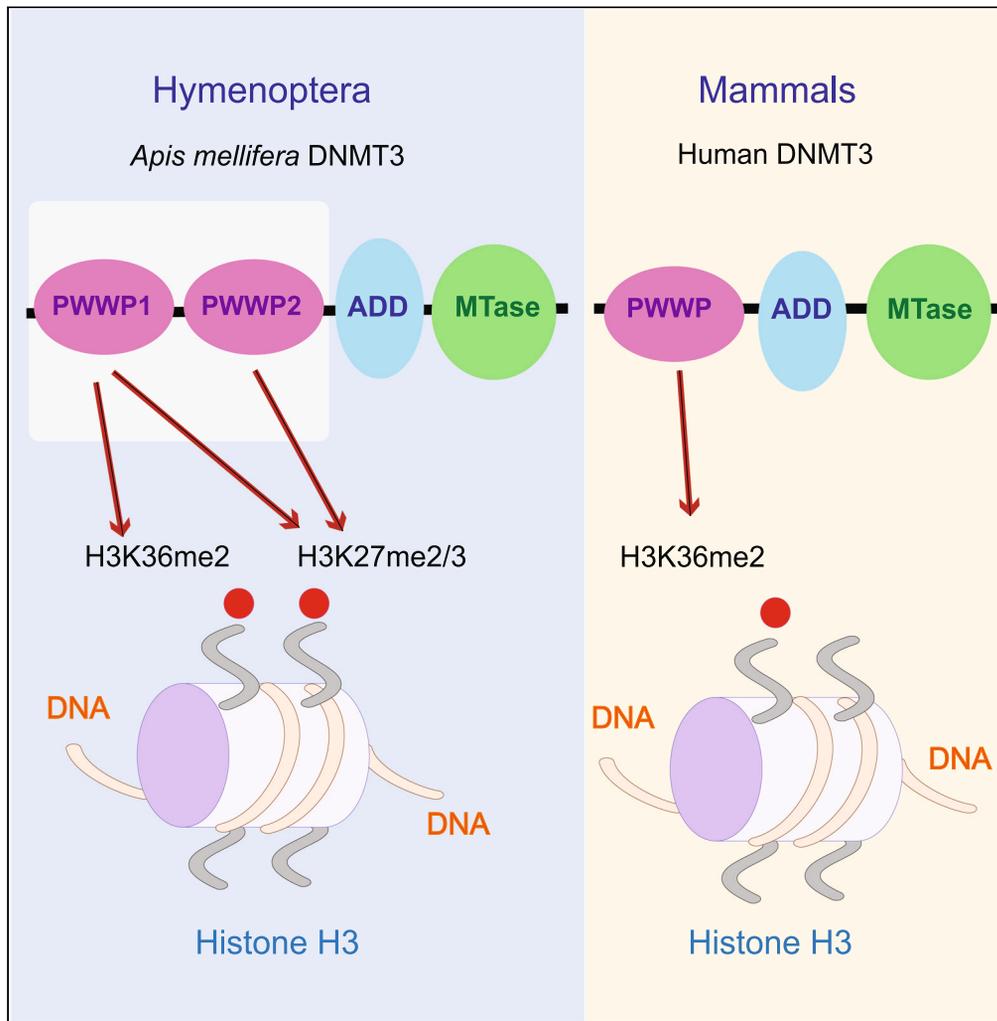**Article**

The PWWP domain and the evolution of unique DNA methylation toolkits in Hymenoptera

Robert Kucharski, Nancy Ellis, Tomasz P. Jurkowski, Paul J. Hurd, Ryszard Maleszka

p.j.hurd@qmul.ac.uk (P.J.H.)
ryszard.maleszka@anu.edu.au (R.M.)

Highlights

Insect DNA methylation remains largely unexplored

The PWWP domain duplication is a unique feature of DNMT3s in Hymenoptera

This study shows the first case of PWWP domains binding to H3K27 chromatin modification

DNA methylation system is highly diverse in Hymenoptera with variable copies of DNMTs

Article

The PWWP domain and the evolution of unique DNA methylation toolkits in Hymenoptera

Robert Kucharski,[1] Nancy Ellis,[2] Tomasz P. Jurkowski,[3] Paul J. Hurd,[2,*] and Ryszard Maleszka[1,4,*]

SUMMARY

DNMT3 in Hymenoptera has a unique duplication of the essential PWWP domain. Using GST-tagged PWWP fusion proteins and histone arrays we show that these domains have gained new properties and represent the first case of PWWP domains binding to H3K27 chromatin modifications, including H3K27me3, a key modification that is important during development. Phylogenetic analyses of 107 genomes indicate that the duplicated PWWP domains separated into two sister clades, and their distinct binding capacities are supported by 3D modeling. Other features of this unique DNA methylation system include variable copies, losses, and duplications of DNMT1 and DNMT3, and combinatorial generations of DNMT3 isoforms including variants missing the catalytic domain. Some of these losses and duplications of are found only in parasitic wasps. We discuss our findings in the context of the crosstalk between DNA methylation and histone methylation, and the expanded potential of epigenomic modifications in Hymenoptera to drive evolutionary novelties.

INTRODUCTION

Our long-term goal is to understand the genesis of gene regulatory networks whose divergence in different phyla generates the potential for evolutionary novelties. Recent progress in delineating the role of epigenetic mechanisms in controlling global gene expression has already provided important clues to the potential of DNA methylation and histone modifications to act as prime movers in adaptability and diversification.[1–4]

DNA methylation is one of the several epigenomic modifications that have been implicated in various cellular and organismal functions in both vertebrates and invertebrates.[5–9] It is a complex highly dynamic process that relies upon DNA methyltransferases (DNMTs), which by recruiting other regulatory proteins and affecting the activities of transcription factors modulate gene expression.[10,11] Although at the sequence level, the proteins required for this process are broadly conserved across the animal kingdom, the apparent diversity of their copy number and domain structure suggests that their present roles are lineage-specific and reflect adaptive events in individual evolutionary histories.[12]

Vertebrates have two types of well characterized DNA methyltransferases, one DNMT1 and three DNMT3s. DNMT1 is a post replication maintenance enzyme involved in the addition of methyl groups to the 5-position of cytosine, largely in CpG dinucleotides. It ensures that the methylation status of DNA is accurately copied on the newly synthesized DNA strand. DNMT3 is the so called *de novo* DNA methyltransferase which adds methyl marks to unmethylated genomic targets.[13,14] However, this catalytic distinction is fluid, and recent data in mammals show that cooperation between DNMTs is needed for the maintenance and *de novo* DNA methylation.[15] Moreover, one of the DNMT3 paralogs in mammals (DNMT3L) has no catalytic activity and acts as an accessory protein.[13,15] Based strictly on sequence similarity, relatives of DNMT1 and DNMT3 have been found in insects, but their taxon-specific distribution is distinct from that in mammals, with some insects having both DNMTs and the others only one copy of DNMT1 and no DNMT3.[16–18] Although some earlier assumptions that DNMTs in social insects are mechanistically similar to their vertebrate counterparts,[19,20] the mosaic distribution of DNMT paralogs, and a very low level of genomic methylation mostly restricted to coding regions of conserved genes suggest that this level of genomic modifications is not only different from that in mammals, but also is driven by lineage-specific mechanisms rather than by a universal invertebrate-type enzymology.[3,16–18,21]

Recently, we have observed that in the honeybee, *Apis mellifera*, DNMT3 has an unusual domain architecture comprising a duplicated PWWP domain.[21] In mammals, the PWWP domain acts as a chromatin methylation reader by simultaneously binding DNA and histone methyl-lysines within the context of the nucleosome, and is essential for targeting enzymes that catalyze these marks to appropriate genomic sites.[22] In mammals, the PWWP domain specifically recognizes H3K36me3, suggesting a functional role for this domain as a H3K36me3 sensor.[22] Mutations in this domain are associated with altered genomic DNA methylation profiles and human disease.[23,24] This connection

Table 1. Variable copy number of DNMTs in the Hymenopteran order

Number of genes					Histone methylation readers with the PWWP domain	
DNMT1	DNMT3	Taxon	PWWP Duplication	TET one gene	ZCWPW1 zinc finger CW-type	ZMYND11 zinc finger MYND-type
2	1	Most common complement of DNMTs found in the large superfamily of *Apoidea*, a small superfamily of *Pompiloidea*, and in a few families belonging to *Chalcidoidea* (*Agaonidae, Aphelinidae, Encyrtidae, Eupelmidae, Eurytomidae, Mymaridae, Ormyridae*)	Yes	Yes	Missing in some *Aculeata* (e.g., in 50% of ants) and in *Chalcidoidea, Braconidae* and *Symphyta*. Duplicated in *Cynipoidea*	Missing in some *Apoidea* and in all *Chalcidoidea* except *Mymaridae* (fairyflies)
1	1	*Formicoidea* (ants) except three *Leptanilla* species that lost DNMT3, superfamily *Vespoidea*, except the *Polistinae* subfamily that has no DNMT3	Yes	Yes		
3	1	Three wasp *Chalcidoidea* families: *Megastigmidae, Torymidae* and *Pteromalinae* (subfamily)	Yes	Yes		
2	2	Only one family *Dryinidae* (parasitic wasps) has this complement of DNMTs	Yes	Yes		
1	0	These two contrasting complements of DNMTs are found in Braconid parasitoid wasps (*Braconidae*)	Yes**	Yes		
0	1					

between two primary epigenomic marks is responsible for mediating much of the crosstalk between DNA methylation and modified histones.[25,26] The patterns of DNA and histone modifications are combinatorial, and are driven by domain architecture of both DNMTs and histone modifying proteins, as well as their accessory proteins. The intricate interactions of chromatin-binding proteins and epigenomic marks create a network of sophisticated interactions that control gene expression by activating or restricting the transcriptional potential of genomic regions.[25] To date only a couple of studies on this important interaction have been conducted in insects, confirming that DNA methylation and histone modifications are two components of a complex multilayer regulatory network.[27–29]

Given the pivotal role of the PWWP domain and the apparent uniqueness of this duplication in the animal Kingdom, we hypothesize that such a structural change in DNMT3 is an original invention specific to this taxon that modulates the selectivity of DNA methylation system in Hymenoptera. To this end we explored over one hundred Hymenopteran genomes deposited to NCBI to compare their DNA methylation toolkits with the focus on DNMTs in Hymenoptera (sawflies, wasps, ants, and bees). Although the copy number of DNMTs is variable, the duplicated PWWP domain appears to be a unifying signature of all DNMT3s in this order. The duplicated domains have diversified early in the evolutionary history of Hymenoptera and phylogenetically can be separated into two sister clades. Importantly, we show that the duplicated PWWP domains have gained novel binding capacity for H3K27. Here we describe the methylation toolkit in Hymenoptera as the most conspicuous example of a remarkable diversity of this important epigenomic modification and debate its functional implications in the evolutionary context.

RESULTS

Widespread diversity of the epigenetic toolkits in Hymenoptera

We have analyzed 107 Hymenopteran genomes representing 33 families, 74 subfamilies and 95 genera available in GenBank with the last check for new entries done in April 2022. Table 1 shows the DNMT1 and DNMT3 gene number in the analyzed taxa. All analyzed species and more details are provided in Table S1. In most families the number of genes encoding DNMT1 varies from one to three, but in some species this gene is missing. The gene encoding DNMT3 is single copy except for a few genera such as *Polistes* in subfamily *Polistinae* that lost this gene (see also ref.[30]), and one family *Dryinidae* that has a rare duplication of DNMT3, so far not reported in other insects. Among the six arrangements uncovered in this study, the most common, consisting of two paralogs of DNMT1 and one copy encoding DNMT3 is
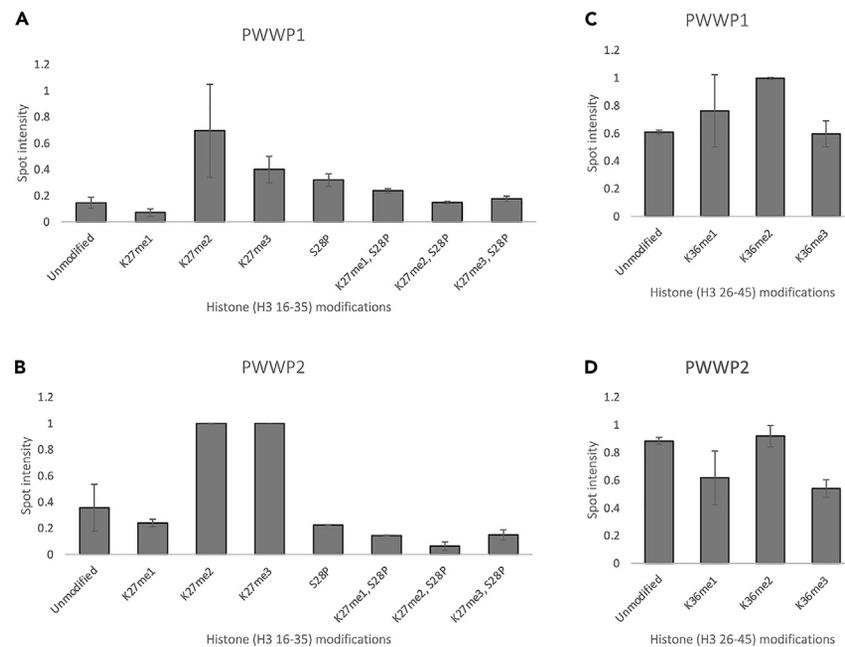
Figure 1. *Apis mellifera* DNMT3 PWWP1 and PWWP 2 domains have contrasting histone modification binding preferences

Quantification of binding of GST-PWWP1 and GST-PWWP2 to MODified histone peptide arrays using anti-GST antibody. Raw spot intensities were normalized to the highest spot signal intensity (=1). For PWWP1 the highest spot intensity is H3K36me2 and for PWWP2 the highest intensity spot is H3K27me2. Bar plots represent the mean of normalized raw spot signal intensities from the left and right-side replicates of the MODified histone peptide arrays. Error bars show one standard deviation.

(A) GST-PWWP1 binding specificity to histone H3 amino acids 16–35.

(B) GST-PWWP2 binding specificity to histone H3 amino acids 16–35.

(C) GST-PWWP1 binding specificity to histone H3 amino acids 26–45.

(D) GST-PWWP2 binding specificity to histone H3 amino acids 26–45. The GST purification of the PWWP domains is shown in Figure S7.

found in the large superfamily of *Apoidea* (bees), in a small superfamily of *Pompiloidea* (includes velvet ants and spider wasps*)*, and in a few families belonging to *Chalcidoidea* (mostly parasitoid wasps) (Table 1). *Formicoidea* (ants) have one copy of each DNMT except the subfamily *Leptanillinae*. Braconid wasps can be divided into two contrasting groups, having either one DNMT1 or one DNMT3 and missing the other DNMT. The loss of DNMT1 in *Braconidae* is unusual especially among insects where most lineages have a single DNMT1 gene.[16,18,31] Indeed, wasps appear to have the most diverged DNA methylation toolkit.

Table 1 also includes two histone methylation readers with the PWWP domain ZCWPW1 and ZMYND11. These two readers are among 75 highly conserved proteins involved in writing, reading and erasing epigenomic marks on histones that we have identified in Hymenoptera, but are the two members of this epi-kit missing in some families (Table S2). In one family *Cynipoidea* that lost DNMT3, there are two copies of ZCWPW1.

The duplicated PWWP domain is present in all Hymenopteran species that have DNMT3. However, the phylogenetic tree shows that the PWWP1 and PWWP2 domains separated into two sister clades early in the evolution of this order (Figure S1 based on alignments in Figure S4). This result suggests that DNMT3s associated with these two clades may have different binding properties (Figures 1 and 3). In total, there are 11 proteins with the PWWP domain in the honeybee genome in comparison to more than 20 in mammals.[22]

In addition, we show that a single copy gene encoding a relative of mammalian TETs (ten-eleven translocation methyl-cytosine dioxygenases) is found in all analyzed species regardless of the number of DNMT genes. While there is some evidence that such as in mammals, TET in honeybees is involved in demethylation by converting 5-methyl-cytosine to 5-hydroxyl-cytosine, it also might perform other regulatory functions similar to those shown in mammals.[32,33]

In attempting to align the bioinformatics and molecular results with the evolutionary history, we have contrasted DNMT-related innovations with a recent and most detailed phylogenetic tree of Hymenoptera.[34] This analysis suggests that DNMT losses, or duplications occurred independently in several lineages, whereas the loss of introns in DNMT3 is found in parasitic wasps, and in one species of willow-galling sawflies, *Euura lappo* (Tenthredinidae). The DNMT-related changes are most frequent in parasitic wasps. One extra exon in DNMT3 is found only in the superfamily of sawflies (*Tenthredinoidea*). The non-canonical 5′ splice sequence GC in DNMT3 intron 3 appears to be restricted to several subfamilies of *Apoidea*. In addition, the unusual distribution of DNMTs in *Braconidae* (Table 1) can be resolved by assigning the DNMT−/DNMT3+ situation to subfamily *Cyclostomes s.l.,* and the DNMT1+/DNMT3− situation to do *Non-Cyclostomes* using phylogeny of *Ichneumonoidea* in ref.[35]
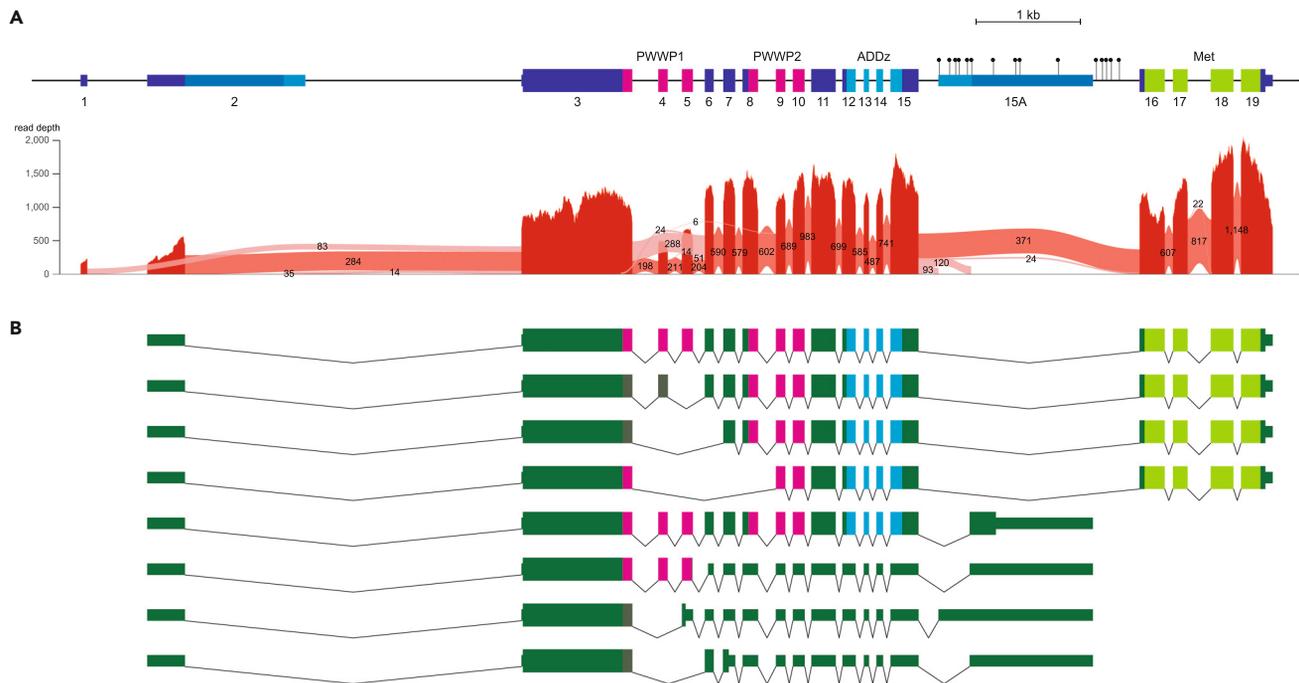
Figure 2. A. mellifera DNMT3 gene structure and combinatorial generation of isoforms

(A) The Sashimi plot aligned with the current A.mellifera DNMT3 gene (Linkage Group LG2; RefSeq accession: NC_037639.1). It shows quantitative comparison of exon usage across three pooled samples from the honeybee pupal brains in diploid females (queens, workers) and haploid males (drone). The black lollipops indicate the methylated CpG dinucleotides. In the latest A.mellifera genome assembly Amel_HAv3.1 this gene is located on chromosome 2 (LG2: 15,815,881-15,827,940).

(B) Examples of alternatively spliced transcripts producing diverse isoforms including a combinatorial usage of the PWWP domain or missing the catalytic domain. Our full analysis of combinatorial splicing patterns is shown in Table S3.

A. mellifera DNMT3 PWWP1 and PWWP2 domains have novel and contrasting histone binding preferences

The PWWP domains of mammalian DNMT3A and DNMT3B have previously been reported to bind H3K36me2 and H3K36me3[26,36] and mutations in these domains are associated with altered genomic DNA methylation profiles and human disease.[23,37,38] Protein alignment of *Apis mellifera* PWWP1 and PWWP2 with mouse and human DNMT3A and DNMT3B PWWP domains reveals that sequence identity between mammalian PWWP and *A. mellifera* PWWP1 and PWWP2 is 33% and 40%, respectively (Figure S8). The PWWP1 domain shows more differences in the region of the aromatic cage involved directly in histone binding. Therefore, to determine the histone binding preferences of honeybee PWWP domains, we used MODified histone peptide arrays containing 384 unique histone post-translational modification combinations.

The PWWP1 and PWWP2 domains of *A. mellifera* DNMT3 show highest affinity for H3K27me2 and H3K27me3 modified peptides (Figure 1). For PWWP1, there is approximately a 5-fold increase in binding to H3K27me2 and a 3-fold increase in binding to H3K27me3, compared to the corresponding unmodified peptide (Figure 1A). For PWWP2, there is approximately a 3-fold increase in binding to both H3K27me2 and H3K27me3 modified peptides compared to unmodified (Figure 1B). Significantly, for both PWWP1 and PWWP2, binding to H3K27me2/me3 is ablated by the phosphorylation of neighboring H3S28 (see Figures 1A and 1B). PWWP1 also shows affinity to H3K36me2, with approximately a 30% increase in binding compared to the corresponding unmodified peptide (Figure 1C). In contrast, PWWP2 shows no significant binding above background (H3 26–45 unmodified peptide) to H3K36 modified peptides (Figure 1D). Taken together, the two PWWP domains of Apis mellifera have contrasting histone modification binding preferences and furthermore, are the first examples of PWWP domains binding to H3K27 chromatin modifications.

Alternative splicing generates multiple isoforms affecting the functional domains in Hymenoptera DNMT3s

The DNMT3 transcription unit in Hymenoptera ranges from ~12kb in *A. mellifera* to over 100kb in some ant species, with multiple exons encoding four conserved domains, two PWWP domains, the ADDz domain and the catalytic domain MTase[14] (Figure 2). In *A. mellifera* this gene has 20 exons and belongs to the category of methylated genes with all methyl-cytosines confined to a repeat-rich region between exons 15 and 16. A predicted nuclear localization signal is located near the N-terminus of the protein.[39] Given the potential of this multiexon gene to generate alternatively spliced variants, we explored both our own and publicly available transcriptomes to determine the extent of DNMT3 protein repertoire. While alternative splicing (AS) is a ubiquitous mechanism utilized in all tissues throughout the entire life cycle, the highest level of splice variants is found in the nervous system.[40] Thus, we have examined this process in more detail using the RNAseq datasets

available for the *A. mellifera* brains in all three phenotypic outcomes, two female castes queens and workers, and haploid males (drones). The results summarized in Figure 2 show the variety of DNMT3 splice variants in the pupal brains during an important stage of nervous system metamorphosis.[41] Our analysis reveals three gene regions affected by AS, the PWWP domains encoding exons, the 5'-end (alternative transcriptional start site), and the methylated intron.

The exons spanning the duplicated region retain the respective ORFs, but in addition to the dominant isoform with two PWWP domains resulting from the splicing of exons 3-4-5, and 8-9-10, two chimeric isoforms with the PWWP domains combined into one unit can be produced, by splicing exons 3-4-10 or 3-9-10. The apparent duplication creates splicing sites very close to each other, and is likely to have an impact on the functioning of the spliceosome/RNA pol II complex.[42,43]

The frequent usage of splicing sites in the Δexon4-6 variant affects the PWWP1 domain and generates transcripts with only the PWWP2 domain. The products with the exon4-5 deletion lack all functional domains. Proteins encoded by transcripts terminating in the methylated intron 15 lose the MTase catalytic domain. Some of the MTase-minus variants have both PWWP domains and the ADDz domain, some have only the PWWP1 domain, and the others have no identifiable functional domains (Figure 2; Table S3). The variants without the catalytic domain resemble the mammalian isoforms that act as accessory proteins.[44] The non-canonical 5' splice sequence of intron 3, involved in AS events, is GC, so far found only in several subfamilies of *Apoidea*. The capacity of isoform generation may be reduced by intron loss as seen in a highly diverse superfamily of parasitic wasps *Ichneumonidea*. All splice variants generating various isoforms affecting the duplicated PWWP domains are predicted to modify the targeting of DNMT3s to specific chromatin regions through altered histone PTM binding.

Although the overall DNMT3 gene structure is similar across Hymenoptera, NCBI database searches reveal that in the superfamily of *Tenthredinoidea* there is an extra exon between exons 7 and 8 which is not included in the consensus model. The alternative "non-coding" exon 15A in Hymenoptera mostly encodes 3'UTR, in contrast to other insects in which exon 15A is spliced in and codes for part of the linker between the ADDz and MTase domains. An additional exon between exons 11 and 12 is present in some insects (Table S3). Table S3 also shows exon lengths consensus in analyzed lineages with lengths of the domains coding exons being more stable (or unchanging) than those coding for linkers. These extra coding exons, in combination with variable exon lengths, change the sizes of linker polypeptides between the adjacent domains, and they may be of special importance because of their potential impact on the DNMT3 function.

PWWP secondary structure

To determine if the distinctive binding capacities of the PWWP domains can be visualized via modeling of their secondary structures, we have aligned the honeybee PWWP domains 1 and 2 core sequences with human DNMT3B (Figures 3A, S4, and S8). The residues through which the PWWP domain acts as a histone methyl-lysine reader form a conserved aromatic cage are denoted with the red rectangles.[22,45] The Trp residues 239 and 263, and Ser270 are conserved, whereas Asp266 and Ile233 are conserved only in PWWP2, but are replaced in PWWP1 with Glu and Cys respectively. The Phe236 residue is replaced with His in PWWP2 and shows a gap in PWWP1. Another important structural feature for histone interactions is the β2—β3 loop that forms one of the walls of the binding pocket.[22,45] This region is less conserved and may be involved in defining ligand binding specificity. Because the PWWP domain binds histones and DNA in separate regions, it can concurrently interact with methylated histone tails and nucleosomal DNA. Combined binding to methylated histone tails and DNA leads to the specific recognition of H3K36me3-containing nucleosomes in mammals. The aromatic cage is encoded by exons 1 and 2 that undergo extensive splicing generating variants affecting the PWWP diversity and consequently, the binding capacity of DNMT3. The aromatic cage of the PWWP domain 1 shows more sequence divergence that might indicate the source of a different specificity. The Ile233 residue is involved in binding preferentially H3K36me3 via hydrophobic contacts surrounding the trimethylated lysine 36. In the the crystal structure of DNMT3B PWWP domain–H3K36me3 complex Pro38 is involved in a hydrophobic contact with the side chain of Ile233.[45] This residue is conserved in the PWWP2 domain, but has cysteine in the PWWP1 domain.

The 3D image (Figure 3B) was generated with the SwissModel server and the PWWP domain of mammalian Dnmt3b structure as template (PDB accession: 1KHC). While this approach is relatively simple, it does show that the cage residues clearly differ between the PWWP1 and PWWP2 domains. The 3D modeling confirms the evolutionary history of this duplication that suggests that PWWP1 and PWWP2 separated early into two sister clades (Figure S1). In addition, their evolutionary rate appears to be faster than that of the catalytic domain, and more in line with the ADDz domain (Figure 3C, also see Figures S1 and S2).

In Figure 4 we show additional evidence that following the duplication of the ancestral PWWP domain, the two derived PWWP domains were under discrete evolutionary pressures that generated their binding distinctiveness. In most Hymenoptera, the PWWP1 and PWWP2 domains have different phylogenetic branch lengths representing amino acid substitutions per site. Interestingly, in one family of *Chalcidoidea* these lengths are similar and have large values. Also, this family shows more pronounced changes in the ADDz domain relative to the other families. This is an important example of the remarkable diversity of the DNA methylation toolkits and their potential for lineage-specific functionality.

DISCUSSION

Our analyses reveal a remarkable variability of the epigenetic proteome in Hymenoptera with great disparities in gene number encoding DNMTs. Although previous studies already uncovered a pool of distinct DNA methylation toolkits in insects,[16,21,34] finding such diversity in one order is surprising. Several unique features of the epigenetic toolkit in such a vastly diverse order make this level of regulation particularly interesting from the mechanistic perspective and the potential for driving evolutionary innovations.
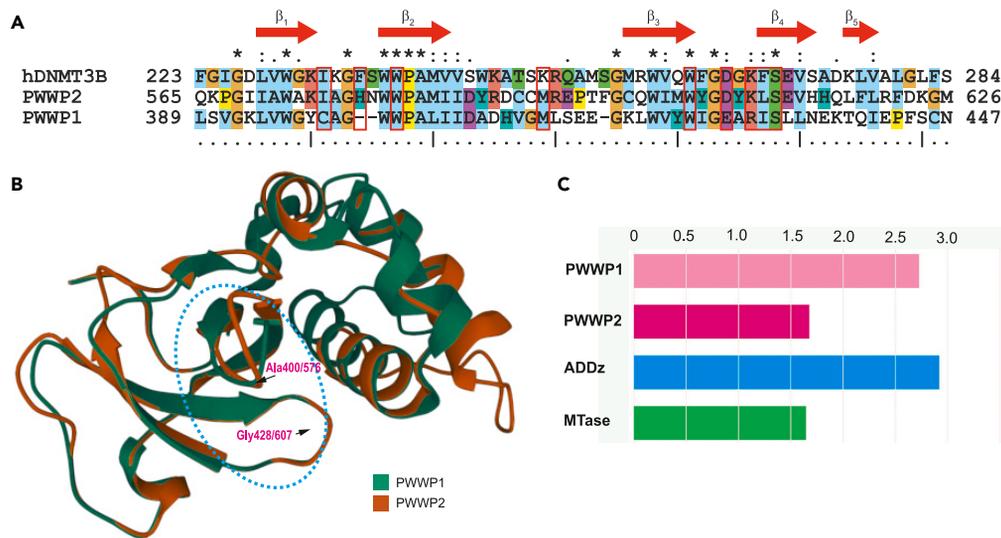
Figure 3. Predicted 3D model of the honeybee PWWP domains

(A) Alignment of the *A. mellifera* PWWP domains with the corresponding region of the human DNMT3B isoform 1 (NP_008823.1). An expanded alignment is shown in Figure S4.

(B) Predicted superimposed 3D models of the PWWP1 and PWWP2 domains. The positions of cage residues occupying the turns between β1 and β2, and β3 and β4 are indicated with a dotted oval.

(C) Evolutionary rate of amino acid substitutions in the PWWP, ADDz and MTase domains. Full alignments of the PWWP, ADDz and MTase domains are shows in Figures S4–S6.

The high level of diversification of the DNA methylation toolkit in Hymenoptera affects two important epigenomic modifiers, DNMT1 and DNMT3, that includes losses, duplications, and structural modifications. Our seminal finding that these domains have a novel binding capacity to recognize H3K27me is of great importance and brings into focus previously unknown targeting potential of the PWWP domain. Together with combinatorial generations of protein isoforms, the duplicated PWWP domains in DNMT3 provide an additional level of flexibility for setting up new interactions with other epigenomic modifiers and regulatory proteins. Interestingly, the PWWP sequence alignments suggest that these duplicated domains diversified into two sister clades at early stages of Hymenopteran evolution, thus additionally enhancing the potential of DNMT3 for molecular interactions. This notion is supported by 3D modeling showing that the cage residues, through which the PWWP domain acts as a histone methyl-lysine reader, differ between the PWWP1 and PWWP2 domains. The aromatic cage of the PWWP2 domain shows more sequence divergence that may indicate the source of a different specificity. In addition, their evolutionary rate appears to be faster than that of the catalytic domain, and more in line with the ADDz domain.

Given this rare but crucial modification, the significance of the PWWP domain for the enzymological engine room of Hymenopteran methylomics cannot be emphasized enough. The implications of such an evolutionary invention are best described by the expanded protein's role as "an element in a network of protein–protein interactions, in which it has a contextual or cellular function within functional modules."[46] It is expected that the phenotypic impact of DNMT3s in various families of Hymenoptera will be affected by their lineage-specific topological position in the complex hierarchical web of molecular interactions.

The extensive generation of alternatively spliced DNMT3 variants in the honeybee is largely driven by the PWWP domains coding region and might be a consequence of the duplicated splicing signals within this region.[47] The usage of the non-coding methylated exon 15A confirms prior reports investigating the impact of DNA methylation on splicing.[48–51] Most intriguing DNMT3 isoforms without the catalytic domain resemble mammalian paralogs acting as accessory proteins. In mammals, the necessity of chromatin components for DNA methylation is well documented and *de novo* DNA methylation by DNMT3A requires the alteration of chromatin structure.[52] One determinant of this process is the ability of accessory protein DNMT3L to bind the N-terminus of histone H3 tail with an unmethylated lysine 4.[53] While the transferability of biological function above the protein sequence level for DNMTs is not straightforward, the expression of non-catalytic variants suggests that these isoforms in Hymenoptera are counterparts of mammalian DNMT3L involved in governing DNA methylation. DNMT3L contains FYVE zinc finger domain belonging to the same superfamily as ADDz. With the addition of the PWWP domain(s) the Hymenopteran counterparts (i.e., MTase-less isoforms) may have additional functionality.

Genome-wide analyses and molecular data have already provided strong support for the role of alternative splicing in driving key evolutionary innovations in brain development, phenotypic plasticity, adaptation, and species divergence (reviewed in[54]). The extent to which the richness of alternatively spliced DNMT3 variants was a contributing factor to evolutionary innovations in Hymenoptera needs to be experimentally examined by comparative genomics and transcriptomics of species within this order.

While the functionality DNMT3-PWWP in Hymenoptera is only beginning to be understood, some important ideas for future studies come from mammalian studies. An interesting study on a mouse mutant carrying a point mutation in the DNMT3A PWWP domain has shown that in
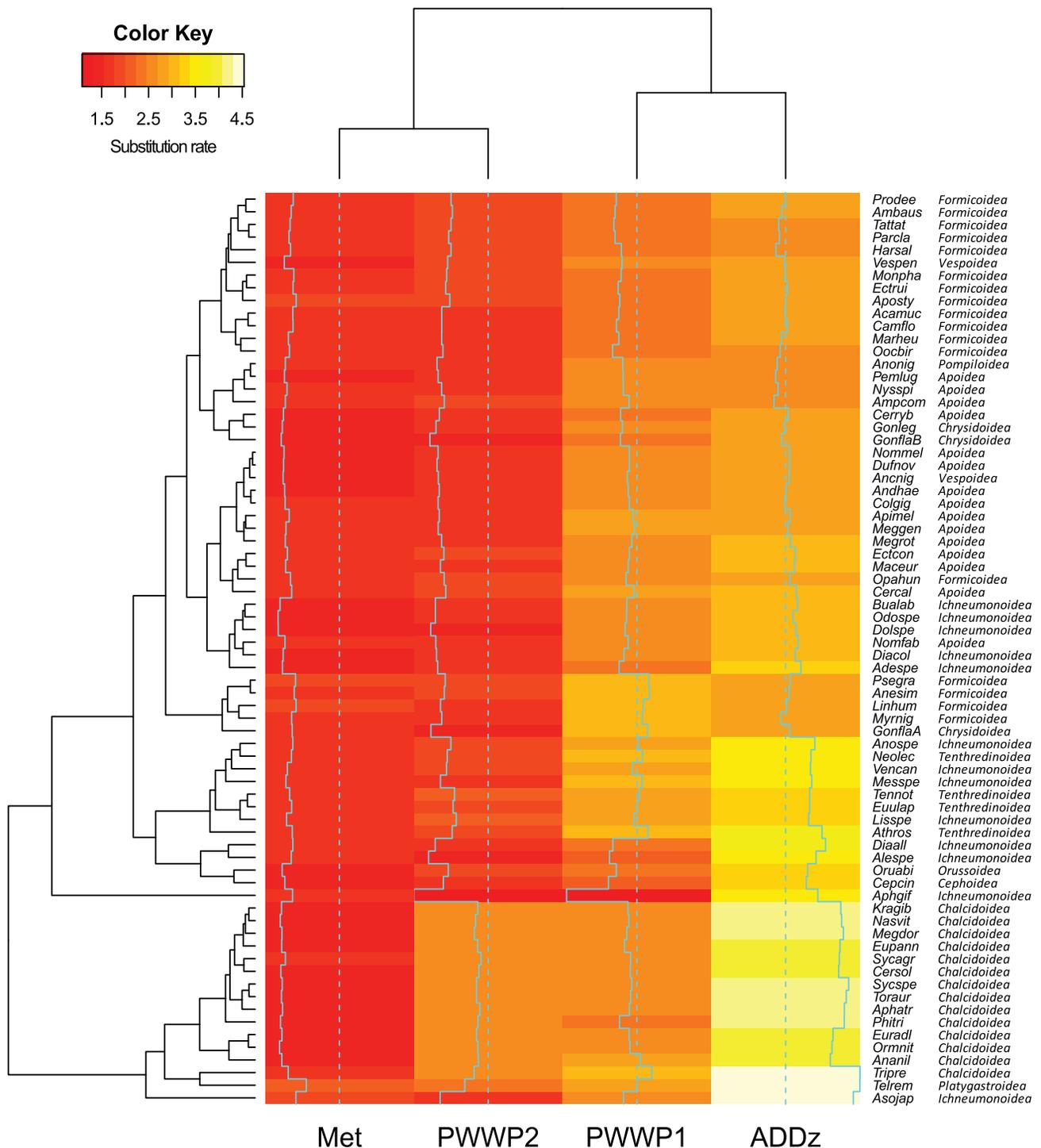
Figure 4. Heatmap showing some interesting patterns in the PWWP domain evolution

In most Hymenopterans the PWWP1 and PWWP2 domains have different branch lengths representing amino acid substitutions per site, but in one family of *Chalcidoidea* these lengths are similar and have large values. Also, this family shows more pronounced changes in the ADDz domain relative to the other families.

addition to postnatal growth retardation this mutation causes hypermethylation across domains marked by H3K27me3 and bivalent chromatin.[55] It is conceivable that the unique and more elaborate mode of reading histone modifications in Hymenoptera is associated with the distribution of 5mCs in this taxon, which is predominantly found in exonic regions of conserved genes.[48,56]

Besides Hymenoptera, there are other insect in families such as Coleoptera and Orthoptera that have a copy of DNMT3 with the consensus domain architecture, although some claims that aphids (Hemiptera) have a conserved DNMT3[57,58] need to be re-evaluated because the available NCBI data suggest that the DNMT3 hallmark domains, PWWP and ADDz are encoded by separate genes. In *Acyrthosiphon pisum*, the PWWP-ADDZ domains are encoded by GenBank: XP_029348651.1, and the MTase domain by GenBank: XP_016662566.1, possibly exemplifying yet another highly diverse DNA methylation toolkit. Genes encoding only a single MTase domain are also found in some nematodes,[12] and together with the aphid genes may represent an intriguing situation in the evolutionary history of these eukaryotic enzymes.[59] Among hundreds of insects with sequenced genomes deposited in the NCBI, the majority have a simplified DNA methylation toolkit and use only DNMT1 that has no PWWP and ADDz domains and must interact differently with modified histones. So far only a limited capacity of DNMT1 to direct readout of the heterochromatic mark H3K9me3 by the RFTS domain (replication focus targeting sequence) has been shown.[60] We also note that the lack of DNA methylation enzymology is more widespread among insects than previously thought. In addition to Diptera[21,61] and Strepsiptera[62] that lost DNA methylation enzymology, we have found that two genomes of *Chrysoperla carnea* and *Chrysopa pallens* in the order of Neuroptera (family Chrysopidae) do not encode proteins with sequence similarity to DNMT1 and DNMT3 (NCBI genome database).

While the proteome involved in histone modifications is highly conserved in Hymenoptera, two histone methylation readers with the PWWP domain ZCWPW1 and ZMYND11 have been lost in some families, notably in ~50% of ant species. Both proteins also have zinc finger domains with topology resembling the PWWP-ADDz tandem in DNMT3s.

Our understanding of the phylogenetic relationships among all major lineages of Hymenoptera has been improved by a study in which 3,256 protein-coding genes in 173 species were used to explore the early history of these insects, as well as the origins and radiation of parasitoids, stinging wasps, and bees.[34] At this stage only broad inferences can be made to link the diversity of the epigenetic machinery with the evolutionary history of this insect order. Hymenoptera are easily distinguishable from other insects, by two pairs of membranous wings (except for some castes that are wingless), the "wasp waist" (except for sawflies), the larger forewings, that are kept together by hooks, and females with hardened ovipositor. These morphological features can evolve using largely conserved pathways but what is actually used will be the result of the evolutionary history of each genome.[63] One example in Hymenoptera is gene reuse that facilitated rapid radiation and independent adaptation to disparate habitats in the Asian honeybee *Apis cerana*.[64] These authors provide evidence that the first, possibly regulatory intron in the neuronal Leucokinin receptor gene was under repeated selection and conclude that gene reuse provides a simple solution to a challenging process of foraging by modulating collective tendency for pollen/nectar collection as an adaptation to floral changes.[64] In our analyses, the most frequent occurrence of either losses or duplications of DNMT1, losses of DNMT3 and intron loss in DNMT3 are associated with parasitoid wasps, a species-rich group with an astonishing biology including unusual ecology, life cycles and extreme miniaturization.[65,66]

In a recent study relevant to our findings, Romiguier and colleagues[67] elucidate some aspects of the ants' evolutionary history concerning their phylogeny and the emergence of eusociality. They conclude that three functional gene categories, histone acetylation, gene silencing by RNA, and autophagy are linked to the main aspects of complex eusociality. They highlight positive selection on histone acetyltransferase, which previously has been linked to functions relevant to eusociality as a key facilitator of caste differentiation.[28,68,69] However, we note that in addition to histone acetyltransferase KAT6B, the list of 110 ant genes under positive selection reported in this article also includes other histone modifying enzymes; histone-lysine N-methyltransferase ash1, histone demethylase UTY-like, methyltransferase activity, histone demethylase activity (H3K27-specific) and epigenetic readers such as bromodomain-containing protein DDB, suggesting the involvement of many epigenetic players in shaping eusociality.

While several authors concluded on the basis of broad genome-wide methylation patterns that DNA methylation does not correlate with sociality in Hymenoptera,[16,70] such notions are premature. We simply do not know how emergent behaviors such as sociality evolve. Evolution can select for distinct life styles and reproductive modes by different initial constraints. Sociality in insects evolved several times[71] and represents a classic example of true degeneracy because the same endpoint was achieved by different mechanisms.[72,73] Insects and other invertebrates were able to have massive, selected forces exerted upon them owing to large numbers (and short life cycles), compared to many vertebrates, which do not even reproduce till later stages of their lives. At the genomic level, it is all about stochasticity and epigenomic modifications in their various forms. The pathways from epigenomic changes to complex phenotypes and behavior are indirect, multi-level, and despite much research are largely unknown. New designs emerge from lower levels of organization without any pre-existing determinations. Until we fully comprehend the tissue and cell-type DNA methylation dynamics in Hymenoptera, especially during early development and in the nervous system,[12] many ideas regarding the role of this epigenomic modification in insect epigenetics remain unsupported. Only then we will be able to tackle the challenge of molecular origins of unique and sophisticated novelties that emerged in this order such as the acquisition of a symbolic communication dialect in a highly eusocial genus *Apis*.[74]

Our findings reinforce the necessity for systematic functional analyses to determine how context-dependent roles of DNMTs have diversified and how their functionality contributed to major evolutionary transitions in Hymenoptera history. With more attention now directed to changes in regulatory circuits, and the role of hierarchical genetic networks in insects[75] we will be able to make progress in understanding how distinct evolutionary traits were shaped by epigenetic inventions.

Limitations of the study

Given the histone modification arrays are based on pre-designed peptides, we cannot exclude the possibility of additional binding preferences in the context of the entire nucleosome nor to modifications that are not currently represented on the array.

Our molecular and evolutionary conclusions carry the proviso that the analyzed sample of 107 genomes is representative of all Hymenopteran species.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - Phylogenetic trees construction
  - Honey bee DNMT3 alternative splicing analysis
  - Cloning, over-expression and purification of PWWP domains
  - Histone peptide arrays and binding
  - 3D modelling of the PWWP domain
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.108193.

AUTHOR CONTRIBUTIONS

Conceptualization, R.K., P.J.H., and R.M. Validation, R.K. Formal analysis, N.E., P.J.H., and R.K. Investigation, R.K. and N.E. Resources, P.J.H., T.P.J., and R.M. Writing – original draft, R.M. Writing – review & editing, R.K., P.J.H., and R.M. Visualization, R.K., R.M., and P.J.H. Supervision, P.J.H. T.P.J., and R.M. Project administration, R.M. and P.J.H. Funding acquisition, P.J.H. and R.M.

DECLARATION OF INTERESTS

Authors declare that they have no competing interests.

REFERENCES

1. Stajic, D., and Jansen, L.E.T. (2021). Empirical evidence for epigenetic inheritance driving evolutionary adaptation. Philos. Trans. R. Soc. Lond. B Biol. Sci. *376*, 20200121. https://doi.org/10.1098/rstb.2020.0121.

2. Sonawane, A.R., DeMeo, D.L., Quackenbush, J., and Glass, K. (2021). Constructing gene regulatory networks using epigenetic data. NPJ Syst. Biol. Appl. *7*, 45. https://doi.org/10.1038/s41540-021-00208-3.

3. Burggren, W.W. (2017). Chapter One - Epigenetics in Insects: Mechanisms, Phenotypes and Ecological and Evolutionary Implications. In Advances in Insect Physiology, H. Verlinden, ed. (Academic Press), pp. 1–30. https://doi.org/10.1016/bs.aiip.2017.04.001.

4. Carroll, S.B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell *134*, 25–36. https://doi.org/10.1016/j.cell.2008.06.030.

5. Feil, R., and Fraga, M.F. (2012). Epigenetics and the environment: emerging patterns and implications. Nat. Rev. Genet. *13*, 97–109. https://doi.org/10.1038/Nrg3142.

6. Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. Nat. Rev. Genet. *14*, 204–220. https://doi.org/10.1038/nrg3354.

7. Day, J.J., and Sweatt, J.D. (2011). Epigenetic Mechanisms in Cognition. Neuron *70*, 813–829. https://doi.org/10.1016/j.neuron.2011.05.019.

8. Lockett, G.A., Wilkes, F., and Maleszka, R. (2010). Brain plasticity, memory and neurological disorders: an epigenetic perspective. Neuroreport *21*, 909–913. https://doi.org/10.1097/WNR.0b013e32833e9288.

9. Kucharski, R., Maleszka, J., Foret, S., and Maleszka, R. (2008). Nutritional control of reproductive status in honeybees via DNA methylation. Science *319*, 1827–1830. https://doi.org/10.1126/science.1153069.

10. Mattei, A.L., Bailly, N., and Meissner, A. (2022). DNA methylation: a historical perspective. Trends Genet. *38*, 676–707. https://doi.org/10.1016/j.tig.2022.03.010.

11. Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F., et al. (2017). Impact of Cytosine Methylation on DNA Binding Specificities of Human Transcription Factors. Science *356*. https://doi.org/10.1126/science.aaj2239.

12. Maleszka, R., and Kucharski, R. (2022). Without mechanisms, theories and models in insect epigenetics remain a black box. Trends Genet. *38*, 1108–1111. https://doi.org/10.1016/j.tig.2022.05.004.

13. Jurkowska, R.Z., and Jeltsch, A. (2016). Enzymology of Mammalian DNA Methyltransferases. Adv. Exp. Med. Biol. *945*, 87–122. https://doi.org/10.1007/978-3-319-43624-1_5.

14. Tajima, S., Suetake, I., Takeshita, K., Nakagawa, A., and Kimura, H. (2016). Domain Structure of the Dnmt1, Dnmt3a, and Dnmt3b DNA Methyltransferases. Adv. Exp. Med.

Biol. *945*, 63–86. https://doi.org/10.1007/978-3-319-43624-1_4.

15. Jeltsch, A., and Jurkowska, R.Z. (2014). New concepts in DNA methylation. Trends Biochem. Sci. *39*, 310–318. https://doi.org/10.1016/j.tibs.2014.05.002.

16. Bewick, A.J., Vogel, K.J., Moore, A.J., and Schmitz, R.J. (2017). Evolution of DNA Methylation across Insects. Mol. Biol. Evol. *34*, 654–665. https://doi.org/10.1093/molbev/msw264.

17. Lyko, F., and Maleszka, R. (2011). Insects as innovative models for functional studies of DNA methylation. Trends Genet. *27*, 127–131. https://doi.org/10.1016/j.tig.2011.01.003.

18. Werren, J.H., Richards, S., Desjardins, C.A., Niehuis, O., Gadau, J., Colbourne, J.K., Nasonia Genome Working Group, Werren, J.H., Richards, S., Desjardins, C.A., et al. (2010). Functional and evolutionary insights from the genomes of three parasitoid Nasonia species. Science *327*, 343–348. https://doi.org/10.1126/science.1178028.

19. Wang, Y., Jorda, M., Jones, P.L., Maleszka, R., Ling, X., Robertson, H.M., Mizzen, C.A., Peinado, M.A., and Robinson, G.E. (2006). Functional CpG methylation system in a social insect. Science *314*, 645–647. https://doi.org/10.1126/science.1135213.

20. Honeybee Genome Sequencing Consortium, Robinson, G.E., Gibbs, R.A., Worley, K.C., Evans, J.D., Maleszka, R., Robertson, H.M., Weaver, D.B., Beye, M., Bork, P., et al. (2006). Insights into social insects from the genome of the honeybee Apis mellifera. Nature *443*, 931–949. https://doi.org/10.1038/Nature05260.

21. Wedd, L., Kucharski, R., and Maleszka, R. (2022). DNA methylation in honey bees and the unresolved questions in insect methylomics. In DNA Methyltransferases - Role and Function, A. Jeltsch and R.Z. Jurkowska, eds. (Springer).

22. Rona, G.B., Eleutherio, E.C.A., and Pinheiro, A.S. (2016). PWWP domains and their modes of sensing DNA and histone methylated lysines. Biophys. Rev. *8*, 63–74. https://doi.org/10.1007/s12551-015-0190-6.

23. Tatton-Brown, K., Loveday, C., Yost, S., Clarke, M., Ramsay, E., Zachariou, A., Elliott, A., Wylie, H., Ardissone, A., Rittinger, O., et al. (2017). Mutations in Epigenetic Regulation Genes Are a Major Cause of Overgrowth with Intellectual Disability. Am. J. Hum. Genet. *100*, 725–736. https://doi.org/10.1016/j.ajhg.2017.03.010.

24. Kibe, K., Shirane, K., Ohishi, H., Uemura, S., Toh, H., and Sasaki, H. (2021). The DNMT3A PWWP domain is essential for the normal DNA methylation landscape in mouse somatic cells and oocytes. PLoS Genet. *17*, e1009570. https://doi.org/10.1371/journal.pgen.1009570.

25. Du, J., Johnson, L.M., Jacobsen, S.E., and Patel, D.J. (2015). DNA methylation pathways and their crosstalk with histone methylation. Nat. Rev. Mol. Cell Biol. *16*, 519–532. https://doi.org/10.1038/nrm4043.

26. Dhayalan, A., Rajavelu, A., Rathert, P., Tamas, R., Jurkowska, R.Z., Ragozin, S., and Jeltsch, A. (2010). The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. J. Biol. Chem. *285*, 26114–26120. https://doi.org/10.1074/jbc.M109.089433.

27. Hunt, B.G., Glastad, K.M., Yi, S.V., and Goodisman, M.A.D. (2013). Patterning and Regulatory Associations of DNA Methylation Are Mirrored by Histone Modifications in Insects. Genome Biol. Evol. *5*, 591–598. https://doi.org/10.1093/gbe/evt030.

28. Wojciechowski, M., Lowe, R., Maleszka, J., Conn, D., Maleszka, R., and Hurd, P.J. (2018). Phenotypically distinct female castes in honey bees are defined by alternative chromatin states during larval development. Genome Res. *28*, 1532–1542. https://doi.org/10.1101/gr.236497.118.

29. Glastad, K.M., Hunt, B.G., and Goodisman, M.A.D. (2015). DNA methylation and chromatin organization in insects: insights from the Ant Camponotus floridanus. Genome Biol. Evol. *7*, 931–942. https://doi.org/10.1093/gbe/evv039.

30. Patalano, S., Vlasova, A., Wyatt, C., Ewels, P., Camara, F., Ferreira, P.G., Asher, C.L., Jurkowski, T.P., Segonds-Pichon, A., Bachman, M., et al. (2015). Molecular signatures of plastic phenotypes in two eusocial insect species with simple societies. Proc. Natl. Acad. Sci. USA *112*, 13970–13975. https://doi.org/10.1073/pnas.1515937112.

31. Wedd, L., Kucharski, R., and Maleszka, R. (2022). DNA Methylation in Honey Bees and the Unresolved Questions in Insect Methylomics. In DNA Methyltransferases - Role and Function, A. Jeltsch and R.Z. Jurkowska, eds. (Springer International Publishing), pp. 159–176. https://doi.org/10.1007/978-3-031-11454-0_7.

32. Hurd, P.J., Grübel, K., Wojciechowski, M., Maleszka, R., and Rössler, W. (2021). Novel structure in the nuclei of honey bee brain neurons revealed by immunostaining. Sci. Rep. *11*, 6852. https://doi.org/10.1038/s41598-021-86078-5.

33. Wojciechowski, M., Rafalski, D., Kucharski, R., Misztal, K., Maleszka, J., Bochtler, M., and Maleszka, R. (2014). Insights into DNA hydroxymethylation in the honeybee from in-depth analyses of TET dioxygenase. Open Biol. *4*, 140110. https://doi.org/10.1098/rsob.140110.

34. Peters, R.S., Krogmann, L., Mayer, C., Donath, A., Gunkel, S., Meusemann, K., Kozlov, A., Podsiadlowski, L., Petersen, M., Lanfear, R., et al. (2017). Evolutionary History of the Hymenoptera. Curr. Biol. *27*, 1013–1018. https://doi.org/10.1016/j.cub.2017.01.027.

35. Sharanowski, B.J., Ridenbaugh, R.D., Piekarski, P.K., Broad, G.R., Burke, G.R., Deans, A.R., Lemmon, A.R., Moriarty Lemmon, E.C., Diehl, G.J., Whitfield, J.B., and Hines, H.M. (2021). Phylogenomics of Ichneumonoidea (Hymenoptera) and implications for evolution of mode of parasitism and viral endogenization. Mol. Phylogenet. Evol. *156*, 107023. https://doi.org/10.1016/j.ympev.2020.107023.

36. Weinberg, D.N., Papillon-Cavanagh, S., Chen, H., Yue, Y., Chen, X., Rajagopalan, K.N., Horth, C., McGuire, J.T., Xu, X., Nikbakht, H., et al. (2019). The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. Nature *573*, 281–286. https://doi.org/10.1038/s41586-019-1534-3.

37. Heyn, P., Logan, C.V., Fluteau, A., Challis, R.C., Auchynnikava, T., Martin, C.A., Marsh, J.A., Taglini, F., Kilanowski, F., Parry, D.A., et al. (2019). Gain-of-function DNMT3A mutations cause microcephalic dwarfism and hypermethylation of Polycomb-regulated regions. Nat. Genet. *51*, 96–105. https://doi.org/10.1038/s41588-018-0274-x.

38. Shirohzu, H., Kubota, T., Kumazawa, A., Sado, T., Chijiwa, T., Inagaki, K., Suetake, I., Tajima, S., Wakui, K., Miki, Y., et al. (2002). Three novel DNMT3B mutations in Japanese patients with ICF syndrome. Am. J. Med. Genet. *112*, 31–37. https://doi.org/10.1002/ajmg.10658.

39. Wedd, L., and Maleszka, R. (2016). DNA Methylation and Gene Regulation in Honeybees: From Genome-Wide Analyses to Obligatory Epialleles. Adv. Exp. Med. Biol. *945*, 193–211. https://doi.org/10.1007/978-3-319-43624-1_9.

40. Raj, B., and Blencowe, B.J. (2015). Alternative Splicing in the Mammalian Nervous System: Recent Insights into Mechanisms and Functional Roles. Neuron *87*, 14–27. https://doi.org/10.1016/j.neuron.2015.05.004.

41. Levine, R.B., and Truman, J.W. (1982). Metamorphosis of the insect nervous system: changes in morphology and synaptic interactions of identified neurones. Nature *299*, 250–252. https://doi.org/10.1038/299250a0.

42. Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., and Oberdoerffer, S. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature *479*, 74–79. https://doi.org/10.1038/Nature10442.

43. Herzel, L., Ottoz, D.S.M., Alpert, T., and Neugebauer, K.M. (2017). Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. Nat. Rev. Mol. Cell Biol. *18*, 637–650. https://doi.org/10.1038/nrm.2017.63.

44. Duymich, C.E., Charlet, J., Yang, X., Jones, P.A., and Liang, G. (2016). DNMT3B isoforms without catalytic activity stimulate gene body methylation as accessory proteins in somatic cells. Nat. Commun. *7*, 11453. https://doi.org/10.1038/ncomms11453.

45. Rondelet, G., Dal Maso, T., Willems, L., and Wouters, J. (2016). Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B. J. Struct. Biol. *194*, 357–367. https://doi.org/10.1016/j.jsb.2016.03.013.

46. Jeong, H., Mason, S.P., Barabási, A.L., and Oltvai, Z.N. (2001). Lethality and centrality in protein networks. Nature *411*, 41–42. https://doi.org/10.1038/35075138.

47. Schwartz, S.H., Silva, J., Burstein, D., Pupko, T., Eyras, E., and Ast, G. (2008). Large-scale comparative analysis of splicing signals and their corresponding splicing factors in eukaryotes. Genome Res. *18*, 88–103. https://doi.org/10.1101/gr.6818908.

48. Foret, S., Kucharski, R., Pellegrini, M., Feng, S., Jacobsen, S.E., Robinson, G.E., and Maleszka, R. (2012). DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. Proc. Natl. Acad. Sci. USA *109*, 4968–4973. https://doi.org/10.1073/pnas.1202392109.

49. Flores, K., Wolschin, F., Corneveaux, J.J., Allen, A.N., Huentelman, M.J., and Amdam, G.V. (2012). Genome-wide association between DNA methylation and alternative splicing in an invertebrate. BMC Genom. *13*, 480. https://doi.org/10.1186/1471-2164-13-480.

50. Bonasio, R., Li, Q., Lian, J., Mutti, N.S., Jin, L., Zhao, H., Zhang, P., Wen, P., Xiang, H., Ding, Y., et al. (2012). Genome-wide and caste-specific DNA methylomes of the ants Camponotus floridanus and Harpegnathos saltator. Curr. Biol. *22*, 1755–1764.

51. Lyko, F., Foret, S., Kucharski, R., Wolf, S., Falckenhayn, C., and Maleszka, R. (2010). The

Honey Bee Epigenomes: Differential Methylation of Brain DNA in Queens and Workers. PLoS Biol. 8, e1000506. https://doi.org/10.1371/journal.pbio.1000506.

52. Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., and Shirakawa, M. (2009). Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. EMBO Rep. 10, 1235–1241. https://doi.org/10.1038/embor.2009.218.

53. Hu, J.L., Zhou, B.O., Zhang, R.R., Zhang, K.L., Zhou, J.Q., and Xu, G.L. (2009). The N-terminus of histone H3 is required for de novo DNA methylation in chromatin. Proc. Natl. Acad. Sci. USA 106, 22187–22192. https://doi.org/10.1073/pnas.0905767106.

54. Wright, C.J., Smith, C.W.J., and Jiggins, C.D. (2022). Alternative splicing as a source of phenotypic diversity. Nat. Rev. Genet. 23, 697–710. https://doi.org/10.1038/s41576-022-00514-4.

55. Sendžikaitė, G., Hanna, C.W., Stewart-Morgan, K.R., Ivanova, E., and Kelsey, G. (2019). A DNMT3A PWWP mutation leads to methylation of bivalent chromatin and growth retardation in mice. Nat. Commun. 10, 1884. https://doi.org/10.1038/s41467-019-09713-w.

56. Foret, S., Kucharski, R., Pittelkow, Y., Lockett, G.A., and Maleszka, R. (2009). Epigenetic regulation of the honey bee transcriptome: unravelling the nature of methylated genes. BMC Genom. 10, 472. https://doi.org/10.1186/1471-2164-10-472.

57. du Preez, P.H., Breeds, K., Burger, N.F.V., Swiegers, H.W., Truter, J.C., and Botha, A.M. (2020). DNA Methylation and Demethylation Are Regulated by Functional DNA Methyltransferases and DnTET Enzymes in Diuraphis noxia. Front. Genet. 11, 452. https://doi.org/10.3389/fgene.2020.00452.

58. Walsh, T.K., Brisson, J.A., Robertson, H.M., Gordon, K., Jaubert-Possamai, S., Tagu, D., and Edwards, O.R. (2010). A functional DNA methylation system in the pea aphid, Acyrthosiphon pisum. Insect Mol. Biol. 19, 215–228. https://doi.org/10.1111/j.1365-2583.2009.00974.x.

59. Jurkowski, T.P., and Jeltsch, A. (2011). On the evolutionary origin of eukaryotic DNA methyltransferases and Dnmt2. PLoS One 6, e28104. https://doi.org/10.1371/journal.pone.0028104.

60. Ren, W., Fan, H., Grimm, S.A., Guo, Y., Kim, J.J., Yin, J., Li, L., Petell, C.J., Tan, X.F., Zhang, Z.M., et al. (2020). Direct readout of heterochromatic H3K9me3 regulates DNMT1-mediated maintenance DNA methylation. Proc. Natl. Acad. Sci. USA 117, 18439–18447. https://doi.org/10.1073/pnas.2009316117.

61. Regev, A., Lamb, M.J., and Jablonka, E. (1998). The role of DNA methylation in invertebrates: Developmental regulation or genome defense? Mol. Biol. Evol. 15, 880–891.

62. Provataris, P., Meusemann, K., Niehuis, O., Grath, S., and Misof, B. (2018). Signatures of DNA Methylation across Insects Suggest Reduced DNA Methylation Levels in Holometabola. Genome Biol. Evol. 10, 1185–1197. https://doi.org/10.1093/gbe/evy066.

63. Miklos, G.L. (1993). Molecules and cognition: the latterday lessons of levels, language, and lac. Evolutionary overview of brain structure and function in some vertebrates and invertebrates. J. Neurobiol. 24, 842–890.

64. Ji, Y., Li, X., Ji, T., Tang, J., Qiu, L., Hu, J., Dong, J., Luo, S., Liu, S., Frandsen, P.B., et al. (2020). Gene reuse facilitates rapid radiation and independent adaptation to diverse habitats in the Asian honeybee. Sci. Adv. 6, eabd3590. https://doi.org/10.1126/sciadv.abd3590.

65. Quicke, D.L. (1997). Parasitic Wasps (Chapman & Hall Ltd).

66. Polilov, A.A. (2012). The smallest insects evolve anucleate neurons. Arthropod Struct. Dev. 41, 29–34. https://doi.org/10.1016/j.asd.2011.09.001.

67. Romiguier, J., Borowiec, M.L., Weyna, A., Helleu, Q., Loire, E., La Mendola, C., Rabeling, C., Fisher, B.L., Ward, P.S., and Keller, L. (2022). Ant phylogenomics reveals a natural selection hotspot preceding the origin of complex eusociality. Curr. Biol. 32, 2942–2947.e4. https://doi.org/10.1016/j.cub.2022.05.001.

68. Libbrecht, R., Nadrau, D., and Foitzik, S. (2020). A Role of Histone Acetylation in the Regulation of Circadian Rhythm in Ants. iScience 23, 100846. https://doi.org/10.1016/j.isci.2020.100846.

69. Spannhoff, A., Kim, Y.K., Raynal, N.J.M., Gharibyan, V., Su, M.B., Zhou, Y.Y., Li, J., Castellano, S., Sbardella, G., Issa, J.P.J., and Bedford, M.T. (2011). Histone deacetylase inhibitor activity in royal jelly might facilitate caste switching in bees. EMBO Rep. 12, 238–243. https://doi.org/10.1038/Embor.2011.9.

70. Glastad, K.M., Arsenault, S.V., Vertacnik, K.L., Geib, S.M., Kay, S., Danforth, B.N., Rehan, S.M., Linnen, C.R., Kocher, S.D., and Hunt, B.G. (2017). Variation in DNA Methylation Is Not Consistently Reflected by Sociality in Hymenoptera. Genome Biol. Evol. 9, 1687–1698. https://doi.org/10.1093/gbe/evx128.

71. Kapheim, K.M., Pan, H., Li, C., Salzberg, S.L., Puiu, D., Magoc, T., Robertson, H.M., Hudson, M.E., Venkat, A., Fischman, B.J., et al. (2015). Genomic signatures of evolutionary transitions from solitary to group living. Science 348, 1139–1143. https://doi.org/10.1126/science.aaa4788.

72. Maleszka, R., Mason, P.H., and Barron, A.B. (2014). Epigenomics and the concept of degeneracy in biological systems. Brief. Funct. Genomics 13, 191–202. https://doi.org/10.1093/bfgp/elt050.

73. Edelman, G.M., and Gally, J.A. (2001). Degeneracy and complexity in biological systems. Proc. Natl. Acad. Sci. USA 98, 13763–13768. https://doi.org/10.1073/pnas.231499798.

74. Menzel, R. (2012). The honeybee as a model for understanding the basis of cognition. Nat. Rev. Neurosci. 13, 758–768. https://doi.org/10.1038/Nrn3357.

75. Maleszka, R. (2018). Beyond Royalactin and a master inducer explanation of phenotypic plasticity in honey bees. Commun. Biol. 1, 8. https://doi.org/10.1038/s42003-017-0004-4.

76. Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., et al. (2007). Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948. https://doi.org/10.1093/bioinformatics/btm404.

77. Minh, B.Q., Dang, C.C., Vinh, L.S., and Lanfear, R. (2021). QMaker: Fast and Accurate Method to Estimate Empirical Models of Protein Evolution. Syst. Biol. 70, 1046–1060. https://doi.org/10.1093/sysbio/syab010.

78. Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., Frandsen, P.B., Ware, J., Flouri, T., Beutel, R.G., et al. (2014). Phylogenomics resolves the timing and pattern of insect evolution. Science 346, 763–767. https://doi.org/10.1126/science.1257570.

79. Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol. Biol. Evol. 38, 3022–3027. https://doi.org/10.1093/molbev/msab120.

80. Boratyn, G.M., Thierry-Mieg, J., Thierry-Mieg, D., Busby, B., and Madden, T.L. (2019). Magic-BLAST, an accurate RNA-seq aligner for long and short reads. BMC Bioinf. 20, 405. https://doi.org/10.1186/s12859-019-2996-x.

81. Lu, S., Wang, J., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Marchler, G.H., Song, J.S., et al. (2020). CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Res. 48, D265–D268. https://doi.org/10.1093/nar/gkz991.

82. Vleurinck, C., Raub, S., Sturgill, D., Oliver, B., and Beye, M. (2016). Linking Genes and Brain Development of Honeybee Workers: A Whole-Transcriptome Approach. PLoS One 11, e0157980. https://doi.org/10.1371/journal.pone.0157980.

83. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. GigaScience 10, giab008. https://doi.org/10.1093/gigascience/giab008.

84. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930. https://doi.org/10.1093/bioinformatics/btt656.

85. Katz, Y., Wang, E.T., Silterra, J., Schwartz, S., Wong, B., Thorvaldsdóttir, H., Robinson, J.T., Mesirov, J.P., Airoldi, E.M., and Burge, C.B. (2015). Quantitative visualization of alternative exon expression from RNA-seq data. Bioinformatics 31, 2400–2402. https://doi.org/10.1093/bioinformatics/btv034.

86. Farris, S., Ward, J.M., Carstens, K.E., Samadi, M., Wang, Y., and Dudek, S.M. (2019). Hippocampal Subregions Express Distinct Dendritic Transcriptomes that Reveal Differences in Mitochondrial Function in CA2. Cell Rep. 29, 522–539.e6. https://doi.org/10.1016/j.celrep.2019.08.093.

87. Qiu, C., Sawada, K., Zhang, X., and Cheng, X. (2002). The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. Nat. Struct. Biol. 9, 217–224. https://doi.org/10.1038/nsb759.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-GST	Abcam	Cat#ab19256; RRID: AB_444809
Goat Anti-Rabbit IgG Antibody, HRP-conjugate	Sigma-Aldrich	Cat#12–-348; RRID: AB_390191
Bacterial and virus strains		
E.coli ER2566	New England Biolabs	Cat#C2566I
Chemicals, peptides, and recombinant proteins		
L(-)-Glutathione (reduced form)	VWR	Cat#0399
Critical commercial assays		
MODified™ Histone peptide array	Active Motif	Cat#13005
Pierce™ ECL developing solution	Thermo Scientific	Cat#32109
Recombinant DNA		
pGEX-6P-2 PWWP1	This paper	N/A
pGEX-6P-2 PWWP2	This paper	N/A
Software and algorithms		
Array Analyze software v16.1	Active Motif	https://www.activemotif.com/catalog/668/modified-histone-peptide-array
ClustalX	Larkin et al.[76]	
IQ-TREE v2.2.0	Minh et al.[77,78]	
MEGA v11	Tamura et al.[79]	
R gplots package (v.3.1.1)	https://shorturl.at/bhJP7	
MagicBLAST	Boratyn et al.[80]	
ClustalX	Larkin et al.[76]	
Samtools v1.3.1	https://www.htslib.org/	
ExPaSy's SwissModel	https://swissmodel.expasy.org/	https://swissmodel.expasy.org/
Other		
AKTA pure chromatography system	Cytiva	https://www.cytivalifesciences.com/en/us/shop/chromatography/chromatography-systems/akta-pure-p-05844
Glutathione Sepharose™ 4B	Sigma-Aldrich	Cat#GE17-0756-01
Superdex® 200 Increase 10/300 GL	Sigma-Aldrich	Cat#GE28-9909-44
FUSION Solo-S Imaging system	Vilber	https://www.vilber.com/fusion-solo-s/
RNA-Seq dataset	BioProject PRJNA193691	RNA-Seq dataset

RESOURCE AVAILABILITY

This study did not generate new resources.

Lead contact

Additional information and requests for resources and reagents should be directed to the lead contact, Ryszard Maleszka (ryszard.maleszka@anu.edu.au).

Materials availability

This study did not generate any unique new reagent. All reagents used in this study are commercially available.

Data and code availability

This paper does not report original code.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

METHOD DETAILS

Phylogenetic trees construction

All analysed species are listed in Table S1. Conserved domain regions identified by searches against CDD database[81] were extracted from protein sequences using fixed-length margins based on the positions of exon boundaries. The sequences were aligned using ClustalX[76] using default options followed by block realignments and manual corrections. Due to low homology of PWWP domain C-termini, the sequences were further truncated to contain only the conserved aromatic cage-encoding region. The alignments were used for maximum likelihood tree construction with IQ-TREE v2.2.0, using Q model[77] estimated for insects,[78] with invariable site plus discrete gamma rate heterogeneity, and ultrafast bootstrap for branch support assessment. The resulting trees were visualised with MEGA v11.[79] To generate heatmap in Figure 4, phylogenetic tree branch data were extracted from IQTree output and converted into table format using MEGA.[79] The resulting tables were imported to MS Excel and total branch lengths for each domain and species were calculated using standard worksheet functions. The values were used to construct a species x branch length matrix and a heatmap was plotted using heatmap.2 function of the R gplots package (v.3.1.1).

Honey bee DNMT3 alternative splicing analysis

Based on sequencing depth and quality we selected honeybee pupa brains RNA-Seq dataset from BioProject PRJNA193691.[82] Reads were mapped to honey bee DNMT3 gene region (NC_037639.1:15,815,881-15,827,940) using MagicBLAST.[80] As for short reads the program only calls canonical GT-AG introns, all CIGAR deletion (D) tags where the value was >= 10 were replaced with alignment gap (N) tags using a custom PERL script. Samtools v1.3.1[83] were used to merge SAM files and to extract base-level exon coverage. Subread package feature-Counts v1.5.1 program[84] was used to extract exon junctions' usage. Junctions representing less than 0.1% of total junctions count were filtered out (except for the exon3-exon9 junction, identified in the form of a cDNA clone), and the remaining data together with exon coverage values were used for quantitative visualisation in the form of a sashimi plot[85] using splicejam v0.0.76.900.[86]

Cloning, over-expression and purification of PWWP domains

DNA sequences corresponding to the two PWWP domains of *Apis mellifera* DNMT3 were codon-optimised for *E. coli* and cloned into pGEX-6P-2 in-frame with Glutathione-S-transferase (GST) using *Bam*HI-*Xho*I cloning sites (GenScript). GST-tagged PWWP fusion proteins were overexpressed in *E. coli* strain ER2566 for 3.5 hours at 24°C, by the addition of 0.1 mM IPTG. Cells were harvested by centrifugation and the cell pellet was resuspended in GST purification buffer (20 mM HEPES pH 7.5, 500 mM KCl, 10 μM ß-mercaptoethanol, 1 mM EDTA and 10% glycerol) and lysed by sonication. GST-tagged PWWP fusion proteins were purified from lysates using Glutathione Sepharose™ 4B beads (Sigma) via gravity flow chromatography and eluted with 10 mM reduced glutathione (VWR). GST-tagged PWWP fusion proteins were then further purified by high performance liquid size exclusion chromatography on a Superdex® 200 Increase 10/300 GL column (Sigma) using an AKTA pure chromatography system (Cytiva). Standard buffers were used (PBS with 5% glycerol, except where the addition of 500 mM NaCl was required). All separations were carried out at 4°C at a flow rate of 0.35 ml/min. Chromatograms were recorded using UV absorbance at 280 nm and 254 nm (Figure S7).

Histone peptide arrays and binding

MODified™ histone peptide arrays (Active Motif) were blocked overnight at 4°C in TBST buffer (10 mM Tris-HCl pH 8.3, 0.05% Tween-20, 150 mM NaCl) containing 5% non-fat dried milk. Subsequently, the array was washed with TBST for 5 min and incubated with 10 μM of GST-tagged PWWP fusion protein, or GST protein alone, at room temperature (RT) for 1 h in binding buffer (20 mM HEPES pH 7.5, 100 mM KCl, 1 mM EDTA, 0.1 mM DTT, 10% glycerol). After washing in TBST, the membrane was incubated for 1 h at RT with a rabbit anti-GST antibody (Abcam; ab19256) at a concentration of 0.2 μg/ml. The array was then washed three times with TBST for 10 min each at RT and incubated with a horseradish peroxidase conjugated anti-rabbit secondary antibody (Sigma) at a concentration of 0.05 ug/ml for 1 h at RT. Finally, the array was submerged in Pierce™ ECL developing solution, imaged (FUSION Solo-S) and the data quantified using array analyser software (Active Motif, v.16.1).

3D modelling of the PWWP domain

The tertiary structure modelling of the honey bee PWWP domains was performed on ExPaSy's SwissModel server (https://swissmodel.expasy.org/) using the crystal structure of the PWWP domain of mammalian DNA methyltransferase DNMT3b as target sequence.[87]

QUANTIFICATION AND STATISTICAL ANALYSIS

Histone peptide arrays: Raw spot intensities were normalised to the highest spot signal intensity (= 1). Bar plots represent the mean of normalised raw spot signal intensities from the left and right-side replicates of the MODified histone peptide arrays. Error bars show one standard deviation.