

OPEN ACCESS

Citation: Li M, Kasan K, Saha Z, Yoon Y, Schmidt-Ott U (2023) Twenty-seven ZAD-ZNF genes of *Drosophila melanogaster* are orthologous to the embryo polarity determining mosquito gene *cucoid*. PLoS ONE 18(1): e0274716. https://doi.org/10.1371/journal.pone.0274716

Editor: René Massimiliano Marsano, University of Bari: Universita degli Studi di Bari Aldo Moro, ITALY

Received: September 1, 2022

Accepted: December 16, 2022

Published: January 3, 2023

Copyright: © 2023 Li et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: ML was the recipient of a fellowship from the Top Student Training Program of Peking University, China, for an International Summer Research Experience for Undergraduates (REU) program of the Biological Sciences Division at the University of Chicago. ZS was the recipient of a Research Foundations in Genetics and Genomics

RESEARCH ARTICLE

Twenty-seven ZAD-ZNF genes of *Drosophila melanogaster* are orthologous to the embryo polarity determining mosquito gene *cucoid*

Muzi Lio, Koray Kasan, Zinnia Saha, Yoseop Yoono*, Urs Schmidt-Ott*

Dept. of Organismal Biology and Anatomy, University of Chicago, Chicago, IL, United States of America

- Eurrent address: Department of Microbiology and Molecular Genetics, School of Medicine, University of California, Irvine, CA, United States of America
- * uschmid@uchicago.edu

Abstract

The C2H2 zinc finger gene *cucoid* establishes anterior-posterior (AP) polarity in the early embryo of culicine mosquitoes. This gene is unrelated to genes that establish embryo polarity in other fly species (Diptera), such as the homeobox gene *bicoid*, which serves this function in the traditional model organism *Drosophila melanogaster*. The *cucoid* gene is a conserved single copy gene across lower dipterans but nothing is known about its function in other species, and its evolution in higher dipterans, including Drosophila, is unresolved. We found that *cucoid* is a member of the ZAD-containing C2H2 zinc finger (ZAD-ZNF) gene family and is orthologous to 27 of the 91 members of this family in *D. melanogaster*, including *M1BP*, *ranshi*, *ouib*, *nom*, *zaf1*, *odj*, *Nnk*, *trem*, *Zif*, and eighteen uncharacterized genes. Available knowledge of the functions of *cucoid* orthologs in *Drosophila melanogaster* suggest that the progenitor of this lineage specific expansion may have played a role in regulating chromatin. We also describe many aspects of the gene duplication history of *cucoid* in the brachyceran lineage of *D. melanogaster*, thereby providing a framework for predicting potential redundancies among these genes in *D. melanogaster*.

1. Introduction

Dipteran insects (true flies) begin embryogenesis with 12 or 13 synchronous nuclear division cycles [1–4]. During this syncytial phase of embryonic development, a uniform blastoderm forms in the cortical layer of the egg, activates the zygotic genome [5–7], and establishes axial polarity [8,9]. Anterior determinants (ADs) establish the embryo's head-to-tail polarity via transcription factor gradients. In the fruit fly *Drosophila melanogaster* (*D. melanogaster*), the AD is encoded by the homeobox gene *bicoid* [10], which has been studied extensively [11–17]. However, genes unrelated to *bicoid* are being used in species of different dipteran lineages for the same developmental task [18,19]. This evolutionary plasticity, along with the simple anatomy of early dipteran embryos and their amenability to experimental perturbation in non-traditional model organisms, set the stage for an attractive experimental system to study the

summer fellowship of the Biological Sciences Collegiate Division at the University of Chicago. YY was the recipient of an award of University of Chicago Henry Hinds Funds for Graduate Student Research in Evolutionary Biology. Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R01GM127366. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

molecular and evolutionary basis of transcriptional network stability and co-option of new central players in embryo development.

What are the molecular mechanisms that guide the co-option of new ADs? Bicoid has many target genes [15,20-22], but it remains unclear how it adopted them. The bicoid gene evolved from a gene duplication of the *Hox3* ortholog of flies (also known as *zerknüllt* or *zen*), more than 145 million years ago [23–25]. The diverged DNA-binding specificity of Bicoid, compared to its closest paralogs, prompted detailed studies on the evolution of its DNA-binding homeodomain using ancestral sequence reconstruction, quantitative in vitro DNA binding assays, and in vivo rescue experiments in Drosophila embryos [26-29]. These studies emphasized the importance of mutations that altered DNA-binding specificity of the Bicoid protein. It was also shown that a feed-forward relay integrates certain regulatory activities of Bicoid and Orthodenticle via shared DNA binding sites [28]. These homeodomain proteins have qualitatively similar DNA affinity and Orthodenticle has a conserved zygotic function in head development, which raised the question of whether Bicoid took over functions of Orthodenticle [29]. However, comparative studies revealed ADs with distinct DNA binding domains and DNA affinities and suggest that the AD of the last common ancestor of dipterans was encoded by pangolin (Tcf) [18]. Therefore, the co-option of new ADs in different fly lineages may not require shared target sites between the old and new ADs.

Why do specific genes adopt the AD function in addition to their other roles? The identification of AD gene orthologs in *Drosophila melanogaster* provides a useful starting point because many of its gene functions have been analyzed. For example, *odd-paired*, the only *zic* (*zinc finger of the cerebellum*) gene family member of flies [30], opens specific chromatin regions to advance the temporal progression of zygotic pattern formation in Drosophila embryos [31,32]. This function appears to be conserved in moth flies, where *odd-paired* additionally adopted the AD function by acquiring a maternal transcription variant [18]. The ability of Odd-paired protein to drive the accessibility of specific chromatin regions, which is also a property of Bicoid [22,33], could have facilitated their convergent co-option as ADs.

In culicine mosquitoes (e.g., Aedes and Culex), a previously uncharacterized C2H2 zinc finger gene, named *cucoid*, adopted the AD function. In these species, three *cucoid* transcript isoforms with alternative 3' ends have been identified in embryos. The shortest isoform is expressed maternally and is localized at the anterior pole of the egg. In culicine mosquitoes, knockdown of *cucoid* by RNAi results in ectopic expression of posterior genes at the anterior and the double abdomen phenotype [18]. However, the function of *cucoid* orthologs in other species is unknown and obscured by a complex evolution of this gene in higher flies, including *Drosophila melanogaster* [18]. Here we show that *cucoid* is a member of the ZAD-ZNF gene family and is orthologous to at least 27 of *D. melanogaster*'s 91 ZAD-ZNF genes [34]. ZAD-ZNF gene family members encode C2H2 zinc finger proteins with an N-terminal Zincfinger-associated domain (ZAD) [34–39]. Most *cucoid* orthologs of *D. melanogaster* have not yet been characterized but those that have been named and studied predominantly function in early development and oogenesis and may affect chromatin states.

2. Materials and methods

2.1 Cucoid structure prediction and identification of *cucoid* orthologs

Protein structure was predicted using AlphaFold2_advanced with default settings [40]. Cucoid orthologs were identified by reciprocal protein BLAST using default E value cut-off threshold of 0.05 while setting maximum target sequences to 5000 to detect all potential orthologs in the target species (https://blast.ncbi.nlm.nih.gov/Blast.cgi) [41]. As queries we used Cucoid sequences from *Culex quinquefasciatus* (*C. quinquefasciatus*; GenBank identifier QFQ59547.1)

and *Aedes aegypti* (*A. aegypti*; GenBank identifier XP_021704552.1). The Cucoid sequence of *C. quinquefasciatus* is referred to as C-isoform in GenBank but corresponds to the non-truncated A-isoform in [18]. The Cucoid sequence of *A. aegypti* is referred to as *myoneurin* in GenBank and correspond to the non-truncated A-isoform in [18]. The name *myoneurin* for *cucoid* in *A. aegypti* appears to be a misnomer due to spurious similarity with non-orthologous *myoneurin* genes in other species. Therefore, *cucoid* and *myoneurin* are not synonyms and the name *myoneurin* should not be used to designate *cucoid* orthologs. Candidate orthologs were searched for conserved domains using NCBI's Conserved Domain Database (CDD) with the server's default E value cut-off of ≤ 0.01 [42]. C2H2 zinc finger proteins with ZAD (also known as zf-AD, smart00868, or pfam07776) were retained for reciprocal BLAST in *A. aegypti* and *C. quinquefasciatus*, using server default E value cut-offs of ≤ 0.05 .

Conservation of Cucoid clade genes of *Drosophila melanogaster* within the Brachycera was assessed by reciprocal protein BLAST in *Drosophila virilis*, *Lucilia cuprina*, *Bactrocera dorsalis*, and *Hermetia illucens*. Syntenies of Cucoid orthologs in these species were examined in Gen-Bank and illustrated using the IBS server [43]. Accession numbers are provided as supplementary material (S1 Table).

Since the assembly of the robber fly *Proctacanthus coquilletti* in GenBank (GenBank identifier GCA_001932985.1) is not annotated, we identified candidate exons of *P. coquilletti cucoid* orthologs (S2 Table), using tblastn with default settings [41] and *Hermetia illucens* (*H. illucens*) Cucoid orthologs and *D. melanogaster* CG9215, CG4424, and CG14711 as queries. Protein sequences of four candidate Cucoid orthologs from the robber fly were assembled manually and used for reciprocal protein BLAST in *H. illucens* and *D. melanogaster*.

2.2 Protein alignment and phylogenetic analysis

The list of D. melanogaster ZAD-ZNF genes has been reported elsewhere [34]. The respective protein sequences were downloaded from GenBank. MAFFT alignments were generated by MAFFT v7.471with the L-INS-i strategy (https://mafft.cbrc.jp/alignment/software/) [44]. Protein alignments were visualized using Geneious Prime 2021.2.2 (https://www.geneious.com/). For the protein tree with 91 ZAD-ZNF sequences, the raw alignments were trimmed using TrimAl v1.3 (http://trimal.cgenomics.org/) [45] with a conservation threshold of 20, a gap threshold of 0.8, and a similarity threshold of 0.05 to remove highly variable positions (columns). The trimmed alignments were further divided into two partitions corresponding to ZAD and ZNF regions. The best molecular substitution model for each partition was selected by partition merging strategy (MFP+-MERGE) using ModelFinder [46] implemented in IQ-TREE v2.1.3 [47], based on Bayesian Information Criterion (BIC). Maximum likelihood trees were then built based on the selected substitution models, with branch support values generated by the implemented ultrafast bootstrap approximation [48], setting replicates to 3000. A majority rule consensus tree was generated form bootstrap trees and visualized by FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/). Trees are unrooted unless otherwise stated. Accession numbers of all sequences used in protein trees (\$1 Table) and full-length alignments of D. melanogaster ZAD-ZNF proteins (\$1 File) and the Cucoid orthologs from D. melanogaster and D. virilis (S2 File) are provided as supporting information.

3. Results and discussion

3.1 Cucoid is a ZAD-ZNF protein with 27 orthologs in *Drosophila* melanogaster

Reciprocal protein BLAST identified single copy *cucoid* orthologs in all major branches of the lower (non-brachyceran) Diptera, including Tipulomorpha, Culicomorpha, Psychodomorpha

[18], and Bibionomorpha (this study), as well as in the insect orders Siphonaptera (fleas) and Lepidoptera (butterflies and moths) (S1 Fig). No *cucoid* orthologs were found in other insect orders, suggesting that *cucoid* evolved during the radiation of holometabolous insects.

The genome of *Drosophila melanogaster* encodes around 300 C2H2 zinc finger proteins [37,39], including multiple candidate orthologs of *cucoid*. To aid in the identification of *cucoid* orthologs in *D. melanogaster*, we searched for diagnostic domains and motifs of Cucoid using protein alignments and protein folding software [49] (**Figs 1 and S1**). The alignment was constructed with previously reported single-copy Cucoid orthologs from the mosquitoes *Culex quinquefasciatus*, *Aedes aegypti*, and *Anopheles gambiae* (Culicidae), the harlequin fly *Chironomus riparius* (Chironomidae), the moth fly *Clogmia albipunctata* (Psychodidae), and the crane fly *Nephrotoma suturalis* (Tipulidae) [18], as well as newly identified single-copy Cucoid orthologs from the gall midge *Contarinia nasturtii* (Cecidomyiidae), the cat flea *Ctenocephalides felis* (Siphonaptera), and the silk moth *Bombyx mori* (Lepidoptera) that we retrieved from sequences deposited in GenBank (**S1 Fig and S1 Table**). The gall midge belongs to the Bibionomorpha, the putative sister taxon of the Brachycera [50], while the cat flea and the silk moth represent close outgroups of the Diptera [51]. We focused on these lineage representatives because of the advanced state of genome resources for these species and because they yielded best matches in reciprocal protein BLAST with our query sequences.

Cucoid proteins typically contain five C2H2 zinc finger domains. However, the Cucoid ortholog of Chironomus lacks zinc fingers 4 and 5, and culicine mosquitoes also express shorter isoforms without the 5th (Culex) or 5th, 4th, and C-terminal half of the 3rd zinc finger domains (Aedes). We found that all these Cucoid orthologs also contain a conserved N-terminal domain, known as Zinc-finger-associated domain (ZAD; **Figs 1 and S1**) [35,38]. The ZAD

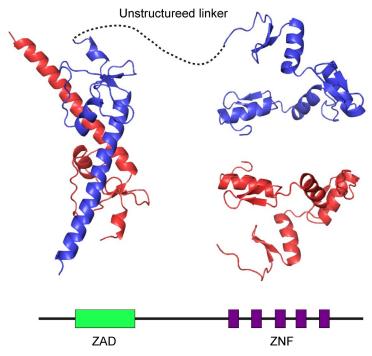


Fig 1. Predicted structure of Cucoid. Structure of a Cucoid homodimer (maternal isoform from *C. quinquefasciatus*, GenBank: QFQ59549.1) as predicted by AlphaFold2 (AlphaFold identifier for Cucoid structure: A0A5P8HWN4) is shown with the two chains colored in red and blue above a simplified sketch of full-length Cucoid protein with the N-terminal end to the left and the C-terminal end to the right, and ZAD and ZNFs marked by colored rectangles.

is stabilized by zinc coordination via four invariant cysteine residues and can drive dimerization [37,52] and nuclear localization [53].

Holometabolous insects evolved many ZAD-ZNF genes through lineage-specific gene duplications [35,36,39], especially in dipterans. For example, 147 ZAD-ZNF proteins have been found in *Anopheles gambiae* [36] and 91 in *Drosophila melanogaster* [34]. To identify the ZAD-ZNF proteins in *D. melanogaster* most similar to Cucoid, we conducted protein BLAST with all 91 ZAD-ZNF proteins of *D. melanogaster* in *A. aegypti* and *C. quinquefasciatus*. The same seventeen Drosophila sequences retrieved *cucoid* in Aedes and Culex (54.1% sequence conservation). The corresponding genes were therefore considered candidate orthologs of *cucoid* (Table 1).

Next, we generated a Maximum Likelihood protein tree with all 91 ZAD-ZNF proteins of *D. melanogaster* and examined the distribution of the candidate Cucoid orthologs on this protein tree (Fig 2). All seventeen candidate orthologs (marked by red triangles in Fig 2) mapped to a monophyletic clade of 27 ZAD-ZNF proteins (henceforth Cucoid clade). The Cucoid clade can be subdivided into two subclades with 9 members (subclade A) and 18 members (subclade B), respectively. All 9 members of subclade A were included in the original list of seventeen candidate Cucoid orthologs (Table 1). The top hit of that list, CG9215, also belongs to subclade A. The 18 members of subclade B experienced on average an elevated substitution

Table 1. D. melanogaster ZAD-ZNF genes in the Cucoid clade and their relationship with mosquito Cucoid in BLAST.

GenBank protein accession number	Gene	Chromosome	E value in Aedes aegypti	E value in Culex quinquefasciatus	%indentiy Aedes/Culex	Rank of Cucoid Aedes/Culex
NP_573045.1	CG9215	X	2.00E-50	3.00E-53	17.8/18.7	1st
NP_650094.1	CG14711	3R	1.00E-42	3.00E-45	20.2/19.7	1st
NP_650859.3	CG4424	3R	4.00E-41	5.00E-45	18.8/20.3	1st
NP_649825.1	M1BP	3R	7.00E-39	3.00E-41	19.3/18.1	1st
NP_649824.1	ranshi	3R	9.00E-38	9.00E-39	17.0/17.6	1st
NP_650860.1	CG4854	3R	9.00E-38	9.00E-38	17.4/17.8	1st
NP_650862.1	CG4936	3R	9.00E-37	3.00E-33	20.5/21.4	1st
NP_650861.1	trem	3R	7.00E-37	1.00E-36	19.8/19.6	1st
NP_649822.2	ouib	3R	2.00E-37	4.00E-39	15.7/16.0	1st
NP_001189188.1	Zif	3R	2.00E-34	2.00E-33	16.3/16.1	1st
NP_649823.2	CG8159	3R	3.00E-34	5.00E-33	12.5/12.7	1st
NP_001262384.1	nom	3R	1.00E-31	3.00E-30	15.2/15.3	1st
NP_001247055.1	Zaf1	3R	6.00E-33	2.00E-32	16.8/17.3	1st
NP_650092.4	CG14710	3R	1.00E-31	4.00E-34	18.9/19.7	1st
NP_731558.1	CG31441	3R	5.00E-31	1.00E-30	16.5/17.4	1st
NP_652712.2	CG18764	3R	2.00E-27	1.00E-27	18.0/17.8	1st
NP_651878.1	CG1792	3R	6.00E-27	2.00E-26	13.2/13.0	1st
NP_650658.1	CG17806	3R	2.00E-22	7.00E-21	14.9/14.0	2nd/19th
NP_650660.2	CG17801	3R	4.00E-21	1.00E-20	15.0/15.7	2nd/6th
NP_650060.1	CG31388	3R	4.00E-19	2.00E-18	13.7/15.0	13th/28th
NP_650659.2	Nnk	3R	7.00E-19	2.00E-17	14.6/15.6	>50th
NP_001163590.	CG6813	3R	1.00E-18	2.00E-18	13.2/13.8	>50th/22nd
NP_650051.1	CG6689	3R	2.00E-17	5.00E-18	10.7/10.8	>50th
NP_650657.2	CG17803	3R	3.00E-15	5.00E-15	10.4/11.5	>50th
NP_001097748.1	CG4820	3R	3.00E-13	3.00E-13	12.7/13.6	5th/4th
NP_650661.1	odj	3R	2.00E-11	8.00E-13	12.4/13.8	>50th
NP_001014607.1	CG14667	3R	8.00E-11	2.00E-12	12.0/13.7	>50th

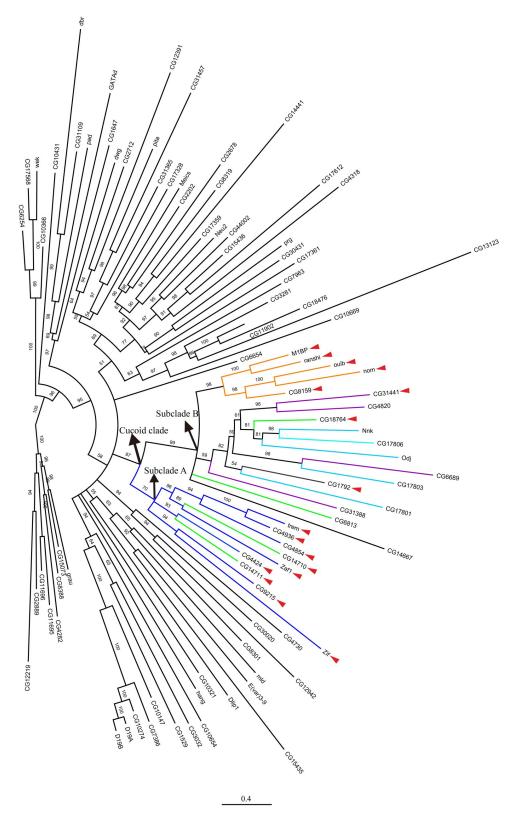


Fig 2. Manually rooted maximum likelihood protein tree of ZAD-ZNF family members in *D. melanogaster*. Orthologs that were identified based on reciprocal protein BLAST (see <u>Table 1</u>) are marked by red arrow heads. Colored lines correspond to chromosomal gene clusters (see <u>Fig 3</u>).

rate. This subclade includes 10 genes that we did not recover using reciprocal protein BLAST. These 10 genes do not form a monophyletic clade but are nested within the Cucoid clade and are therefore probably true Cucoid orthologs. We therefore conclude that at least 27 of *D. melanogaster*'s 91 ZAD-ZNF genes are orthologous to *cucoid*.

3.2 Genomic organization and relationship of genes of the Cucoid clade

All proteins of the Cucoid clade, except CG9215, are encoded by genes on the right arm of chromosome 3 and are organized in five gene complexes of 4–5 genes and three isolated singletons (Fig 3). These 26 genes share intron positions among each other and with the *cucoid* orthologs from lower dipterans (Fig 4), consistent with an evolutionary origin by DNA-based tandem gene duplications. *CG9215* is an intron-less gene on the X-chromosome that may have evolved by retro-transposition from the singleton *zif*, its most likely parent gene (Fig 2). We denoted each gene cluster in *D. melanogaster* by the member that gave the smallest E value in reciprocal BLAST with Cucoid (Table 1), that is: M1BP cluster (orange), CG4424 cluster (blue), CG14711 cluster (green), and CG31441 cluster (purple). One cluster (light blue) did not include any of the genes that we identified by reciprocal BLAST and was named after a previously described gene, *oddjob* (*odj*) [54].

The M1BP cluster genes form a monophyletic clade that can be traced to a single M1BP-like precursor gene, preserved in other schizophoran flies such as blow flies and tephritid fruit flies (see below). The first duplication of the M1BP-like precursor gave birth to M1BP/ranshi and nom/ouib/CG8159. The precursor of nom/ouib/CG8159 duplicated twice, first generating nom/ouib and CG8159 and then generating nom and ouib (Fig 2). These duplications occurred before the split of the D. virilis and D. melanogaster lineages [55]. M1PB/ranshi duplicated after the split of D. melanogaster and Drosophila ananassae [55].

The other *cucoid*-related gene clusters of *D. melanogaster* do not form monophyletic clades. These incongruences between our protein tree and clustering in the *D. melanogaster* genome

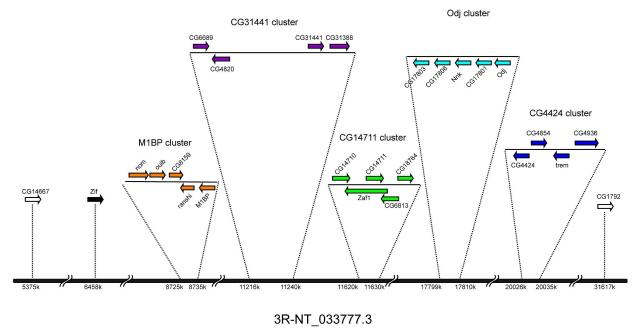


Fig 3. Synteny of genes in the Cucoid clade. Genes of the Cucoid clade are distributed in clusters on the right arm of chromosome 3, except *CG9215* (not shown) which is located on the X chromosome. M1BP cluster (orange), CG14711 cluster (green), CG4424 cluster (blue), Odj cluster (light blue), CG31441 cluster (purple), *Zif* (black), other dispersed genes (outlined).

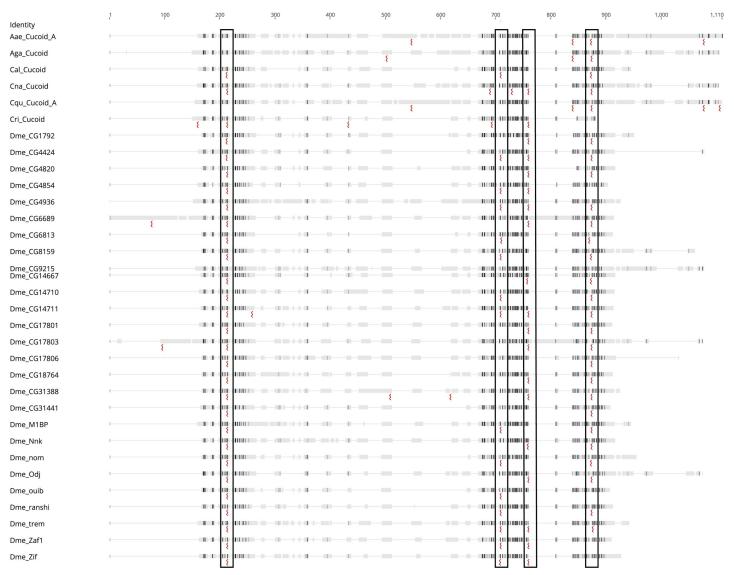


Fig 4. Conservation of Cucoid introns. Multiple sequence alignment of Cucoid orthologs from lower dipterans and Cucoid clade members of *D. melanogaster*. Conserved intron positions are boxed, and the red zigzags represent the splicing points. Similarity of aligned amino acids was assessed using the Blosum62 matrix with black representing 100% similarity, dark grey 80–100% similarity, light grey 60–80% similarity, and white less than 60% similarity.

https://doi.org/10.1371/journal.pone.0274716.g004

could have resulted from limitations of the phylogenetic inference methods that we used to build the protein tree, such as model choice and long-branch attraction [56], or non-allelic recombination (gene conversion) within the ZAD-ZNF family [57,58]. However, in *D. virilis*, genes related to members of the CG14711 cluster (green), the CG31441 cluster (purple), and the Odj cluster (light blue) form a single gene complex with a different gene order, and this gene order is consistent with the inferred close relationship of neighboring Cucoid clade genes (Fig 5A). Based on synteny in *D. virilis* and phylogenetic inference with *D. virilis* and *D. melanogaster* orthologs (Fig 5B), CG18764 of the CG14711 cluster (green) is paralogous to all genes of the Odj cluster (light blue) as well as CG6689 of the CG31441 cluster (purple). Additionally, we infer that CG6689 is a paralog of CG17803 of the Odj cluster, even though a gene-specific N-terminal THAP (Thanatos Associated Proteins) domain [59–61] that CG6689 inherited from its precursor, CG6689/CG17803, is not preserved in CG17803 (S2 Fig). Finally, we infer that *D. virilis* lost the Nnk/

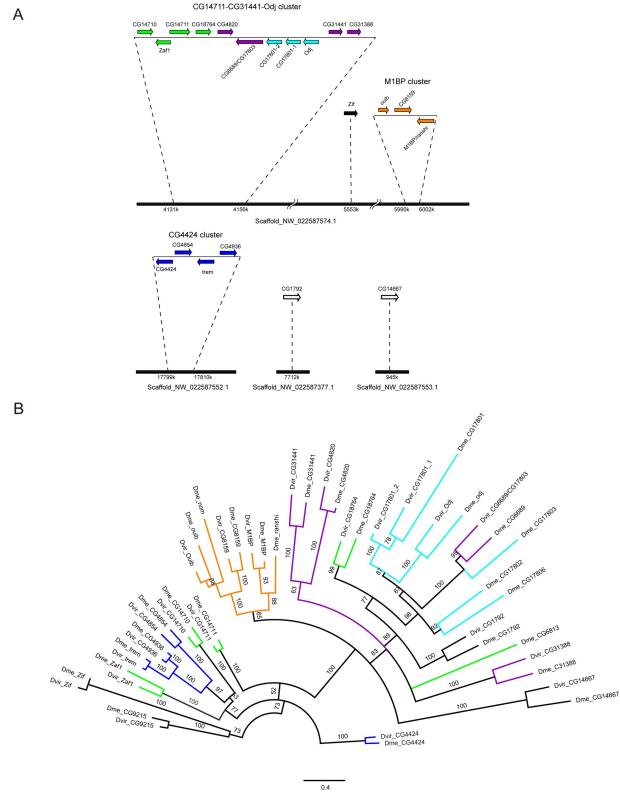


Fig 5. Cucoid clade protein tree is consistent with gene synteny in *D. virilis*. (A) Synteny of *cucoid* orthologs in *D. virilis*. Genes are color coded to indicate their relationship to gene clusters in *D. melanogaster* (see **Fig 3**). (B) Manually rooted maximum likelihood protein tree of Cucoid orthologs from *D. melanogaster* and *D. virilis*. Note that *D. virilis* has two CG17801 orthologs.

CG17806 precursor because a CG17806-like precursor gene existed before the split of *D. melanogaster* and *D. virilis* but was not found in *D. virilis*. Duplication of the Nnk/CG17806 precursor occurred after the split of the *D. willistoni* lineage from *D. melanogaster* lineage [55].

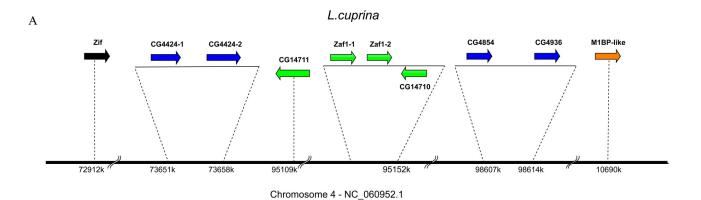
3.3 The Cucoid clade in Drosophila outgroups

Lineage-specific gene family expansions may reflect innovations or adaptations [38], but it is unknown why the number of ZAD-ZNF genes independently increased so much in multiple lineages of the Holometabola. To better understand when and how the Cucoid clade expanded, we searched for orthologs of the Cucoid clade members in representatives of other schizophoran fly species, including a blow fly (*Lucilia cuprina*) and a tephritid fruit fly (*Bactrocera dorsalis*), and in two representatives of the lower Brachycera, including a soldier fly (*Hermetia illucens*) and a robber fly (*Proctacanthus coquilleti*). While non-brachyceran dipterans have single *cucoid* orthologs (see section 3.1), we identified multiple *cucoid* orthologs in all the brachyceran species, albeit in lower numbers than in Drosophila.

In Lucilia and Bactrocera, we identified a single *M1BP*-like gene (orange), and several genes related to the CG4424 cluster (dark blue) and the CG14711 cluster (green), respectively, as well as putative orthologs of *zif* and *CG9215* (Fig 6A and 6B). The presence of a *M1BP*-like gene in these unrelated species (they represent paraphyletic/parallel lineages of the Schizophora [62]) suggests that the M1BP cluster expanded during, rather than before the radiation of the Schizophora in the Tertiary epoch [63]. Whether the expansion of the M1BP cluster within the Schizophora resulted in subfunctionalization or the acquisition of new gene functions or a mix of both remains unknown, due to the lack of functional comparisons of the M1BP-like gene in lower Schizophora with their multiple orthologs in Drosophila.

Two additional features of the genomic organization of Cucoid clade genes in *Bactrocera dorsalis* deserve attention. First, the *M1BP*-like gene of this species is in the immediate vicinity of the CG14711 cluster (green, Fig 6A). This finding may suggest that the founder gene of the M1BP cluster originated as an offshoot of the CG14711 cluster, even though this is not apparent in the protein tree (Fig 2). Second, *CG14711* and *CG14710* of *B. dorsalis* have merged; the predicted protein has two linear ZAD-ZNFs structures that correspond to CG14710 and CG14711, respectively. Whether these genes resulted from the same duplication is unclear. Phylogenetic analysis suggests that *CG14711* is more closely related to *CG4424* than to *CG14710*. However, since CG14711 and CG4424 are more similar to Cucoid than CG14710 and other members of the CG14711 and CG4424 clusters, the inferred close relationship of CG14711 and CG4424 might reflect their less diverged status rather than their duplication history.

In the genomes of lower Brachycera, we identified five *cucoid* loci on chromosome 6 of the soldier fly *Hermetia illucens* (*Hil_cucoid_1-5*, GenBank accession numbers: XP_037925088.1, XP_037925165.1, XP_037924715.1, CAD7093451.1, XP_037922166.1) [64] (Fig 6C) and four *cucoid* loci in the robber fly *Proctacanthus coquilleti* (*Pco_cucoid_1-4*) [65] (S2 Table), which seem to be orthologous to Hermetia *cucoid* orthologs 1, 2, 3, and 4/5, respectively (S3 Fig). The lower brachyceran Cucoid proteins 1 and 2 are closely related to CG9215 judged by protein BLAST E value but retain conserved introns and are therefore potentially orthologous to *CG9215/Zif*, whereas the lower brachyceran Cucoid orthologs 3 and 4 are closely related to CG4424 and CG14711, respectively. Therefore, the last common ancestor of soldier flies, robber flies, and Drosophila may have had at least three *cucoid* orthologs, including a *CG4424*-like member, *CG14711*-like member, and a CG9215-like ortholog of the Zif/CG9215 precursor. No *Hil_5* ortholog was found in *D. melanogaster* and the robber fly *Proctacanthus coquilletti*. Its location within an intron of *Hil_4* suggests that it was born by lineage-specific duplication of *Hil_4*. Thus, *Hil_5* may also be orthologous to *CG14711*.



B B.dorsalis

M1BP-like Zaf1

CG4424

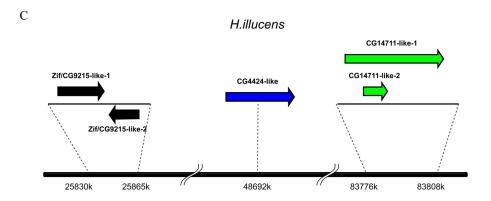
Zif

CG44710_14711 complex

CG4936/trem CG4854

2808k 2816k 495k

Scaffold_NW_011876392.1 Scaffold_NW_011876251.1



Chromosome 6 - NC_051854.1

Fig 6. Syntenies of *cucoid* orthologs in Lucilia, Bactrocera, and Hermetia. Genes are color coded to indicate their relationship to gene clusters in *D. melanogaster*. (A) Synteny of *cucoid* orthologs in *Lucilia cuprina*. (B) Synteny of *cucoid* orthologs in *Bactrocera dorsalis*. (C) Synteny of *cucoid* orthologs in *Hermetia illucens*.

https://doi.org/10.1371/journal.pone.0274716.g006

4. Conclusions

D. melanogaster contains at least 27 *cucoid* orthologs, that is, almost one third of the 91 ZAD ZNF genes of this species. Reciprocal BLAST, phylogenetic inference, and genomic

organization suggest that the Cucoid clade of *D. melanogaster* expanded gradually in the brachyceran lineage (Fig 7), while its founder gene was already present in the last common ancestor of butterflies, fleas, and flies. The last common ancestor of the brachyceran species that we analyzed may have had three *cucoid* orthologs that encoded proteins similar to CG9215/zif, CG4424, and CG14711. We infer this because in protein BLAST against Drosophila proteins, *H. illucens* and *P. coquilletti* Cucoid proteins 1 and 2 recovered CG9215 as the best hit, and their Cucoid orthologs 3 and 4 recovered CG4424 and CG14711 as the best hits, respectively. The founder of the monophyletic M1BP-cluster originated before the radiation of the Schizophora. All other clusters of *D. melan*ogaster may not have monophyletic origins.

Our study was motivated by the question of what is known about *cucoid* orthologs in *Drosophila melanogaster*. Most of the 27 *cucoid* orthologs of *D. melanogaster* that we identified in this study did not affect viability when downregulated in previous large-scale screens (Table 2) [34,66–68]. However, several orthologs have been characterized in greater depth and display diverse, essential functions. For example, M1BP binds core promoters of thousands of genes and functions during transcription activation and polymerase II pausing while promoting chromatin accessibility surrounding the transcription start sites [69,70]. Other genes in the M1BP cluster show more specialized functions: *ranshi* regulates oocyte differentiation [71], *nom* functions in muscle development, and *ouib* is necessary during ecdysteroid synthesis by regulating *spookier* [72,73]. The closely related genes *odj* and *Nnk*

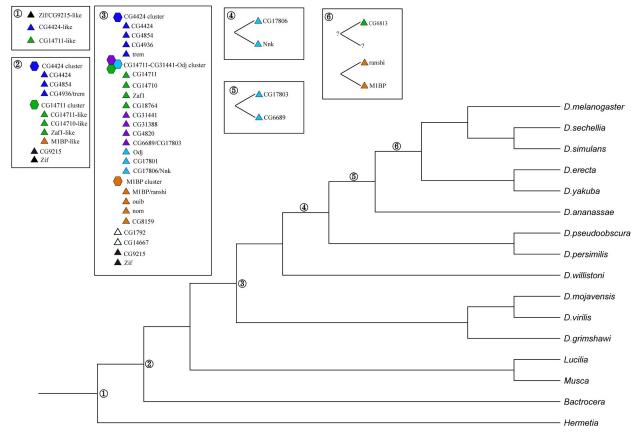


Fig 7. Evolution of the Cucoid clade family in higher flies. Inferred gene duplications in the Cucoid clade based on data in this study and *D. melanogaster* gene ages reported elsewhere [55]. For details see text. Gene clusters (hexagons) and gene loci (triangles) are indicated. For color code see Fig 2.

Table 2. Functions of D. melanogaster cucoid orthologs.

Gene	Functions	Loss-of-function	References	
M1BP	activates transcription; orchestrates RNA polymerase II pausing	viable	[69,70]	
Zaf1	architectural protein, chromosome insulator	viable	[66-68,74,78]	
Nnk	heterochromatin binding	lethal	[34,66,79]	
odj	heterochromatin binding	lethal	[34,80]	
ranshi	involves in oocyte differentiation	sterile	[71]	
ouib	regulates ecdysteroid synthesis; stimulates transcription of spookier	lethal	[72,73]	
nom	regulates embryonic muscle morphogenesis	viable	[66-68,81]	
Zif	regulates neuroblast differentiation with aPKC	lethal	[66,67,76,77]	
trem	required for DSB formation in meiosis	sterile	[75]	
CG14667	uncharacterized	viable	[68]	
CG14710	uncharacterized	viable	[34]	
CG14711	uncharacterized	unknown		
CG17801	uncharacterized	viable	[34,68]	
CG17803	uncharacterized	viable	[34,66-68]	
CG17806	uncharacterized	viable	[34,67]	
CG1792	uncharacterized	viable	[34,67,68]	
CG18764	uncharacterized	viable	[66-68]	
CG31388	uncharacterized	viable	[66,67]	
CG31441	uncharacterized	lethal	[66]	
CG4424	uncharacterized	viable	[66-68]	
CG4820	uncharacterized	viable	[66-68]	
CG4854	uncharacterized	viable	[66-68]	
CG4936	uncharacterized	lethal	[66,67,82]	
CG6689	uncharacterized	lethal	[66]	
CG6813	uncharacterized	viable	[34]	
CG8159	uncharacterized	viable	[66]	
CG9215	uncharacterized	viable	[66-68]	

https://doi.org/10.1371/journal.pone.0274716.t002

have essential functions in heterochromatin regulation [34], *zaf1* is a chromosome architecture protein that serves as insulator in *Drosophila melanogaster* [74], *trem* is required for binding Mei-P22 on meiotic chromosomes to initiate double strand breaks for homologous recombination [75], and *Zif* is required for the expression and asymmetric localization of aPKC in neuroblast cells to regulate their polarity and self-renewal [76,77]. All other genes of the Cucoid clade remain uncharacterized. Taken together, our study suggests that many *cucoid* orthologs of *D. melanogaster* function in oogenesis and embryogenesis and several of them modify chromatin states. It will be interesting to find out whether single copy Cucoid orthologs from lower dipterans function in similar ways to some *D. melanogaster* orthologs and what structural and/or regulatory features enable Cucoid in culicine mosquitoes to regulate early zygotic segmentation genes.

Supporting information

S1 Table. Accession numbers. (XLSX)

S2 Table. Locations of *cucoid* orthologs in *P. coquiletti* inferred from tblastn. (XLSX)

S1 File. Full length alignment of 91 D. *melanogaster* ZAD-ZNF protiens. (AFA)

S2 File. Full length alignment of *D. melanogaster* and *D. virilis* Cucoid orthologs. (AFA)

S1 Fig. Cucoid protein alignment and prediction of Cucoid homodimer. (A) Multiple sequence alignment of Cucoid orthologs from *Aedes aegypti* (Aae), *Anopheles gambiae* (Aga), *Bombyx mori* (Bmo), *Chironomus riparius* (Cri), *Clogmia albipunctata* (Cal), *Contarinia nasturtii* (Cna), *Ctenocephalides felis* (Cfe), *Culex quinquefasciatus* (Cqu), and *Nephrotoma suturalis* (Nsu). (B) A plot of the predicted alignment error of the best model acquired from AlphaFold2 output which estimates the distance error for every pair of residues. Both axes represent the positions on the dimer of Cucoid maternal isoform (499 aa) from *C. quinquefasciatus*. The color key is measured in angstrom. Very low position errors are found for the overlapping of residues in the ZAD dimer as well as between zinc fingers on the same strand, indicating true packing of these domains. (C) The plot of predicted local distance difference test (pLDDT) per position gives a confidence level between 0–100 for each residue. All models predict ZAD and ZNF domain with very high confidence, whereas the highly variable linker regions get deficient support. (TIF)

S2 Fig. CG6689 acquired a DNA-binding THAP domain. The THAP domains from 9 THAP-containing proteins in *D. melanogaster* and a THAP-like fragment from CG17803 are shown in alignment here. The color code for each column is based on similarity of aligned amino acids, with black representing high similarity and white representing no similarity. The N-terminal THAP domain of CG6689 is absent in all other Cucoid orthologs including its most recent paralog, CG17803, which has incomplete THAP features. THAP is a zinc-coordinating DNA binding domain with a conserved C2CH structure and shares features with the DNA binding domain of the P element transposase [59,60]. THAP-domain-containing proteins have been found in human, *D. melanogaster*, and *C. elegans* [59]. In the nine *D. melanogaster* proteins that have this domain, only CG6689 and CG10431 belong to the ZAD-ZNF family. CG10431 is only distantly related to CG6689 and located on a different chromosome (2L), suggesting that even within the ZAD-ZNF family THAP domains evolved *de novo*. The THAP domain of CG6689 is encoded by the first two exons of this gene, which are only conserved in *CG17803* (Fig 4). (PDF)

S3 Fig. Phylogeny of Cucoid orthologs in *H. illucnes* **and** *P. coquilletti***.** A phylogenetic tree with Cucoid orthologs in *H. illucnes*, *P. coquilletti* and lower flies was constructed based on an untrimmed alignment using 3 partitions inlucing ZAD, ZNF, and the other regions. Regions outside the ZAD and ZNF domains include diagnostic features useful for inferring orthology. This tree suggests that *Hil_cucoid_1* to *Hil_cucoid_4* are orthologous to *Pco_cucoid_1* to *Pco_cucoid_4*, respectively. (TIF)

Acknowledgments

We thank our colleagues Dr. Phoebe Rice and Dr. Manyuan Long for helpful discussions, and Dr. Shengqian Xia and Dylan Sosa in the Long lab for technical assistance.

Author Contributions

Conceptualization: Muzi Li, Koray Kasan, Urs Schmidt-Ott.

Formal analysis: Muzi Li, Koray Kasan, Zinnia Saha.

Funding acquisition: Muzi Li, Zinnia Saha, Yoseop Yoon, Urs Schmidt-Ott.

Investigation: Muzi Li, Koray Kasan, Zinnia Saha, Yoseop Yoon, Urs Schmidt-Ott.

Supervision: Urs Schmidt-Ott.

Visualization: Muzi Li.

Writing - original draft: Muzi Li, Koray Kasan, Urs Schmidt-Ott.

Writing - review & editing: Muzi Li, Koray Kasan, Yoseop Yoon, Urs Schmidt-Ott.

References

- Foe VE, Odell GM, Edgar BA. Mitosis and morphogenesis in the Drosophila embryo. In: Bate M, Martinez Arias A, editors. The development of *Drosophila melanogaster*. 1: Cold Spring Harbor Laboratory Press; 1993. p. 149–300.
- Jimenez-Guri E, Wotton KR, Gavilan B, Jaeger J. A staging scheme for the development of the moth midge Clogmia albipunctata. PLoS One. 2014; 9(1):e84422. https://doi.org/10.1371/journal.pone.0084422 PMID: 24409296
- Wotton KR, Jimenez-Guri E, Garcia Matheu B, Jaeger J. A staging scheme for the development of the scuttle fly Megaselia abdita. PLoS One. 2014; 9(1):e84421. https://doi.org/10.1371/journal.pone.0084421 PMID: 24409295
- Lemke S, Kale G, Urbansky S. Comparing gastrulation in flies: Links between cell biology and the evolution of embryonic morphogenesis. Mechanisms of Development. 2020;164.
- Schulz KN, Harrison MM. Mechanisms regulating zygotic genome activation. Nat Rev Genet. 2019; 20 (4):221–34. https://doi.org/10.1038/s41576-018-0087-x PMID: 30573849
- Harrison MM, Eisen MB. Transcriptional Activation of the Zygotic Genome in Drosophila. Curr Top Dev Biol. 2015; 113:85–112. https://doi.org/10.1016/bs.ctdb.2015.07.028 PMID: 26358871
- Larson ED, Marsh AJ, Harrison MM. Pioneering the developmental frontier. Mol Cell. 2021; 81 (8):1640–50. https://doi.org/10.1016/j.molcel.2021.02.020 PMID: 33689750
- 8. Ma J, He F, Xie G, Deng WM. Maternal AP determinants in the Drosophila oocyte and embryo. Wiley Interdiscip Rev Dev Biol. 2016; 5(5):562–81. https://doi.org/10.1002/wdev.235 PMID: 27253156
- Schmidt-Ott U, Yoon Y. Evolution and loss of beta-catenin and TCF-dependent axis specification in insects. Curr Opin Insect Sci. 2022; 50:100877.
- 10. Berleth T, Burri M, Thoma G, Bopp D, Richstein S, Frigerio G, et al. The role of localization of bicoid RNA in organizing the anterior pattern of the Drosophila embryo. Embo j. 1988; 7(6):1749–56. https://doi.org/10.1002/j.1460-2075.1988.tb03004.x PMID: 2901954
- Porcher A, Dostatni N. The bicoid morphogen system. Curr Biol. 2010; 20(5):R249–54. https://doi.org/10.1016/j.cub.2010.01.026 PMID: 20219179
- Ali-Murthy Z, Kornberg TB. Bicoid gradient formation and function in the Drosophila pre-syncytial blastoderm. eLife. 2016; 5:e13222. https://doi.org/10.7554/eLife.13222 PMID: 26883601
- Struhl G, Struhl K, Macdonald PM. The gradient morphogen bicoid is a concentration-dependent transcriptional activator. Cell. 1989; 57(7):1259–73. https://doi.org/10.1016/0092-8674(89)90062-7 PMID: 2567637
- 14. Rivera-Pomar R, Niessing D, Schmidt-Ott U, Gehring WJ, Jäckle H. RNA binding and translational suppression by bicoid. Nature. 1996; 379(6567):746–9. https://doi.org/10.1038/379746a0 PMID: 8602224
- Schroeder MD, Pearce M, Fak J, Fan H, Unnerstall U, Emberly E, et al. Transcriptional Control in the Segmentation Gene Network of Drosophila. PLOS Biology. 2004; 2(9):e271. https://doi.org/10.1371/ journal.pbio.0020271 PMID: 15340490
- Huang A, Amourda C, Zhang S, Tolwinski NS, Saunders TE. Decoding temporal interpretation of the morphogen Bicoid in the early Drosophila embryo. eLife. 2017; 6:e26258. https://doi.org/10.7554/eLife. 26258 PMID: 28691901
- Hannon CE, Blythe SA, Wieschaus EF. Concentration dependent chromatin states induced by the bicoid morphogen gradient. eLife. 2017; 6:e28275. https://doi.org/10.7554/eLife.28275 PMID: 28891464

- Yoon Y, Klomp J, Martin-Martin I, Criscione F, Calvo E, Ribeiro J, et al. Embryo polarity in moth flies and mosquitoes relies on distinct old genes with localized transcript isoforms. eLife. 2019; 8:e46711. https://doi.org/10.7554/eLife.46711 PMID: 31591963
- Klomp J, Athy D, Kwan CW, Bloch NI, Sandmann T, Lemke S, et al. Embryo development. A cysteineclamp gene drives embryo polarity in the midge Chironomus. Science. 2015; 348(6238):1040–2. https://doi.org/10.1126/science.aaa7105 PMID: 25953821
- Chen H, Xu Z, Mei C, Yu D, Small S. A system of repressor gradients spatially organizes the boundaries of Bicoid-dependent target genes. Cell. 2012; 149(3):618–29. https://doi.org/10.1016/j.cell.2012.03.018 PMID: 22541432
- Ochoa-Espinosa A, Yucel G, Kaplan L, Pare A, Pura N, Oberstein A, et al. The role of binding site cluster strength in Bicoid-dependent patterning in Drosophila. Proc Natl Acad Sci U S A. 2005; 102 (14):4960–5. https://doi.org/10.1073/pnas.0500373102 PMID: 15793007
- Hannon CE, Blythe SA, Wieschaus EF. Concentration dependent chromatin states induced by the bicoid morphogen gradient. Elife. 2017;6. https://doi.org/10.7554/eLife.28275 PMID: 28891464
- Stauber M, Jäckle H, Schmidt-Ott U. The anterior determinant bicoid of Drosophila is a derived Hox class 3 gene. Proc Natl Acad Sci U S A. 1999; 96(7):3786–9. https://doi.org/10.1073/pnas.96.7.3786
 PMID: 10097115
- Stauber M, Taubert H, Schmidt-Ott U. Function of bicoid and hunchback homologs in the basal cyclorrhaphan fly Megaselia (Phoridae). Proceedings of the National Academy of Sciences. 2000; 97 (20):10844–9. https://doi.org/10.1073/pnas.190095397 PMID: 10995461
- Stauber M, Prell A, Schmidt-Ott U. A single Hox3 gene with composite bicoid and zerknullt expression characteristics in non-Cyclorrhaphan flies. Proc Natl Acad Sci U S A. 2002; 99(1):274–9. https://doi.org/ 10.1073/pnas.012292899 PMID: 11773616
- Liu Q, Onal P, Datta RR, Rogers JM, Schmidt-Ott U, Bulyk ML, et al. Ancient mechanisms for the evolution of the bicoid homeodomain's function in fly development. Elife. 2018;7. https://doi.org/10.7554/eLife.34594 PMID: 30298815
- Onal P, Gunasinghe HI, Umezawa KY, Zheng M, Ling J, Azeez L, et al. Suboptimal Intermediates Underlie Evolution of the Bicoid Homeodomain. Mol Biol Evol. 2021; 38(6):2179–90. https://doi.org/10.1093/molbev/msab051 PMID: 33599280
- Datta RR, Ling J, Kurland J, Ren X, Xu Z, Yucel G, et al. A feed-forward relay integrates the regulatory activities of Bicoid and Orthodenticle via sequential binding to suboptimal sites. Genes Dev. 2018; 32 (9–10):723–36. https://doi.org/10.1101/gad.311985.118 PMID: 29764918
- Lynch J, Desplan C. Evolution of development: beyond bicoid. Curr Biol. 2003; 13(14):R557–9. https://doi.org/10.1016/s0960-9822(03)00472-x PMID: 12867048
- Houtmeyers R, Souopgui J, Tejpar S, Arkell R. The ZIC gene family encodes multi-functional proteins essential for patterning and morphogenesis. Cellular and molecular life sciences: CMLS. 2013; 70 (20):3791–811. https://doi.org/10.1007/s00018-013-1285-5 PMID: 23443491
- Soluri IV, Zumerling LM, Payan Parra OA, Clark EG, Blythe SA. Zygotic pioneer factor activity of Oddpaired/Zic is necessary for late function of the Drosophila segmentation network. Elife. 2020;9. https:// doi.org/10.7554/eLife.53916 PMID: 32347792
- **32.** Koromila T, Gao F, Iwasaki Y, He P, Pachter L, Gergen JP, et al. Odd-paired is a pioneer-like factor that coordinates with Zelda to control gene expression in embryos. Elife. 2020;9. https://doi.org/10.7554/elife.59610 PMID: 32701060
- Haines JE, Eisen MB. Patterns of chromatin accessibility along the anterior-posterior axis in the early Drosophila embryo. PLoS Genet. 2018; 14(5):e1007367. https://doi.org/10.1371/journal.pgen.1007367 PMID: 29727464
- Kasinathan B, Colmenares SU 3rd, McConnell H, Young JM, Karpen GH, Malik HS. Innovation of heterochromatin functions drives rapid evolution of essential ZAD-ZNF genes in Drosophila. eLife. 2020; 9: e63368. https://doi.org/10.7554/eLife.63368 PMID: 33169670
- Chung H-R, Schäfer U, Jäckle H, Böhm S. Genomic expansion and clustering of ZAD-containing C2H2 zinc-finger genes in Drosophila. EMBO Rep. 2002; 3(12):1158–62. https://doi.org/10.1093/embo-reports/kvf243 PMID: 12446571
- Chung H-R, Löhr U, Jäckle H. Lineage-specific expansion of the Zinc Finger Associated Domain ZAD. Molecular Biology and Evolution. 2007; 24(9):1934–43. https://doi.org/10.1093/molbev/msm121 PMID: 17569752
- Jauch R, Bourenkov GP, Chung H-R, Urlaub H, Reidt U, Jäckle H, et al. The Zinc Finger-Associated Domain of the Drosophila Transcription Factor Grauzone Is a Novel Zinc-Coordinating Protein-Protein Interaction Module. Structure. 2003; 11(11):1393–402. https://doi.org/10.1016/j.str.2003.09.015 PMID: 14604529

- Lespinet O, Wolf YI, Koonin EV, Aravind L. The role of lineage-specific gene family expansion in the evolution of eukaryotes. Genome research. 2002; 12(7):1048–59. https://doi.org/10.1101/gr.174302 PMID: 12097341
- Krystel J, Ayyanathan K. Global analysis of target genes of 21 members of the ZAD transcription factor family in Drosophila melanogaster. Gene. 2013; 512(2):373–82. https://doi.org/10.1016/j.gene.2012. 09.114 PMID: 23085320
- Mirdita M, Schutze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. Nat Methods. 2022; 19(6):679–82. https://doi.org/10.1038/s41592-022-01488-1 PMID: 35637307
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of Molecular Biology. 1990; 215(3):403–10. https://doi.org/10.1016/S0022-2836(05)80360-2 PMID: 2231712
- Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, et al. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Research. 2011; 39 (suppl_1):D225–D9. https://doi.org/10.1093/nar/gkq1189 PMID: 21109532
- Liu W, Xie Y, Ma J, Luo X, Nie P, Zuo Z, et al. IBS: an illustrator for the presentation and visualization of biological sequences. Bioinformatics. 2015; 31(20):3359–61. https://doi.org/10.1093/bioinformatics/ btv362 PMID: 26069263
- Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution. 2013; 30(4):772–80. https://doi.org/10.1093/molbev/mst010 PMID: 23329690
- 45. Sánchez R, Serra F, Tárraga J, Medina I, Carbonell J, Pulido L, et al. Phylemon 2.0: a suite of web-tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing. Nucleic Acids Res. 2011; 39(Web Server issue):W470–4. https://doi.org/10.1093/nar/gkr408 PMID: 21646336
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nature Methods. 2017; 14(6):587–9. https://doi.org/10.1038/nmeth.4285 PMID: 28481363
- 47. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Molecular Biology and Evolution. 2020; 37(5):1530–4. https://doi.org/10.1093/molbev/msaa015 PMID: 32011700
- **48.** Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Molecular Biology and Evolution. 2017; 35(2):518–22.
- 49. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021; 596(7873):583–9. https://doi.org/10.1038/s41586-021-03819-2 PMID: 34265844
- Wiegmann BM, Trautwein MD, Winkler IS, Barr NB, Kim JW, Lambkin C, et al. Episodic radiations in the fly tree of life. Proc Natl Acad Sci U S A. 2011; 108(14):5690–5. https://doi.org/10.1073/pnas.1012675108 PMID: 21402926
- Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, et al. Phylogenomics resolves the timing and pattern of insect evolution. Science. 2014; 346(6210):763–7. https://doi.org/10.1126/science. 1257570 PMID: 25378627
- Zolotarev N, Fedotova A, Kyrchanova O, Bonchuk A, Penin AA, Lando AS, et al. Architectural proteins Pita, Zw5,and ZIPIC contain homodimerization domain and support specific long-range interactions in Drosophila. Nucleic Acids Research. 2016; 44:7228–41. https://doi.org/10.1093/nar/gkw371 PMID: 27137890
- 53. Zolotarev NA, Maksimenko OG, Georgiev PG, Bonchuk AN. ZAD-Domain Is Essential for Nuclear Localization of Insulator Proteins in Drosophila melanogaster. Acta Naturae. 2016; 8(3):97–102. PMID: 27795848
- 54. Swenson JM, Colmenares SU, Strom AR, Costes SV, Karpen GH. The composition and organization of Drosophila heterochromatin are heterogeneous and dynamic. eLife. 2016; 5:e16096. https://doi.org/10.7554/eLife.16096 PMID: 27514026
- Zhang YE, Vibranovski MD, Krinsky BH, Long M. Age-dependent chromosomal distribution of malebiased genes in Drosophila. Genome research. 2010; 20(11):1526–33. https://doi.org/10.1101/gr. 107334.110 PMID: 20798392
- 56. Kapli P, Yang Z, Telford MJ. Phylogenetic tree building in the genomic age. Nature Reviews Genetics. 2020; 21(7):428–44. https://doi.org/10.1038/s41576-020-0233-0 PMID: 32424311
- 57. Hahn MW. Distinguishing Among Evolutionary Models for the Maintenance of Gene Duplicates. Journal of Heredity. 2009; 100(5):605–17. https://doi.org/10.1093/jhered/esp047 PMID: 19596713

- Magadum S, Banerjee U, Murugan P, Gangapur D, Ravikesavan R. Gene duplication as a major force in evolution. J Genet. 2013; 92(1):155–61. https://doi.org/10.1007/s12041-013-0212-8 PMID: 23640422
- 59. Roussigne M, Kossida S, Lavigne AC, Clouaire T, Ecochard V, Glories A, et al. The THAP domain: a novel protein motif with similarity to the DNA-binding domain of P element transposase. Trends Biochem Sci. 2003; 28(2):66–9. https://doi.org/10.1016/S0968-0004(02)00013-0 PMID: 12575992
- 60. Clouaire T, Roussigne M, Ecochard V, Mathe C, Amalric F, Girard J-P. The THAP domain of THAP1 is a large C2CH module with zinc-dependent sequence-specific DNA-binding activity. Proceedings of the National Academy of Sciences. 2005; 102(19):6907–12. https://doi.org/10.1073/pnas.0406882102 PMID: 15863623
- Sanghavi HM, Mallajosyula SS, Majumdar S. Classification of the human THAP protein family identifies an evolutionarily conserved coiled coil region. BMC Structural Biology. 2019; 19(1):4. https://doi.org/10.1186/s12900-019-0102-2 PMID: 30836974
- Wiegmann BM, Trautwein MD, Winkler IS, Barr NB, Kim J-W, Lambkin C, et al. Episodic radiations in the fly tree of life. Proceedings of the National Academy of Sciences. 2011; 108(14):5690–5. https://doi. org/10.1073/pnas.1012675108 PMID: 21402926
- 63. Grimaldi D, Engel MS. Evolution of the Insects 1st ed: Cambridge University Press; 2005 2005.
- 64. Generalovic TN, McCarthy SA, Warren IA, Wood JMD, Torrance J, Sims Y, et al. A high-quality, chromosome-level genome assembly of the Black Soldier Fly (Hermetia illucens L.). G3 (Bethesda). 2021; 11(5). https://doi.org/10.1093/g3journal/jkab085 PMID: 33734373
- 65. Dikow RB, Frandsen PB, Turcatel M, Dikow T. Genomic and transcriptomic resources for assassin flies including the complete genome sequence of Proctacanthus coquilletti (Insecta: Diptera: Asilidae) and 16 representative transcriptomes. PeerJ. 2017; 5:e2951. https://doi.org/10.7717/peerj.2951 PMID: 28168115
- 66. Mummery-Widmer JL, Yamazaki M, Stoeger T, Novatchkova M, Bhalerao S, Chen D, et al. Genome-wide analysis of Notch signalling in Drosophila by transgenic RNAi. Nature. 2009; 458(7241):987–92. https://doi.org/10.1038/nature07936 PMID: 19363474
- Schnorrer F, Schönbauer C, Langer CCH, Dietzl G, Novatchkova M, Schernhuber K, et al. Systematic genetic analysis of muscle morphogenesis and function in Drosophila. Nature. 2010; 464(7286):287– 91. https://doi.org/10.1038/nature08799 PMID: 20220848
- 68. Neely GG, Hess A, Costigan M, Keene AC, Goulas S, Langeslag M, et al. A Genome-wide Drosophila Screen for Heat Nociception Identifies α2δ3 as an Evolutionarily Conserved Pain Gene. Cell. 2010; 143 (4):628–38.
- 69. Li J, Gilmour DS. Distinct mechanisms of transcriptional pausing orchestrated by GAGA factor and M1BP, a novel transcription factor. The EMBO Journal. 2013; 32(13):1829–41. https://doi.org/10.1038/emboj.2013.111 PMID: 23708796
- Bag I, Chen S, Rosin LF, Chen Y, Liu CY, Yu GY, et al. M1BP cooperates with CP190 to activate transcription at TAD borders and promote chromatin insulator activity. Nat Commun. 2021; 12(1):4170. https://doi.org/10.1038/s41467-021-24407-y PMID: 34234130
- Lewandowski JP, Sheehan KB, Bennett PE Jr., Boswell RE. Mago Nashi, Tsunagi/Y14, and Ranshi form a complex that influences oocyte differentiation in Drosophila melanogaster. Dev Biol. 2010; 339 (2):307–19. https://doi.org/10.1016/j.ydbio.2009.12.035 PMID: 20045686
- 72. Niwa YS, Niwa R. Ouija board: A transcription factor evolved for only one target in steroid hormone biosynthesis in the fruit fly Drosophila melanogaster. Transcription. 2016; 7(5):196–202. https://doi.org/10.1080/21541264.2016.1210370 PMID: 27434771
- Komura-Kawa T, Hirota K, Shimada-Niwa Y, Yamauchi R, Shimell M, Shinoda T, et al. The Drosophila Zinc Finger Transcription Factor Ouija Board Controls Ecdysteroid Biosynthesis through Specific Regulation of spookier. PLoS Genet. 2015; 11(12):e1005712. https://doi.org/10.1371/journal.pgen.1005712 PMID: 26658797
- Maksimenko O, Kyrchanova O, Klimenko N, Zolotarev N, Elizarova A, Bonchuk A, et al. Small Drosophila zinc finger C2H2 protein with an N-terminal zinc finger-associated domain demonstrates the architecture functions. Biochim Biophys Acta Gene Regul Mech. 2020; 1863(1):194446. https://doi.org/10.1016/j.bbagrm.2019.194446 PMID: 31706027
- Lake CM, Nielsen RJ, Hawley RS. The Drosophila zinc finger protein trade embargo is required for double strand break formation in meiosis. PLoS Genet. 2011; 7(2):e1002005. https://doi.org/10.1371/journal.pgen.1002005 PMID: 21383963
- Chang KC, Garcia-Alvarez G, Somers G, Sousa-Nunes R, Rossi F, Lee YY, et al. Interplay between the transcription factor Zif and aPKC regulates neuroblast polarity and self-renewal. Dev Cell. 2010; 19 (5):778–85. https://doi.org/10.1016/j.devcel.2010.10.007 PMID: 21074726

- 77. Burguete AS, Francis D, Rosa J, Ghabrial A. The regulation of cell size and branch complexity in the terminal cells of the Drosophila tracheal system. Developmental Biology. 2019; 451(1):79–85. https://doi.org/10.1016/j.ydbio.2019.02.001 PMID: 30735663
- 78. Zolotarev N, Georgiev P, Maksimenko O. Removal of extra sequences with I-Scel in combination with CRISPR/Cas9 technique for precise gene editing in Drosophila. BioTechniques. 2019; 66(4):198–201. https://doi.org/10.2144/btn-2018-0147 PMID: 30987444
- 79. Chen S, Zhang YE, Long M. New genes in Drosophila quickly become essential. Science. 2010; 330 (6011):1682–5. https://doi.org/10.1126/science.1196380 PMID: 21164016
- 80. Schertel C, Albarca M, Rockel-Bauer C, Kelley NW, Bischof J, Hens K, et al. A large-scale, in vivo transcription factor screen defines bivalent chromatin as a key property of regulatory factors mediating Drosophila wing development. Genome Res. 2015; 25(4):514–23. https://doi.org/10.1101/gr.181305.114
 PMID: 25568052
- 81. Dobi Krista C, Halfon Marc S, Baylies Mary K. Whole-Genome Analysis of Muscle Founder Cells Implicates the Chromatin Regulator Sin3A in Muscle Identity. Cell Reports. 2014; 8(3):858–70. https://doi.org/10.1016/j.celrep.2014.07.005 PMID: 25088419
- 82. Zhang J, Schulze KL, Hiesinger PR, Suyama K, Wang S, Fish M, et al. Thirty-One Flavors of Drosophila Rab Proteins. Genetics. 2007; 176(2):1307–22. https://doi.org/10.1534/genetics.106.066761 PMID: 17409086