Signatures of DNA Methylation across Insects Suggest Reduced DNA Methylation Levels in Holometabola

Panagiotis Provataris[1], Karen Meusemann[1,2,3], Oliver Niehuis[1,2], Sonja Grath[4,*], and Bernhard Misof[1,*]

[1]Center for Molecular Biodiversity Research, Zoological Research Museum Alexander Koenig, Bonn, Germany

[2]Evolutionary Biology and Ecology, Institute of Biology I (Zoology), Albert Ludwig University Freiburg, Freiburg (Brsg.), Germany

[3]Australian National Insect Collection, CSIRO National Research Collections Australia, Acton, Australian Capital Territory, Australia

[4]Division of Evolutionary Biology, Faculty of Biology, Ludwig-Maximilians-Universität München, Planegg, Germany

*Corresponding authors: E-mails: grath@bio.lmu.de; b.misof@zfmk.de.

Abstract

It has been experimentally shown that DNA methylation is involved in the regulation of gene expression and the silencing of transposable element activity in eukaryotes. The variable levels of DNA methylation among different insect species indicate an evolutionarily flexible role of DNA methylation in insects, which due to a lack of comparative data is not yet well-substantiated. Here, we use computational methods to trace signatures of DNA methylation across insects by analyzing transcriptomic and genomic sequence data from all currently recognized insect orders. We conclude that: 1) a functional methylation system relying exclusively on DNA methyltransferase 1 is widespread across insects. 2) DNA methylation has potentially been lost or extremely reduced in species belonging to springtails (Collembola), flies and relatives (Diptera), and twisted-winged parasites (Strepsiptera). 3) Holometabolous insects display signs of reduced DNA methylation levels in protein-coding sequences compared with hemimetabolous insects. 4) Evolutionarily conserved insect genes associated with housekeeping functions tend to display signs of heavier DNA methylation in comparison to the genomic/transcriptomic background. With this comparative study, we provide the much needed basis for experimental and detailed comparative analyses required to gain a deeper understanding on the evolution and function of DNA methylation in insects.

Key words: DNA methylation, DNMT1, DNMT3, Tet, Hexapoda, CpG o/e.

Introduction

Methylation of cytosine residues constitutes a common epigenetic modification among eukaryotes. It is functionally associated with the regulation of expression of genomic elements (Zemach et al. 2010). For example, promoter-proximal methylation is linked to the transcriptional repression of associated genes (Jones 2012; Schübeler 2015). Methylation of repetitive noncoding DNA elements also has a repressive effect, limiting the expression, and thus, the genomic expansion of these elements (Schübeler 2015). In contrast, intragenic methylation is associated with active transcription (Feng et al. 2010; Zemach et al. 2010; Jones 2012), but a "cause and effect" relationship has not been established in this context (Schübeler 2015). Although DNA methylation is widely present among eukaryotes, the levels (i.e., the proportion of methylated cytosines or CpG sites in a given genome), patterns, and genomic targets of DNA methylation are not evolutionarily conserved. In vertebrates, and especially in mammals, CpG dinucleotides are heavily methylated genome-wide, with the exception of CpG islands. CpG islands typically overlap with promoter regions and remain mostly unmethylated (Schübeler 2015). In contrast, invertebrates show intermediate or even negligible levels of DNA methylation at CpG sites, which is typically targeted to a subset of gene bodies (the term gene body refers to the transcribed part of a gene, comprised of exons and introns; Suzuki and Bird 2008; Feng et al. 2010; Zemach et al. 2010; but see Wang et al. 2014; Kao et al. 2016).

In insects, the levels of gene body methylation vary considerably (Glastad et al. 2011). On one hand, model organisms like the fruit fly, *Drosophila melanogaster*, and the red flour beetle, *Tribolium castaneum*, do not display notable DNA methylation levels in their genomes (Zemach et al. 2010; Bewick et al. 2017). On the other hand, nutritionally regulated

levels of DNA methylation contribute to the ontogenetic establishment of alternative castes in the honey bee, *Apis mellifera* (Kucharski et al. 2008; Foret et al. 2012). This observation has supported the hypothesis that DNA methylation in insects is associated with caste development and the evolution of (eu)sociality, but recent empirical evidence from research on eusocial Hymenoptera (wasps, ants, and bees) suggests that this association is not universal (Bonasio et al. 2012; Patalano et al. 2015; Kapheim et al. 2015; Libbrecht et al. 2016; Standage et al. 2016). Obviously, a taxonomically representative description of the levels and patterns of DNA methylation is one major prerequisite to improve our understanding of the evolution and, eventually, the function of DNA methylation in insects. Therefore, we conducted a comparative analysis of DNA methylation patterns in insects by making use of recently published, extensive transcriptomic (Misof et al. 2014) and publicly available genomic sequence data covering all extant insect orders.

Two types of DNA methyltransferases (DNMTs), DNMT1 and DNMT3, are responsible for DNA methylation in animals (Goll and Bestor 2005). In mammals, DNMTs carry out de novo (DNMT3) and maintenance (DNMT1) methylation, with functional overlap (Jeltsch and Jurkowska 2014). Another, noncanonical member of the DNMT family, TRDMT1 (tRNA aspartic acid methyltransferase 1, most commonly known as DNMT2), long considered a DNMT, has seemingly shifted substrate and is now known to methylate tRNA, not DNA (Goll et al.2006; Lyko 2017). It is generally assumed that these functions are conserved in insects (Wang et al. 2006). This assumption is supported by the observation that the absence of DNMT1 and DNMT3 is associated with the loss or extreme reduction of DNA methylation in *D. melanogaster* (Raddatz et al. 2013). However, in contrast to mammals, the DNMT toolkit of insects that show substantial levels of DNA methylation in their genomes is not conserved. For example, the silk moth, *Bombyx mori*, has empirically determined DNA methylation, but lacks copies of DNMT3 homologs from its genome (Xiang et al. 2010; Bewick et al. 2017). Thus, functional methylation systems in insects can be realized in the absence of DNMT3. The frequency of DNMT3 loss in different lineages is, however, unknown due to a lack of extensive comparative data.

Whereas DNMTs are responsible for generating methylcytosine (5mC) residues, Tet dioxygenases are shown to convert 5mC to hydroxymethylcytosine (5hmC) in many animal species (Pastor et al 2013). In contrast to mammals which harbor three Tet paralogs, invertebrate species, including some insects, seem to encode a single Tet homolg without a characterized function in the majority of cases (Pastor et al 2013; Wojciechowski et al 2014). In the honey bee, it was recently shown that the single Tet enzyme is capable of converting 5mC to 5hmC (Wojciechowski et al 2014). However, Tet enzymes may display functional promiscuity in insects, because a Tet homolg seems to mediate N6-methyladenine

demethylation and 5mC demethylation in the *D. melanogaster* DNA and mRNA, respectively (Zhang et al. 2015; Delatte et al. 2016). The distribution of Tet enzymes in insects and its relationship to the presence of 5mC is currently not known.

Comparative analyses using experimental data have shown that the levels of gene body methylation of the silk moth and the honey bee are substantially lower compared with other invertebrates (sea squirt, *Ciona intestinalis*, and sea anemone, *Nematostella vectensis*; Sarda et al. 2012). These results have fueled the hypothesis that DNA methylation was reduced in the ancestors of insects (Glastad et al. 2014). However, experimental and computational evidence from research on hemimetabolous lineages point to significantly elevated DNA methylation levels in species belonging to Orthoptera, Phasmatodea, and Isoptera compared with certain Hymenoptera and *B. mori* (Krauss et al. 2009; Falckenhayn et al. 2013; Glastad et al. 2013, 2016; Terrapon et al. 2014). Thus, the validity of the proposed hypothesis on the ancestral state of DNA methylation in insects is questionable.

The methylomic profiling of insects, mostly representing Hymenoptera, and to a lesser extent Lepidoptera and Coleoptera, revealed largely similar patterns of DNA methylation, primarily targeted to exons of protein-coding genes (Lyko et al. 2010; Xiang et al. 2010; Bonasio et al. 2012; Wang et al. 2013; Cunningham et al. 2015; Patalano et al. 2015; Libbrecht et al. 2016; Rehan et al. 2016; Standage et al. 2016). Additionally, genes targeted by DNA methylation were ubiquitously expressed among various tissue types (Foret et al. 2012; Xiang et al. 2010), among different morphs in ants (Bonasio et al. 2012; Libbrecht et al. 2016), and among developmental stages in the parasitoid wasp, *Nasonia vitripennis* (Wang et al. 2013). Gene ontology annotations showed that the majority of these genes mostly serves basic cellular functions, exhibit a highly methylated state among species, and are highly conserved at sequence level (Elango et al. 2009; Lyko et al. 2010; Hunt et al. 2013; Wang et al. 2013; Cunningham et al. 2015; Rehan et al. 2016). These patterns are even found when comparing orthologous genes among distantly related invertebrates (Sarda et al. 2012). These findings strongly imply that the targeting of DNA methylation in insect genomes is nonrandom, but a solid explanation for this observation remains elusive.

The aim of the present study is to improve our understanding of the evolution of DNA methylation in insects. Specifically, we focused on the hypothesis stating that DNA methylation has been reduced in the ancestors of insects (Glastad et al. 2014). For this purpose, we analyzed whole-body transcriptomes and genomic data (protein-coding sequences and predicted proteins) of 143 insect species, representing all 32 currently recognized insect orders, and eleven outgroup species. First, we document the presence or absence of DNA methyltransferases

(DNMT1, DNMT3). Second, we use the normalized CpG dinucleotide content (CpG observed/expected or simply CpG o/e) to predict the occurrence and estimate the levels of DNA methylation in protein-coding sequences. The last approach provided the means to assess the relationship between DNA methylation and the evolutionary conservation of genes across insects.

We found that, unlike in vertebrates, the phylogenetic distribution of DNMT1 in insects is much wider compared with DNMT3. On the basis of the patterns of CpG o/e distributions, our data suggest that DNA methylation is widespread among insect orders. More importantly, we estimate DNA methylation levels of protein-coding sequences to be significantly higher in hemimetabolous insects than in Holometabola. Finally, we show that single-copy genes present across insects tend to display signs of heavy DNA methylation compared with the genomic background. Our analyses point to a complex DNA methylation landscape in insects and set the basis for large scale comparative analyses using direct measurements of DNA methylation.

Materials and Methods

Data Acquisition

We identified DNMTs, Tet dioxygenases, and calculated CpG o/e ratios of 102 transcript assemblies from the 1KITE project (www.1kite.org) representing species of all extant insect orders (Misof et al. 2014); we used the latest version of all 1KITE assemblies (supplementary table S1, Supplementary Material online). Details concerning sequencing and assembly are described by Misof et al. (2014) and Mayer et al. (2016). We appended the 1KITE data with additional transcriptomic and genomic (CDS and predicted proteins) data of 53 arthropod species obtained from public and other resources (supplementary table S2, Supplementary Material online). For orthology assessment (see below), we used the 1KITE species, the aforementioned published transcriptomes, and 14 arthropod official gene sets previously used by Misof et al. (2014, supplementary tables S2, S4; in this study, supplementary table S3, Supplementary Material online).

Identification of DNMTs and Tet Dioxygenases

To search for DNMT1, DNMT3, TRDMT1, and Tet homologs in the transcriptomes and genomes presented previously, we constructed profile Hidden Markov Models (pHMMs) for the proteins in question (all pHMMs are available at: doi: 10.17632/8y5wm8887b.3). Amino-acid sequences of arthropod DNMTs and Tet proteins were downloaded from OrthoDB using the text-based search option (Kriventseva et al. 2015; Zdobnov et al 2017). Subsequently, we aligned each group of orthologous sequences using MAFFT L-INSI (Katoh and Standley 2013) and generated pHMMs from

each alignment using HMMER 3.1b1 (www.hmmer.org). We translated transcript sequences into all six possible reading frames with Exonerate, version 2.2.0 (Slater and Birney 2005) and searched with each pHMM the translated transcriptome and genome (predicted proteins) data using hmmsearch with default options (HMMER 3.1b1).

Because DNMT1, DNMT3, and TRDMT1 share a homologous DNA methylase domain (Pfam-accession no. PF00145), some sequences were identified as common candidates among these three proteins. Consequently, we removed redundant candidate sequences by keeping the ones with lowest e-value. Furthermore, we excluded all candidate sequences with an e-value $>10^{-5}$ from downstream analyses. To determine whether or not the candidate sequences were properly annotated as DNMT1, DNMT3, or TRDMT1, we introduced the following levels of control. First, we used blastp (BLAST+ v 2.2.28, Camacho et al. 2009) to search candidate sequences against N. vitripennis OGS v 2.0 (Munoz-Torres et al. 2011). We selected N. vitripennis as reference because it possesses a well-characterized DNMT toolkit (Werren et al. 2010). We excluded all candidate sequences that did not match a corresponding Nasonia DNMT as a best hit. Second, we scanned all remaining candidate sequences with a Nasonia match against Pfam-A pHMM library (version 27, Finn et al. 2014) and kept only the ones that did contain a characteristic DNA methylase or DNMT1-RFD domain (PF12047 which is a unique DNMT1 domain).

To search for Tet proteins, we compared candidate amino-acid sequences with the Pfam-A pHMM library (version 27, Finn et al. 2014) and retained only the ones that contained an annotated Tet-JBP domain (Pfam-accession no. PF12851) (for a detailed process on the identification of DNMTs and Tet, see Supplementary Material sections 1 and 2).

Calculation of Normalized CpG Dinucleotide Content (CpG o/e)

The normalized CpG dinucleotide content can serve as a proxy for the presence of DNA methylation, because cytosines targeted by DNA methylation are prone to spontaneous deamination into thymines, leading to a gradual reduction of CpG dinucleotides, termed CpG depletion. Therefore, in genomic regions that are subject to intense germline methylation over evolutionary time, CpGs are underrepresented. In contrast, regions with limited germline methylation maintain a high CpG content (Bird 1980). In insects and other invertebrates with considerable levels of DNA methylation in their genomes, two classes of genes are present, one with low CpG o/e (high germline DNA methylation) and another with high CpG o/e (low germline DNA methylation). Thus, a bimodal CpG o/e distribution typically occurs in such cases. In contrast, in species with very low or no DNA methylation, only one class of genes is expected, signified by a unimodal CpG o/e distribution and lack of CpG depletion.

We calculated the normalized CpG dinucleotide content using the following equation:

$$CpGo/e = \frac{P_{CpG}}{P_C * P_G}$$

where P_{CpG}, P_C, and P_G are the frequencies of 5'-CpG-3 dinucleotides, C nucleotides and G nucleotides, respectively, estimated from each sequence. In addition, we plotted distributions of the normalized GpC content, to control for causative factors unrelated to DNA methylation, like GC content (Fryxell and Moon 2005). We excluded sequences containing <200 nucleotides or containing >5% ambiguous nucleotides (N) from the calculation of normalized dinucleotide content. Furthermore, we excluded all nucleotide sequences with a normalized dinucleotide content equal to zero from any downstream analyses. All analyses were carried out using custom-made Perl and R (R Core Team 2016) scripts.

Inferring the Presence of DNA Methylation Based on CpG o/e Distributions

Species like the honeybee, *A. mellifera*, and the pea aphid, *Acyrthosiphon pisum*, in which DNA methylation has been experimentally verified, display clear bimodal CpG o/e distributions in protein-coding sequences with two distinct components, one with low CpG o/e and one with high CpG o/e values. A bimodal CpG o/e distribution may thus serve as an indication for the presence of DNA methylation. However, species with experimentally verified DNA methylation, like the branchiopod *Daphnia pulex*, the silk moth *B. mori*, and the beetle *Nicrophorus vespilloides*, lack clearly defined bimodality in protein-coding sequences, but the presence of DNA methylation is indicated due to an extensive tail spanning towards the low CpG o/e part of their distributions (Glastad et al. 2011; Sarda et al. 2012; Cunningham et al. 2015). In contrast, species like *D. melanogaster* and *T. castaneum* in which DNA methylation in protein-coding sequences is extremely reduced or absent, display a unimodal, almost normal CpG o/e distribution, with a mean around one (*D. melanogaster* ∼0.89, *T. castaneum* ∼1.1; Elango et al. 2009). Using these empirically well-documented cases, we defined a set of criteria to infer the presence of DNA methylation based on the modality of CpG o/e distributions. To test the modality of CpG o/e distributions, we used the Gaussian mixture modeling software package mclust (v 5.2) similar to Park et al. (2011) and fitted two Gaussian distributions in the CpG o/e and GpC o/e distributions of each species in question. We consider the following criteria as sufficient evidence for the presence of germline DNA methylation in protein-coding sequences of a species:

1. a CpG o/e distribution is bimodal, with one class of genes showing signs of CpG depletion. To identify bimodality,

we expect the absolute difference of the means of the two fitted Gaussian distributions to be 0.25 or higher, provided that one of the means is <0.7. Furthermore, the proportion of data belonging to the smallest of the fitted components should be >0.1. These criteria of bimodality should not be fulfilled by the GpC o/e distribution, which is unaffected by DNA methylation. A CpG o/e distribution fulfilling this set of criteria is described as "bimodal depleted" (fig. 1a).

2. In the absence of clearly defined bimodality, as observed in *B. mori* and *D. pulex*, we do not expect the criteria of bimodality to apply. However, in both these species a large proportion of data belongs to the smallest of the two fitted distributions (0.36 in *B. mori* and 0.43 in *D. pulex*). If we apply such criteria, we can identify species with similar CpG o/e distributions which, based on empirical evidence, should indicate the presence of DNA methylation. Therefore, we set the threshold for the proportion of smallest of the fitted normal distributions to 0.36 or higher (equal to that of *B. mori* or higher). This should not apply to the corresponding GpC o/e distribution. The CpG o/e distributions of these species are described as "unimodal, indicative of DNA methylation" (fig. 1b and c).

If the above criteria did not apply, we considered the evidence as insufficient to infer the presence of DNA methylation. The CpG o/e distributions of these species are described as "unimodal, not indicative of DNA methylation" (fig. 1d). We acknowledge that these criteria are conservative. However, we think that missing true positives is likely less misleading than building conclusions based on false positives.

Phylogenetic Generalized Least Squares (PGLS) Analysis

We used PGLS to correlate estimations of DNA methylation in protein-coding sequences (continuous dependent variable; obtained from Bewick et al. 2017) to the mode of development (categorical independent variable, binary coded as hemimetabolism or holometabolism) in 26 holometabolous and 14 hemimetabolous insect species (supplementary table S1 in Bewick et al. 2017). The multilocus coalescent tree estimated by Bewick et al. (fig. 1 in Bewick et al. 2017) was used to control for statistical nonindependence between species traits. To perform PGLS, we used the R packages ape (Paradis et al. 2004) and nlme (Pinheiro et al. 2017).

Orthology Assessment

We used an ortholog set of 1,478 protein-coding genes that are single-copy in twelve reference species (Misof et al. 2014). We used Orthograph version 0.5.4 (Petersen et al. 2017) to identify the protein-coding sequences of orthologs of the 1,478 single-copy genes in 129 additional species (see
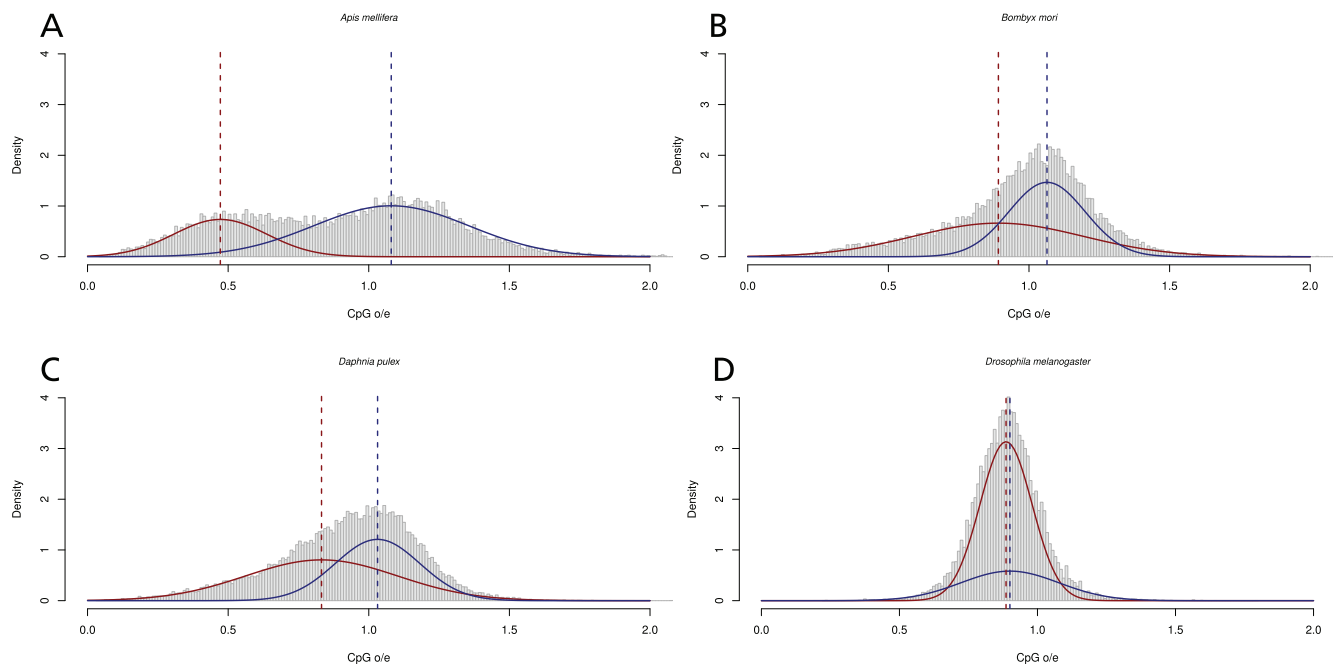
Fig. 1.—Distinct types of CpG o/e distributions in protein-coding sequences of four insect species. A mixture of two Gaussian distributions was fitted to the data using mclust (v. 5.2). Dark red and dark blue dashed lines correspond to the means of each fitted distribution ($mean_{low}$ and $mean_{high}$, respectively). (a) *Apis mellifera* displays a clearly bimodal CpG o/e distribution, with one component displaying low CpG o/e values (sequences mostly affected by CpG depletion) and the other one high (sequences less affected by CpG depletion). We describe the CpG o/e distribution of *A. mellifera* as "bimodal depleted" since difference between the component means is >0.25 ($mean_{high}$ - $mean_{low}$ = 0.61), whereas the low CpG o/e component has a mean <0.7 ($mean_{low}$ = 0.47). (b, c) *Bombyx mori* and *Daphnia pulex* lack clearly defined bimodality ($mean_{high}$ - $mean_{low}$ < 0.25 and $mean_{low}$ > 0.7 in both cases), but their low CpG o/e component displays a characteristic extensive tail, which contains a significant proportion of data (0.36 and 0.43, respectively). We describe distributions that lack clearly defined bimodality similar to *B. mori* and *D. pulex*, but their smallest component contains a significant proportion of data ($proportion_{low}$ = 0.36 or higher) as "unimodal, indicative of DNA methylation." (d) Finally, *Drosophila melanogaster*, which is almost devoid of DNA methylation from protein-coding sequences, displays a clearly unimodal CpG o/e distribution with two component means being almost identical ($mean_{high}$ - $mean_{low}$ = 0.004), show no signs of significant CpG depletion ($mean_{low}$ = 0.886), and the proportion of data belonging to the smallest component is very low ($proportion_{low}$ = 0.087 < 0.36). We describe the CpG o/e distribution of *D. melanogaster* as "unimodal, not indicative of DNA methylation."

supplementary tables S1 and S3, Supplementary Material online in this study; supplementary tables S1, S2, and S4, Supplementary Material online in Misof et al. 2014). We applied a relaxed setting for the reciprocal best hit search to any of the reference species included in the ortholog set. In all identified orthologs (see supplementary table S4, Supplementary Material online), we subsequently masked stop codons and Seleocysteine with X in the predicted amino-acid sequences and with NNN in the coding nucleotide sequences (CDS). We then aligned all orthologous amino-acid sequences as outlined by Misof et al (2014), including check for suspiciously aligned outlier sequences, alignment–refinement of identified outliers, and exclusion of persistent outliers. Subsequently, we generated corresponding multiple sequence alignments (MSAs) on nucleotide level with the software pal2nal (Suyama et al. 2006), using the amino-acid MSAs as blueprint. Finally, the 1,478 MSAs on nucleotide level served as basis for CpG o/e calculations (see Supplementary Material, section 3, Supplementary Material online).

Results

DNMT1 Homologs Are Likely Indispensable for Maintaining a Functional Methylation System in Insects

We characterized the occurrence of DNMTs and Tet proteins in the investigated insect and outgroup species by using pHMMs constructed from orthologous protein sequences of arthropods for each of the proteins in question. With these pHMMs at hand, we searched transcriptomes representing all insect orders, crustaceans and myriapods. Transcriptomic data were complemented by genomic data (protein predictions) of species belonging to nine insect orders (Collembola, Isoptera, Hemiptera, Psocodea, Hymenoptera, Strepsiptera, Coleoptera, Lepidoptera, and Diptera) plus crustaceans, myriapods, and a chelicerate (see Materials and Methods).

We identified homologous sequences of DNMT1 in species belonging to all insect orders and outgroups, except Collembola (seven species, including three with sequenced genomes), Diptera (13 species, including three with

sequenced genomes), and Strepsiptera (two species, including one with a sequenced genome; fig. 2; supplementary table S5, Supplementary Material online). DNMT3 homologs were not identified in species belonging to these three orders either, which apparently lack all currently known cytosine-specific DNMTs. In contrast to DNMT1, DNMT3 was sparsely found in insects, being present in species belonging to only seven out of 32 insect orders (Hemimetabola: Diplura, Orthoptera, Isoptera, Hemiptera, and Thysanoptera; Holometabola: Hymenoptera and Coleoptera), plus species of crustaceans and myriapods (fig. 2; supplementary table S5, Supplementary Material online). Within hemimetabolous insects, DNMT3 was absent from Palaeoptera (seven species) and the polyneopteran clade formed by Mantophasmatodea, Grylloblattodea, Embioptera, and Phasmatodea (eight species). Within Holometabola, DNMT3 was lacking from Neuropterida (eight species) and Mecopterida (40 species; fig. 2).

The tRNA methyltransferase TRDMT1 was the most commonly found enzyme in our data set being present in species belonging to 31 out of 32 insect orders (140/154 species possessed putative TRDMT1 homologs). TRDMT1 was absent from the transcriptome of the only representative of Zoraptera in our data set, *Zorotypus caudeli* (supplementary table S5, Supplementary Material online).

We identified homologous sequences of Tet dioxygenases in species belonging to 25 out of 32 insect orders, plus species belonging to all three outgroups (supplementary table S6, Supplementary Material online). Within hemimetabolous insects, Tet homologs are apparently missing in Arachaeognatha (two species), in the polyneopteran clade formed by Mantophasmatodea, Grylloblattodea, Embioptera, and Phasmatodea (eight species), and in Mantodea (three species). Within Holometabola, only Strepsiptera lack Tet homologs (two species). We have to note that Tet homologs were consistently identified in genomes (28/30), but not in transcriptomes (52/124).

CpG o/e Patterns Suggest DNA Methylation Being Taxonomically Widespread in Winged Insects

In order to infer the occurrence of DNA methylation in insects, we calculated CpG o/e ratios of protein-coding sequences in 143 species covering all insect orders and eleven additional outgroup species (see Materials and Methods). CpG o/e has been widely used as a proxy for estimating the patterns and levels of DNA methylation in various species of invertebrates (Suzuki et al. 2007; Elango et al. 2009; Glastad et al. 2013) with high concordance to empirical measurements (Glastad et al. 2011; Sarda et al. 2012).

Applying a set of stringent criteria (see Materials and Methods), we identified CpG o/e distributions pointing to the presence of DNA methylation in species belonging to 24 out of 32 total insect orders (fig. 2; supplementary
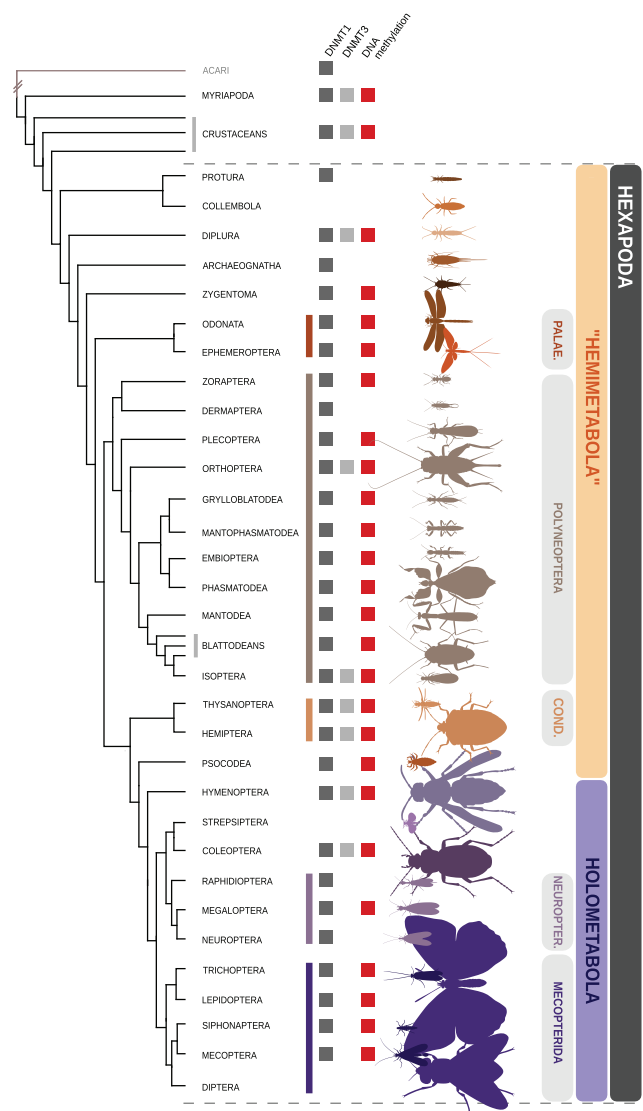

Fig. 2.—Occurrence of DNA methyltransferases and DNA methylation in investigated species. We plotted the presence of DNA methyltransferases (DNMT1, DNMT3) on a phylogram representing the phylogenetic relationships among all investigated species. Additionally, we plotted the presence of DNA methylation as inferred by the CpG o/e distributions of investigated species on this phylogram (DNMT1: dark gray, DNMT3: light gray, DNA methylation: red). The phylogenetic relationships of depicted insect orders and outgroups are congruent with the proposed relationships in Misof et al. (2014). DNMT1 is found in species belonging to all insect orders except in Collembola, Strepsiptera, and Diptera. DNMT3 was only identified in seven insect orders. Methylation-indicative CpG o/e distributions were identified in species belonging to 24 insect orders plus crustacean and myriapod species. PALAE, Palaeoptera; COND, Condylognatha; NEUROPTER, Neuropterida.

fig. S1; table S7, Supplementary Material online. Furthermore, our data suggest that DNA methylation is applied by close relatives of insects, as we found signatures of DNA methylation in crustaceans (four out of seven species),

including the only representative of remipedes (the proposed sister group of insects; Misof et al. 2014) *Xibalbanus tulumensis*, and in the diplopod, *Glomeris pustulata* (fig. 2; supplementary fig. S1; table S7, Supplementary Material online). Interestingly, however, CpG o/e distributions pointing to the presence of DNA methylation were not consistently observed in apterygote insect orders, as only species of Diplura (one out of two species) and Zygentoma (two out of three species), but not Protura (one species), Collembolla (seven species), or Archaeognatha (two species) showed signs strongly suggesting the occurrence of DNA methylation. In contrast, we found consistent evidence for the occurrence of DNA methylation in winged hemimetabolous insects, including all representatives of Palaeoptera (all seven species), all polyneopteran orders, except Dermaptera (24 out of 27 species), and many representatives of Condylognatha (Hemiptera [ten out of 16 species], Thysanoptera [all three species]; fig. 2; supplementary fig. S1; table S7, Supplementary Material online). CpG o/e distributions strongly suggesting the presence of DNA methylation are comparatively sparse in Holometabola (17 out of 70 species in total). Representatives of Diptera (15 species), Neuroptera (four species), Raphidioptera (two species), and Strepsiptera (two species) showed no signs of DNA methylation. These results show that CpG o/e distributions pointing to the presence of germline DNA methylation in protein-coding sequences can be easily tracked in the majority of hemimetabolous insects, but are largely absent from holometabolous species.

We did not identify CpG o/e distributions pointing to the presence of DNA methylation in any of the species belonging to eight insect orders (i.e., Protura, Collembola, Archaeognatha, Dermaptera, Neuroptera, Raphidioptera, Strepsiptera, and Diptera). However, in certain species belonging to Archaeognatha, Collembola, Diptera, and Protura, unimodal CpG o/e distributions displayed low mean values (below 0.9 and as low as ~0.7) whereas corresponding GpC o/e distributions displayed mean values close to the expected ones under random chance (mean ~0.9 or higher; supplementary table S7, Supplementary Material online). These mean CpG o/e values are lower than the ones observed in species with extremely reduced or no DNA methylation (*Aedes aegypti* ~1.1, *Anopheles gambiae* ~1.0, *D. melanogaster* ~0.9, *T. castaneum* ~1.1).

Normalized CpG Content Points to Lower Levels of DNA Methylation in Holometabola

Normalized CpG content constitutes a powerful means for drawing conclusions not only for the patterns, but also for the levels of genomic DNA methylation (Yi and Goodisman 2009). Thus, we calculated the mean CpG o/e value of each transcriptome included in our analysis. First, we compared mean CpG o/e values of holometabolous insects (52 species), to those of hemimetabolous insects (67 species) and outgroup species (six crustacean and two myriapod species; fig. 3a). Holometabolous insect species exhibited higher overall mean CpG o/e values (lower mean germline DNA methylation) in protein-coding sequences compared with both hemimetabolous insects and outgroups (Kruskal–Wallis H test, $P < 0.001$; ignoring phylogenetic relatedness). Subsequently, we compared mean CpG o/e values of insect species separated by order (fig. 3c). The majority of species belonging to hemimetabolous insect orders shows lower mean CpG o/e values than species belonging to holometabolous orders, except Dermaptera and Psocodea. Species belonging to Zygentoma, Odonata, and most polyneopteran orders (excluding Dermaptera) consistently display very low mean CpG o/e values, with Mantodea representing the most extreme example. Condylognathan species (i.e., Hemiptera and Thysanoptera) tend to display higher mean CpG o/e values than most Polyneoptera, but still clearly lower values than species belonging to orders of Holometabola. Proturan, collembolan, and dipluran species exhibit higher mean values than species of Palaeoptera and Polyneoptera, with the exception of Dermaptera. In conclusion, mean CpG o/e values suggest lower levels of germline DNA methylation in the protein-coding sequences of Holometabola and their closest relatives, Psocodea (Misof et al. 2014).

Despite offering a decent first approximation on the levels of DNA methylation within genes (Sarda et al. 2012), CpG o/e is also suggested to be influenced by other factors, such as local GC content (Fryxell and Moon 2005) and recombination or gene conversion (Kent et al. 2012), for which we cannot currently control. Furthermore, certain insect lineages (most commonly Hymenoptera) are known to possess high mean CpG o/e values, genome-wide (Simola et al. 2013). Thus, we tested whether our observation that levels of DNA methylation are lower in protein-coding sequences of Holometabola compared with hemimetabolous insects still holds when using experimental DNA methylation data. To do that, we exploited the recently published and most comprehensive to date insect DNA methylation data set, encompassing holometabolous species from four orders (Hymenoptera, Coleoptera, Lepidoptera, and Diptera) and hemimetabolous species from three orders (Isoptera, Blattodea, and Hemiptera) published by Bewick et al. (2017).

We performed a PGLS analysis, to measure the strength of phylogenetic signal (following the definition by Revell et al 2008) between DNA methylation in protein-coding sequences and the mode of insect development (hemimetabolism or holometabolism). To measure phylogenetic signal we used Pagel's lambda (λ; Pagel 1999). In brief, a λ equal to one (λ_1) corresponds to traits being as similar among species as expected from the phylogenetic tree, assuming a Brownian motion model of evolution. In contrast, a λ equal to zero (λ_0) suggests species traits evolving independently from the phylogenetic tree. We estimated weak phylogenetic signal between DNA methylation and the mode of insect
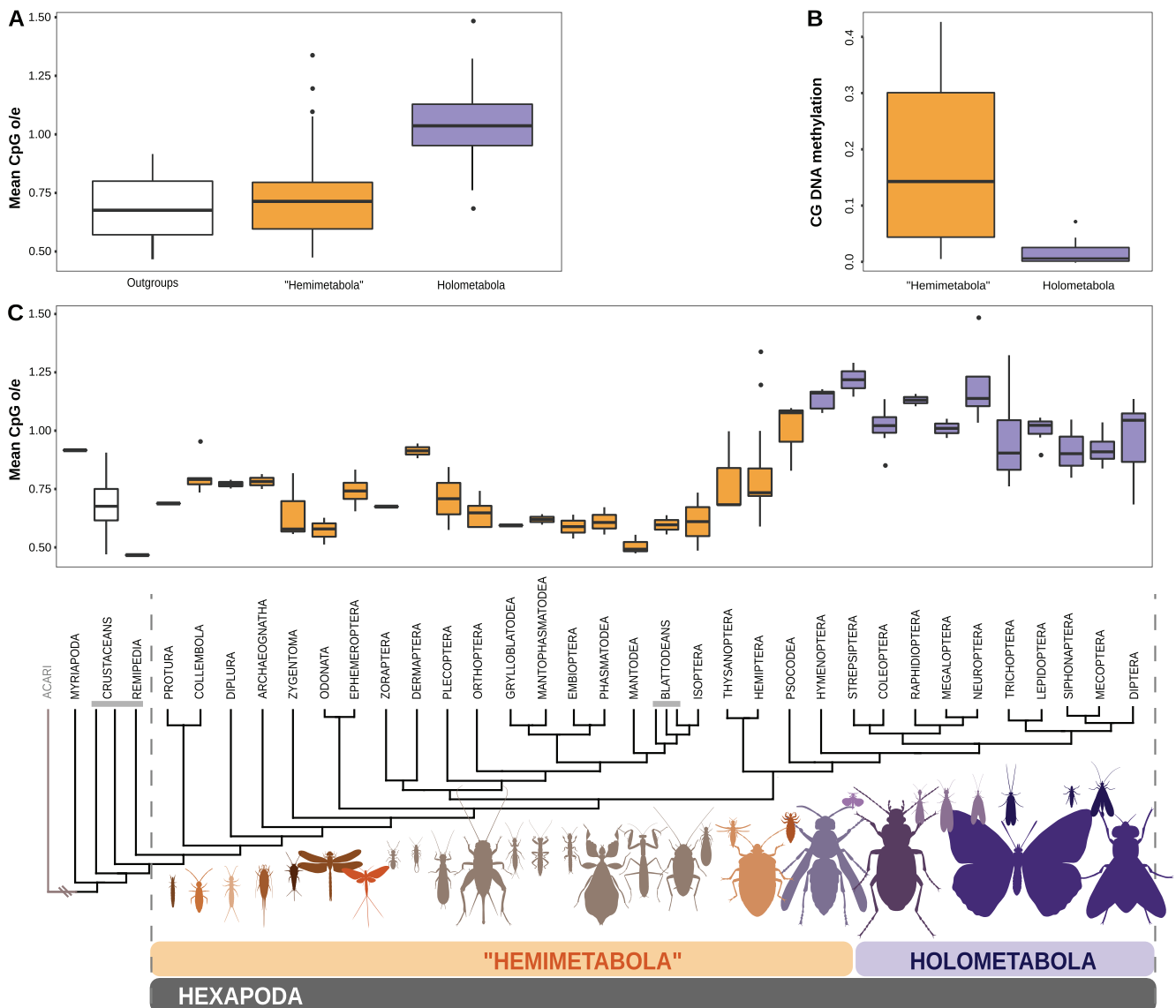
Fig. 3.—(a) Comparison of mean normalized CpG dinucleotide content (CpG o/e) among species belonging to Holometabola (52 species, violet box plot), Hemimetabola (67 species, orange box plot), and other arthropod outgroups (8 species, white box plot) based on investigated transcriptomes (127 species in total, representing all currently recognized insect orders plus crustacean and myriapod outgroups). We tested whether the difference of mean CpG o/e values among Hemimetabola, Holometabola, and outgroups was significant with a Kruskal–Wallis H test ($P < 0.001$). (b) Comparison of CG DNA methylation levels between the protein-coding sequences of 14 hemimetabolous and 26 holometabolous insect species. Holometabolous species display lower levels of DNA methylation in protein-coding sequences compared with hemimetabolous species (Mann–Whitney U test $P < 001$). (c) Comparison of mean CpG o/e values of species described in (a) separated by insect order. The CpG o/e levels strongly vary among insect orders, but orders of Holometabola show higher overall mean CpG o/e values than orders belonging to hemimetabolous insects.

development ($\lambda_{ml} = 0.047$). Most importantly, λ_{ml} was significantly different from λ_1, but not significantly different from λ_0 (supplementary table S8, Supplementary Material online). Thus, we can directly compare DNA methylation values between holometabolous and hemimetabolous insects as the traits in this data set are independent from the given phylogeny. Similar to our CpG o/e comparisons, we found that holometabolous insects tend to display significantly lower DNA methylation levels in protein-coding sequences compared

with hemimetabolous insects (Mann–Whitney U test, $P < 0.001$; fig. 3b).

Single-Copy Genes across All Insect Orders Show Signs of High DNA Methylation

Sarda et al. (2012) showed that most evolutionarily conserved genes tend to be highly methylated among four distantly related invertebrates. We investigated whether there is a

congruent pattern among insects. For this purpose, we analyzed a set of 1, 478 clusters of nuclear-encoded protein-coding genes that have been retained in single-copy across insects and whose DNA sequences we obtained from the genomes and transcriptomes of 141 species representing all insect orders and other arthropods (Misof et al. 2014). For each transcriptome/official gene set we compared the CpG o/e distribution of all transcripts/genes with the CpG o/e distribution of the corresponding set of single-copy genes. We found that in species that possess methylation-indicative CpG o/e distributions, these single-copy genes tend to be overrepresented among low CpG o/e genes (supplementary fig. S2, Supplementary Material online). To clearly display this relationship, we compared the median CpG o/e value of all transcripts/genes to the median CpG o/e value of the single-copy gene set of each species. Specifically, we selected a conservative set of taxa that according to our analysis and/or empirical evidence do not display signs of DNA methylation (i.e., lack of DNMT1 and DNMT3 accompanied by a CpG o/e distribution that does not indicate the presence of DNA methylation, or experimentally verified lack of CG DNA methylation from protein-coding sequences), namely Collembola, Strepsiptera, and Diptera (see Discussion), plus two beetles (Coleoptera), *T. castaneum* and *Dendroctonus ponderosae*, and calculated a linear regression between the median CpG o/e values of all transcriptomes/official gene sets and the corresponding set of single-copy genes. Using these taxa as reference, we found that in a number of species the calculated median CpG o/e value of the set of single-copy genes is significantly lower than the median CpG o/e value of the corresponding transcriptome/official gene set (fig. 4; supplementary table S9, Supplementary Material online). Overall, we found that genes that are consistently present across diverse insect lineages and possess highly conserved amino-acid sequences tend to exhibit low CpG o/e values, thus, high historical levels of germline DNA methylation.

Discussion

The Taxonomic Distribution of DNMTs in Insects

Our results suggest that DNMT1 was present in the last common ancestor of all insects and the last common ancestor of each extant insect order, except Collembola, Diptera, and Strepsiptera. Furthermore, our results are in agreement with previously published work on species of Diptera (reviewed by Glastad et al. 2011; Falckenhayn et al. 2016; Bewick et al. 2017) and Strepsiptera (Niehuis et al. 2012). The losses of DNMT1 in Collembola, Strepsiptera, and Diptera are certainly evolutionarily independent phenomena, because phylogenetic reconstructions rule out a close relationship among these lineages (Misof et al. 2014). We conclude that the loss of DNMT1 in insects is an evolutionarily rare event. In contrast, DNMT3 has been possibly lost numerous times
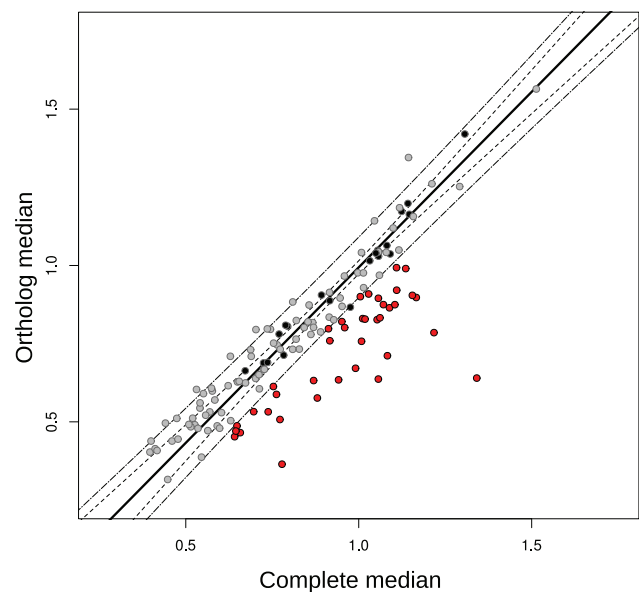

Fig. 4.—Comparison of the median CpG o/e value of all transcripts/ genes of a transcriptome/official gene set (complete median) with the median CpG o/e value of a subset of 1,478 single-copy genes with orthologs across 141 insect and other arthropod species (ortholog median). Black dots indicate species with no signs of DNA methylation according to our analysis and/or experimental evidence (species from the orders Collembola, Strepsiptera, Diptera, plus two beetles, *Dendroctonus ponderosae* and *Tribolium castaneum*). On the basis of the median CpG o/e values of these species, we calculated a linear regression (black solid line). The black dashed lines indicate the confidence intervals and the black dash-dotted lines indicate the prediction intervals that were calculated based on this regression. Species in which the median CpG o/e of single-copy genes is significantly lower than the median CpG o/e of the transcriptomic/genomic background are colored red (dots below the lower dash-dotted line). The remaining species are shown in gray.

during the evolutionary history of insects. Independent DNMT3 gains constitute an unlikely scenario for insects (Bewick et al. 2017). We did not identify DNMT3 in major insect groups such as Mecopterida, Palaeoptera, Neuropterida, and most Polyneoptera (except Orthoptera and Isoptera). However, the absence of DNMT3 from the inspected transcriptomes could be attributed to low or no expression of the corresponding gene. For example, we did not find DNMT3 in the transcriptome of the brown planthopper, *Nilaparvata lugens*, although it was shown that DNMT3 is weakly expressed in all life stages, but the mated and gravid females of this species (Zhang et al. 2015). In Mecopterida, our dense taxonomic sampling (40 species) combined with the availability of sequenced genomes provide congruent evidence for the loss of DNMT3 in this clade (Misof et al. 2014). Furthermore, Bewick et al. (2017) did not identify DNMT3 in the genomes of two palaeopteran species, in congruence with our results. The case is less clear in Neuropterida and Polyneoptera (excluding Isoptera and Orthoptera). In these clades, our species sampling per order is comparatively low

and sequenced genomes were not yet published. To conclude, DNMT1 and DNMT3 do not constitute an indispensable functional pair in insects (in contrast to vertebrates), because the insect DNMT toolkit seems to be mainly comprised of DNMT1 homologs.

CpG o/e Patterns When DNMTs Are Present

On the basis of CpG o/e distributions, it is reasonable to assume that species belonging to Trichoptera, Siphonaptera, Lepidoptera, Mecoptera (all belong to Mecopterida), Odonata, and Ephemeroptera (together form Palaeoptera) possess functional methylation systems despite the apparent loss of DNMT3. DNA methylation occurs in species belonging to 20 additional insect orders based on indicative CpG o/e distributions and DNMT3 complemented DNMT1 in just seven of them. Thus, our data indicate that DNA methylation is established and maintained without DNMT3 homologs in a possibly wide range of insect taxa. In Protura, Archaeognatha, Dermaptera, Raphidioptera, and Neuroptera, only copies of DNMT1 were found in at least one species per order, but the corresponding CpG o/e distributions are not unequivocally pointing to the presence of DNA methylation. However, some insect species with experimentally verified DNA methylation at protein-coding sequences lack bimodal CpG o/e distributions despite the presence of either DNMT1 or both DNMT1 and DNMT3 (Glastad et al. 2011; Oxley et al. 2014; Libbrecht et al. 2016). Therefore, DNA methylation probably occurs at an even higher number of insect orders than the ones specified here.

The likely presence of CG methylation in protein-coding sequences of multiple insect taxa despite the absence of DNMT3 homologs, shows that the definition of a functional methylation toolkit needs to be redefined in insects. In certain species, like *B. mori* or the paper wasp *Polistes canadensis*, which possess a single DNMT1 homolog as their only identified DNMT (Xiang et al. 2010; Patalano et al. 2015), it is possible that DNA methylation is introduced and maintained by this one enzyme (Maleszka 2016). However, in some insects, including multiple Hymenoptera and the human body louse, *P. humanus*, which also lacks DNMT3, more than one DNMT1 homologs are present (Glastad et al. 2011; Lyko and Maleszka 2011). Thus, certain DNMT1 paralogs may have shifted their function and are able to methylate de novo and/or in contexts other than CG, similar to vertebrate DNMT3 enzymes. Another scenario is that a novel and currently unknown enzymatic machinery may be able to carry out DNA methylation in insects (Glastad et al. 2011; Maleszka 2016).

CpG o/e Patterns When DNMT1 and DNMT3 Are Absent

It has been shown that the absence of DNMT1 and DNMT3 from the genomes of invertebrate species, including the dipteran insects, *A. aegypti*, *Aedes albopictus*, *A. gambiae*, and *D. melanogaster*, the nematode, *Caenorhabditis elegans*, and the trematode *Schistostoma mansonii*, is correlated with the absence or extreme reduction of DNA methylation (Simpson et al. 1986; Raddatz et al. 2013; Falckenhayn et al. 2016; Bewick et al. 2017). In line with this observation, we did not identify DNMT1, DNMT3, or methylation-indicative CpG o/e distributions in protein-coding sequences of species belonging to Collembola, Diptera, and Strepsiptera. Thus, because DNA methylation is predominantly found in CG context at protein-coding sequences across insects (Bewick et al. 2017), it is highly probable that species in these three orders lack or show extremely low levels of DNA methylation. Only TRDMT1 homologs were identified in these species, reflecting the predicted absence of DNA methylation. The potential losses or extreme reductions of DNA methylation and its accompanying machinery in species belonging to three phylogenetically distinct insect lineages support the notion that DNA methylation might not be vital for the proper ontogenetic development of various insect species (Lyko and Maleszka 2011; Raddatz et al. 2013).

The Taxonomic Distribution of Tet Dioxygenases in Insects

Our results show that Tet dioxygenases are widely distributed across insects, because we identified homologs in species belonging to most insect orders. The underrepresentation of putative Tet homologs in transcriptomes compared with genomes can be attributed to low or no expression of the Tet gene and hence its absence from the analyzed transcriptomes. The identification of Tet homologs in the genome, but not the transcriptome of the springtail *Folsomia candida* or the mountain pine beetle *D. ponderosae* substantiate this idea. The presence of Tet homologs in species belonging to Collembola and Diptera, in which according to our analyses and/or experimental evidence (Bewick et al. 2017) DNA methylation is extremely reduced or absent is in line with the proposed multifunctional role of Tet enzymes in insect genomes (Maleszka 2016). In Collembola, Diptera, or other insects in which DNA methylation is extremely reduced or absent, Tet homologs may act as 6mA DNA demethylases and/or 5mC mRNA demethylases, similar to their roles in *D. melanogaster* (Zhang et al. 2015; Delatte et al. 2016). Thus, the presence of Tet enzymes in insects may not be strictly correlated to its most designated function, that is, 5mC DNA demethylation.

The Presence of DNA Methylation Is Ancestral to Insects

The identification of a complete DNMT toolkit and the presence of methylation-indicative CpG o/e distributions in crustaceans show that DNA methylation is probably ancestral to insects. The absence of DNMTs from the transcriptome of the remipede *X. tulumensis* should be considered a limitation of this specific transcriptomic data set, because the species shows signs of heavy CpG depletion of protein-coding sequences, while no other remipede species was examined.

Thus, the potential losses or extreme reductions of DNA methylation and its machinery from insect groups are secondary, lineage-specific events (Glastad et al. 2011). This pattern shows that DNA methylation is a dispensable epigenetic mechanism for insects and its function may be compensated by other molecular mechanisms (Glastad et al. 2011; Raddatz et al. 2013).

DNA Methylation Has Been Reduced in Holometabola

The sparse presence of DNA methylation observed in holometabolous species and comparative analyses between two holometabolous insects (*A. mellifera* and *B. mori*) and two other invertebrates (*N. vectensis* and *C. intestinallis*; Sarda et al. 2012) led to the hypothesis that the levels of DNA methylation may have been reduced in the ancestors of insects (Glastad et al. 2014). However, our comparative analysis, combined with experimental evidence from single-species studies, point to a different scenario: The heavy CpG depletion of protein-coding sequences observed in the majority of species belonging to Zygentoma, Palaeoptera, Polyneoptera, and to a lesser extent Condylognatha, suggests that DNA methylation levels have been reduced in the ancestors of Holometabola, while there is no indication that DNA methylation levels were already reduced in the ancestors of insects. Our analysis of published empirical methylation data (Bewick et al. 2017) backs this hypothesis. Furthermore, empirical evidence obtained from direct measurements of DNA methylation in Orthoptera (*Schistocerca gregaria*, *Locusta migratoria*), Phasmatodea (*Medauroidea extradentata*), and Isoptera (*Zootermopsis nevadensis*) and computational evidence from analyzing Isoptera (*Zootermopsis nevadensis*, *Coptotermes lacteus*, *Reticulitermes flavipes*) support this conclusion. These polyneopteran species are, in comparison to holometabolous insects, characterized by significantly elevated levels of DNA methylation (Krauss et al. 2009; Falckenhayn et al. 2013; Glastad et al. 2013, 2016; Terrapon et al. 2014; Wang et al. 2014). Alternatively, the high mean CpG o/e values of Psocodea, the proposed sister group of Holometabola (Misof et al. 2014), suggest a reduction in the levels of DNA methylation that already occurred in the last common ancestor of Psocodea and Holometabola.

Evolutionary Conservation of Genes Is Strongly Associated with DNA Methylation in Insects

We showed that a set of single-copy genes that are associated with housekeeping functions (Misof et al. 2014) and have orthologs in all insects tend to display signatures of heavy DNA methylation in species with evident historical germline methylation. Our result is in line with those of previous investigations showing that the majority of orthologs among four distantly related invertebrates is extensively methylated (Sarda et al. 2012) and reveals that most evolutionarily conserved

housekeeping genes have been strongly methylated throughout insect evolution.

The evolutionary interconnection between DNA methylation and housekeeping genes may have a functional explanation. Bird (1995) conjectured that intragenic methylation may reduce transcriptional noise (high transcript variability) by suppressing spurious transcription initiation in vertebrate genomes. Both points of this hypothesis have recently received support by studies on mammalian systems. First, Huh and colleagues found that transcriptional noise is reduced in heavily methylated human genes (Huh et al. 2013). Second, Neri et al. (2017) showed that DNMT3-dependent intragenic DNA methylation acts to prevent spurious transcription initiation in mouse cells. Reducing transcriptional noise could be especially beneficial for constitutively expressed housekeeping genes (Suzuki et al. 2007). Thus, it is likely that intragenic DNA methylation acts to reduce transcriptional noise on evolutionarily conserved housekeeping genes in insects, perhaps with a mechanism similar to the one described by Neri et al. (2017). However, because many insect species that show signs of intragenic DNA methylation seem to lack DNMT3 homologs, a DNMT3-independent enzymatic machinery would contribute to a noise reduction mechanism in certain insects.

Conclusions

Our results provide an invaluable resource for experimental studies designed towards continuing this line of work. Experimental tests designed for investigating the functional role of DNMT1 homologs should be applied, by employing, for example, RNAi and/or CRISPR/Cas based methods, especially in DNMT3-deficient species. Additionally, large scale comparative studies using direct measurements of DNA methylation, such as whole genome bisulfite sequencing, should be conducted. Applying such approaches will not only aid in estimating the levels of DNA methylation in certain lineages, but also in determining the genomic targets of DNA methylation with accuracy, which in turn may provide important insights towards understanding its function in insects.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

Literature Cited

Bewick AJ, Vogel KJ, Moore AJ, Schmitz RJ. 2017. Evolution of DNA methylation across insects. Mol Biol Evol. 34(3):654–665.

Bird AP. 1980. DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res. 8(7):1499–1504.

Bird AP. 1995. Gene number, noise reduction and biological complexity. Trends Genet. 11(3):94–100.

Bonasio R, et al. 2012. Genome-wide and caste-specific DNA methylomes of the ants *Camponotus floridanus* and *Harpegnathos saltator*. Curr Biol. 22(19):1755–1764.

Camacho C, et al. 2009. BLAST plus: architecture and applications. BMC Bioinformatics. 10(1):421.

Cunningham CB, et al. 2015. The genome and methylome of a beetle with complex social behavior, *Nicrophorus vespilloides* (Coleoptera: silphidae). Genome Biol Evol. 7(12):3383–3396.

Delatte B, et al. 2016. Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. Science 351(6270):282–285.

Elango N, Hunt BG, Goodisman MAD, Yi SV. 2009. DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*. Proc Natl Acad Sci U S A. 106(27):11206–11211.

Falckenhayn C, et al. 2013. Characterization of genome methylation patterns in the desert locust *Schistocerca gregaria*. J Exp Biol. 216(Pt 8):1423–1429.

Falckenhayn C, et al. 2016. Comprehensive DNA methylation analysis of the *Aedes aegypti* genome. Sci Rep. 6:36444.

Feng S, et al. 2010. Conservation and divergence of methylation patterning in plants and animals. Proc Natl Acad Sci U S A. 107(19):8689–8694.

Finn RD, et al. 2014. Pfam: the protein families database. Nucleic Acids Res. 42(D1):D222–D230.

Foret S, et al. 2012. DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. Proc Natl Acad Sci U S A. 109(13):4968–4973.

Fryxell KJ, Moon WJ. 2005. CpG mutation rates in the human genome are highly dependent on local GC content. Mol Biol Evol. 22(3):650–658.

Glastad KM, Gokhale K, Liebig J, Goodisman MAD. 2016. The caste- and sex-specific DNA methylome of the termite *Zootermopsis nevadensis*. Sci Rep. 6(1):37110.

Glastad KM, Hunt BG, Goodisman MAD. 2014. Evolutionary insights into DNA methylation in insects. Curr Opin Insect Sci. 1:25–30.

Glastad KM, Hunt BG, Goodisman MAD. 2013. Evidence of a conserved functional role for DNA methylation in termites. Insect Mol Biol. 22(2):143–154.

Glastad KM, Hunt BG, Yi SV, Goodisman MAD. 2011. DNA methylation in insects: on the brink of the epigenomic era. Insect Mol Biol. 20(5):553–565.

Goll MG, Bestor TH. 2005. Eukaryotic cytosine methyltransferases. Annu Rev Biochem. 74:481–514.

Goll MG, et al. 2006. Methylation of tRNA. Science 311(5759):395–398.

Huh I, Zeng J, Park T, Yi SV. 2013. DNA methylation and transcriptional noise. Epigenet Chromatin. 6(1):9.

Hunt BG, Glastad KM, Yi SV, Goodisman MAD. 2013. Patterning and regulatory associations of DNA methylation are mirrored by histone modifications in insects. Genome Biol Evol. 5(3):591–598.

Jeltsch A, Jurkowska RZ. 2014. New concepts in DNA methylation. Trends Biochem Sci. 39(7):310–318.

Jones PA. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. 13(7):484–492.

Kao D, et al. 2016. The genome of the crustacean *Parhyale hawaiensis*, a model for animal development, regeneration, immunity and lignocellulose digestion. eLIFE 5:e20062.

Kapheim KM, et al. 2015. Genomic signatures of evolutionary transitions from solitary to group living. Science 348(6239):1139–1143.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780.

Kent CF, Minaei S, Harpur BA, Zayed A. 2012. Recombination is associated with the evolution of genome structure and worker behavior in honey bees. Proc Natl Acad Sci. 109:18012–18017.

Krauss V, Eisenhardt C, Unger T. 2009. The genome of the stick insect *Medauroidea extradentata* is strongly methylated within genes and repetitive DNA. PLoS ONE. 4(9):e7223.

Kriventseva EV, et al. 2015. OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. Nucleic Acids Res. 43(D1):D250–D256.

Kucharski R, Maleszka J, Foret S, Maleszka R. 2008. Nutritional control of reproductive status in honeybees via DNA methylation. Science 319(5871):1827–1830.

Libbrecht R, Oxley PR, Keller L, Kronauer DJC. 2016. Robust DNA methylation in the clonal raider ant brain. Curr Biol. 26(3):391–395.

Lyko F. 2017. The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. Nat Rev Genet. 19(2):81–92.

Lyko F, et al. 2010. The honey bee epigenomes: differential methylation of brain DNA in queens and workers. PLoS Biol. 8(11): e1000506.

Lyko F, Maleszka R. 2011. Insects as innovative models for functional studies of DNA methylation. Trends Genet. 27(4):127–131.

Maleszka R. 2016. Epigenetic code and insect behavioural plasticity. Curr Opin Insect Sci. 15:45–52.

Mayer C, et al. 2016. BaitFisher: A software package for multi-species target DNA enrichment probe design. Mol Biol Evol. 33(7):1875–1886.

Misof B, et al. 2014. Phylogenomics resolves the timing and pattern of insect evolution. Science 346(6210):763–767.

Munoz-Torres MC, et al. 2011. Hymenoptera Genome Database: integrated community resources for insect species of the order Hymenoptera. Nucleic Acids Res. 39(Database):D658–D662.

Neri F, et al. 2017. Intragenic DNA methylation prevents spurious transcription initiation. Nature 543(7643):72–77.

Niehuis O, et al. 2012. Genomic and morphological evidence converge to resolve the enigma of Strepsiptera. Curr Biol. 22(14):1309–1313.

Oxley PR, et al. 2014. The genome of the Clonal raider ant *Cerapachys Biroi*. Curr Biol. 24(4):451–458.

Pagel M. 1999. Inferring the historical patterns of biological evolution. Nature 401:877–884.

Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20(2):289–290.

Park J, et al. 2011. Comparative analyses of DNA methylation and sequence evolution using Nasonia genomes. Mol Biol Evol. 28(12):3345–3354.

Pastor WA, Aravind L, Rao A. 2013. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. Nat Rev Mol Cell Biol. 14(6):341–356.

Patalano S, et al. 2015. Molecular signatures of plastic phenotypes in two eusocial insect species with simple societies. Proc Natl Acad Sci U S A. 112(45):13970–13975.

Petersen M, et al. 2017. Orthograph: a versatile tool for mapping coding nucleotide sequences to clusters of orthologous genes. BMC Bioinformatics. 18(1):1–10.

Pinheiro JC, Bates DJ, DebRoy SD, Sarkar D, and R Core Team. 2017. nlme: linear and nonlinear mixed effects models. R package version 3.1-131, https://CRAN.R-project.org/package=nlme, last accessed November 2017.

R Core Team (2016). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Raddatz G, et al. 2013. Dnmt2-dependent methylomes lack defined DNA methylation patterns. Proc Natl Acad Sci U S A. 110(21): 8627–8631.

Rehan SM, Glastad KM, Lawson SP, Hunt BG. 2016. The genome and methylome of a subsocial small carpenter bee, Ceratina calcarata. Genome Biol Evol. 8(5):1401–1410.

Revell LJ, Harmon LJ, Collar DC, Oakley T. 2008. Phylogenetic signal, evolutionary process, and rate. Syst Biol. 57(4):591–601.

Sarda S, Zeng J, Hunt BG, Yi SV. 2012. The evolution of invertebrate gene body methylation. Mol Biol Evol. 29(8):1907–1916.

Schübeler D. 2015. Function and information content of DNA methylation. Nature 517(7534):321–326.

Simola DF, et al. 2013. Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality. Genome Res. 23(8):1235–1247.

Simpson VJ, Johnson TE, Hammen RF. 1986. Caenorhabditis elegans DNA does not contain 5-methylcytosine at any time during development or aging. Nucleic Acids Res. 14(16):6711–6719.

Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics. 6:31.

Standage DS, et al. 2016. Genome, transcriptome, and methylome sequencing of a primitively eusocial wasp reveal a greatly reduced DNA methylation system in a social insect. Mol Ecol. 25(8): 1769–1784.

Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 34(Web Server):W609–W612.

Suzuki MM, Bird A. 2008. DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet. 9(6):465–476.

Suzuki MM, Kerr ARW, De Sousa D, Bird A. 2007. CpG methylation is targeted to transcription units in an invertebrate genome. Genome Res. 17(5):625–631.

Terrapon N, et al. 2014. Molecular traces of alternative social organization in a termite genome. Nat Commun. 5:3636.

Wang X, et al. 2013. Function and evolution of DNA methylation in Nasonia vitripennis. PLoS Genet. 9(10):e1003872.

Wang X, et al. 2014. The locust genome provides insight into swarm formation and long-distance flight. Nat Commun. 5:2957.

Wang Y, et al. 2006. Functional CpG methylation system in a social insect. Science 314(5799):645–647.

Werren JH, et al. 2010. Functional and evolutionary insights from the genomes of three parasitoid Nasonia species. Science 327(5963): 343–348.

Wojciechowski M, et al. 2014. Insights into DNA hydroxymethylation in the honeybee from in-depth analyses of TET dioxygenase. Open Biol. 4:140110.

Xiang H, et al. 2010. Single base-resolution methylome of the silk moth reveals a sparse epigenomic map. Nat Biotechnol. 28(5):516–520.

Yi SV, Goodisman MAD. 2009. Computational approaches for understanding the evolution of DNA methylation in animals. Epigenetics 4(8):551–556.

Zdobnov EM, et al. 2017. OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. Nucleic Acids Res 45:744–749.

Zemach A, McDaniel IE, Silva P, Zilberman D. 2010. Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science 328(5980):916–919.

Zhang G, et al. 2015. N6-methyladenine DNA modification in Drosophila. Cell 161(4):893–906.

Associate editor: Gunter Wagner