. *Nucleic Acids Res.* 36, D88–D92. <https://doi.org/10.1093/nar/gkm964>.
- Wolf, F., Angerer, P., Theis, F., 2018. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* 19, 15. <https://doi.org/10.1186/s13059-017-1382-0>.
- Wu, S., Liu, Y., Zheng, Y., Dong, J., Pan, D., 2008. The TEAD/TEF family protein Scalloped mediates transcriptional output of the Hippo growth-regulatory pathway. *Dev. Cell* 14, 388–398. <https://doi.org/10.1016/j.devcel.2008.01.007>.
- Yamanaka, N., Rewitz, K.F., O'Connor, M.B., 2013. Ecdysone Control of developmental transitions: lessons from *Drosophila* research. *Annu. Rev. Entomol.* 58, 497–516. <https://doi.org/10.1146/annurev-ento-120811-153608>.
- Yang, H., Jaime, M., Polihronakis, M., Kanegawa, K., Markow, T., Kaneshiro, K., Oliver, B., 2018. Re-annotation of eight *Drosophila* genomes. *Life Sci. Alliance* 1. <https://doi.org/10.26508/lsa.201800156>.
- Yao, J.-G., Weasner, B.M., Wang, L.-H., Jang, C.-C., Weasner, B., Tang, C.-Y., Salzer, C.L., Chen, C.-H., Hay, B., Sun, Y.H., Kumar, J.P., 2008. Differential requirements for the Pax6(5a) genes eyegone and twin of eyegone during eye development in *Drosophila*. *Dev. Biol.* 315, 535–551. <https://doi.org/10.1016/j.ydbio.2007.12.037>.
- Yu, J., Wu, H., Wen, Y., Liu, Y., Zhou, T., Ni, B., Lin, Y., Dong, J., Zhou, Z., Hu, Z., Guo, X., Sha, J., Tong, C., 2015. Identification of seven genes essential for male fertility through a genome-wide association study of non-obstructive azoospermia and RNA interference-mediated large-scale functional screening in *Drosophila*. *Hum. Mol. Genet.* 24, 1493–1503. <https://doi.org/10.1093/hmg/ddu557>.
- Zaret, K.S., 2020. Pioneer transcription factors initiating gene network changes. *Annu. Rev. Genet.* 54, 367–385. <https://doi.org/10.1146/annurev-genet-030220-015007>.
- Zhang, R.-J., Chen, J., Jiang, L.-Y., Qiao, G.-X., 2019. The genes expression difference between winged and wingless bird cherry-oat aphid *Rhopalosiphum padi* based on transcriptomic data. *Sci. Rep.* 9, 4754. <https://doi.org/10.1038/s41598-019-41348-1>.
- Zhang, T., Zhou, Q., Ogmundsdottir, M.H., Möller, K., Siddaway, R., Larue, L., Hsing, M., Kong, S.W., Goding, C.R., Palsson, A., Steingrimsdottir, E., Pignoni, F., 2015. Mitf is a master regulator of the v-ATPase, forming a control module for cellular homeostasis with v-ATPase and TORC1. *J. Cell Sci.* 128, 2938–2950. <https://doi.org/10.1242/jcs.173807>.