



HAL
open science

Comprehensive annotation of transcription factors, chromatin-associated factors, and basal transcription machinery in the pea aphid, *Acyrtosiphon pisum*

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.. Comprehensive annotation of transcription factors, chromatin-associated factors, and basal transcription machinery in the pea aphid, *Acyrtosiphon pisum*. 2024. hal-04792881

HAL Id: hal-04792881

<https://hal.science/hal-04792881v1>

Preprint submitted on 20 Nov 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.

1 **Title:** Comprehensive annotation of transcription factors, chromatin-associated factors, and
2 basal transcription machinery in the pea aphid, *Acyrtosiphon pisum*, and development of the
3 ATFdb database, a resource for studies of transcriptional regulation

4

5 **Author names**

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

8

9 **Author affiliations**

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

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

13

14 **Co-corresponding authors**

15 nicolas.parisot@insa-lyon.fr

16 federica.calevro@insa-lyon.fr

17 patrick.callaerts@kuleuven.be

18

19

20

21

22

23 **Abstract**

24 The pea aphid, *Acyrtosiphon pisum*, is an emerging model system in functional and
25 comparative genomics, in part due to the availability of new genomic approaches and the
26 different sequencing and annotation efforts that the community has dedicated to this important
27 crop pest insect. The pea aphid is also used as a model to study fascinating biological traits of
28 aphids, such as their extensive polyphenisms, their bacteriocyte-confined nutritional symbiosis,
29 or their adaptation to the highly unbalanced diet represented by phloem sap. To get insights into
30 the molecular basis of all these processes, it is important to have an appropriate annotation of
31 transcription factors (TFs), which would enable the reconstruction/inference of gene regulatory
32 networks in aphids. All available annotations of pea aphid TFs are based on the first versions
33 of the genome assembly and annotation, that are now replaced by a new reference assembly.
34 We took advantage of this new resource, which represents the first chromosome-level *A. pisum*
35 genome, to annotate the complete repertoire of *A. pisum* TFs, and to complement this
36 information by annotating genes encoding chromatin-associated and basal transcription
37 machinery proteins. These annotations were done using information from the model *Drosophila*
38 *melanogaster*, for which we also provide a revisited list of these proteins. The comparison
39 between the two model systems allowed the identification of major losses or expansions in each
40 genome, while a deeper analysis was made of ZNF TFs (with certain families expanded in the
41 pea aphid), and the Hox gene cluster (showing reorganization in gene position in the pea aphid
42 compared to *D. melanogaster*). All annotations are made available to the community the
43 different annotations we produced storing all this information in the Aphid Transcription
44 Factors database (ATFdb), a resource for gene regulation studies in aphids.

45

46 **Keywords:** Hemiptera; pea aphid; transcription factors; chromatin; basal transcription
47 machinery; TF database

48

49 **1. Introduction**

50 Changes in gene expression arise in response to internal and external signals and are essential
51 for decoding genotypes into cellular, tissue and organismal phenotypes. Gene expression
52 alterations are observed, for example, during development (Martín et al., 2016; Spitz and
53 Furlong, 2012), upon cellular differentiation (Peñalosa-Ruiz et al., 2019), in metabolic
54 homeostasis and physiological regulations (Desvergne et al., 2006), and in response to
55 environmental stimuli (Han and Kaufman, 2017; Song et al., 2016). Gene expression is
56 controlled to an important extent by Transcription Factors (TFs), which bind accessible cis-
57 regulatory DNA elements to positively or negatively regulate the expression of target genes
58 (Hammonds et al., 2013; Hens et al., 2011; Kim and Wysocka, 2023; Lambert et al., 2018). The
59 accessibility of DNA depends on chromatin state and can be enhanced by TFs (Zaret, 2020).
60 Sets of TFs and their target genes are organized in so-called Gene Regulatory Networks (GRNs)
61 that govern cellular gene expression (Aerts, 2012; Davidson and Levine, 2008; Singh et al.,
62 2018). Altered specificity or activity of TFs or changes in cis-regulatory sequences impact the
63 structure or activity of GRNs and represent an important source of phenotypic diversity and
64 evolutionary adaptation (Carroll, 2008; Schember and Halfon, 2022).

65 TFs are a large group of evolutionary conserved regulatory proteins. Their modular
66 structure determines their ability to interact with DNA sequences through DNA binding
67 domains (DBDs), to exert their function as positive or negative regulators of target gene
68 expression through transcriptional regulatory domains, or even to associate with each other or
69 with co-activating/repressing proteins through oligomerization (Gonzalez, 2016; Lemon and

70 Tjian, 2000; Näär et al., 2001). TFs are classified into several superfamilies, families and sub-
71 families, based on structural similarities, primarily of their DBDs (Luscombe et al., 2000).
72 Three major superfamilies are the C₂H₂ zinc finger, homeodomain and helix-loop-helix, which
73 together account for more than 80% of all predicted human TFs (Vaquerizas et al., 2009) and
74 around 70% of TFs in the fruit fly *Drosophila melanogaster* (Gramates et al. (2022)).

75 For over a century, *D. melanogaster* has been used as a model system to dissect and
76 understand genetics, developmental biology, organ specification and function, physiology and
77 metabolism (Bellen and Yamamoto, 2015; Dow et al., 2022; Fischer et al., 2023; Helfand and
78 Rogina, 2003; Mohr and Perrimon, 2019). Many sophisticated biochemical, molecular, and
79 cellular approaches have been developed in *D. melanogaster* and were used to functionally
80 validate TFs (Shokri et al., 2019). Consequently, the *Drosophila* genome is one of the best
81 characterized metazoan genomes in terms of functionally annotated genes and regulatory
82 elements (Celniker and Rubin, 2003; Chen et al., 2014; Yang et al., 2018). In recent years, the
83 increasing number of sequenced genomes has accelerated TF predictions and annotations in
84 diverse insects beyond *Drosophila* (i5K Consortium, 2013; Sproul et al., 2023). The pea aphid
85 *Acyrtosiphon pisum* has emerged as an important model system in functional and comparative
86 genomics, in part due to the availability of new genomic approaches (Calevro et al., 2019; Le
87 Trionnaire et al., 2019; Sapountzis et al., 2014; Tagu et al., 2016) and the high annotation
88 quality of its genome (Fernández et al., 2020; Li et al., 2019; The International Aphid Genomics
89 Consortium, 2010). Aphids are a great model to study extensive polyphenisms, including
90 parthenogenetic and sexual reproduction and the presence of winged and non-winged morphs
91 (Davis et al., 2021; Ogawa and Miura, 2014). They are one of the best studied models of
92 nutritional symbioses and have bacteriocytes, a specialized cell type that houses the obligate
93 endosymbiont *Buchnera aphidicola* (Baumann, 2005; Calevro et al., 2023; Simonet et al.,
94 2018). Not only the developmental origin of this novel cell type remains unknown, but recent

95 work has shown extensive plasticity and transcriptional reprogramming of these cells in
96 response to environmental challenges (Colella et al., 2018; Ribeiro Lopes et al., 2022). Finally,
97 aphids are phloem sap-sucking insects that are highly adapted to this unbalanced diet (Douglas,
98 2015). We propose that all these processes are supported by the interaction of TFs with their
99 target genes and by GRNs that are either modified from preexisting GRNs or are novel and
100 possibly aphid specific.

101 To facilitate studies on gene regulation and GRN inference in aphids, we made a
102 comprehensive annotation of the TFs, chromatin-associated proteins, and basal transcription
103 machinery in the pea aphid, at both the family and gene levels. A first effort by Shigenobu and
104 colleagues (2010) made use of the first pea aphid genome assembly but only precisely annotated
105 developmentally important TFs, limiting the rest of the annotation to a DBD-based
106 classification. Other analyses, focusing on the basic helix-loop-helix TF family in *A. pisum*
107 (Dang et al., 2011) or on the annotation of TFs and chromatin-related proteins in different
108 hemipteran species including *A. pisum* (Vidal et al., 2016) have also been made on the first pea
109 aphid genome assembly. Since then, the availability of the first chromosome-scale assembly
110 for the pea aphid (Li et al., 2019) resulted in a significant adjustment of predicted protein-coding
111 gene numbers, making a new and comprehensive annotation a necessity. Taking advantage of
112 the knowledge accumulated in *Drosophila*, we predicted 854 TF-coding genes in *A. pisum*. In
113 addition, we annotated 230 chromatin-associated genes and 67 genes belonging to the basal
114 transcription machinery. All annotations are compiled in a dedicated database, the Aphid
115 Transcription Factors database (ATFdb) (<http://atf.cycadsys.org>). This complete annotation is
116 expected to be of use to the community studying aphids and other hemipterans, prominent crop
117 pest insects, as well as to researchers studying insect development and evolution.

118

119 **2. Materials and methods**

120 **2.1 Definition of a list of *bona fide* *D. melanogaster* transcription factors (TFs)**

121 *D. melanogaster* TF annotations were retrieved from Hens et al. (2011) and Hammonds et al.
122 (2013). These lists were merged and then filtered to remove splicing variants that had been
123 annotated as independent genes, or genes that have since been (re)annotated as not being TFs
124 or as being a component of the basal transcription machinery or associated with chromatin. The
125 remaining TFs were compared to the “Transcription factors” gene group (FBgg0000745)
126 available in FlyBase release 6.55 (Gramates et al., 2022), to the precompiled *Drosophila* TF
127 predictions from the DBD database (Kummerfeld and Teichmann, 2006), and to the TF lists
128 extracted from the relevant databases FlyMine (Lyne et al., 2007) and FlyNet (Tian et al., 2009).
129 *De novo* predictions were made based on the *Drosophila melanogaster* genome release 6.55
130 using the transcription factor prediction tools from PlantTFDB (Jin et al., 2017) and Pfam using
131 TF domains of the DBD database (Mistry et al., 2021; Wenger, 2018; Wilson et al., 2008). The
132 refinement strategy is summarized in the upper part of Figure 1.

133 **2.2 Prediction of *Acyrtosiphon pisum* transcription factors**

134 All predicted proteins of *A. pisum* were downloaded from the NCBI datasets using the RefSeq
135 Annotation release 103 (assembly ID: GCF_005508785.2), also published in Li et al. (2019).
136 A first BLASTP (Altschul et al., 1990) analysis was performed against the *A. pisum* proteins
137 using our new list of *bona fide* *D. melanogaster* TF protein sequences as a query with an e-
138 value cutoff of 1e-5. BLASTP results were manually curated. To confirm the homology
139 relationships and to remove proteins that are not homologs, reverse BLASTP analyses (*i.e.*
140 using the *A. pisum* candidate protein as query against *D. melanogaster*) were performed using
141 as query the *A. pisum* proteins corresponding to the top ten first BLASTP hits against all *D.*
142 *melanogaster* proteins. When both BLASTP analyses yielded the same *A. pisum/D.*
143 *melanogaster* homologous protein pairs (one to one relationships) the *A. pisum* orthologous

144 protein was annotated as such according to *D. melanogaster* nomenclature. When multiple *A.*
145 *pisum* TFs gave the same *Drosophila* TF as the top hit, then the *A. pisum* TF sequences were
146 aligned pairwise using BLAST Global Alignment to identify identical protein-coding
147 sequences, i.e., putative splicing variants in *A. pisum* or almost identical copies homologous to
148 one *Drosophila* TF. Nearly identical sequences in *A. pisum* were annotated based on *Drosophila*
149 nomenclature, with chromosomal localization used to discriminate between splicing variants
150 (one localization) or paralogs (several localizations). In cases where the divergence between *A.*
151 *pisum* and *D. melanogaster* hits did not allow an orthologous relationship to be defined, the
152 full-length amino acid sequences were aligned and a manual multi-criteria expert judgement
153 was used to distinguish the paralogs, which were then numbered and named according to the
154 *Drosophila* nomenclature. In most cases, sequence bordering the DNA-binding domain
155 contains unique residues that allow unambiguous identification. If the divergence from the *D.*
156 *melanogaster* sequence was too strong, but the DNA-binding domain identified it as belonging
157 to the same TF family, the suffix "-like" was added to the end of the gene name symbol. Finally,
158 any remaining sequences that were too short or too divergent and lacked an identifiable DNA-
159 binding domain were removed. Since the approach described above would not identify TFs that
160 are not homologous to *Drosophila*, a *de novo* prediction was done in parallel on all 17,681
161 *A. pisum* protein-coding genes using the same prediction tools as for *Drosophila* (i.e.,
162 PlantTFDB and Pfam).

163 The global strategy we used to annotate the *A. pisum* TFs is summarized in the lower part of
164 Figure 1.

165 **2.3 Prediction of chromatin associated proteins and basal transcription machinery**

166 Analysis of the gene lists provided by Hens et al. (2011) and Hammonds et al. (2013) revealed
167 some overlap between TFs and chromatin-associated proteins, or proteins belonging to the basal

168 transcription machinery. This stimulated us to produce a complete list of the latter, for which
169 an exhaustive single repository was not available, using information derived from FlyBase, and
170 expert literature. A master list of *Drosophila* chromatin-associated proteins was thus compiled
171 and used to search for homologs in the *A. pisum* genome, as detailed above (see § 2.2 and Figure
172 1). For chromatin-associated proteins also functioning as TFs, they were double-listed (as TFs
173 and as chromatin-associated proteins). Lastly, the basal transcription machinery of *A. pisum*
174 was annotated following the same procedure. All identified TFs, chromatin-associated genes
175 and genes involved in the transcription machinery were classified into families and subfamilies
176 based on the available information listed in § 2.1.

177 **2.4 Annotation of HNF4 homologs in Hemiptera**

178 Following the identification of a HNF4 homolog in *A. pisum*, the search was extended to 14
179 additional species: six aphid species (*Aphis gossypii*, *Diuraphis noxia*, *Melanaphis sacchari*,
180 *Myzus persicae*, *Rhopalosiphum maidis* and *Sipha flava*, all members of the Aphidoidea
181 family), two aphid-related species (*Adelges cooleyi*, member of the Adelgidae family and
182 *Daktulosphaira vitifoliae*, member of the Phylloxeridae family) and six other hemipterans
183 (*Bemisia tabaci*, *Cimex lectularius*, *Diaphorina citri*, *Halyomorpha halys*, *Homalodisca*
184 *vitripennis* and *Nilavarpata lugens*) (Table 1). Among all available hemipteran genomes, we
185 specifically selected those with a NCBI RefSeq annotation, which guarantees a quality
186 reference for genome annotation and gene identification. The protein sequences of *A. pisum*
187 HNF4 (identified in this study) and *D. melanogaster* HNF4 were used as query to perform
188 BLASTP searches against the RefSeq proteome of the 14 insect species listed above. For each
189 species, reverse BLASTP searches (*i.e.* using the candidate protein as query against the *D.*
190 *melanogaster* and *A. pisum* protein sets) were performed on the top ten BLASTP hits.
191 Additionally, candidate sequences were analyzed with InterProScan (v5.59-91.0) to identify

192 functional domains (Paysan-Lafosse et al., 2023). Candidate proteins were considered *bona fide*
193 HNF4 homologs only if the reverse BLASTP analyses yielded HNF4 in the top ten hits and if
194 the InterProScan analyses revealed the presence of a DBD (InterPro signature IPR049636)
195 and/or ligand binding domain (InterPro signature IPR049635) that are typical of HNF4 TFs.
196 Alignments were performed using the Kalign multiple sequence alignment web service from
197 EMBL-EBI (Madeira et al., 2022) with default parameters. Graphical representations of the
198 alignments were performed using the ESPript v3.0 web service (Robert and Gouet, 2014).

199 **2.5 Hox cluster gene annotation in Hemiptera**

200 To better understand the evolution of the Hox gene cluster in hemipterans, genes belonging to
201 this cluster were annotated in the 14 additional species listed in Table 1 (see § 2.4). For four of
202 them (*A. gossypii*, *R. maidis*, *H. vitripennis* and *N. lugens*) chromosome-level assemblies were
203 available, and for one (*D. vitifoliae*) genes belonging to the Hox gene cluster were located on
204 the same genomic scaffold, which permitted cross-species comparison of the chromosomal
205 organization of the Hox gene cluster. Protein sequences encoded by the *A. pisum* Hox cluster
206 identified in this study were used as query to perform BLASTP searches against the NCBI
207 RefSeq proteomes of the 14 selected insect species. For each species, the BLASTP hits with
208 the 10% highest identity scores were subsequently blasted back against the *A. pisum* reference
209 proteins. When both BLASTP analyses yielded the same homologous protein pairs as top hits,
210 these were annotated as such. When there were discrepancies between the two BLASTP results,
211 semi-manual curation was performed by (i) repeating the analysis with the *D. melanogaster*
212 Hox cluster protein sequences as query and (ii) looking at the relative position of each Hox
213 cluster gene relative to the others on the genome. Systematic BLASTP searches were performed
214 to compare the protein sequences of each predicted Hox gene with the complete set of homologs

215 identified in hemipterans or model species, enabling the level of inter- and intra-specific
216 divergence within the Hox cluster to be assessed (see § 3, Results and Discussion).
217 This annotation strategy was validated by using it to reannotate the canonical and non-canonical
218 Hox genes of the red flour beetle *Tribolium castaneum* and the domestic silk moth *Bombyx*
219 *mori*, insect models representative of the orders Coleoptera and Lepidoptera, respectively, using
220 the latest NCBI RefSeq genome sequences available (GCF_000002335.3 and
221 GCF_014905235.1). Our results were consistent with the annotation of the Hox cluster genes
222 available in previous studies: we identified the same genes and confirmed their relative position
223 to each other in the Hox cluster (Chai et al., 2008; Mulhair and Holland, 2022; Pace et al.,
224 2016).

225

226 **2.6 Database construction and web interface**

227 All predicted TF sequences and annotations (*i.e.* *A. pisum* TFs, transcription basal machinery
228 and chromatin-related genes, and HNF4 and the Hox gene cluster in 14 additional hemipteran
229 genomes) were stored in a MySQL relational database on a Linux server. Queries to the
230 database were implemented in PHP scripts running in an Apache/PHP environment. ATFdb is
231 accessible online (<https://atf.cycadsys.org/>) and allows users to browse by TF family. Users can
232 search ATFdb by gene ID, gene name, TF family or by keywords. Users can also run BLAST
233 searches against all the sequences available in the database. The lists of TFs as well as their
234 nucleotide and protein sequences are available for download.

235 **3. Results and Discussion**

236 **3.1 A new annotation of *D. melanogaster* TFs, chromatin-associated and basal** 237 **transcription machinery-associated genes**

238 The analysis of the 755 putative TFs of Hens et al. (2011) revealed that they correspond to 747
239 independent genes. The difference is due to the independent listing of splicing variants and the
240 duplicate listing of a few TFs. The intersection of these 747 TFs with the 707 from Hammonds
241 et al. (2013) is 643 TFs, with 104 putative TFs that were listed only in Hens et al. (2011) and
242 64 only in Hammonds et al. (2013). By removing genes that had since been (re)annotated as
243 not being TFs or as being a component of the basal transcription machinery or associated with
244 chromatin, we obtained a final list of 703 putative *D. melanogaster* TFs. This list was then
245 complemented with annotation datasets from FlyBase (FBgg0000745, 628 candidates),
246 FlyMine (142), FlyNet (633) and TranscriptionFactor.org (529). *De novo* predictions were also
247 performed on the 13,962 protein-coding genes of the *D. melanogaster* genome using the Pfam
248 domain (536) and the PlantTFDB (483) prediction tools. These analyses yielded an additional
249 32 candidate TFs (upper part of Figure 1). Thus, the complete list of *bona fide D. melanogaster*
250 TFs that was used to annotate the *A. pisum* genome comprised 735 unique TF-coding genes
251 (Suppl. File 1).

252 We also annotated 195 unique *D. melanogaster* chromatin-associated genes (Suppl. File 2).
253 Thirty-two of those genes function as TFs and are also listed in our *D. melanogaster* TF list
254 (Suppl. File 1). Lastly, we report in Suppl. File 3 our annotation of the 51 *D. melanogaster*
255 genes involved in the basal transcription machinery. Two of them are double-listed in Suppl.
256 File 2, as they also belong to the chromatin-associated machinery.

257 **3.2 The *A. pisum* repertoire of TFs, chromatin-associated and basal transcription** 258 **machinery-associated genes**

259 Using the curated TF list from *D. melanogaster* as query in combination with our *de novo*
260 prediction approach, we identified 854 putative TFs in the pea aphid genome (Suppl. File 1),
261 including 16 chromatin-associated genes. Those were double-listed in Suppl. File 2, which

262 contains the 230 unique *A. pisum* chromatin-associated genes. We also annotated 67 basal
263 transcription machinery associated genes (Suppl. File 3), two of which also belong to the class
264 of chromatin associated genes. All this information is available in ATFdb. The basal
265 transcription machinery did not display important differences between *D. melanogaster* and *A.*
266 *pisum*. The TFs and the chromatin-associated genes did, however, show remarkable differences.
267 While some families are well conserved relative to the ones present in the *D. melanogaster*
268 genome, others show independent expansions and losses (Table 2) and are discussed below.

269 Among the 854 unique *A. pisum* TF genes, 215 were exclusively found in *A. pisum* and did not
270 have homologs in *D. melanogaster*.

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

281 **3.3 The expansion of distinct groups of chromatin-associated genes suggests an important** 282 **role for epigenetic regulation in the pea aphid**

283 The annotation of the chromatin-associated genes yielded a total of 230 genes in the *A. pisum*
284 genome compared to 195 in the *D. melanogaster* genome. The expansions in *A. pisum* include

285 multiple homologs of the Polycomb-repressive Complex 1 (PRC1) associated genes *Kdm2*
286 (histone demethylase specific for H3 lysine 36; Lagarou et al., 2008), *Sce* (*Sex combs extra*, E3
287 ubiquitin ligase that monoubiquitinates H2A; Gutiérrez et al., 2012) and *Scm* (*Sex combs on*
288 *midleg*, enables PRC1 binding; Bornemann et al., 1996), as well as the SET-domain lysine
289 methyltransferase *egg* (*eggless*, histone H3-lysine(9) N-trimethyltransferase; Wang et al.,
290 2011). In addition, we found expansions of histone lysine acetyltransferase 6 complex genes
291 (*enok - enoki mushroom* (Scott et al., 2001), *Ing5 - Inhibitor of growth family member 5* (Huang
292 et al., 2016)) and histone deacetylase genes (*HDAC1*; Mannervik and Levine, 1999). Lastly,
293 the heterochromatin protein 1 (HP1) family, which binds histone H3 tails, is also larger in *A.*
294 *pisum* (Mendez et al., 2011). Combined, these results suggest that epigenetic mechanisms may
295 play a prominent role in gene and transcriptional regulation in the pea aphid, a hypothesis also
296 supported by studies suggesting that gene expression in *A. pisum* is regulated by multiple
297 histone acetyltransferase and histone deacetylase genes (Kirfel et al., 2020).

298 **3.4 TF family-specific differences in *A. pisum* and *D. melanogaster***

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

302 The most prominent differences were found in the zinc finger (ZNF) TFs. Using a previous
303 version of the *A. pisum* genome assembly, Shigenobu et al., (2010) described significantly more
304 zinc finger containing TFs of the C₂H₂ (also called classical ZNF) GATA, BED, and MIZ
305 families. We have now systematically reannotated all these families and assigned each ZNF TF
306 in *A. pisum* to its precise family and observed species-specific expansions of ZNF TFs. The
307 most striking expansions concern the following families : “ZNF-classical transcription factors”
308 (with 269 genes in *A. pisum* and 149 genes in *Drosophila*), ZNF-BED (29 genes in *A. pisum*
309 vs. six in *D. melanogaster*), ZNF-THAP (17 genes in *A. pisum* vs. seven in *D. melanogaster*),

310 ZNF-SP1/KLF (19 genes in *A. pisum* vs. 13 in *D. melanogaster*) and ZNF-RING/FYBVE/PHD
311 (also known as ZNF-MIZ, 38 genes in *A. pisum* vs. 14 in *D. melanogaster*). By contrast, the
312 family ZNF-AD displays an expansion in *D. melanogaster* (83 TFs). Several of the ZNF-AD
313 TFs are lacking in the *A. pisum* genome (29 TFs), yet we noted seven *A. pisum* homologs for
314 the *D. melanogaster Meics* gene. There is no notable difference in the ZNF-GATA family.
315 Concerning the ZNF-classical family, we were able to subdivide this expansion in *A. pisum*
316 based on homology with known *D. melanogaster* genes: 18 homologs for CG12299, 12 for
317 *clamp (clp)*, 135 for *crooked legs (crol)*, and seven for *Phaser*. Overall, these results suggest
318 that an independent evolution of ZNF TFs occurred in the two species. Panfilio *et al.* (2019)
319 described a different expansion of the C₂H₂ ZNF TFs in the milkweed bug *Oncopeltus fasciatus*
320 therefore suggesting that this family of TFs is particularly prone to expansions. Whether and
321 how this is reflected functionally remains to be discovered. However, similar expansions of
322 ZNF TFs have been observed in vertebrates suggesting that the modularity of the zinc fingers
323 and the possibility of changing amino acids that interact with DNA are two factors that
324 contribute to the fact that they are the largest group of TFs (Nardelli *et al.*, 1991; Vaquerizas *et*
325 *al.*, 2009). The different classes of ZNF TFs in the *A. pisum* genome are listed in Table 2.
326 Other less prominent differences were found. In *A. pisum* we annotated larger numbers of TFs
327 of the following families: ARID, MAD homology domain, and MADS. On the other hand, in
328 the *D. melanogaster* genome, we found an overall higher number of TFs belonging to the
329 following TF families: bHLH, Homeobox-TALE, HTH-Psq and T-box.

330 The families that are not explicitly listed here comprised the same or nearly the same numbers
331 of TFs in *A. pisum* and *D. melanogaster*. However, in some instances, and even though the
332 overall numbers were the same in a TF family, expansions of some subfamilies were seen in
333 either species, e.g. for MADF-BESS. In addition, there were numerous cases where a single
334 homolog was present in *A. pisum* and two homologs in *D. melanogaster* (Suppl. File 1). This

335 was the case for *Met/gce* (*Methoprene tolerant/germ cell-expressed bHLH-PAS*; Juvenile
336 Hormone signaling; Baumann et al., 2010; Miura et al., 2005), *Rbf/Rbf2* (retinoblastoma family
337 protein; cell cycle regulation; Stevaux et al., 2002), *slp1/slp2* (*sloppy paired 1* and 2;
338 segmentation; Grossniklaus et al., 1992), *gcm/gcm2* (*glial cell missing 1* and 2; glia and
339 plasmacyte development; Chotard et al., 2005), B-H1/B-H2 (Bar; eye and leg development;
340 Higashijima et al., 1992), *en/inv* (*engrailed/invented*; segmentation; Cheng et al., 2014), *lbe/lbl*
341 (*ladybird early/late*; heart development; Jagla et al., 1997), *Vsx1/Vsx2* (*visual system homeobox*
342 *1* and 2; Erelik et al., 2008), *bab1/bab2* (*bric-a-brac*; Lours et al., 2003), *tsh/tio* (*teashirt/tiptop*;
343 head and trunk development; Datta et al., 2011), and *eyg/toe* (*eyegone/twin of eyegone*; eye
344 development; Yao et al., 2008). The presence of a single *engrailed/invented* homolog is likely
345 due to secondary loss since it was previously shown that they arose through tandem duplication
346 prior to the radiation of hexapods (Peel et al., 2006).

347 **3.5 Differences and similarities in TFs regulating early development and organ** 348 **development of *D. melanogaster* and *A. pisum***

349 We confirm the absence of *A. pisum* homologs for the early developmental genes *bicoid* (*bcd*),
350 *huckebein* (*hkb*), *buttonhead* (*btd*) and *giant* (*gt*), as originally reported by Shigenobu et al.
351 (2010). *Bicoid* is only found in higher Diptera (Gregor et al., 2008). Its absence in the *A. pisum*
352 genome was therefore expected. By contrast, the absence of homologs for *huckebein*,
353 *buttonhead* and *giant* was less expected. *Giant* is conserved in many insects and has been shown
354 to act as a gap gene in the hemipteran *O. fasciatus* (Ben-David and Chipman, 2010). While we
355 did not find a *bona fide* homolog for *buttonhead* in the *A. pisum* genome, we did identify an
356 Sp1 homolog. Given the close relatedness of *buttonhead* and Sp1 and similar functions in
357 appendage development in *Drosophila*, *T. castaneum* and *O. fasciatus* (Schaeper et al., 2009),
358 we propose that the presence of Sp1 in *A. pisum* likely reflects the fact that it retained this role

359 as a regulator of appendage development. *Huckebein* is present in *O. fasciatus* where its
360 expression pattern suggests a role different from what was described in *Drosophila* where it is
361 a terminal gap gene (Weisbrod et al., 2013).

362 Key TFs involved in the development of various organs are conserved in *A. pisum*. These
363 include development of (i) wings: *apterous* (*ap*; Cohen et al., 1992) and *vestigial* (*vg*; Williams
364 et al., 1991); (ii) salivary gland: *forkhead* (*fkh*; Mach et al., 1996); (iii) muscle: *Mef2* (Kelly et
365 al., 2002); (iv) heart: *tinman* (*tin*; Bodmer, 1993). For eye (and head) development we found
366 that many of the genes that constitute the retinal determination pathway in *Drosophila*
367 (Domínguez and Casares, 2005; Kumar, 2010), are conserved in *A. pisum*, namely *so*, *lz*, *ey*,
368 *toy*, *eyg*, *eya*, *hth*, *exd*, *oc*. At the same time, the absence of *dan*, *danr*, *toe* and the presence of
369 a single *ara/caup/mirror* homolog indicates that the pathway has likely undergone
370 modifications. A final interesting observation was the complete absence of homologs for *Doc1*,
371 *Doc2* and *Doc3*, which are involved in the amnioserosa development in *Drosophila*
372 (Hamaguchi et al., 2004). Hemimetabolous insects, including the pea aphid, have two
373 extraembryonic membranes (amnion and serosa) that are involved in blastokinesis (comprising
374 anatrepsis and katatrepsis) (Panfilio, 2008; Schmidt-Ott, 2000). It contrasts with *Drosophila*
375 where extraembryonic development is extremely reduced, thus making it likely that the
376 underlying genetic mechanisms will also show considerable differences.

377 **3.6 TFs regulating developmental transitions are conserved**

378 Ecdysone and juvenile hormone (JH) are essential hormones regulating developmental
379 transitions. All TFs implicated therein in *Drosophila* are conserved in *A. pisum*. *Ecdysone*
380 *Receptor* (*EcR*) and *Ultraspiracle* (*Usp*) for ecdysone signaling (Yamanaka et al., 2013), and
381 *Met/Gce*, *Taiman*, *Kr-h1* and *Eip93F* for JH signaling (Jindra et al., 2015).

382 **3.7 Key TFs of signaling pathways are conserved, but TCF and MAD TFs show expansion**
383 **in *A. pisum***

384 The activity of signaling pathways ultimately leads to transcriptional changes mediated by
385 specific TFs. Not unexpectedly, we identified homologs of all these TFs in the *A. pisum*
386 genome. *Yan/Aop* (*anterior open*; Rebay and Rubin, 1995) and *Pointed* (O'Neill et al., 1994)
387 are central negative and positive transcriptional regulators downstream of receptor tyrosine
388 kinases (EGFR, FGFR, PVR). Others include *Foxo* (insulin signaling; Puig et al., 2003), *cubitus*
389 *interruptus* (*ci*) (hedgehog signaling; Alexandre et al., 1996), *scalloped* (Wu et al., 2008) and
390 *yorkie* (Hippo signaling; Huang et al., 2005), *Su(H)* (Notch signaling; Schweisguth and
391 Posakony, 1992), and *Jra/kay = JUN/cFos* (JNK signaling; Perkins et al., 1990). Interestingly,
392 we observed expansions of TFs associated with Wnt signaling (pangolin and TCF TFs; Brunner
393 et al., 1997), JAK-STAT signaling (three homologs of *Stat92E*; Hou et al., 1996) and TGF-beta
394 signaling (SMAD TFs; Sekelsky et al., 1995). The expansion of TCF TFs is intriguing in light
395 of previous observations of duplications of the Wnt pathway component armadillo/b-catenin in
396 *O. fasciatus* and other hemimetabolous insects (Bao et al., 2012; Panfilio et al., 2019). An
397 expansion of SMAD TFs has been previously reported in *O. fasciatus* where it was suggested
398 to be *Oncopeltus*-specific (Panfilio et al., 2019). The fact that we found an expansion in the pea
399 aphid seems to suggest that the amplification may be more ancient in insect evolution. bHLH
400 TFs of the E(Spl) complex acting in Notch signaling show a strongly reduced number as well
401 as an altered organization of the complex. Dearden (2015) described and discussed these
402 differences for the E(Spl) complex in arthropods.

403 **3.8 Identification of an *A. pisum* homolog of HNF4, a key TF regulating metabolic**
404 **homeostasis**

405 Previously it was reported that there was no HNF4 homolog in *A. pisum* (Shigenobu et al.,
406 2010). Given its central importance in regulating metabolic homeostasis (Palanker et al., 2009),
407 this was unexpected. We here report the identification of a *bona fide* HNF4 homolog in *A.*
408 *pisum* (corresponding to different isoforms, listed in Table 3). This annotation was supported
409 by the presence of two domains characteristic of HNF4 TFs, namely an N-terminal ZNF-
410 GATA-type DNA binding domain (DBD) (InterPro signature IPR049636; Hepatocyte Nuclear
411 Factor 4-like, DNA binding domain) and a C-terminal nuclear receptor-type ligand binding
412 domain (LBD) (InterPro signature IPR049635; Hepatocyte nuclear factor 4, ligand-binding
413 domain). The isoforms encoded by the *D. melanogaster* HNF4 gene are longer than those of
414 the pea aphid (between 666 and 732 amino acids vs. between 375 and 420 amino acids,
415 respectively). Nonetheless, *D. melanogaster* and *A. pisum* HNF4 are highly similar, with 65%
416 identity over the whole sequence and 72% when only functional domains are considered. We
417 next searched the sequence of HNF4 in 14 additional hemipteran species (see Table 1 and §
418 2.4). At least one copy of HNF4 was found in each of these species (Table 3), with very high
419 conservation of functional domains, especially the DNA binding domain (Figure 2). Three
420 species (*B. tabaci*, *D. citri* and *D. noxia*) appeared to have a duplication of HNF4. In *B. tabaci*,
421 we identified three genes that have a size consistent with HNF4 and have DBD and LBD that
422 were successfully identified by InterProScan. Importantly, the sequence of the *B. tabaci*
423 proteins encoded by the LOC109043749 and LOC109038961 genes differ significantly from
424 those of the other HNF4 homologs. On the one hand, the position of LOC109043749 in the *B.*
425 *tabaci* genome of *B. tabaci*, in tandem with LOC109043751, suggests that this gene could have
426 appeared through tandem duplication, and diverged afterwards. On the other hand,
427 LOC109038961 probably appeared through duplication of LOC109043749. Two HNF4
428 homologs were also identified for *D. citri* and *D. noxia*, but the presence of two genes seems to
429 be the result of assembly errors. In the case of *D. citri*, one of the genes codes for a very short

430 protein of 90 amino acids (LOC103508991), not compatible with a functional HNF4. In the
431 case of *D. noxia*, one sequence encodes a protein with only a LBD (LOC107163700), and the
432 other sequence corresponds to a single gene on a very short contig, encoding a protein with
433 only a DBD (LOC107172291). In the latter case, it is possible that the two genes identified as
434 homologs actually correspond to a single gene that was not reconstructed during assembly.

435 The conservation of HNF4 in all hemipteran genomes we analyzed here supports the
436 hypothesis, emerging from recent studies (Cheng et al., 2020; Wang et al., 2023) that HNF4
437 has important roles in hemipteran physiology as well. Moreover, the observation that the length
438 of HNF4 proteins encoded in all these hemipteran genomes is closer to the size of isoforms we
439 found in *A. pisum* and shorter than the isoforms we found in *D. melanogaster*, suggests the
440 presence of potential lineage-specific features.

441 **3.9 TFs regulating essential cellular and organismal processes**

442 We also analyzed TFs with central roles in important cellular or organismal processes in *D.*
443 *melanogaster*. For the majority of these we identified homologs in *A. pisum*. Specifically, we
444 found homologs for *SREBP* (HLH106 - lipid homeostasis; Theopold et al., 1996), *Atf6*
445 (endoplasmic reticulum stress response; Allen and Seo, 2018), *cnc* (oxidative stress response;
446 Sykiotis and Bohmann, 2008), *Xbp1* (UPR pathway; Huang et al., 2017), *REPTOR/REPTOR-*
447 *bp* (TOR mediator complex; Tiebe et al., 2015). Intriguingly, we were not able to identify
448 homologs for *Myc* (cell growth, cell competition, proliferation; Gallant et al., 1996) and *Mitf*
449 (regulator of V-ATPase and lysosomal – autophagic pathway; Zhang et al., 2015). We did find
450 a distant relative of *Myc*, namely *Mnt* (Loo et al., 2005) to be conserved. Since both are
451 dimerization partners for *Max*, it is possible that one or both functionally replace *Myc*.

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

455 **3.10 *A. pisum* has no homolog for *doublesex*, a key TF in the *Drosophila* sex determination** 456 **pathway**

457 The genome of *D. melanogaster* contains four related TFs (*dmrt11E*, *dmrt93B*, *dmrt99B* and
458 *dsx*) of the DM domain family. *doublesex* (*dsx*) encodes a key TF in the *Drosophila* sex
459 determination pathway (Coschigano and Wensink, 1993), while *dmrt11E* is required in testis
460 somatic cells for male fertility (Yu et al., 2015). The two other genes have not been functionally
461 characterized. We did not find direct homologs for *dsx* and *dmrt11E*, but homologs for
462 *Drosophila* *dmrt93b* and *dmrt99b* are present in *A. pisum*. The absence of a *dsx* homolog may
463 reflect the differences in sex determination between *A. pisum* and *D. melanogaster*. Whether
464 the *A. pisum* homologs *dmrt93b* and *dmrt99b* have a role in spermatogenesis remains to be
465 investigated. Consistent with a different organization of sex determination is the observation
466 that the *A. pisum* genome does not encode a *fruitless* homolog, a BTB transcription factor that
467 regulates male sexual behavior in *Drosophila* (Just et al., 2023; Laslo et al., 2023).

468 **3.11 The *A. pisum* Hox cluster is reorganized and lacks a *bona fide* *fushi tarazu* homolog**

469 The insect Hox cluster is thought to have been composed, in a bilaterian ancestor, of 10 genes:
470 *labial* (*lab*), *proboscipedia* (*pb*), *zerknüllt* (*zen*), *Deformed* (*Dfd*), *Sex combs reduced* (*Scr*),
471 *fushi tarazu* (*ftz*), *Antennapedia* (*Antp*), *Ultrabithorax* (*Ubx*), *abdominal-A* (*abdA*), and
472 *Abdominal-B* (*AbdB*), (listed here in the order they occupy in the cluster in a metazoan
473 hypothetical ancestor). Unlike other arthropods (*e.g.* Crustacea), that can lack some of them,
474 insects appear to have conserved all eight canonical Hox genes (*lab*, *pb*, *Dfd*, *Scr*, *Antp*, *Ubx*,

475 *abdA*, *AbdB*), essential in determining positional identity along the body axis (Mulhair and
476 Holland, 2022). The other two genes of the cluster, *zen* and *ftz*, which had ancestral homeotic
477 functions, evolved novel roles in insects (extraembryonic membrane patterning for *zen* and
478 segmentation for *ftz*), and are often referred to as the non-canonical Hox genes. We focused on
479 the Hox cluster since it is one of only few gene clusters in Arthropod genomes and, given the
480 availability of the chromosomal assignment of the *A. pisum* genome, we analyzed the structure
481 of the cluster and determined whether the linear arrangement is conserved in *A. pisum* or
482 whether it is rearranged.

483 The *A. pisum* Hox gene cluster is localized on chromosome A1 and contains the eight canonical
484 Hox genes. The fact that all insects have retained these eight could have functional implications:
485 segment number and tagmatization are constant in insects, despite their high species radiation,
486 leaving no place for redundancy or loss for these genes essential for the establishment of the
487 insect body plan (Mulhair and Holland, 2022). In Figure 3 we compare the *A. pisum* Hox cluster
488 organization to the ones present in three model species, *D. melanogaster*, *T. castaneum* and *B.*
489 *mori*. This shows several differences in the *A. pisum* Hox cluster when compared to other
490 insects: (i) the genomic size of the cluster is four times bigger than in *D. melanogaster* or *T.*
491 *castaneum*, with longer average gene length and intergenic distances, (ii) the split between the
492 anterior and the posterior parts of the cluster observed in *D. melanogaster* is not present, and a
493 shorter split separates the first three genes from the others, (iii) while the respective positions
494 of *Dfd*, *Scr*, *Antp*, *Ubx*, *abd-A* and *Abd-B* are the same as in the three other species, the first two
495 genes of the cluster are reorganized, with inversion of the respective positions of the *lab* and *pb*
496 genes. These observations are in agreement with the work of Mulhair and Holland (2022), who
497 analyzed the Hox cluster arrangements in 243 insect genomes and showed that the most anterior
498 genes of the cluster have undergone several rearrangements and that intergenic distances can
499 vary greatly.

500 Concerning the non-canonical Hox, we have identified a unique *HoxB4-like* gene in *A. pisum*
501 instead of the two *zen* genes (*zen* and *zen-2*) present in Diptera and Coleoptera (Figure 3). The
502 *A. pisum* genome also lacks another Hox-related gene, *bicoid* (see also § 3.5), which is
503 considered as a Diptera-specific duplication of *zen* (Stauber et al., 1999). Moreover, it was
504 difficult to identify a *ftz* homolog. A homeobox-containing gene is present at the usual position
505 for this gene, located between *Scr* and *Antp*, but it encodes a shorter protein (237 predicted
506 amino acids vs. 410 in *D. melanogaster*) (Figure 3 and Table 4B). Moreover, blasting it back
507 to the *Drosophila* genome gave *Scr* as best hit, and not *ftz*. This difficulty in identifying a clear
508 homolog of *ftz* was already reported in the first annotation of TFs in the pea aphid genome
509 (Shigenobu et al., 2010), and it was attributed to the lack of a chromosome-level assembly for
510 the *A. pisum* genome. Our analysis has been made on the first chromosome-level genome
511 assembly for this species (Li et al., 2019), which excludes that this problematic annotation of
512 *ftz* could be due to the quality of the genome assembly. Given the divergence of the *A. pisum*
513 *ftz* gene versus its homolog in *D. melanogaster*, we propose to refer to it as *ftz-like* instead.

514 **3.12 The Hox cluster in Hemiptera: organization and sequence homology/divergence**

515 In their recent analysis of the insect Hox cluster, Mulhair and Holland (2022) took into
516 consideration 243 insect genomes. As this analysis included only two hemipteran genomes, we
517 extended our analysis of the Hox cluster genes to all the Hemiptera for which a NCBI RefSeq
518 annotation was available (Table 1). The lengths of the predicted Hox genes and proteins we
519 have annotated are listed in Tables 4A and 4B, respectively. This analysis shows a great
520 diversity in Hox gene lengths, that do not correlate with the genome sizes (Tables 1 and 4). We
521 were not able to identify a *zen* homolog in *H. halis* and *H. vitripennis*, but there was one in all
522 aphid species (Table 4). The identification of a *ftz* homolog was as difficult as in *A. pisum* and
523 the putative homolog we identified in the different hemipteran genomes always corresponded

524 to proteins shorter than the *ftz* homologs present in the genomes of *D. melanogaster*, *T.*
525 *castaneum* or *B. mori*.

526 To gain a better understanding of the levels of intra- and inter-specific divergence within the
527 Hox cluster in Hemiptera, we compared the protein sequences of the genes that make up this
528 cluster in the species listed in Table 1 through systematic BLASTP. Results are presented in
529 Suppl. File 4 and include the percentages of coverage and identity for each pairwise
530 comparison. This latter analysis shows that some genes are more conserved than others. For
531 example, when we compare the Hox genes of *D. melanogaster* with those of the pea aphid, we
532 find coverage and identity percentages of 56% and 61%, respectively, for *abd-A* versus 30%
533 and 56% for *ftz*. When considering the average values of coverage and identity for each group
534 of homologs, we find low coverage rates, averaging 27% (Suppl. File 5). Significant identity is
535 primarily observed in the homeodomain sequences. Identity percentages vary between 56%
536 (using *pb* proteins used as queries) and 66% (using *Antp* as queries), on average. These values
537 reflect the level of divergence between proteins encoded by Hox cluster genes, despite the
538 presence of the conserved homeodomain. When only orthologs are compared, coverage
539 percentages are higher, with a minimum of 48% (57% when only canonical Hox proteins are
540 considered) and a maximum of 88%. The *pb* and *lab* proteins are the most divergent, with
541 coverage and identity percentages averaging no more than 60% and 63% respectively.
542 Conversely, *Abd-B*, *abd-A*, *Ubx*, *Antp* and *Scr* proteins have percentages of coverage and
543 identity that go up to 88% and 81% on average. These results are intriguing in light of the fact
544 that most hemipteran Hox clusters present a split between the posterior region of the cluster
545 (from *Abd-B* to *Scr*) and the anterior region of the cluster (comprising *lab* and *pb*), suggesting
546 that the evolutionary constraints are higher on the former while *lab* and *pb* diverge more rapidly.
547 *Dfd*, which can be associated to either the posterior or anterior region depending on the species,

548 has an intermediate status, with coverage and identity averaging 65% and 66% respectively. *ftz*
549 and *zen* homologs appear to be highly divergent compared to canonical Hox gene products.

550 A comparison of genomic organization and genomic distances between genes of the Hox
551 clusters was possible for a small group of Hemiptera, for which all the Hox genes were on the
552 same chromosome/scaffold (Figure 4). This analysis shows how, in the Hemiptera group, the
553 general organization of the Hox cluster, in terms of relative position of the genes and the splits,
554 can vary greatly from each other. General observations made by Mulhair and Holland (2022)
555 in other insects, for instance that intergenic distances can be large in the anterior part of the
556 cluster, are reduced in the middle and become even larger in the posterior part of the Hox cluster
557 do not apply to the hemipteran genomes we have analyzed here (Figure 4). Moreover, contrary
558 to what has been found in other insect genomes, *Ubx*, *abd-A* and *Abd-B* are not always present
559 in the cluster in this specific order in hemipteran genomes.

560 Concerning the function of TFs in *A. pisum*, very little information is available in the literature.
561 A few studies have focused on the Hox genes and have shown that they evolved new roles in
562 aphids, related to their symbiotic status or certain polyphenisms. Immunohistochemical studies
563 have demonstrated that *Ubx* and *AbdA/B* are localized in the bacteriocytes of aphid embryos
564 and nymphs, which supports the hypothesis that they are associated with the early development
565 and the differentiation of these symbiont-containing cells (Braendle et al., 2003). The important
566 role of *Ubx* in the development of insect bacteriocytes has subsequently been confirmed in the
567 hemipteran *Nysius plebeius* (Matsuura et al., 2015), but no study has since been performed in
568 *A. pisum*. Other roles for the Hox TFs seem to be related to the wing polyphenisms, as *Scr*,
569 *Antp* and *Ubx* are up-regulated in apterae vs. alate aphids and are differentially expressed
570 between the two morphs during development (Zhang et al., 2019). Future availability of high-
571 quality insect genomes with chromosomal annotation will permit further comparative studies

572 of the Hox cluster, not only to infer its evolutionary origin and dynamics, but also to determine
573 what - if any - the developmental and functional consequences are of the rearrangements.

574 **Conclusion**

575 In the present work, we have (re)annotated all TFs, chromatin-associated genes and genes
576 associated with the basal transcriptional machinery in the model aphid *A. pisum* using the latest
577 genomic data. The procedure we used for the annotation, combining homology-based and
578 manually curation of *D. melanogaster* TFs and *de novo* predictions, is described in the
579 manuscript. The results of these annotations are made available to the scientific community
580 through the ATF database. This approach may be of interest to researchers who want to annotate
581 other aphids to expand the ATFdb database, and to all researchers with an interest in
582 transcriptional regulatory networks in aphids and other insects.

583 Despite the evolutionary divergence between the two species, homologs for most *D.*
584 *melanogaster* TF families have been found in *A. pisum*. This includes the identification of an
585 *A. pisum* homolog of HNF4, a key TF regulating metabolic homeostasis and that had not been
586 previously described in aphids. We propose that the major differences in development,
587 physiology and reproduction between the two species can be explained by the numerous
588 differences they display in their TF repertoires. In particular, the expansion of distinct sets of
589 chromatin-associated genes suggests an important role for epigenetic regulation in the pea
590 aphid. Finally, we performed and discussed in-depth analyses of the ZNF TFs with certain
591 families expanded in the pea aphid, and of the Hox gene cluster, which shows a reorganization
592 of gene position in the pea aphid compared to *Drosophila* and other insect models.

593

594 **Figures and Tables**

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

600 **Figure 2. Alignment of HNF4 transcription factor homologs in hemipteran insects and**
601 **model species representative of Diptera, Coleoptera and Lepidoptera.** Amino acids
602 conserved in at least 90% of all sequences are indicated below the alignment (consensus >90).
603 The regions corresponding to the DNA binding domain (DBD) and ligand binding domain
604 (LBD) are indicated above the alignment by blue and green bars, respectively.

606 **Figure 3. Comparison of the Hox gene cluster in *Acyrtosiphon pisum* with the Hox**
607 **clusters of model species representative of Diptera, Coleoptera and Lepidoptera.** The
608 orientation of each arrow indicates the transcriptional orientation of each canonical and non-
609 canonical Hox gene. Genes are represented with their actual lengths and genomic distances. In
610 *B. mori*, the « *Shx* genes » box refers to 12 genes obtained through extensive tandem gene
611 duplication of *zen* (Chai et al., 2008). In *T. castaneum*, orthologs of *Antp*, *Scr* and *pb* are often
612 referred to as *ptl* (*prothoraxless*, Brown et al., 2002), *Cx* (*cephalothorax*, Curtis et al., 2001),
613 and *mvp* (*maxillopedia*, Shippy et al., 2000), respectively.

615 **Figure 4. Genomic organization and gene orientation across hemipteran Hox clusters. (A)**
616 Order and transcriptional orientation of canonical and non-canonical Hox genes in each species.
617 Splits within the Hox cluster are denoted by double black lines and inversions with respect to
618 the predicted ancestral Hox cluster are annotated with a black border around the gene. (B)
619 Organization of the Hox cluster per species, shown using actual genomic distances. Each line
620 represents a Hox gene as it occurs in the genome. Genomic distances are shown in Megabases
621 (Mb).

622 623 **Tables**

624 **Table 1. Genomic information (assembly and annotation numbers) used for the**
625 **annotation of HNF4 and the Hox cluster genes in hemipteran insects and model species**
626 **representative of Diptera, Coleoptera and Lepidoptera.**

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

630 **Table 3. Annotation of HNF4 transcription factor homologs in hemipteran insects and**
631 **model species representative of Diptera, Coleoptera and Lepidoptera.** For each NCBI gene
632 ID, all corresponding protein entries present in the database are listed. Each protein identifier
633 corresponds to a protein isoform encoded by a unique predicted messenger RNA. Some
634 isoforms are identical but are encoded by mRNAs that differ in their UTR regions, suggesting
635 differences in the regulation of their expression.

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

639 **Supplementary data**

640 **Suppl. File 1. Transcription factor annotation in *Acyrtosiphon pisum* (green), and**
641 **comparison with the reference organism, *Drosophila melanogaster* (blue).**

642 **Suppl. File 2. Chromatin-associated factor annotation in *Acyrtosiphon pisum* (green),**
643 **and comparison with the reference organism, *Drosophila melanogaster* (blue).**

644 **Suppl. File 3. Basal transcription machinery annotation in *Acyrtosiphon pisum* (green),**
645 **and comparison with the reference organism, *Drosophila melanogaster* (blue).**

646 **Suppl. File 4. Pairwise alignment statistics between the protein sequences of canonical and**
647 **non-canonical Hox genes in hemipteran insects and model species representative of**
648 **Diptera, Coleoptera and Lepidoptera, determined through systematic BLASTP. Coverage**
649 **and identity values above 80 are highlighted in red. Abbreviations: id, identity percentage; cov,**
650 **coverage percentage.**

651 **Suppl. File 5. Average coverage and identity values obtained through systematic BLASTP**
652 **between predicted orthologs of canonical and non-canonical Hox genes in insects,**
653 **hemipterans and aphids. Coverage and identity percentages are represented with a color**
654 **gradient from white to green for coverage and from white to red for identity. Darker colors**
655 **indicate higher percentages. Abbreviations: id, identity percentage; cov, coverage percentage.**
656

657 **CRediT authorship contribution statement**

658 **Nicolas Parisot:** conceptualization, methodology, software, formal analysis, investigation,
659 writing - original draft, writing – review & editing, visualization; **Mélanie Ribeiro Lopes:**
660 methodology, formal analysis, investigation, writing - original draft; **Sergio Peignier:**
661 validation, resources, data curation, writing – review & editing; **Patrice Baa-Puyoulet:**
662 software, data curation, writing – review & editing; **Hubert Charles:** validation, formal
663 analysis, writing-review & editing; **Federica Calevro:** conceptualization, investigation,
664 writing - original draft , writing – review & editing, supervision, funding acquisition ; **Patrick**
665 **Callaerts:** conceptualization, investigation, writing - original draft, writing – review & editing,
666 supervision, funding acquisition

667 **Acknowledgements**

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

670 **Funding**

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

675 **References**

- 676 Aerts, S., 2012. Chapter five - Computational strategies for the genome-wide identification of
677 cis-regulatory elements and transcriptional targets, in: Plaza, S., Payre, F. (Eds.),
678 Current Topics in Developmental Biology, Transcriptional Switches During
679 Development. Academic Press, pp. 121–145. <https://doi.org/10.1016/B978-0-12-386499-4.00005-7>
- 681 Alexandre, C., Jacinto, A., Ingham, P.W., 1996. Transcriptional activation of hedgehog target
682 genes in *Drosophila* is mediated directly by the cubitus interruptus protein, a member
683 of the GLI family of zinc finger DNA-binding proteins. *Genes & Development* 10,
684 2003–2013. <https://doi.org/10.1101/gad.10.16.2003>
- 685 Allen, D., Seo, J., 2018. ER Stress Activates the TOR Pathway through Atf6. *Journal of*
686 *Molecular Signaling*.13, 1. <https://doi.org/10.5334/1750-2187-13-1>
- 687 Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment
688 search tool. *Journal of Molecular Biology* 215, 403–410.
689 [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- 690 Bao, R., Fischer, T., Bolognesi, R., Brown, S.J., Friedrich, M., 2012. Parallel duplication and
691 partial subfunctionalization of β -Catenin/Armadillo during insect evolution. *Molecular*
692 *Biology and Evolution* 29, 647–662. <https://doi.org/10.1093/molbev/msr219>
- 693 Baumann, A., Fujiwara, Y., Wilson, T.G., 2010. Evolutionary divergence of the paralogs
694 Methoprene tolerant (Met) and germ cell expressed (gce) within the genus *Drosophila*.
695 *Journal of Insect Physiology, Insect Molecular Physiology - Basic Science to*
696 *Application, A Special Issue in Honour of Dr. Judith H. Willis* 56, 1445–1455.
697 <https://doi.org/10.1016/j.jinsphys.2010.05.001>
- 698 Baumann, P., 2005. Biology of bacteriocyte-associated endosymbionts of plant sap-sucking
699 Insects. *Annual Review of Microbiology* 59, 155–189.
700 <https://doi.org/10.1146/annurev.micro.59.030804.121041>
- 701 Bellen, H.J., Yamamoto, S., 2015. Morgan's legacy: fruit flies and the functional annotation
702 of conserved genes. *Cell* 163, 12–14. <https://doi.org/10.1016/j.cell.2015.09.009>
- 703 Ben-David, J., Chipman, A.D., 2010. Mutual regulatory interactions of the trunk gap genes
704 during blastoderm patterning in the hemipteran *Oncopeltus fasciatus*. *Developmental*
705 *Biology* 346, 140–149. <https://doi.org/10.1016/j.ydbio.2010.07.010>

706 Bodmer, R., 1993. The gene tinman is required for specification of the heart and visceral
707 muscles in *Drosophila*. *Development* 118, 719–729.
708 <https://doi.org/10.1242/dev.118.3.719>

709 Bornemann, D., Miller, E., Simon, J., 1996. The *Drosophila* Polycomb group gene Sex comb
710 on midleg (Scm) encodes a zinc finger protein with similarity to polyhomeotic protein.
711 *Development* 122, 1621–1630. <https://doi.org/10.1242/dev.122.5.1621>

712 Braendle, C., Miura, T., Bickel, R., Shingleton, A.W., Kambhampati, S., Stern, D.L., 2003.
713 Developmental origin and evolution of bacteriocytes in the aphid–*Buchnera*
714 Symbiosis. *PLOS Biology* 1, e21. <https://doi.org/10.1371/journal.pbio.0000021>

715 Brown, S.J., Fellers, J.P., Shippy, T.D., Richardson, E.A., Maxwell, M., Stuart, J.J., Denell,
716 R.E., 2002. Sequence of the *Tribolium castaneum* homeotic complex: The region
717 corresponding to the *Drosophila melanogaster* Antennapedia Complex. *Genetics* 160,
718 1067–1074. <https://doi.org/10.1093/genetics/160.3.1067>

719 Brunner, E., Peter, O., Schweizer, L., Basler, K., 1997. pangolin encodes a Lef-1 homologue
720 that acts downstream of Armadillo to transduce the Wingless signal in *Drosophila*.
721 *Nature* 385, 829–833. <https://doi.org/10.1038/385829a0>

722 Calevro, F., Callaerts, P., Matsuura, Y., Michalik, A., 2023. Editorial: Symbiotic organs in
723 insects: development, metabolism, and physiological regulation. *Frontiers in*
724 *Physiology* 14.

725 Calevro, F., Tagu, D., Callaerts, P., 2019. *Acyrtosiphon pisum*. *Trends in Genetics, The*
726 *Nucleolus* 35, 781–782. <https://doi.org/10.1016/j.tig.2019.07.003>

727 Carroll, S.B., 2008. Evo-devo and an expanding evolutionary synthesis: A genetic theory of
728 morphological evolution. *Cell* 134, 25–36. <https://doi.org/10.1016/j.cell.2008.06.030>

729 Celniker, S.E., Rubin, G.M., 2003. The *Drosophila Melanogaster* Genome. *Annual Review*
730 *of Genomics and Human Genetics* 4, 89–117.
731 <https://doi.org/10.1146/annurev.genom.4.070802.110323>

732 Chai, C.-L., Zhang, Z., Huang, F.-F., Wang, X.-Y., Yu, Q.-Y., Liu, B.-B., Tian, T., Xia, Q.-
733 Y., Lu, C., Xiang, Z.-H., 2008. A genomewide survey of homeobox genes and
734 identification of novel structure of the Hox cluster in the silkworm, *Bombyx mori*.
735 *Insect Biochemistry and Molecular Biology, Special Issue on the Silkworm Genome*
736 38, 1111–1120. <https://doi.org/10.1016/j.ibmb.2008.06.008>

737 Chen, Z.-X., Sturgill, D., Qu, J., Jiang, H., Park, S., Boley, N., Suzuki, A.M., Fletcher, A.R.,
738 Plachetzki, D.C., FitzGerald, P.C., Artieri, C.G., Atallah, J., Barmina, O., Brown, J.B.,
739 Blankenburg, K.P., Clough, E., Dasgupta, A., Gubbala, S., Han, Y., Jayaseelan, J.C.,
740 Kalra, D., Kim, Y.-A., Kovar, C.L., Lee, S.L., Li, M., Malley, J.D., Malone, J.H.,
741 Mathew, T., Mattiuzzo, N.R., Munidasa, M., Muzny, D.M., Onger, F., Perales, L.,
742 Przytycka, T.M., Pu, L.-L., Robinson, G., Thornton, R.L., Saada, N., Scherer, S.E.,
743 Smith, H.E., Vinson, C., Warner, C.B., Worley, K.C., Wu, Y.-Q., Zou, X., Cherbas,
744 P., Kellis, M., Eisen, M.B., Piano, F., Kionte, K., Fitch, D.H., Sternberg, P.W., Cutter,
745 A.D., Duff, M.O., Hoskins, R.A., Graveley, B.R., Gibbs, R.A., Bickel, P.J., Kopp, A.,
746 Carninci, P., Celniker, S.E., Oliver, B., Richards, S., 2014. Comparative validation of
747 the *D. melanogaster* modENCODE transcriptome annotation. *Genome Res.* 24, 1209–
748 1223. <https://doi.org/10.1101/gr.159384.113>

749 Cheng, Y., Brunner, A.L., Kremer, S., DeVido, S.K., Stefaniuk, C.M., Kassis, J.A., 2014. Co-
750 regulation of invected and engrailed by a complex array of regulatory sequences in
751 *Drosophila*. *Developmental Biology* 395, 131–143.
752 <https://doi.org/10.1016/j.ydbio.2014.08.021>

753 Cheng, Y., Li, Y., Li, W., Song, Y., Zeng, R., Lu, K., 2020. Effect of hepatocyte nuclear
754 factor 4 on the fecundity of *Nilaparvata lugens*: Insights from RNA interference

755 combined with transcriptomic analysis. *Genomics* 112, 4585–4594.
756 <https://doi.org/10.1016/j.ygeno.2020.08.002>

757 Chotard, C., Leung, W., Salecker, I., 2005. Glial cells missing and *gcm2* cell autonomously
758 regulate both glial and neuronal development in the visual system of *Drosophila*.
759 *Neuron* 48, 237–251. <https://doi.org/10.1016/j.neuron.2005.09.019>

760 Cohen, B., McGuffin, M.E., Pfeifle, C., Segal, D., Cohen, S.M., 1992. *apterous*, a gene
761 required for imaginal disc development in *Drosophila* encodes a member of the LIM
762 family of developmental regulatory proteins. *Genes Dev.* 6, 715–729.
763 <https://doi.org/10.1101/gad.6.5.715>

764 Colella, S., Parisot, N., Simonet, P., Gaget, K., Duport, G., Baa-Puyoulet, P., Rahbé, Y.,
765 Charles, H., Febvay, G., Callaerts, P., Calevro, F., 2018. Bacteriocyte reprogramming
766 to cope with nutritional stress in a phloem sap feeding hemipteran, the pea aphid
767 *Acyrtosiphon pisum*. *Frontiers in Physiology* 25:9:1498.
768 <https://doi.org/10.3389/fphys.2018.01498>. eCollection 2018.

769 Coschigano, K.T., Wensink, P.C., 1993. Sex-specific transcriptional regulation by the male
770 and female doublesex proteins of *Drosophila*. *Genes & Development.* 7, 42–54.
771 <https://doi.org/10.1101/gad.7.1.42>

772 Curtis, C.D., Brisson, J.A., DeCamillis, M.A., Shippy, T.D., Brown, S.J., Denell, R.E., 2001.
773 Molecular characterization of *Cephalothorax*, the *Tribolium* ortholog of Sex combs
774 reduced. *genesis* 30, 12–20. <https://doi.org/10.1002/gene.1027>

775 Dang, C.-W., Wang, Y., Chen, K.-P., Yao, Q., Zhang, D., Guo, M., 2011. The basic helix-
776 loop-helix transcription factor family in the pea aphid, *Acyrtosiphon pisum*. *Journal*
777 *of Insect Science* 11, 84. <https://doi.org/10.1673/031.011.8401>

778 Datta, R.R., Weasner, B.P., Kumar, J.P., 2011. A dissection of the *teashirt* and *tiptop* genes
779 reveals a novel mechanism for regulating transcription factor activity. *Developmental*
780 *Biology* 360, 391–402. <https://doi.org/10.1016/j.ydbio.2011.09.030>

781 Davidson, E.H., Levine, M.S., 2008. Properties of developmental gene regulatory networks.
782 *Proceedings of the National Academy of Sciences* 105, 20063–20066.
783 <https://doi.org/10.1073/pnas.0806007105>

784 Davis, G.K., Brisson, J.A., Bickel, R.D., 2021. Evo-devo lessons learned from aphids, in:
785 Nuño de la Rosa, L., Müller, G.B. (Eds.), *Evolutionary Developmental Biology: A*
786 *Reference Guide*. Springer International Publishing, Cham, pp. 817–829.
787 https://doi.org/10.1007/978-3-319-32979-6_182

788 Dearden, P.K., 2015. Origin and evolution of the enhancer of split complex. *BMC Genomics*
789 16, 712. <https://doi.org/10.1186/s12864-015-1926-1>

790 Desvergne, B., Michalik, L., Wahli, W., 2006. Transcriptional regulation of metabolism.
791 *Physiological Reviews* 86, 465–514. <https://doi.org/10.1152/physrev.00025.2005>

792 Domínguez, M., Casares, F., 2005. Organ specification–growth control connection: New in-
793 sights from the *Drosophila* eye–antennal disc. *Developmental Dynamics* 232, 673–
794 684. <https://doi.org/10.1002/dvdy.20311>

795 Douglas, A.E., 2015. Multiorganismal insects: diversity and function of resident
796 microorganisms. *Annual Review of Entomology* 60, 17–34.
797 <https://doi.org/10.1146/annurev-ento-010814-020822>

798 Dow, J.A.T., Simons, M., Romero, M.F., 2022. *Drosophila melanogaster*: a simple genetic
799 model of kidney structure, function and disease. *Nature Reviews Nephrology* 18, 417–
800 434. <https://doi.org/10.1038/s41581-022-00561-4>

801 Erclik, T., Hartenstein, V., Lipshitz, H.D., McInnes, R.R., 2008. Conserved role of the *Vsx*
802 genes supports a monophyletic origin for bilaterian visual systems. *Current Biology*
803 18, 1278–1287. <https://doi.org/10.1016/j.cub.2008.07.076>

804 Fernández, R., Marcet-Houben, M., Legeai, F., Richard, G., Robin, S., Wucher, V.,
805 Pegueroles, C., Gabaldón, T., Tagu, D., 2020. Selection following gene duplication
806 shapes recent genome evolution in the pea aphid *Acyrtosiphon pisum*. *Molecular*
807 *Biology and Evolution* 37, 2601–2615. <https://doi.org/10.1093/molbev/msaa110>
808 Fischer, F.P., Karge, R.A., Weber, Y.G., Koch, H., Wolking, S., Voigt, A., 2023. *Drosophila*
809 *melanogaster* as a versatile model organism to study genetic epilepsies: An overview.
810 *Frontiers in Molecular Neuroscience* 16 - 2023 |
811 <https://doi.org/10.3389/fnmol.2023.1116000>
812 Gallant, P., Shiio, Y., Cheng, P.F., Parkhurst, S.M., Eisenman, R.N., 1996. Myc and Max
813 homologs in *Drosophila*. *Science* 274, 1523–1527.
814 <https://doi.org/10.1126/science.274.5292.1523>
815 Gibert, J.-M., Peronnet, F., 2021. The paramount role of *Drosophila melanogaster* in the
816 study of epigenetics: From simple phenotypes to molecular dissection and higher-
817 order genome organization. *Insects* 12, 884. <https://doi.org/10.3390/insects12100884>
818 Gonzalez, D.H., 2016. Chapter 1 - Introduction to transcription factor structure and function,
819 in: Gonzalez, D.H. (Ed.), *Plant Transcription Factors*. Academic Press, Boston, pp. 3–
820 11. <https://doi.org/10.1016/B978-0-12-800854-6.00001-4>
821 Gopalan, S., Wang, Y., Harper, N.W., Garber, M., Fazio, T.G., 2021. Simultaneous profiling
822 of multiple chromatin proteins in the same cells. *Molecular Cell* 81, 4736-4746.e5.
823 <https://doi.org/10.1016/j.molcel.2021.09.019>
824 Gramates, L.S., Agapite, J., Attrill, H., Calvi, B.R., Crosby, M.A., dos Santos, G., Goodman,
825 J.L., Goutte-Gattat, D., Jenkins, V.K., Kaufman, T., Larkin, A., Matthews, B.B.,
826 Millburn, G., Strelets, V.B., the FlyBase Consortium, 2022. FlyBase: a guided tour of
827 highlighted features. *Genetics* 220, iyac035. <https://doi.org/10.1093/genetics/iyac035>
828 Gregor, T., McGregor, A.P., Wieschaus, E.F., 2008. Shape and function of the Bicoid
829 morphogen gradient in dipteran species with different sized embryos. *Developmental*
830 *Biology* 316, 350–358. <https://doi.org/10.1016/j.ydbio.2008.01.039>
831 Grossniklaus, U., Pearson, R.K., Gehring, W.J., 1992. The *Drosophila* sloppy paired locus
832 encodes two proteins involved in segmentation that show homology to mammalian
833 transcription factors. *Genes & Development* 6, 1030–1051.
834 <https://doi.org/10.1101/gad.6.6.1030>
835 Gutiérrez, L., Oktaba, K., Scheuermann, J.C., Gambetta, M.C., Ly-Hartig, N., Müller, J.,
836 2012. The role of the histone H2A ubiquitinase Sce in Polycomb repression.
837 *Development* 139, 117–127. <https://doi.org/10.1242/dev.074450>
838 Hamaguchi, T., Yabe, S., Uchiyama, H., Murakami, R., 2004. *Drosophila* Tbx6-related gene,
839 Dorsocross, mediates high levels of Dpp and Scw signal required for the development
840 of amnioserosa and wing disc primordium. *Developmental Biology* 265, 355–368.
841 <https://doi.org/10.1016/j.ydbio.2003.09.034>
842 Hammonds, A.S., Bristow, C.A., Fisher, W.W., Weiszmann, R., Wu, S., Hartenstein, V.,
843 Kellis, M., Yu, B., Frise, E., Celniker, S.E., 2013. Spatial expression of transcription
844 factors in *Drosophila* embryonic organ development. *Genome Biology* 14, R140.
845 <https://doi.org/10.1186/gb-2013-14-12-r140>
846 Han, J., Kaufman, R.J., 2017. Physiological/pathological ramifications of transcription factors
847 in the unfolded protein response. *Genes & Development* 31, 1417–1438.
848 <https://doi.org/10.1101/gad.297374.117>
849 Hardin, P.E., 2011. Chapter 5 - Molecular genetic analysis of circadian timekeeping in
850 *Drosophila*, in: Brody, S. (Ed.), *Advances in Genetics, The Genetics of Circadian*
851 *Rhythms*. Academic Press, pp. 141–173. [https://doi.org/10.1016/B978-0-12-387690-](https://doi.org/10.1016/B978-0-12-387690-4.00005-2)
852 [4.00005-2](https://doi.org/10.1016/B978-0-12-387690-4.00005-2)

853 Helfand, S.L., Rogina, B., 2003. Genetics of aging in the fruit fly, *Drosophila melanogaster*.
854 Annual Review of Genetics 37, 329–348.
855 <https://doi.org/10.1146/annurev.genet.37.040103.095211>

856 Hens, K., Feuz, J.-D., Isakova, A., Iagovitina, A., Massouras, A., Bryois, J., Callaerts, P.,
857 Celniker, S.E., Deplancke, B., 2011. Automated protein-DNA interaction screening of
858 *Drosophila* regulatory elements. Nature Methods 8, 1065–1070.
859 <https://doi.org/10.1038/nmeth.1763>

860 Higashijima, S., Kojima, T., Michiue, T., Ishimaru, S., Emori, Y., Saigo, K., 1992. Dual Bar
861 homeo box genes of *Drosophila* required in two photoreceptor cells, R1 and R6, and
862 primary pigment cells for normal eye development. Genes & Development 6, 50–60.
863 <https://doi.org/10.1101/gad.6.1.50>

864 Hou, X.S., Melnick, M.B., Perrimon, N., 1996. marelle acts downstream of the *Drosophila*
865 HOP/JAK kinase and encodes a protein similar to the mammalian STATs. Cell 84,
866 411–419. [https://doi.org/10.1016/S0092-8674\(00\)81286-6](https://doi.org/10.1016/S0092-8674(00)81286-6)

867 Huang, F., Saraf, A., Florens, L., Kusch, T., Swanson, S.K., Szerszen, L.T., Li, G., Dutta, A.,
868 Washburn, M.P., Abmayr, S.M., Workman, J.L., 2016. The Enok acetyltransferase
869 complex interacts with Elg1 and negatively regulates PCNA unloading to promote the
870 G1/S transition. Genes & Development 5;30(10):1198-210
871 <https://doi.org/10.1101/gad.271429.115>

872 Huang, H.-W., Zeng, X., Rhim, T., Ron, D., Ryoo, H.D., 2017. The requirement of IRE1 and
873 XBP1 in resolving physiological stress during *Drosophila* development. Journal of
874 Cell Science 130, 3040–3049. <https://doi.org/10.1242/jcs.203612>

875 Huang, J., Wu, S., Barrera, J., Matthews, K., Pan, D., 2005. The Hippo signaling pathway
876 coordinately regulates cell proliferation and apoptosis by inactivating Yorkie, the
877 *Drosophila* homolog of YAP. Cell 122, 421–434.
878 <https://doi.org/10.1016/j.cell.2005.06.007>

879 i5K Consortium, 2013. The i5K Initiative: Advancing arthropod genomics for knowledge,
880 human health, agriculture, and the environment. Journal of Heredity 104, 595–600.
881 <https://doi.org/10.1093/jhered/est050>

882 Jagla, K., Frasch, M., Jagla, T., Dretzen, G., Bellard, F., Bellard, M., 1997. ladybird, a new
883 component of the cardiogenic pathway in *Drosophila* required for diversification of
884 heart precursors. Development 124, 3471–3479.
885 <https://doi.org/10.1242/dev.124.18.3471>

886 Jin, J., Tian, F., Yang, D.-C., Meng, Y.-Q., Kong, L., Luo, J., Gao, G., 2017. PlantTFDB 4.0:
887 toward a central hub for transcription factors and regulatory interactions in plants.
888 Nucleic Acids Research 45, D1040–D1045. <https://doi.org/10.1093/nar/gkw982>

889 Jindra, M., Bellés, X., Shinoda, T., 2015. Molecular basis of juvenile hormone signaling.
890 Current Opinion in Insect Science, Global change biology * Molecular physiology 11,
891 39–46. <https://doi.org/10.1016/j.cois.2015.08.004>

892 Just, J., Laslo, M., Lee, Y.J., Yarnell, M., Zhang, Z., Angelini, D.R., 2023. Distinct
893 developmental mechanisms influence sexual dimorphisms in the milkweed bug
894 *Oncopeltus fasciatus*. Proceedings of the Royal Society B: Biological Sciences 290,
895 20222083. <https://doi.org/10.1098/rspb.2022.2083>

896 Kelly, K.K., Meadows, S.M., Cripps, R.M., 2002. *Drosophila* MEF2 is a direct regulator of
897 Actin57B transcription in cardiac, skeletal, and visceral muscle lineages. Mechanisms
898 of Development 110, 39–50. [https://doi.org/10.1016/S0925-4773\(01\)00586-X](https://doi.org/10.1016/S0925-4773(01)00586-X)

899 Kim, S., Wysocka, J., 2023. Deciphering the multi-scale, quantitative cis-regulatory code.
900 Molecular Cell, Reimagining the Central Dogma 83, 373–392.
901 <https://doi.org/10.1016/j.molcel.2022.12.032>

902 Kirfel, P., Skaljic, M., Grotmann, J., Kessel, T., Seip, M., Michaelis, K., Vilcinskas, A., 2020.
903 Inhibition of histone acetylation and deacetylation enzymes affects longevity,
904 development, and fecundity in the pea aphid (*Acyrtosiphon pisum*). Archives of
905 Insect Biochemistry and Physiology 103, e21614. <https://doi.org/10.1002/arch.21614>
906 Kumar, J.P., 2010. Chapter one - Retinal Determination: The beginning of eye development,
907 in: Cagan, R.L., Reh, T.A. (Eds.), Current Topics in Developmental Biology,
908 Invertebrate and Vertebrate Eye Development. Academic Press, pp. 1–28.
909 <https://doi.org/10.1016/B978-0-12-385044-7.00001-1>
910 Kummerfeld, S.K., Teichmann, S.A., 2006. DBD: a transcription factor prediction database.
911 Nucleic Acids Research 34, D74–D81. <https://doi.org/10.1093/nar/gkj131>
912 Lagarou, A., Mohd-Sarip, A., Moshkin, Y.M., Chalkley, G.E., Bezstarosti, K., Demmers,
913 J.A.A., Verrijzer, C.P., 2008. dKDM2 couples histone H2A ubiquitylation to histone
914 H3 demethylation during Polycomb group silencing. Genes Dev. 22, 2799–2810.
915 <https://doi.org/10.1101/gad.484208>
916 Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale,
917 J., Hughes, T.R., Weirauch, M.T., 2018. The human transcription factors. Cell 172,
918 650–665. <https://doi.org/10.1016/j.cell.2018.01.029>
919 Laslo, M., Just, J., Angelini, D.R., 2023. Theme and variation in the evolution of insect sex
920 determination. Journal of Experimental Zoology Part B: Molecular and
921 Developmental Evolution 340, 162–181. <https://doi.org/10.1002/jez.b.23125>
922 Le Trionnaire, G., Tanguy, S., Hudaverdian, S., Gleonnec, F., Richard, G., Cayrol, B.,
923 Monsion, B., Pichon, E., Deshoux, M., Webster, C., Uzest, M., Herpin, A., Tagu, D.,
924 2019. An integrated protocol for targeted mutagenesis with CRISPR-Cas9 system in
925 the pea aphid. Insect Biochemistry and Molecular Biology 110, 34–44.
926 <https://doi.org/10.1016/j.ibmb.2019.04.016>
927 Lecona, E., Rojas, L.A., Bonasio, R., Johnston, A., Fernández-Capetillo, O., Reinberg, D.,
928 2013. Polycomb protein SCML2 regulates the cell cycle by binding and modulating
929 CDK/CYCLIN/p21 complexes. PLOS Biology 11, e1001737.
930 <https://doi.org/10.1371/journal.pbio.1001737>
931 Lemon, B., Tjian, R., 2000. Orchestrated response: a symphony of transcription factors for
932 gene control. Genes & Development 14, 2551–2569.
933 <https://doi.org/10.1101/gad.831000>
934 Li, Y., Park, H., Smith, T.E., Moran, N.A., 2019. Gene family evolution in the pea aphid
935 based on chromosome-level genome assembly. Molecular Biology and Evolution 36,
936 2143–2156. <https://doi.org/10.1093/molbev/msz138>
937 Loo, L.W.M., Secombe, J., Little, J.T., Carlos, L.-S., Yost, C., Cheng, P.-F., Flynn, E.M.,
938 Edgar, B.A., Eisenman, R.N., 2005. The transcriptional repressor dMnt is a regulator
939 of growth in *Drosophila melanogaster*. Molecular and Cellular Biology 25, 7078–
940 7091. <https://doi.org/10.1128/MCB.25.16.7078-7091.2005>
941 Lours, C., Bardot, O., Godt, D., Laski, F.A., Couderc, J., 2003. The *Drosophila melanogaster*
942 BTB proteins bric à brac bind DNA through a composite DNA binding domain
943 containing a pipsqueak and an AT-Hook motif. Nucleic Acids Research 31, 5389–
944 5398. <https://doi.org/10.1093/nar/gkg724>
945 Luscombe, N.M., Austin, S.E., Berman, H.M., Thornton, J.M., 2000. An overview of the
946 structures of protein-DNA complexes. Genome Biology 1, reviews001.1.
947 <https://doi.org/10.1186/gb-2000-1-1-reviews001>
948 Lyne, R., Smith, R., Rutherford, K., Wakeling, M., Varley, A., Guillier, F., Janssens, H., Ji,
949 W., McLaren, P., North, P., Rana, D., Riley, T., Sullivan, J., Watkins, X., Woodbridge,
950 M., Lilley, K., Russell, S., Ashburner, M., Mizuguchi, K., Micklem, G., 2007.

951 FlyMine: an integrated database for *Drosophila* and *Anopheles* genomics. *Genome*
952 *Biology* 8, R129. <https://doi.org/10.1186/gb-2007-8-7-r129>

953 Mach, V., Ohno, K., Kokubo, H., Suzuki, Y., 1996. The *Drosophila* fork head factor directly
954 controls larval salivary gland-specific expression of the glue protein gene Sgs3.
955 *Nucleic Acids Research* 24, 2387–2394. <https://doi.org/10.1093/nar/24.12.2387>

956 Madeira, F., Pearce, M., Tivey, A.R.N., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N.,
957 Kolesnikov, A., Lopez, R., 2022. Search and sequence analysis tools services from
958 EMBL-EBI in 2022. *Nucleic Acids Research* 50, W276–W279.
959 <https://doi.org/10.1093/nar/gkac240>

960 Mannervik, M., Levine, M., 1999. The Rpd3 histone deacetylase is required for segmentation
961 of the *Drosophila* embryo. *Proceedings of the National Academy of Sciences* 96,
962 6797–6801. <https://doi.org/10.1073/pnas.96.12.6797>

963 Martín, M., Organista, M.F., de Celis, J.F., 2016. Structure of developmental gene regulatory
964 networks from the perspective of cell fate-determining genes. *Transcription* 7, 32–37.
965 <https://doi.org/10.1080/21541264.2015.1130118>

966 Matsuura, Y., Kikuchi, Y., Miura, T., Fukatsu, T., 2015. Ultrabithorax is essential for
967 bacteriocyte development. *Proceedings of the National Academy of Sciences* 112,
968 9376–9381. <https://doi.org/10.1073/pnas.1503371112>

969 Mendez, D.L., Kim, D., Chruszcz, M., Stephens, G.E., Minor, W., Khorasanizadeh, S., Elgin,
970 S.C.R., 2011. The HP1a disordered C terminus and chromo shadow domain cooperate
971 to select target peptide partners. *ChemBioChem* 12, 1084–1096.
972 <https://doi.org/10.1002/cbic.201000598>

973 Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L.,
974 Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., Finn, R.D., Bateman, A., 2021.
975 Pfam: The protein families database in 2021. *Nucleic Acids Research* 49, D412–D419.
976 <https://doi.org/10.1093/nar/gkaa913>

977 Miura, K., Oda, M., Makita, S., Chinzei, Y., 2005. Characterization of the *Drosophila*
978 Methoprene -tolerant gene product. *The FEBS Journal* 272, 1169–1178.
979 <https://doi.org/10.1111/j.1742-4658.2005.04552.x>

980 Mohr, S.E., Perrimon, N., 2019. *Drosophila melanogaster*: a simple system for understanding
981 complexity. *Disease Models & Mechanisms* 12, dmm041871.
982 <https://doi.org/10.1242/dmm.041871>

983 Mulhair, P.O., Holland, P.W.H., 2022. Evolution of the insect Hox gene cluster: Comparative
984 analysis across 243 species. *Seminars in Cell & Developmental Biology, Hox genes:*
985 *The Original Body Builders* 152–153, 4–15.
986 <https://doi.org/10.1016/j.semcd.2022.11.010>

987 Näär, A.M., Lemon, B.D., Tjian, R., 2001. Transcriptional coactivator complexes. *Annual*
988 *Review of Biochemistry* 70, 475–501.
989 <https://doi.org/10.1146/annurev.biochem.70.1.475>

990 Nardelli, J., Gibson, T.J., Vesque, C., Charnay, P., 1991. Base sequence discrimination by
991 zinc-finger DNA-binding domains. *Nature* 349, 175–178.
992 <https://doi.org/10.1038/349175a0>

993 Ogawa, K., Miura, T., 2014. Aphid polyphenisms: trans-generational developmental
994 regulation through viviparity. *Frontiers in Physiology* 5: 1.
995 <https://doi.org/10.3389/fphys.2014.00001>

996 O'Neill, E.M., Rebay, I., Tjian, R., Rubin, G.M., 1994. The activities of two Ets-related
997 transcription factors required for drosophila eye development are modulated by the
998 Ras/MAPK pathway. *Cell* 78, 137–147. [https://doi.org/10.1016/0092-8674\(94\)90580-](https://doi.org/10.1016/0092-8674(94)90580-0)
999 0

1000 Pace, R.M., Grbić, M., Nagy, L.M., 2016. Composition and genomic organization of
1001 arthropod Hox clusters. *EvoDevo* 7, 11. <https://doi.org/10.1186/s13227-016-0048-4>

1002 Palanker, L., Tennessen, J.M., Lam, G., Thummel, C.S., 2009. *Drosophila* HNF4 regulates
1003 lipid mobilization and β -oxidation. *Cell Metabolism* 9, 228–239.
1004 <https://doi.org/10.1016/j.cmet.2009.01.009>

1005 Panfilio, K.A., 2008. Extraembryonic development in insects and the acrobatics of
1006 blastokinesis. *Developmental Biology* 313, 471–491.
1007 <https://doi.org/10.1016/j.ydbio.2007.11.004>

1008 Panfilio, K.A., Vargas Jentzsch, I.M., Benoit, J.B., Erezyilmaz, D., Suzuki, Y., Colella, S.,
1009 Robertson, H.M., Poelchau, M.F., Waterhouse, R.M., Ioannidis, P., Weirauch, M.T.,
1010 Hughes, D.S.T., Murali, S.C., Werren, J.H., Jacobs, C.G.C., Duncan, E.J., Armisen,
1011 D., Vreede, B.M.I., Baa-Puyoulet, P., Berger, C.S., Chang, C., Chao, H., Chen, M.-
1012 J.M., Chen, Y.-T., Childers, C.P., Chipman, A.D., Cridge, A.G., Crumière, A.J.J.,
1013 Dearden, P.K., Didion, E.M., Dinh, H., Doddapaneni, H.V., Dolan, A., Dugan, S.,
1014 Extavour, C.G., Febvay, G., Friedrich, M., Ginzburg, N., Han, Y., Heger, P., Holmes,
1015 C.J., Horn, T., Hsiao, Y., Jennings, E.C., Johnston, J.S., Jones, T.E., Jones, J.W.,
1016 Khila, A., Koelzer, S., Kovacova, V., Leask, M., Lee, S.L., Lee, C.-Y., Lovegrove,
1017 M.R., Lu, H., Lu, Y., Moore, P.J., Munoz-Torres, M.C., Muzny, D.M., Palli, S.R.,
1018 Parisot, N., Pick, L., Porter, M.L., Qu, J., Refki, P.N., Richter, R., Rivera-Pomar, R.,
1019 Rosendale, A.J., Roth, S., Sachs, L., Santos, M.E., Seibert, J., Sghaier, E., Shukla,
1020 J.N., Stancliffe, R.J., Tidswell, O., Traverso, L., van der Zee, M., Viala, S., Worley,
1021 K.C., Zdobnov, E.M., Gibbs, R.A., Richards, S., 2019. Molecular evolutionary trends
1022 and feeding ecology diversification in the Hemiptera, anchored by the milkweed bug
1023 genome. *Genome Biology* 20, 64. <https://doi.org/10.1186/s13059-019-1660-0>

1024 Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B.L., Salazar, G.A.,
1025 Bileschi, M.L., Bork, P., Bridge, A., Colwell, L., Gough, J., Haft, D.H., Letunić, I.,
1026 Marchler-Bauer, A., Mi, H., Natale, D.A., Orengo, C.A., Pandurangan, A.P., Rivoire,
1027 C., Sigrist, C.J.A., Sillitoe, I., Thanki, N., Thomas, P.D., Tosatto, S.C.E., Wu, C.H.,
1028 Bateman, A., 2023. InterPro in 2022. *Nucleic Acids Research* 51, D418–D427.
1029 <https://doi.org/10.1093/nar/gkac993>

1030 Peel, A.D., Telford, M.J., Akam, M., 2006. The evolution of hexapod engrailed-family genes:
1031 evidence for conservation and concerted evolution. *Proceedings of the Royal Society*
1032 *B: Biological Sciences* 273, 1733–1742. <https://doi.org/10.1098/rspb.2006.3497>

1033 Peñalosa-Ruiz, G., Bright, A.R., Mulder, K.W., Veenstra, G.J.C., 2019. The interplay of
1034 chromatin and transcription factors during cell fate transitions in development and
1035 reprogramming. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory*
1036 *Mechanisms* 1862, 194407. <https://doi.org/10.1016/j.bbagr.2019.194407>

1037 Perkins, K.K., Admon, A., Patel, N., Tjian, R., 1990. The *Drosophila* Fos-related AP-1
1038 protein is a developmentally regulated transcription factor. *Genes Dev.* 4, 822–834.
1039 <https://doi.org/10.1101/gad.4.5.822>

1040 Puig, O., Marr, M.T., Ruhf, M.L., Tjian, R., 2003. Control of cell number by *Drosophila*
1041 FOXO: downstream and feedback regulation of the insulin receptor pathway. *Genes &*
1042 *Development* 17, 2006–2020. <https://doi.org/10.1101/gad.1098703>

1043 Rebay, I., Rubin, G.M., 1995. Yan functions as a general inhibitor of differentiation and is
1044 negatively regulated by activation of the Ras1/MAPK pathway. *Cell* 81, 857–866.
1045 [https://doi.org/10.1016/0092-8674\(95\)90006-3](https://doi.org/10.1016/0092-8674(95)90006-3)

1046 Ribeiro Lopes, M., Gaget, K., Renoz, F., Dupont, G., Balmand, S., Charles, H., Callaerts, P.,
1047 Calevro, F., 2022. Bacteriocyte plasticity in pea aphids facing amino acid stress or
1048 starvation during development. *Frontiers in Physiology* 13: 982920.
1049 <https://doi.org/10.3389/fphys.2022.982920>

- 1050 Robert, X., Gouet, P., 2014. Deciphering key features in protein structures with the new
1051 ENDscript server. *Nucleic Acids Research* 42, W320-324.
1052 <https://doi.org/10.1093/nar/gku316>
- 1053 Sapountzis, P., Duport, G., Balmand, S., Gaget, K., Jaubert-Possamai, S., Febvay, G.,
1054 Charles, H., Rahbé, Y., Colella, S., Calevro, F., 2014. New insight into the RNA
1055 interference response against cathepsin-L gene in the pea aphid, *Acyrtosiphon pisum*:
1056 Molting or gut phenotypes specifically induced by injection or feeding treatments.
1057 *Insect Biochemistry and Molecular Biology* 51, 20–32.
1058 <https://doi.org/10.1016/j.ibmb.2014.05.005>
- 1059 Schaeper, N.D., Prpic, N.-M., Wimmer, E.A., 2009. A conserved function of the zinc finger
1060 transcription factor Sp8/9 in allometric appendage growth in the milkweed
1061 bug *Oncopeltus fasciatus*. *Developmental Genes and Evolution* 219, 427–435.
1062 <https://doi.org/10.1007/s00427-009-0301-0>
- 1063 Schember, I., Halfon, M.S., 2022. Common themes and future challenges in understanding
1064 gene regulatory network evolution. *Cells* 11, 510.
1065 <https://doi.org/10.3390/cells11030510>
- 1066 Schmidt-Ott, U., 2000. The amnioserosa is an apomorphic character of cyclorrhaphan flies.
1067 *Developmental Genes and Evolution* 210, 373–376.
1068 <https://doi.org/10.1007/s004270000068>
- 1069 Schweisguth, F., Posakony, J.W., 1992. Suppressor of Hairless, the *Drosophila* homolog of
1070 the mouse recombination signal-binding protein gene, controls sensory organ cell
1071 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)
- 1072 Scott, E.K., Lee, T., Luo, L., 2001. enok encodes a *Drosophila* putative histone
1073 acetyltransferase required for mushroom body neuroblast proliferation. *Current*
1074 *Biology* 11, 99–104. [https://doi.org/10.1016/S0960-9822\(01\)00020-3](https://doi.org/10.1016/S0960-9822(01)00020-3)
- 1075 Sekelsky, J.J., Newfeld, S.J., Raftery, L.A., Chartoff, E.H., Gelbart, W.M., 1995. Genetic
1076 characterization and cloning of mothers against dpp, a gene required for
1077 decapentaplegic function in *Drosophila melanogaster*. *Genetics* 139, 1347–1358.
1078 <https://doi.org/10.1093/genetics/139.3.1347>
- 1079 Shigenobu, S., Bickel, R.D., Brisson, J.A., Butts, T., Chang, C., Christiaens, O., Davis, G.K.,
1080 Duncan, E.J., Ferrier, D.E.K., Iga, M., Janssen, R., Lin, G.-W., Lu, H.-L., McGregor,
1081 A.P., Miura, T., Smagghe, G., Smith, J.M., Van Der Zee, M., Velarde, R.A., Wilson,
1082 M.J., Dearden, P.K., Stern, D.L., 2010. Comprehensive survey of developmental
1083 genes in the pea aphid, *Acyrtosiphon pisum*: frequent lineage-specific duplications
1084 and losses of developmental genes. *Insect Molecular Biology* 19, 47–62.
1085 <https://doi.org/10.1111/j.1365-2583.2009.00944.x>
- 1086 Shippy, T.D., Brown, S.J., Denell, R.E., 2000. maxillopedia is the *Tribolium* ortholog of
1087 proboscipedia. *Evolution & Development* 2, 145–151. <https://doi.org/10.1046/j.1525-142x.2000.00055.x>
- 1089 Shokri, L., Inukai, S., Hafner, A., Weinand, K., Hens, K., Vedenko, A., Gisselbrecht, S.S.,
1090 Dainese, R., Bischof, J., Furger, E., Feuz, J.-D., Basler, K., Deplancke, B., Bulyk,
1091 M.L., 2019. A comprehensive *Drosophila melanogaster* transcription factor
1092 interactome. *Cell Reports* 27, 955-970.e7. <https://doi.org/10.1016/j.celrep.2019.03.071>
- 1093 Simonet, P., Gaget, K., Balmand, S., Ribeiro Lopes, M., Parisot, N., Buhler, K., Duport, G.,
1094 Vulsteke, V., Febvay, G., Heddi, A., Charles, H., Callaerts, P., Calevro, F., 2018.
1095 Bacteriocyte cell death in the pea aphid/ *Buchnera* symbiotic system. *Proceedings of*
1096 *the National Academy of Sciences* 115, E1819–E1828.
1097 <https://doi.org/10.1073/pnas.1720237115>

- 1098 Singh, A.J., Ramsey, S.A., Filtz, T.M., Kioussi, C., 2018. Differential gene regulatory
1099 networks in development and disease. *Cell. Mol. Life Sci.* 75, 1013–1025.
1100 <https://doi.org/10.1007/s00018-017-2679-6>
- 1101 Song, L., Huang, S.C., Wise, A., Castanon, R., Nery, J.R., Chen, H., Watanabe, M., Thomas,
1102 J., Bar-Joseph, Z., Ecker, J.R., 2016. A transcription factor hierarchy defines an
1103 environmental stress response network. *Science* 354, aag1550.
1104 <https://doi.org/10.1126/science.aag1550>
- 1105 Spitz, F., Furlong, E.E.M., 2012. Transcription factors: from enhancer binding to
1106 developmental control. *Nature Reviews Genetics* 13, 613–626.
1107 <https://doi.org/10.1038/nrg3207>
- 1108 Sproul, J.S., Hotaling, S., Heckenhauer, J., Powell, A., Marshall, D., Larracuente, A.M.,
1109 Kelley, J.L., Pauls, S.U., Frandsen, P.B., 2023. Analyses of 600+ insect genomes
1110 reveal repetitive element dynamics and highlight biodiversity-scale repeat annotation
1111 challenges. *Genome Research* 33, 1708–1717. <https://doi.org/10.1101/gr.277387.122>
- 1112 Stauber, M., Jäckle, H., Schmidt-Ott, U., 1999. The anterior determinant bicoid of *Drosophila*
1113 is a derived Hox class 3 gene. *Proceedings of the National Academy of Sciences* 96,
1114 3786–3789. <https://doi.org/10.1073/pnas.96.7.3786>
- 1115 Stevaux, O., Dimova, D., Frolov, M.V., Taylor-Harding, B., Morris, E., Dyson, N., 2002.
1116 Distinct mechanisms of E2F regulation by *Drosophila* RBF1 and RBF2. *The EMBO*
1117 *Journal* 21, 4927–4937. <https://doi.org/10.1093/emboj/cdf501>
- 1118 Sykiotis, G.P., Bohmann, D., 2008. Keap1/Nrf2 signaling regulates oxidative stress tolerance
1119 and lifespan in *Drosophila*. *Developmental Cell* 14, 76–85.
1120 <https://doi.org/10.1016/j.devcel.2007.12.002>
- 1121 Tagu, D., Calevro, F., Colella, S., Gabaldon, T., Sugio, A., 2016. Functional and evolutionary
1122 genomics in aphids, in: Vilcinskis, A. (Ed.), *Biology and Ecology of Aphids*. CRC
1123 Press, Taylor and Francis Group, pp. 52–88.
- 1124 Tataroglu, O., Emery, P., 2014. Studying circadian rhythms in *Drosophila melanogaster*.
1125 *Methods, Drosophila developmental biology methods* 68, 140–150.
1126 <https://doi.org/10.1016/j.ymeth.2014.01.001>
- 1127 The International Aphid Genomics Consortium, 2010. Genome sequence of the pea
1128 aphid *Acyrtosiphon pisum*. *PLOS Biology* 8, e1000313.
1129 <https://doi.org/10.1371/journal.pbio.1000313>
- 1130 Theopold, U., Ekengren, S., Hultmark, D., 1996. HLH106, a *Drosophila* transcription factor
1131 with similarity to the vertebrate sterol responsive element binding protein.
1132 *Proceedings of the National Academy of Sciences* 93, 1195–1199.
1133 <https://doi.org/10.1073/pnas.93.3.1195>
- 1134 Tian, F., Shah, P.K., Liu, X., Negre, N., Chen, J., Karpenko, O., White, K.P., Grossman, R.L.,
1135 2009. Flynet: a genomic resource for *Drosophila melanogaster* transcriptional
1136 regulatory networks. *Bioinformatics* 25, 3001–3004.
1137 <https://doi.org/10.1093/bioinformatics/btp469>
- 1138 Tiebe, M., Lutz, M., De La Garza, A., Buechling, T., Boutros, M., Teleman, A.A., 2015.
1139 REPTOR and REPTOR-BP regulate organismal metabolism and transcription
1140 downstream of TORC1. *Developmental Cell* 33, 272–284.
1141 <https://doi.org/10.1016/j.devcel.2015.03.013>
- 1142 Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., Luscombe, N.M., 2009. A census of
1143 human transcription factors: function, expression and evolution. *Nature Reviews*
1144 *Genetics* 10, 252–263. <https://doi.org/10.1038/nrg2538>
- 1145 Vidal, N.M., Graziotin, A.L., Iyer, L.M., Aravind, L., Venancio, T.M., 2016. Transcription
1146 factors, chromatin proteins and the diversification of Hemiptera. *Insect Biochemistry*

1147 and Molecular Biology, Special issue: *Rhodnius prolixus* genome 69, 1–13.
1148 <https://doi.org/10.1016/j.ibmb.2015.07.001>

1149 Wang, K., Yang, Z., Li, X., Liu, S., Wang, L., Zhang, H., Yu, H., 2023. A hepatocyte nuclear
1150 factor BtabHNF4 mediates desiccation tolerance and fecundity in whitefly (*Bemisia*
1151 *tabaci*). *Environmental Entomology* 52, 138–147. <https://doi.org/10.1093/ee/nvac103>

1152 Wang, X., Pan, L., Wang, S., Zhou, J., McDowell, W., Park, J., Haug, J., Staehling, K., Tang,
1153 H., Xie, T., 2011. Histone H3K9 trimethylase eggless controls germline stem cell
1154 maintenance and differentiation. *PLOS Genetics* 7, e1002426.
1155 <https://doi.org/10.1371/journal.pgen.1002426>

1156 Weisbrod, A., Cohen, M., Chipman, A.D., 2013. Evolution of the insect terminal patterning
1157 system—Insights from the milkweed bug, *Oncopeltus fasciatus*. *Developmental*
1158 *Biology* 380, 125–131. <https://doi.org/10.1016/j.ydbio.2013.04.030>

1159 Wenger, Y., 2018. Pfam 18 IDs from <http://www.transcriptionfactor.org>.
1160 <https://doi.org/10.6084/m9.figshare.6610229.v1>

1161 Williams, J.A., Bell, J.B., Carroll, S.B., 1991. Control of *Drosophila* wing and haltere
1162 development by the nuclear vestigial gene product. *Genes Dev.* 5, 2481–2495.
1163 <https://doi.org/10.1101/gad.5.12b.2481>

1164 Wilson, D., Charoensawan, V., Kummerfeld, S.K., Teichmann, S.A., 2008. DBD—
1165 taxonomically broad transcription factor predictions: new content and functionality.
1166 *Nucleic Acids Research* 36, D88–D92. <https://doi.org/10.1093/nar/gkm964>

1167 Wu, S., Liu, Y., Zheng, Y., Dong, J., Pan, D., 2008. The TEAD/TEF family protein Scalloped
1168 mediates transcriptional output of the Hippo growth-regulatory pathway.
1169 *Developmental Cell* 14, 388–398. <https://doi.org/10.1016/j.devcel.2008.01.007>

1170 Yamanaka, N., Rewitz, K.F., O’Connor, M.B., 2013. Ecdysone Control of developmental
1171 transitions: lessons from *Drosophila* research. *Annual Review of Entomology* 58,
1172 497–516. <https://doi.org/10.1146/annurev-ento-120811-153608>

1173 Yang, H., Jaime, M., Polihronakis, M., Kanegawa, K., Markow, T., Kaneshiro, K., Oliver, B.,
1174 2018. Re-annotation of eight *Drosophila* genomes. *Life Science Alliance* 1.
1175 <https://doi.org/10.26508/lsa.201800156>

1176 Yao, J.-G., Weasner, B.M., Wang, L.-H., Jang, C.-C., Weasner, B., Tang, C.-Y., Salzer, C.L.,
1177 Chen, C.-H., Hay, B., Sun, Y.H., Kumar, J.P., 2008. Differential requirements for the
1178 Pax6(5a) genes eyegone and twin of eyegone during eye development in *Drosophila*.
1179 *Developmental Biology* 315, 535–551. <https://doi.org/10.1016/j.ydbio.2007.12.037>

1180 Yu, J., Wu, H., Wen, Y., Liu, Y., Zhou, T., Ni, B., Lin, Y., Dong, J., Zhou, Z., Hu, Z., Guo,
1181 X., Sha, J., Tong, C., 2015. Identification of seven genes essential for male fertility
1182 through a genome-wide association study of non-obstructive azoospermia and RNA
1183 interference-mediated large-scale functional screening in *Drosophila*. *Human*
1184 *Molecular Genetics* 24, 1493–1503. <https://doi.org/10.1093/hmg/ddu557>

1185 Zaret, K.S., 2020. Pioneer transcription factors initiating gene network changes. *Annual*
1186 *Review of Genetics* 54, 367–385. <https://doi.org/10.1146/annurev-genet-030220-015007>

1188 Zhang, R.-J., Chen, J., Jiang, L.-Y., Qiao, G.-X., 2019. The genes expression difference
1189 between winged and wingless bird cherry-oat aphid *Rhopalosiphum padi* based on
1190 transcriptomic data. *Sci Rep* 9, 4754. <https://doi.org/10.1038/s41598-019-41348-1>

1191 Zhang, T., Zhou, Q., Ogmundsdottir, M.H., Möller, K., Siddaway, R., Larue, L., Hsing, M.,
1192 Kong, S.W., Goding, C.R., Palsson, A., Steingrimsson, E., Pignoni, F., 2015. Mitf is a
1193 master regulator of the v-ATPase, forming a control module for cellular homeostasis
1194 with v-ATPase and TORC1. *J Cell Sci* 128, 2938–2950.
1195 <https://doi.org/10.1242/jcs.173807>

1196