Received: 8 March 2022 | Accepted: 30 June 2022

DOI: 10.1111/imb.12800

ORIGINAL ARTICLE

Insect Molecular Biology

The interaction between the Dbf4 ortholog Chiffon and Gcn5 is conserved in Dipteran insect species

Smitha George[1] | Hannah R. Blum[1] | Eliana F. Torres-Zelada[1] | Grace N. Estep[1] | Youssef A. Hegazy[1] | Gina M. Speer[1] | Vikki M. Weake[1,2]

[1]Department of Biochemistry, Purdue University, West Lafayette, Indiana, USA

[2]Purdue University Center for Cancer Research, Purdue University, West Lafayette, Indiana, USA

Correspondence
Vikki M. Weake, Department of Biochemistry, Purdue University, West Lafayette, IN 47907, USA.
Email: vweake@purdue.edu

Present address
Smitha George, Van Andel Institute, Grand Rapids, Michigan, USA

Funding information
Division of Molecular and Cellular Biosciences, Grant/Award Number: 1930237

Abstract

Chiffon is the sole *Drosophila* ortholog of Dbf4, the regulatory subunit for the cell-cycle kinase Cdc7 that initiates DNA replication. In *Drosophila*, the *chiffon* gene encodes two polypeptides with independent activities. Chiffon-A contains the conserved Dbf4 motifs and interacts with Cdc7 to form the Dbf4-dependent Kinase (DDK) complex, which is essential for a specialized form of DNA replication. In contrast, Chiffon-B binds the histone acetyltransferase Gcn5 to form the Chiffon histone acetyltransferase (CHAT) complex, which is necessary for histone H3 acetylation and viability. Previous studies have shown that the Chiffon-B region is only present within insects. However, it was unclear how widely the interaction between Chiffon-B and Gcn5 was conserved among insect species. To examine this, we performed yeast two-hybrid assays using Chiffon-B and Gcn5 from a variety of insect species and found that Chiffon-B and Gcn5 interact in Diptera species such as Australian sheep blowfly and yellow fever mosquito. Protein domain analysis identified that Chiffon-B has features of acidic transcriptional activators such as Gal4 or VP16. We propose that the CHAT complex plays a critical role in a biological process that is unique to Dipterans and could therefore be a potential target for pest control strategies.

KEYWORDS
acetylation, CHAT, Chiffon, Diptera, *Drosophila*, Gcn5

INTRODUCTION

The *Drosophila melanogaster chiffon* gene was first identified in a screen for female sterile mutations on the second chromosome and is named after the phenotype of the mutant embryos with their thin, fragile chorion (eggshell) that resembles the fabric of the same name (Landis & Tower, 1999; Schupbach & Wieschaus, 1991). Originally, *chiffon* was shown to be the ortholog of Dbf4 (Dumbbell former 4 protein), which is the regulatory subunit of the Cdc7 kinase (cell division cycle 7) that initiates DNA replication by phosphorylating components of the minichromosome maintenance protein complex (MCM) helicase complex (Bousset & Diffley, 1998; Landis & Tower, 1999; Matthews & Guarné, 2013). Dbf4 is highly conserved in fungi and metazoans but has not been identified in plants (Scofield et al., 2014). In *Drosophila*, *chiffon* is necessary for a specialized form of DNA replication termed gene amplification that is required for the production of the chorion in ovary follicle cells (Landis & Tower, 1999). We previously showed that the N-terminal domain of Chiffon (hereafter referred to as Chiffon-A) directly binds and activates *Drosophila* Cdc7 like other Dbf4 orthologs (Stephenson et al., 2015). However, in insects Chiffon contains a long C-terminal extension that is absent from fungi or vertebrate Dbf4 orthologs (Torres-Zelada et al., 2019). Mass spectrometry studies demonstrated that this insect-specific C-terminal

domain of Chiffon (hereafter referred to as Chiffon-B) interacts with the histone acetyltransferase Gcn5/KAT2A (lysine acetyltransferases 2A) and three partner proteins, Ada2b-PA, Sgf29 and Ada3 to nucleate formation of the Chiffon histone acetyltransferase (CHAT) complex (Torres-Zelada et al., 2019). Surprisingly, this second insect-specific function of Chiffon is essential for viability in *Drosophila*, while its conserved Cdc7-binding activity is not (Torres-Zelada et al., 2019). Moreover, *chiffon* appears to be a dicistronic gene that encodes Chiffon-A and Chiffon-B independently, likely via translational mechanisms, because nonsense mutations that disrupt Chiffon-A activity still result in the production of Chiffon-B protein (Torres-Zelada et al., 2019).

Gcn5/KAT2A was the first nuclear histone acetyltransferase to be identified, and functions as part of multi-subunit complexes that have key roles in gene expression and are widely conserved in yeast, *Drosophila*, plants and mammals (Torres-Zelada & Weake, 2021). Recently, gene expression profiling of *Drosophila* embryos revealed that CHAT has an important role in regulating the expression of developmental genes, albeit partially redundant with other Gcn5-containing complexes (Torres-Zelada et al., 2022). Intriguingly, there is a switch between the expression of the Chiffon-A (Cdc7 binding) and Chiffon-B (CHAT) proteins during the early stages of embryogenesis in *Drosophila*. Chiffon-A is expressed during and is maternally required for, the early nuclear cycles, similar to Cdc7 (Seller & O'Farrell, 2018; Torres-Zelada et al., 2022). In contrast, Chiffon-B only begins to appear in embryos just prior to zygotic genomic activation and is essential for zygotic development but is not maternally required for embryogenesis (Torres-Zelada et al., 2022). These data suggest that the unique gene structure of *chiffon* provides a mechanism in which to temporally separate DNA replication from transcription activation via histone acetylation in developing embryos. Because Chiffon-B appears only to be present in insects, we hypothesized that the CHAT complex and this unique *chiffon* gene structure might play a specialized biological role that is unique to insects.

To investigate this possibility, in this study, we examined a wide variety of insect species to test if the interaction between Chiffon-B and Gcn5 was indeed conserved in all insect orders. A secondary goal of this project was to provide a course-based research experience (CURE) to early-stage undergraduate students to enhance their academic performance; undergraduate researchers played key roles in performing and analysing the yeast two-hybrid assays as part of the CURE and are fully listed in the acknowledgements section. These studies identify interactions between Chiffon-B and Gcn5 within Diptera species, suggesting that the CHAT complex might not be conserved within all insect species as previously thought. Moreover, phylogenetic analysis coupled with protein motif searches identify domains in Chiffon-B that are reminiscent of strong acidic activators such as the classical transcription factors Gal4, MyoD and VP16. To our knowledge, these studies represent the first characterization of potential Gcn5 interacting partners in insects outside of the well-studied *Drosophila* model.

RESULTS AND DISCUSSION

Insect Chiffon proteins have diverged from fungi and vertebrate Dbf4 orthologs

Drosophila Chiffon consists of 1695 amino acids (aa), nearly double the size of vertebrate Dbf4 orthologs (Figure 1a, Figure S1). The first 400 aa in the N-terminal region of Chiffon (Chiffon-A) are highly conserved with other Dbf4 orthologs and include the N (45–89 aa), M (200–240 aa) and C (309–349 aa) motifs that are essential for binding and activating Cdc7 to form the Dbf4-dependent kinase (DDK) complex (Masai & Arai, 2000; Stephenson et al., 2015). In addition, motif scan searching identified a DBF4-DBF Zinc finger (307–356 aa) overlapping with the C motif and a potential AT-hook DNA binding domain motif just after the N-terminal region (493–505 aa: PRGRGRPPNQVDS) (Sigrist et al., 2010) (Figure 1a). Importantly, the first 400 amino acids of Chiffon are necessary and sufficient for the Cdc7-binding activity of Chiffon and its role in gene amplification (Stephenson et al., 2015; Torres-Zelada et al., 2019). Our previous studies have shown that the C-terminal domain of Chiffon (1243–1695 aa; Chiffon-B) directly binds to Gcn5 and nucleates the formation of the CHAT complex. CHAT is required in *Drosophila* for histone H3 acetylation and viability, but not for DNA replication (Torres-Zelada et al., 2019). There is a single ~5 kB exon that contains all of the coding regions for the *chiffon* gene (Torres-Zelada et al., 2019). Although there are four splice isoforms listed for *chiffon* on FlyBase (Chiffon-RA, -RB, -RD and -RE) and another gene shares the *chiffon* promoter but encodes a distinct protein (*Hyls1*) (Figure 1b), there is no evidence of alternative splicing in *chiffon* that could account for distinct production of the Chiffon-A and Chiffon-B proteins. Instead, uncharacterized translational mechanisms are thought to underly the switch between Chiffon-A and Chiffon-B expression (Torres-Zelada et al., 2022). Chiffon interacts with the histone acetyltransferase Gcn5 and three partner proteins, Ada2b-PA, Sgf29 and Ada3 to nucleate formation of the CHAT and a schematic illustrating the subunit composition of the *Drosophila* DDK and CHAT complexes are shown in Figure 1c (Torres-Zelada et al., 2019; Torres-Zelada & Weake, 2021).

To examine Chiffon-B function in other insect species, we first compared Chiffon homologues in various insect species with the Dbf4 orthologs in selected vertebrate species and the yeast *Saccharomyces cerevisiae*. Whereas yeast and insects each have only one Dbf4/Chiffon ortholog, there are two paralogs of Dbf4 (Dbf4_A and Dbf4_B) in most vertebrates (Figure 2, Figure S1). When we aligned Dbf4/Chiffon across these species, we observed that as previously found, the insect Chiffons share an extended C-terminal domain that is absent from the mammalian and yeast Dbf4 orthologs (Figure S1, Supplementary file 1, accession numbers provided in Table S1). Moreover, consistent with other reports, Dbf4/Chiffon proteins are grouped into three distinct classes: Class I Fungi (They have motif C at the extreme C terminus of the protein and include a long linker connecting motifs M and C together); Class II Animals (Similar length to class I but have a

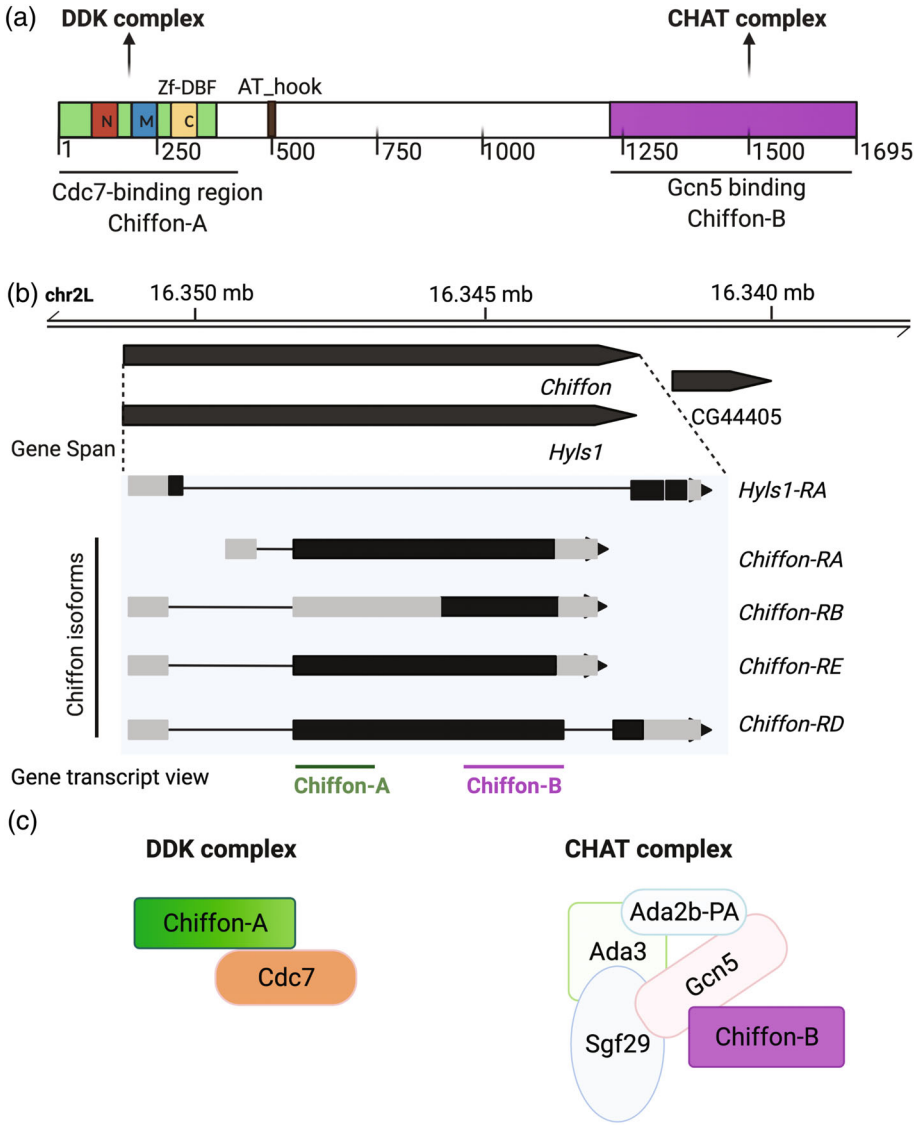

FIGURE 1 *Drosophila* Chiffon encodes two distinct polypeptides with independent activities. (a) Schematic representation of *Drosophila melanogaster* Chiffon illustrating the domain architecture, indicated by coloured boxes. Amino acid positions are indicated. The N-terminal 400 aa are necessary and sufficient to bind and activate Cdc7, forming the DDK complex and are referred to as Chiffon-A. The C-terminal amino acids from 1243–1695 aa are necessary and sufficient to bind Gcn5, forming the CHAT complex and are referred to as Chiffon-B. The exact size of the Chiffon-A and Chiffon-B proteins has not yet been determined, but mass spectrometry analysis of peptides from N- and C-terminal purifications of Chiffon suggests break-points in the region between 400–500 aa. (b) Diagram of the *D. melanogaster chiffon* gene structure. Exons are indicated by boxes, with coding regions shown in black and untranslated regions in grey. Coding regions corresponding to the minimal Chiffon-A and Chiffon-B proteins are underlined. (c) Schematic illustrating the subunit composition of the *Drosophila* DDK and CHAT complexes

short linker between motifs M and C); and Class III Insects (They are twice the size of Dbf4 from other classes, due to the presence of a long C-terminal extension. This region contains three additional conserved motifs) (Matthews & Guarné, 2013). In addition, we observed a clear separation in the phylogenetic analysis for Chiffon between the Diptera relative to other insect orders such as Hymenoptera, Coleoptera, Hemiptera and Lepidoptera (Figure 2). Thus, while Chiffon shares a C-terminal extension in all insect species, its sequence has diverged in different orders, suggesting that the function of this C-terminal region might have distinct roles in particular groups of insects (Figure S1, Supplementary file 1).

Divergence of Gcn5 homologues in Diptera relative to other insects

Because Chiffon-B directly interacts with Gcn5 in *Drosophila* as demonstrated by yeast two-hybrid, recombinant protein binding and immunoprecipitation of endogenous proteins (Torres-Zelada et al., 2019), we next compared Gcn5 homologues in a variety of insect species with vertebrates and yeast. Gcn5, also known as KAT2A, is highly conserved in fungi, plants, insects and vertebrates and exists as part of four multi-subunit complexes in *Drosophila*: Spt-Ada-Gcn5 acetyltransferase, Ada2a-containing complex, CHAT and

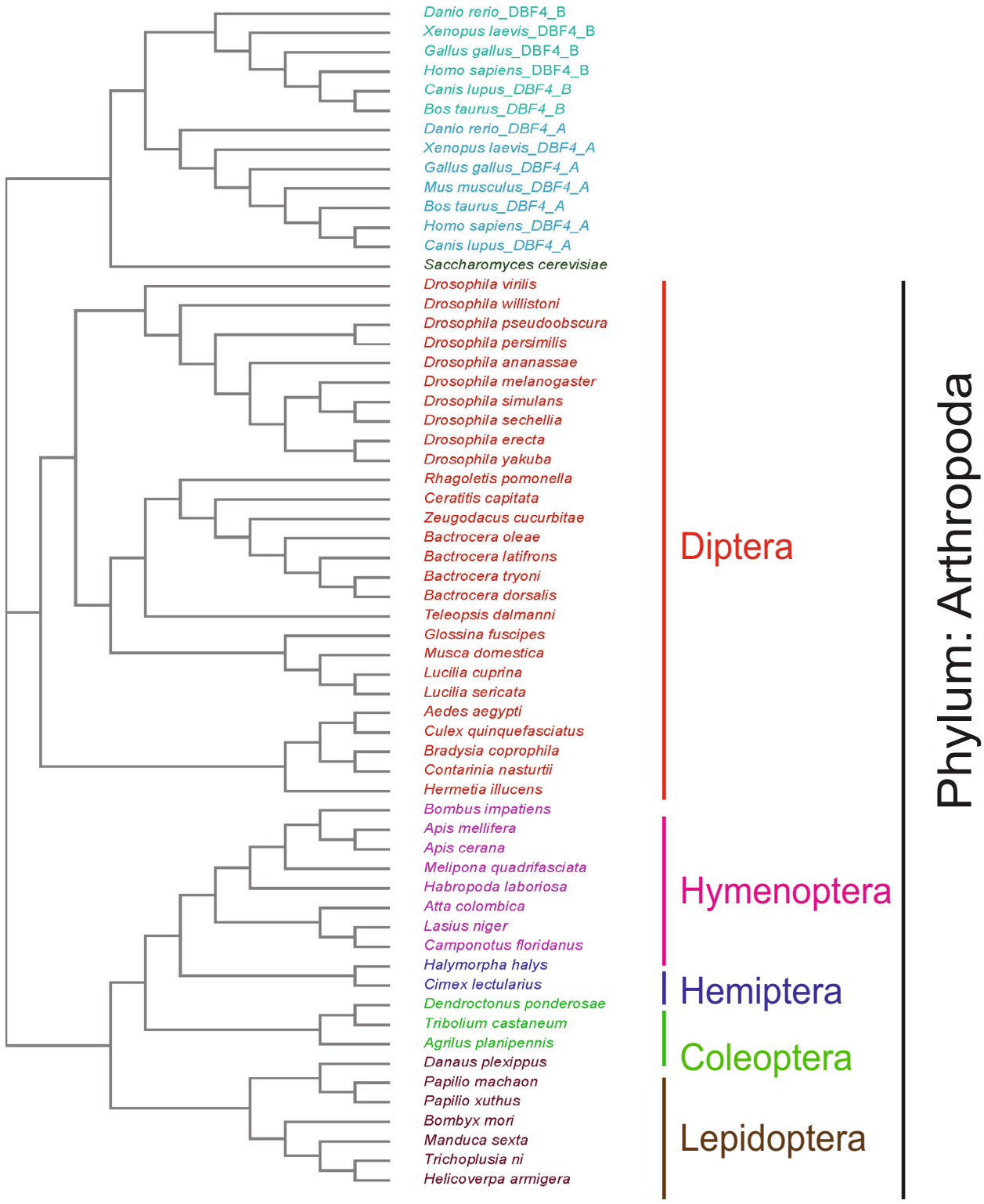

FIGURE 2 Insect Chiffon proteins have diverged from fungi and vertebrate Dbf4 orthologs. Phylogenetic analysis of Dbf4/Chiffon proteins from the indicated species. The phylogenetic tree was generated using the neighbour-joining method without distance corrections based on the Clustal-Omega multiple sequence alignment of protein sequences. Accession numbers used to generate the alignment are provided in Table S2. The vertebrate paralogs of Dbf4 are designated DBF4_A and DBF4_B. Insect orders are marked using different colour codes and indicated to the right of the vertical lines.

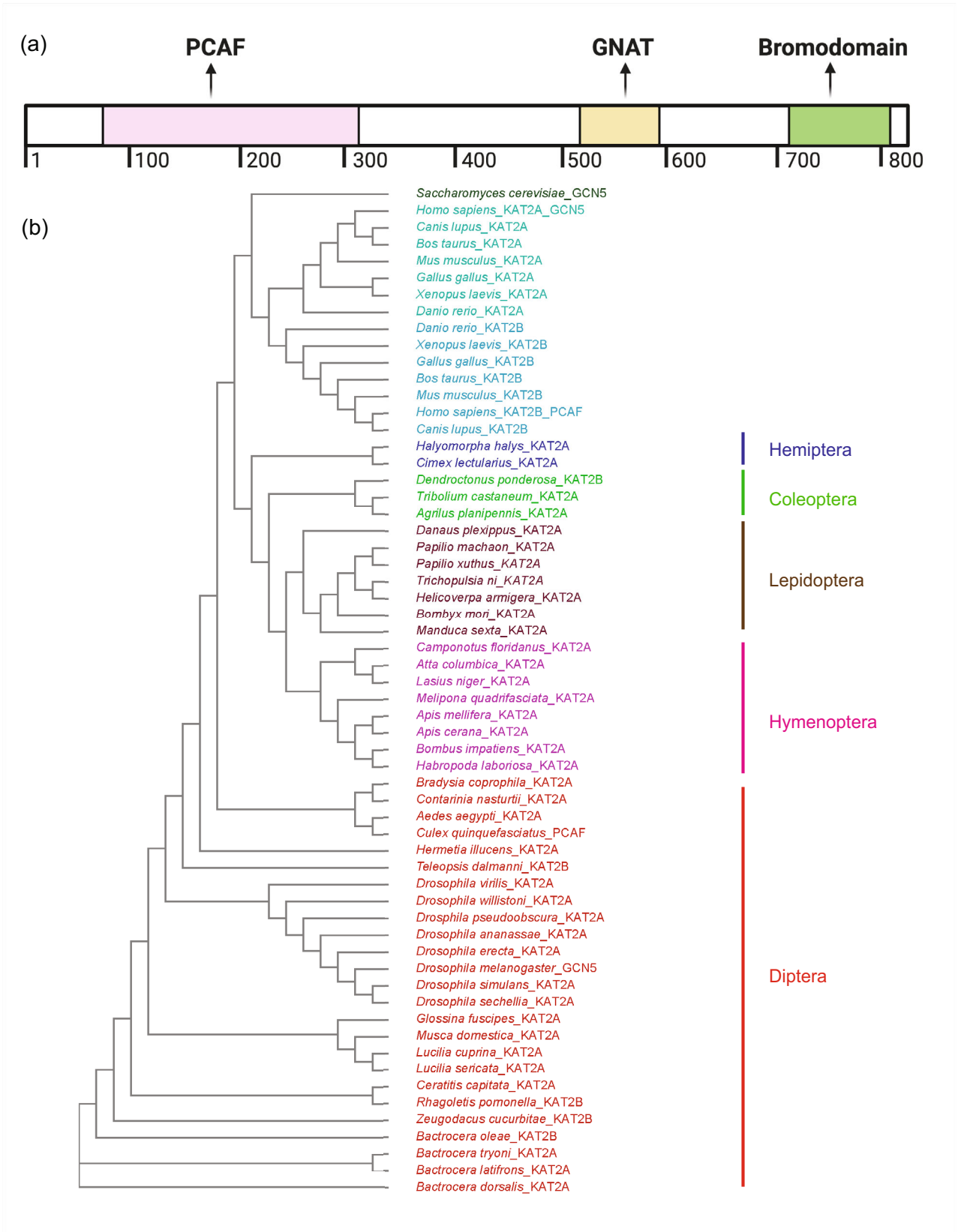

FIGURE 3 Gcn5 has diverged in Diptera relative to other insect species. (a) Schematic representation of *Drosophila* Gcn5 Chiffon illustrating the domain architecture, indicated by coloured boxes. Amino acid positions are indicated. (b) Phylogenetic analysis of GCN5 orthologs from the indicated species, as described in Figure 2. Vertebrates have two Gcn5 orthologs, KAT2A and KAT2B (PCAF and Gcn5), whereas insects possess one ortholog. Insect orders are marked using different colour codes.

TABLE 1 Predicted PADDLE *Z*-scores of selected organisms

Z-score	
	Drosophila melanogaster
4.56	LEYDILNIHLQSKDHELFAKNSDNFLALDTLIQSSADVNRFLEEEPVESELDM
5.99	EGREERLVNTPQTPPPTDCFTSEFDLCDLIMGSAGSGDDDEDVSRGNPPGSGR
	Drosophila simulans
4.563	LEYDILNIHLQSKDHELFAKNSDNFLALDTLIQSSADVNRFLEEEPVESELDM
4.694	EERLVNTPQTPPPTDCFTSEFDLCDLIMGSGGSGDDDEDVSRGNPPGRRISNL
	Drosophila sechellia
4.563	LEYDILNIHLQSKDHELFAKNSDNFLALDTLIQSSADVNRFLEEEPVESELDM
4.694	EERLVNTPQTPPPTDCFTSEFDLCDLIMGSGGSGDDDEDVSRGNPPGRRISNL
	Lucilia cuprina
2.72	IALDSLIHSSADVNKFLDENRKLEIKTELEEMEVDHESLMCIEQDPLQNSENE
5.83	AINDVNLNTPPATDYFSSDFDLCDLINQSSERLQEDGDNGLPHDSSSKKFSIF
	Musca domestica
6.1	EWQIDKFLTLTPPSGLDVDNDLFELNGSTLEPSVNHQSSQLTTRQEQDMPTAS
6.25	SSSILNTPPSTDYFSSDFDLCDLINQDLENGQKDADDGRAGQDAKAKKFSIFN
	Aedes aegypti
2.84	KFDPPVLEDNFVNYKVMCQELDVFLNEMSEGAEEDEAFFEMDKKSVPIIARME
	Apis mellifera
5.91	EVEKMQFSFEAVPKSEPWYQTYQRQDKGAEFWHYFTEGDSEKPFLLPYEIENF
6.89	NLELVESFEPDIMQMELNLYESNIPRGPPPNLLELLDSCHEIVNYLENSSCAS
	Tribolium castaneum
4.88	YNFSFESVPKNEAWYQTFQRRDKGEEKYVFLSDDNYWDPLILPYQLDYVPPLD
	Dendroctonus ponderosae
3.744	KLIGEDTNYIALNGYIHADVGIESLLSTGIDAIDFEDFAPKKKRRNMPRTRAS
	Cimex lectularius
4.26	EALLSSSLQGPSFEIEESHPPNHKLFSNLRYVNIDSCASSECSDFFGFLDDGR
	*Danio rerio*_DBF4_B
2.844	NTTSHLYPMQNNCPDAFDHDKAGVYFTCTTKVNEFQPDVLCMDLIQHSVSSID
5.571	PVESRVTTHSSELDLPQLVPFYYEDNKLQRHQYNDPPELFPFFFQTPEDNSQA
	*Danio rerio*_DBF4_A
	None
	*Xenopus laevis*_DBF4_B
4.689	VGEHHFRFVSNPLSYKMIDDLAAQLTCDLMELPFGSPTSPEAERSSQNEDWDL
2.639	CGEIPLANIEVDVYNVCSFDQPVVTCTMELPDVSAEGKIHSNLLGSTVGDERV
	*Xenopus laevis*_DBF4_A
6.249	KLKEQKKHGYCECCLKKYDDLESHILSPQHKNFSESAYYQVVDDLISTFDFDF
5.151	SHKVSVPQQEDDSKFPSETLLALFESSEDKTEFFGFAGSPAYESCSMDDGDTP
	*Mus musculus*_DBF4_A
4.634	LDKKRTEYLPAHEDRTCGSPVQSLLDLFQTSEEKSEFLGFTGYTENSGICDVL
	Helicoverpa armigera
2.297	NTDYEDAALHRRSPHHLAFVRDHTNFLALDSLIGSGADVSSFLDKATPLNGER
3.768	FLSEVFEDAADYDVSDEVMDGDAAVTTSTNMPDVSSLVSECENCDLIRNEIKE
	Bombyx mori
	None
	Danaus plexippus

(Continues)

TABLE 1 (Continued)

Z-score	
3.878	PEVSPLREEPSEKCLKWEDGRLKYTPAVEQLEFAFESVPQSEPWFETFKRQDQ
5.198	TKSTASEETVSDTHSNTKVESVLSETQEASERLQQFLSEVFEEPTDYDIYDQL
	*Homo sapiens*_DBF4_A
3.57	RNRKENLEPNAEFDKRTEFITQEENRICSSPVQSLLDLFQTSEEKSEFLGFTS
	*Homo sapiens*_DBF4_B
	None
	Saccharomyces cerevisiae
	None

Ada2/Gcn5/Ada3 transcription activator (Guelman et al., 2006; Kusch et al., 2003; Muratoglu et al., 2003; Soffers et al., 2019; Torres-Zelada et al., 2019). *Drosophila* Gcn5 shares the domains that are common to all Gcn5 homologues, including the PCAF (P300/CBP-associated factor) domain (73–323 aa), Gcn5-N-Acetyltransferase domain (514–598 aa) and bromodomain (717–795 aa), which binds acetylated lysine (Figure 3a) (Carré et al., 2005; Nagy & Tora, 2007). In addition, in *Drosophila* and other insects, Gcn5 shares a similar domain architecture with the mammalian Gcn5 paralogs that include the metazoan-specific N-terminal domain (Torres-Zelada et al., 2022) (Figure S2, Supplementary file 2).

Similar to Dbf4, Gcn5 has two paralogs in most vertebrate species (Gcn5 and PCAF) but only one homologue in insects. We aligned Gcn5 orthologs using Clustal (Figure 3b, Figure S2, Supplementary file 2, accession numbers provided in Table S2) and found that like Chiffon, Gcn5 in the Diptera order of insects appears to have diverged from all other insects and vertebrates. Together, these analyses indicate that both Chiffon-B and Gcn5 differ in the Diptera order of insects, which includes the *Drosophila* genus in which the CHAT complex was identified, relative to other insect species. These data raise the possibility that the interaction between Chiffon-B and Gcn5 that leads to the formation of the CHAT complex might not be conserved within all insect species but might instead be unique to Dipterans.

The C-terminal region of Chiffon (Chiffon-B) binds to Gcn5 in Dipteran insects

To examine this possibility in more detail, we aligned the insect-specific C-terminal region of Chiffon that corresponded to the minimal region of *Drosophila* Chiffon that has been shown to bind Gcn5 in vitro (1243–1695 aa). Using this approach, we observed high sequence similarity between this Chiffon-B region in Diptera (47%–97%) with less conservation in other insect species (9%–20%) (Figure 4; Supplementary file 3, accession numbers provided in Table S3). In particular, we observed conserved blocks of acidic amino acids in the Chiffon-B proteins (Figure 4). These regions of acidic amino acids were highly reminiscent of acidic activation domains that are characteristic features of several well-studied transcription activators such as Gal4, VP16 and MyoD (Hirai et al., 2010; Sadowski et al., 1988). Acidic activation domains are usually characterized by bulky hydrophobic amino acids interspersed with acidic amino acid residues and have been shown to interact with transcriptional coactivators such as Gcn5 and p300 (Drysdale et al., 1995; Melcher, 2000; Raj & Attardi, 2017; Sanborn et al., 2021). In general, these acidic domains remain relatively unstructured until binding to their respective coactivator protein and do not have a high level of sequence conservation (Sanborn et al., 2021). Since the presence of an acidic activation domain could suggest that Chiffon-B functions as a transcription activator, we sought to examine this possibility further by performing Predictor of Activation Domains using Deep Learning in Eukaryotes (PADDLE) (https://paddle.sites.stanford.edu/). PADDLE is a deep convolutional neural network that predicts acidic transcriptional activation domains (TADs) based on protein sequence (A score of 0 indicates no activation; scores greater than 4 are considered significant and scores greater than 6 are considered strongly significant) (Sanborn et al., 2021). PADDLE analysis of the *Drosophila* Chiffon-B region resulted in a *Z*-score of 5.99 for 1325–1378 aa strongly supporting a potential activation domain in this region (Table 1, Figure 4). In addition, we also observed a *Z*-score of 4.6 at the N-terminal region of *Drosophila* Chiffon-A (319–372 aa), suggesting the possibility that this region could also contain an activating domain (AD). Together, these analyses suggest that Chiffon-B shares features of classical transcription factors that act as strong acidic activators and directly bind histone acetyltransferases and other transcription coactivators. DNA-binding transcription factors generally contain a minimum of two domains: a DNA-binding domain (DBD) and a TAD (Hirai et al., 2010). Because there is a potential DNA binding domain in Chiffon that lies just after the Chiffon-A region (AT-hook domain, 493–505 aa, PRGRGRPPNQVDS), it is possible that Chiffon-B could bind DNA and therefore direct CHAT activity to chromatin targets; however, this has not been tested in the current study.

Yeast two-hybrid analysis demonstrates interactions between Chiffon-B and Gcn5 in Diptera

Because insects in the Diptera order appeared to have diverged Chiffon-B and Gcn5 sequences relative to other insects, we next examined the interaction between Chiffon-B and Gcn5 in selected

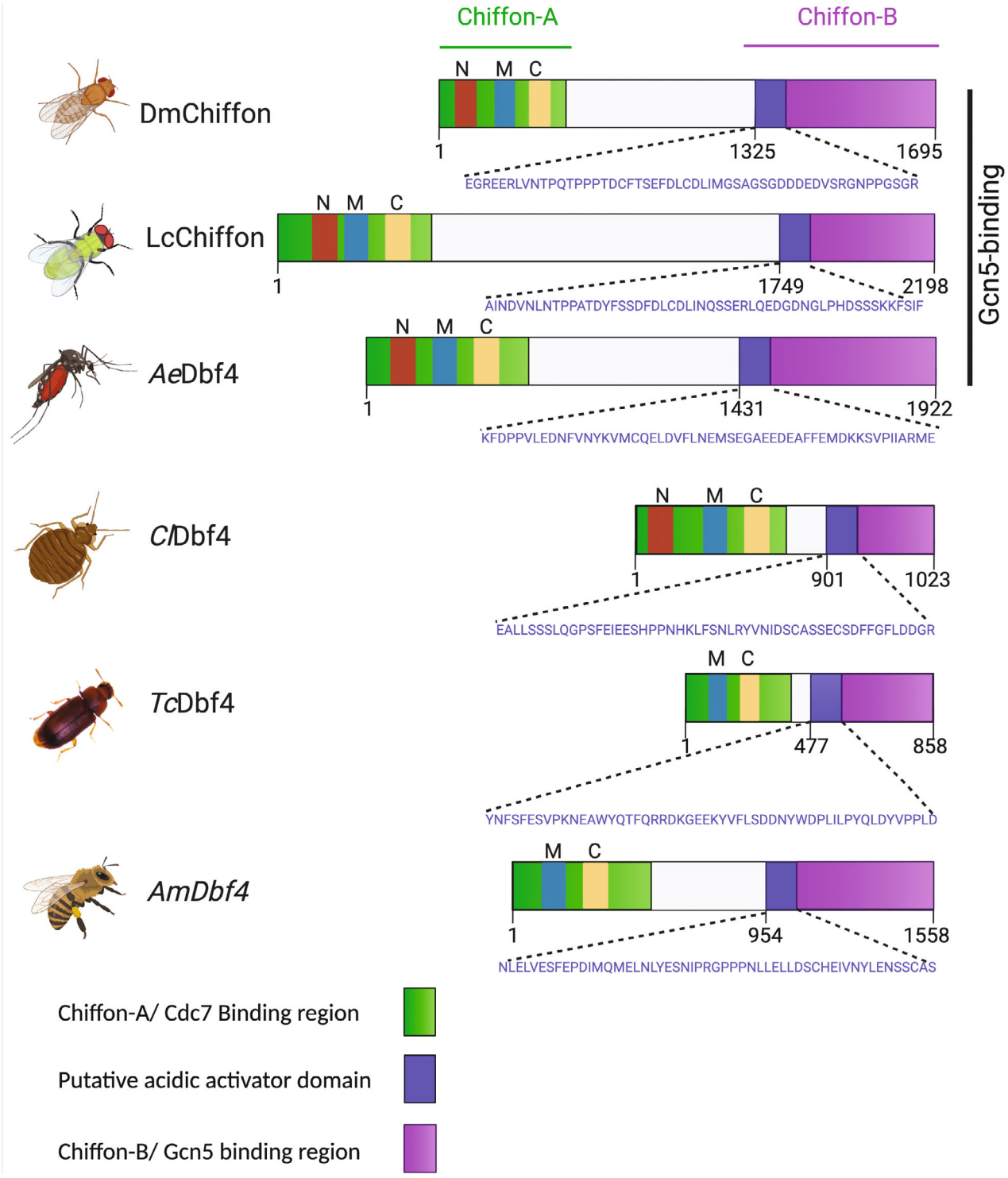

FIGURE 4 The Chiffon-B domain that binds Gcn5 is highly conserved within insects in the Diptera order. The Chiffon-B regions that correspond to 1243–1695 aa in Drosophila Chiffon were aligned, as described in Figure 2, and a full version of the alignment is provided in Supplementary file 1. A schematic illustrating conserved regions in Chiffon-B within different insect species are shown, with boxes indicating regions of high sequence similarity. A putative acidic activation domain was identified between 1325–1378 aa in *Drosophila* Chiffon and other insect-Dbf4 homologues using PADDLE. The PADDLE prediction *Z*-score values of selected organisms are listed in Table 1. PADDLE predicts the activation *Z*-score for any 53 amino acid-long protein sequences, ranging from approximately −1 to 12: A score of 0 indicates no activation, scores greater than 4 are considered significant and scores greater than 6 are considered strongly significant (Sanborn et al., 2021). Ae, *Aedes aegypti* (all three belongs to order Diptera); Am, *Apis mellifera* (Hymenoptera); Cl, *Cimex lectularius* (Hemiptera); Dm, *Drosophila melanogaster*; Lc, *Lucilia cuprina*; Tc, *Tribolium castaneum* (Coleoptera)

TABLE 2 Primers used in this study

	Insect/primer	Sequence (5′–3′)
1	*Cimex lectularius* DBF4_DBD_EcoRI_F	GAC CTG CAT ATG GCC ATG GAG GCC GAA TTC ATG AAA GTT AAA CAG GAA GTT GAA G
2	*C. lectularius* DBF4_DBD_BamHI_R	GTTATGCGGCCGCTGCAGGTCGACGGATCCCTAGTTTACGCTATTTGTG
3	*Tribolium castaneum* DBF4_DBD_EcoRI_F	GAC CTG CAT ATG GCC ATG GAG GCC GAA TTC ATG TCA TAT CTT TCT CCT GTT CCT ATG
4	*T. castaneum* DBF4_DBD_BamHI_R	GTTATGCGGCCGCTGCAGGTCGACGGATCCTTATCTCTCTCGATACCAGCG
5	*Apis mellifera* DBF4_DBD_EcoRI_F	GAC CTG CAT ATG GCC ATG GAG GCC GAA TTC ATG GGA GAT TCG GAA AAA CCA TTT CTT
6	*A. mellifera* DBF4_DBD_BamHI_R	GTT ATG CGG CCG CTG CAG GTC GAC GGA TCC TCA GCT ACT TTT TAA CCA TCT TCC G
7	*Lucilia cuprina* DBF4_DBD_EcoRI_F	GAC CTG CAT ATG GCC ATG GAG GCC GAA TTC ATG AAG ACG GGT TCT AAG AAG GGA G
8	*L. cuprina* DBF4_DBD_BamHI_R	GTTATGCGGCCGCTGCAGGTCGACGGATCCCTATCTACGATAACGAAAACG
9	*Aedes aegypti* DBF4_DBD_EcoRI_F	GAC CTG CAT ATG GCC ATG GAG GCC GAA TTC ATG AAT CTG CCA CCG CCA AG
10	*A. aegypti* DBF4_DBD_BamHI_R	GTT ATG CGG CCG CTG CAG GTC GAC GGA TCC CTA TCG TTC CCT GTA CCA CCG
11	*C. lectularius* Gcn5_AD_EcoRI_F	GCT CAT ATG GCC ATG GAG GCC AGT GAA TTC ATG AAC TCA TCC GGC CCG
12	*C. lectularius* Gcn5_AD_BamHI_R	ATT CAT CTG CAG CTC GAG CTC GAT GGA TCC TTA TTT GTC CCA TAG ATT GAG ATC GCG
13	*T. castaneum* Gcn5_AD_EcoRI_F	GCT CAT ATG GCC ATG GAG GCC AGT GAA TTC ATG TCA GAA CAG GGC CAG TTG
14	*T. castaneum* Gcn5_AD_BamHI_R	ATT CAT CTG CAG CTC GAG CTC GAT GGA TCC TTA TTT GTC CCA CAA ACC GAC TTC
15	*A. mellifera* Gcn5_AD_EcoRI_F	GCTCATATGGCCATGGAGGCCAGTGAATTCATGACAACAGAAGAAG
16	*A. mellifera* Gcn5_AD_BamHI_R	ATTCATCTGCAGCTCGAGCTCGATGGATCCCTAAAGATATCCAATTTCTTTC
17	*L. cuprina* Gcn5_AD_EcoRI_F	GCTCATATGGCCATGGAGGCCAGTGAATTCATGTCAGGGGGCTCTCTTATAACGT
18	*L. cuprina* Gcn5_AD_BamHI_R	ATTCATCTGCAGCTCGAGCTCGATGGATCCTCATTTATCCCATAAACCTAATTC
19	*A. aegypti* Gcn5_AD_EcoRI_F	GCTCATATGGCCATGGAGGCCAGTGAATTCATGAATGACCATTCGAAACATTCGG
20	*A. aegypti* Gcn5_AD_BamHI_R	ATTCATCTGCAGCTCGAGCTCGATGGATCCTCATTTATCCCATAGTCCGATTTCC

insects from a range of orders. To do this, we used a yeast two-hybrid approach to screen for interactions between Chiffon-B and Gcn5 proteins from a selected group of insects representing a variety of orders: Coleopteran model *Tribolium castaneum* (red flour beetle); Hemipteran model, *Cimex lectularius* (Bed bug); Hymenopteran model, *Apis mellifera* (Honey bee); Dipteran insects *Lucilia cuprina*, (Australian sheep blowfly) and *Aedes aegypti* (Yellow fever mosquito). We have previously shown an interaction between *D. melanogaster* Chiffon-B (1243–1695 aa) and Gcn5 by yeast two-hybrid (Torres-Zelada et al., 2019), demonstrating that this approach is able to capture the protein interaction between these partners. Briefly, we co-transformed yeast with plasmids expressing Gcn5 fused to the Gal4 activating domain and Chiffon-B fused to the Gal4 DNA binding domain. If Gcn5 and Chiffon-B interact in this system, they reconstitute the Gal4 transcription activator resulting in the expression of auxotrophic reporters for growth on media lacking adenine and histidine (Figure 5a). In this system, yeast growth on media lacking leucine and tryptophan demonstrates the presence of both plasmids (Figure 5a). As controls, we also tested interactions between Gcn5 or Chiffon-B and the empty plasmid containing only the DNA binding or activating domain. Using this yeast two-hybrid approach, we observed an interaction between *D. melanogaster* Chiffon-B and Gcn5 (Figure 5b), as previously reported (Torres-Zelada et al., 2019). We also identified interactions between Chiffon-B and Gcn5 from the Australian sheep blowfly and yellow fever mosquito, but not from any other insect species tested (Figure 5b). We were not able to determine

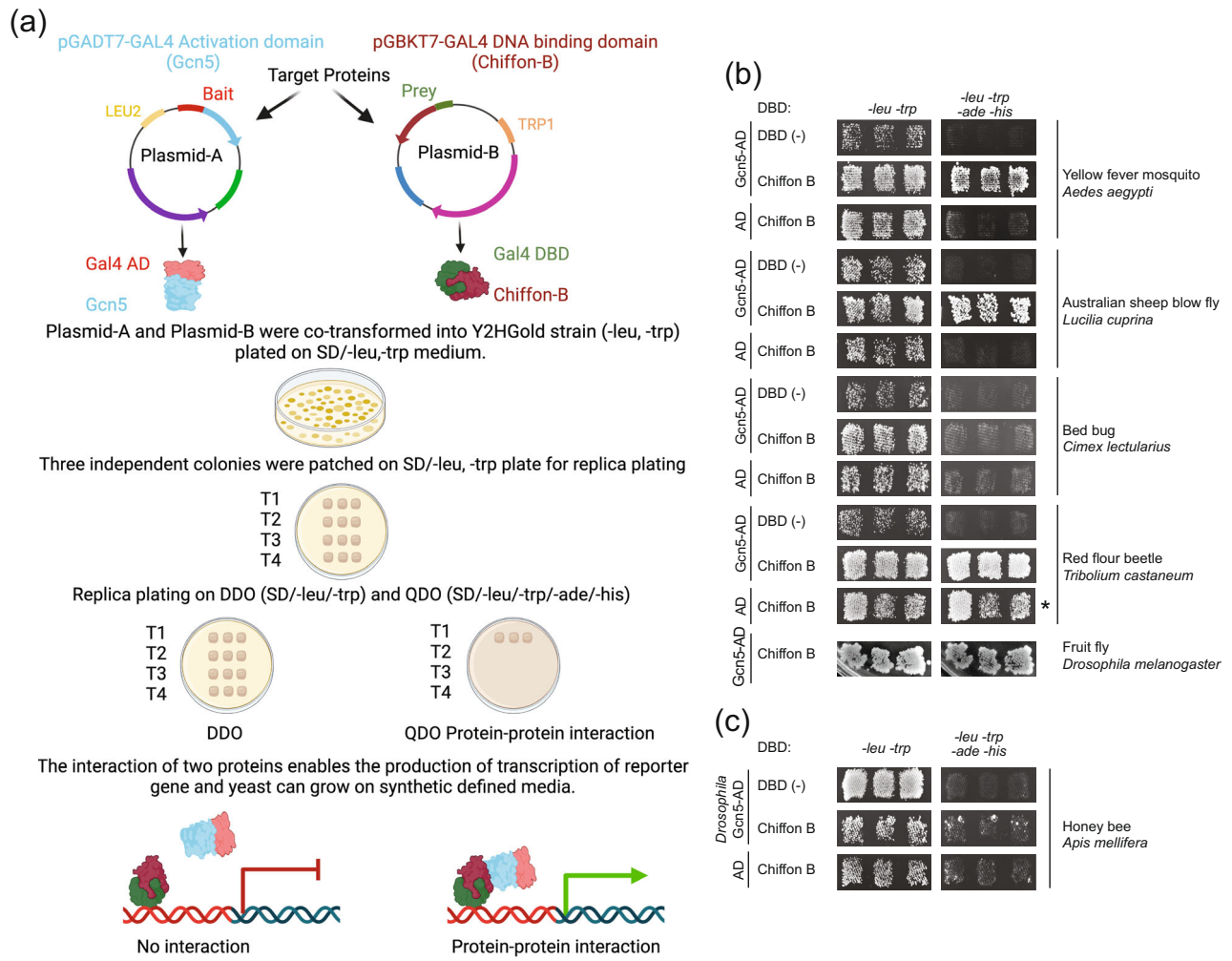

FIGURE 5 Yeast two-hybrid analysis demonstrates interactions between Chiffon-B and Gcn5 in Diptera. (a) Schematic representation of the yeast two-hybrid assay to test the interaction between Chiffon-B and Gcn5 from different insect species. The Gal4 activating domain (AD) or the Gal4 DNA binding domain (DBD) were fused to Gcn5 or Chiffon-B, respectively. Empty plasmids expressing only the AD or DBD were used to test for auto-activation of each protein (*, auto-activation). Three independent transformed yeast colonies were patched on media lacking leucine and tryptophan to test for the presence of the AD and DBD plasmids, and on media lacking leucine, tryptophan, adenine and histidine to test for interaction. (b) Yeast two-hybrid analysis of Chiffon-B and Gcn5 interactions in the indicated insect species, as outlined in panel a. Growth on media lacking leucine and tryptophan indicates the presence of both plasmids and growth on media lacking leucine, tryptophan, adenine, and histidine demonstrates an interaction. (c) Yeast two-hybrid analysis of Chiffon-B from honey bee with Drosophila Gcn5, as described in panel b

if Chiffon-B and Gcn5 interact in the red flour beetle because we observed growth on selective media when Chiffon-B was co-expressed with the activating domain alone (Figure 5b). These data suggested that *Tribolium* Chiffon-B autoactivates expression in the yeast two-hybrid assay, indicating that Chiffon-B might be able to directly bind yeast coactivator proteins. Unfortunately, we were not able to PCR amplify and clone Gcn5 from honey bees; thus, instead, we tested an interspecies interaction between honey bee Chiffon-B and *Drosophila* Gcn5 (Figure 5c). Using this approach, we did not observe a positive interaction with honey bee Chiffon-B. Thus, although all insect Chiffon homologues share a long C-terminal extension, we were only able to detect interactions between Chiffon-B and Gcn5 in *L. cuprina* and *A. aegypti* (Figure 4). These experimental data combined with the phylogenetic analyses suggest that the CHAT complex might be unique to the Diptera and could therefore have a specialized role in gene regulation in these insects. It remains possible that Chiffon-B has other chromatin-related interacting partners in different insect species that could play a similar role in transcription. Supporting this idea, *Tribolium* Chiffon-B showed autoactivation in the yeast two-hybrid assay and also contained a putative acidic activator domain (Figure 4, Figure 5b).

CHAT complex formation may be limited to the Diptera: What might CHAT possibly do uniquely in Dipteran insects?

Compared to other insect orders, Dipterans have recently evolved and are considered as long germ band insects meaning that all segments are specified almost simultaneously within the blastoderm, that is, before gastrulation (Davis & Patel, 2002; Liu & Kaufman, 2005).

The long germ band type is mainly observed in those orders of insects that utilize nurse cells during oogenesis, and the provisioning of maternal information by nurse cells is thought to have been a prerequisite for the evolution of long germ band embryogenesis (Davis & Patel, 2002). Chiffon-A is required for DNA replication in *Drosophila* egg chamber follicle cells (Stephenson et al., 2015), and must also be maternally supplied for the early nuclear divisions during embryogenesis (Torres-Zelada et al., 2022). Interestingly, there is a temporal switch in protein expression between Chiffon-A and Chiffon-B during early embryonic development in *Drosophila* with Chiffon-A levels decreasing towards the later nuclear cycles just as Chiffon-B expression begins. Because segment specification is determined relatively early in long germ band insects, the zygotic genes required for these molecular patterning events must be expressed quickly, synchronously and robustly in the appropriate regions of the developing embryo. We propose that the switch between Chiffon-A and Chiffon-B expression could be part of the mechanisms that coordinate the precise timing of these transcriptional events in Diptera. In this model, as Chiffon-A levels decrease in the syncytial embryo, DDK activity would decline to result in delayed or reduced initiation of DNA replication. At the same time, the increase in Chiffon-B levels would nucleate the formation of CHAT, leading to increased histone acetylation and activation of gene expression. Dissecting the translational mechanisms that control the switch between Chiffon-A and Chiffon-B expression may help provide insight into the unique role played by this interesting Dbf4 ortholog in Diptera in segmentation and long germ development.

Together, our results suggest that the CHAT complex plays a critical role in a biological process that is unique to Dipterans and could therefore be a potential target for pest control strategies. Identifying novel protein complexes in insects will provide better structural insights and contribute to a better understanding of their functions and mechanisms.

EXPERIMENTAL PROCEDURES

Insect species and source

L. cuprina cDNA was donated by Max Scott (Entomology and Plant Pathology, North Carolina State University). *T. castaneum* was generously supplied by Dr. Dieudonné Baributsa; *A. mellifera* by Dr. Brock Harpur; *C. lectularius* by Dr. Ameya Gondhalekar and *A. aegypti* by Dr. Catherine Hill (Entomology, Purdue University). *D. melanogaster* constructs were previously described in (Torres-Zelada et al., 2019).

RNA extraction, cDNA preparation and cloning

Total RNA was extracted from whole adult insects using the Direct-zol RNA kit according to the manufacturer's protocol (Zymo Research, Irvine, CA, USA). RNA (1 μg) was converted to cDNA using Super-Script III reverse transcriptase (Invitrogen, Carlsbad, CA, USA) and used as a template for PCR amplification of the specific region of interest of the Chiffon and Gcn5 orthologs in each insect using the indicated primers (Table 2). Briefly, the coding regions for Chiffon-B or Gcn5 were inserted into the pGADT7 (AD) or pGBKT7 (DBD) yeast two-hybrid vectors (Takara BIO, USA) as previously described (Stephenson et al., 2015). Constructs were confirmed by DNA sequencing.

Yeast two-hybrid assays

Plasmids were co-transformed into Y2HGold *S. cerevisiae* and yeast two-hybrid assays performed as described in the Matchmaker Gold Yeast Two-Hybrid System (Clontech Lab. Inc). Briefly, transformations were plated on SD/-leu/-trp at 30°C and then patched onto SD/-leu/-trp and replica plated onto SD/-leu/-trp and SD/-ade/-his/-leu/-trp media to test for interaction. Three independent transformed colonies were replica plated for all interactions tested.

Phylogenetic analysis

Phylogenetic trees were generated using the neighbour-joining method without distance corrections based on the Clustal-Omega multiple sequence alignment of protein sequences from the National Center for Biotechnology Information (NCBI) (Waterhouse et al., 2009). Accession numbers for the Chiffon/Dbf4 and Gcn5 proteins are provided in Tables S1 and S2, respectively. Schematic diagrams were generated using BioRender.com.

AUTHOR CONTRIBUTIONS

Smitha George: Conceptualization (equal); investigation (equal); writing – original draft (equal). **Hannah R Blum:** Investigation (equal); writing – original draft (supporting). **Eliana F Torres-Zelada:** Investigation (equal). **Grace N Estep:** Investigation (equal). **Youssef A Hegazy:** Investigation (equal); methodology (equal). **Gina M Speer:** investigation (equal). **Vikki Marie Weake:** Conceptualization (equal); data curation (equal); funding acquisition (equal); supervision (equal); writing – original draft (equal); writing – review and editing (equal).

ACKNOWLEDGEMENTS

The following undergraduate students contributed to the yeast two-hybrid studies as part of the BCHM101 laboratory: Harrison Clark, Nick Craft, Grace Estep, Malerie Kleiman, Jarrett Klusmeier, Adam Meroueh, David Mitchell, Mario Perez-Ahuatl, Lauryn Rhodes, Nick Roberts, Thomas Sheeley, Hern Tan, Nathan Tuchscherer and Payne Turney. Thanks to Stephanie Hathaway (Dr. Harpur Lab) for providing extra honey bee samples. Information from the FlyBase was used in this study (FB2021_05).

FUNDING INFORMATION

This work was supported by National Science Foundation (NSF) grant 1930237 to VMW.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supplementary material of this article.

ORCID

Gina M. Speer https://orcid.org/0000-0002-3766-6995
Vikki M. Weake https://orcid.org/0000-0002-5933-9952

REFERENCES

Bousset, K. & Diffley, J.F. (1998) The Cdc7 protein kinase is required for origin firing during S phase. *Genes & Development*, 12(4), 480–490. https://doi.org/10.1101/gad.12.4.480

Carré, C., Szymczak, D., Pidoux, J. & Antoniewski, C. (2005) The histone H3 acetylase dGcn5 is a key player in *Drosophila melanogaster* metamorphosis. *Molecular and Cellular Biology*, 25(18), 8228–8238. https://doi.org/10.1128/MCB.25.18.8228-8238.2005

Davis, G.K. & Patel, N.H. (2002) Short, long, and beyond: molecular and embryological approaches to insect segmentation. *Annual Review of Entomology*, 47(1), 669–699. https://doi.org/10.1146/annurev.ento.47.091201.145251

Drysdale, C.M., Dueñas, E., Jackson, B.M., Reusser, U., Braus, G.H. & Hinnebusch, A.G. (1995) The transcriptional activator GCN4 contains multiple activation domains that are critically dependent on hydrophobic amino acids. *Molecular and Cellular Biology*, 15(3), 1220–1233. https://doi.org/10.1128/MCB.15.3.1220

Guelman, S., Suganuma, T., Florens, L., Swanson, S.K., Kiesecker, C.L., Kusch, T. et al. (2006) Host cell factor and an uncharacterized SANT domain protein are stable components of ATAC, a novel dAda2A/dGcn5-containing histone acetyltransferase complex in *Drosophila*. *Molecular and Cellular Biology*, 26(3), 871–882. https://doi.org/10.1128/MCB.26.3.871-882.2006

Hirai, H., Tani, T. & Kikyo, N. (2010) Structure and functions of powerful transactivators: VP16, MyoD and FoxA. *The International Journal of Developmental Biology*, 54(11–12), 1589–1596. https://doi.org/10.1387/ijdb.103194hh

Kusch, T., Guelman, S., Abmayr, S.M. & Workman, J.L. (2003) Two Drosophila Ada2 homologues function in different multiprotein complexes. *Molecular and Cellular Biology*, 23(9), 3305–3319. https://doi.org/10.1128/MCB.23.9.3305-3319.2003

Landis, G. & Tower, J. (1999) The Drosophila chiffon gene is required for chorion gene amplification, and is related to the yeast Dbf4 regulator of DNA replication and cell cycle. *Development (Cambridge, England)*, 126(19), 4281–4293.

Liu, P.Z. & Kaufman, T.C. (2005) Short and long germ segmentation: unanswered questions in the evolution of a developmental mode. *Evolution and Development*, 7(6), 629–646. https://doi.org/10.1111/j.1525-142X.2005.05066.x

Masai, H. & Arai, K. (2000) Dbf4 motifs: conserved motifs in activation subunits for Cdc7 kinases essential for S-phase. *Biochemical and Biophysical Research Communications*, 275(1), 228–232. https://doi.org/10.1006/bbrc.2000.3281

Matthews, L.A. & Guarné, A. (2013) Dbf4. *Cell Cycle*, 12(8), 1180–1188. https://doi.org/10.4161/cc.24416

Melcher, K. (2000) The strength of acidic activation domains correlates with their affinity for both transcriptional and non-transcriptional proteins. *Journal of Molecular Biology*, 301(5), 1097–1112. https://doi.org/10.1006/jmbi.2000.4034

Muratoglu, S., Georgieva, S., Pápai, G., Scheer, E., Enünlü, I., Komonyi, O. et al. (2003) Two different Drosophila ADA2 homologues are present in distinct GCN5 histone acetyltransferase-containing complexes. *Molecular and Cellular Biology*, 23(1), 306–321. https://doi.org/10.1128/MCB.23.1.306-321.2003

Nagy, Z. & Tora, L. (2007) Distinct GCN5/PCAF-containing complexes function as co-activators and are involved in transcription factor and global histone acetylation. *Oncogene*, 26(37), 5341–5357. https://doi.org/10.1038/sj.onc.1210604

Raj, N. & Attardi, L.D. (2017) The transactivation domains of the p53 protein. *Cold Spring Harbor Perspectives in Medicine*, 7(1), a026047. https://doi.org/10.1101/cshperspect.a026047

Sadowski, I., Ma, J., Triezenberg, S. & Ptashne, M. (1988) GAL4-VP16 is an unusually potent transcriptional activator. *Nature*, 335(6190), 563–564. https://doi.org/10.1038/335563a0

Sanborn, A.L., Yeh, B.T., Feigerle, J.T., Hao, C.V., Townshend, R.J.L., Aiden, E.L. et al. (2021) Simple biochemical features underlie transcriptional activation domain diversity and dynamic, fuzzy binding to mediator. *eLife*, 10, e68068. https://doi.org/10.7554/eLife.68068

Schupbach, T. & Wieschaus, E. (1991) Female sterile mutations on the second chromosome of *Drosophila melanogaster*. II. Mutations blocking oogenesis or altering egg morphology. *Genetics*, 129(4), 1119–1136.

Scofield, S., Jones, A. & Murray, J.A.H. (2014) The plant cell cycle in context. *Journal of Experimental Botany*, 65(10), 2557–2562. https://doi.org/10.1093/jxb/eru188

Seller, C.A. & O'Farrell, P.H. (2018) Rif1 prolongs the embryonic S phase at the Drosophila mid-blastula transition. *PLoS Biology*, 16(5), e2005687. https://doi.org/10.1371/journal.pbio.2005687

Sigrist, C.J.A., Cerutti, L., de Castro, E., Langendijk-Genevaux, P.S., Bulliard, V., Bairoch, A. et al. (2010) PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Research*, 38-(Supplement 1), D161–D166. https://doi.org/10.1093/nar/gkp885

Soffers, J.H.M., Li, X., Saraf, A., Seidel, C.W., Florens, L., Washburn, M.P. et al. (2019) Characterization of a metazoan ADA acetyltransferase complex. *Nucleic Acids Research*, 47(7), 3383–3394. https://doi.org/10.1093/nar/gkz042

Stephenson, R., Hosler, M.R., Gavande, N.S., Ghosh, A.K. & Weake, V.M. (2015) Characterization of a Drosophila ortholog of the Cdc7 kinase: a role for Cdc7 in endoreplication independent of Chiffon. *The Journal of Biological Chemistry*, 290(3), 1332–1347. https://doi.org/10.1074/jbc.M114.597948

Torres-Zelada, E.F. & Weake, V.M. (2021) The Gcn5 complexes in *Drosophila* as a model for metazoa. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1864(2), 194610. https://doi.org/10.1016/j.bbagrm.2020.194610

Torres-Zelada, E.F., Stephenson, R.E., Alpsoy, A., Anderson, B.D., Swanson, S.K., Florens, L. et al. (2019) The Drosophila Dbf4 ortholog Chiffon forms a complex with Gcn5 that is necessary for histone acetylation and viability. *Journal of Cell Science*, 132(2). https://doi.org/10.1242/jcs.214072

Torres-Zelada, E.F., George, S., Blum, H.R. & Weake, V.M. (2022) Chiffon triggers global histone H3 acetylation and expression of developmental genes in *Drosophila* embryos. *Journal of Cell Science*, 135(2), jcs259132. https://doi.org/10.1242/jcs.259132

Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. & Barton, G.J. (2009) Jalview version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9), 1189–1191. https://doi.org/10.1093/bioinformatics/btp033

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Figure S1. A schematic illustrating conserved regions in Chiffon A and Chiffon-B within selected organisms is shown. Accession numbers used to generate the alignment are provided in Table S1. Alignments are available in Supplementary file 1.

Figure S2. A schematic illustrating conserved regions in Gcn5 within selected organisms is shown. Accession numbers used to generate the alignment are provided in Table S2. Alignments are available in Supplementary file 2.

File S1. Chiffon/Dbf4 alignment.

File S2. Gcn5 alignment.

File S3. Chiffon-B region alignment.

Table S1. Accession numbers for Dbf4/Chiffon.

Table S2. Accession numbers for KAT2A/Gcn5/Pcaf.

Table S3. Accession numbers for Chiffon from selected insect species.

How to cite this article: George, S., Blum, H.R., Torres-Zelada, E.F., Estep, G.N., Hegazy, Y.A., Speer, G.M. et al. (2022) The interaction between the Dbf4 ortholog Chiffon and Gcn5 is conserved in Dipteran insect species. *Insect Molecular Biology*, 31(6), 734–746. Available from: https://doi.org/10.1111/imb.12800