**TITLE:** Sociality and DNA methylation are not evolutionary dependent

**RUNNING TITLE:** Independence of sociality and DNA methylation

Adam J. Bewick[1*], Kevin J. Vogel[2], Allen J. Moore[1], Robert J. Schmitz[1]

[1]Department of Genetics, University of Georgia, Athens, GA 30602, USA

[2]Department of Entomology, University of Georgia, Athens, GA 30602, USA

**CO-CORRESPONDENCE:** Adam J. Bewick (bewickaj@uga.edu) and Robert J. Schmitz (schmitz@uga.edu)

**KEY WORDS:** DNA methylation, whole genome bisulfite sequencing, social behavior, phylogenetic comparative methods

## ABSTRACT

Despite widespread conservation across the tree of life, little is known about how DNA methylation contributes to the evolution of complex traits. In particular, DNA methylation has been hypothesized to allow the evolution of highly flexible traits, such as sociality. We sought to better understand patterns of DNA methylation and its association between the expression of advanced social behavior across insects. DNA methylation in insects is widespread and found in social and solitary species from all orders, except Diptera (flies). Solitary species within Blattodea (cockroaches, termites) had the highest levels of DNA methylation. The presence/absence of underlying methyltransferases corroborates most patterns observed, but alternative DNA methylation pathways may exist. Furthermore, we found no evidence that supports evolutionary dependency between either advanced social behavior or division of labor and DNA methylation within insects using phylogenetically corrected comparisons. These results suggest that DNA methylation is not the driver of social behavior.

## INTRODUCTION

DNA methylation is recognized as an important chromatin modification providing structural integrity and proper regulation of the genome for many species. In animals DNA methylation typically occurs at CG sites, established *de novo* by DNA methyltransferase 3 (DNMT3), and maintained by the maintenance methyltransferase DNMT1 [1–3]. Homologous to DNMT1 and 3 is DNMT2: a $tRNA^{Asp}$ DNA methyltransferase, which does not contribute to the DNA methylome [4]. Despite CG DNA methylation conservation across the tree of life [5–7], our understanding of its contribution to complex traits is limited. Although, it has been argued that heritable epigenetic variation is an important contributor to evolutionary change, direct evidence for this hypothesis is limited. In the model plant species *Arabidopsis thaliana*, the use of epigenetic recombinant inbred lines (epiRILs) has allowed researchers to quantitatively assess the epigenetic basis of complex traits [8]. In insects, methylation is implicated in behavioral plasticity and especially social behavior [9,10]. Unfortunately, this is difficult to test because the generation of epiRILs outside of plants remains elusive. As an alternative, comparative methods have been

used to test for conserved function of DNA methylation [11,12]. In insects, comparisons between pairs of species within the order Hymenoptera (ants, bees and wasps) have been used to suggest that DNA methylation is important for the expression of social behavior [11,13]. By and large taxonomic sampling is sparse, and a test for the evolutionary relationship between sociality and DNA methylation has yet to be performed.

Here we used comparative phylogenetic approaches to test the hypothesis that DNA methylation is correlated with the evolution of sociality in 123 insect species from 11 orders. Specifically, we examined the phylogenetic distribution of DNA methylation and social behavior and then tested for a correlation between these two traits while controlling for non-independence of species relationships. We tested this at two levels; the evolution of any level of sociality (from solitary to eusocial) and the evolution of division of labor associated with eusociality. We collected and generated the largest sampling of DNA methylation-related data from publically available transcriptome and genome assemblies (123 species) in addition to previously published and newly generated DNA methylomes (41 species) to test our hypothesis. Our results revealed diverse patterns of DNA methylation across the insect tree of life, but no support for a relationship between the evolution of sociality and DNA methylation in insects.

## RESULTS AND DISCUSSION

**Presence of DNA methylation across the insect tree of life.** The distribution of $CpG_{O/E}$ of 123 insect species from orders Odonata, Ephemeroptera, Blattodea, Thysanoptera, Hemiptera, Phthiraptera, Hymenoptera, Coleoptera, Trichoptera, Lepidoptera, and Diptera were investigated for signatures of DNA methylation. DNA methylation, signified by a bimodal distribution of $CpG_{O/E}$ values, was identified in insect species from all orders except Diptera (Fig. 1A and B). Species belonging to the order Diptera comprised ~45.8% (55/120) of the total number of species investigated, which was the largest number of species sampled from one order. The absence of bimodal distribution from members of both the Nematocera and Brachycera sub-orders suggests that DNA methylation was lost early on in Dipteran evolution. The evolution of DNA methylation is deeply rooted in insects, going as far

back as the Ordovician and the divergence of Diplura (~489.84-436.68 MYA [14]), but is most likely older.

DNA methylation was found in insects exhibiting the complete range of social behavior based on modality of $CpG_{O/E}$ (Table S1). This includes insects with solitary, sub-social (parental care but no cooperative brood care), presocial (lacking one eusocial characteristic) and eusocial behavior (cooperative brood care, generational overlap and reproductive castes). For example, the solitary thrip *Frankliniella occidentalis* had an extremely pronounced bimodal $CpG_{O/E}$ distribution compared to the eusocial honey bee *Apis mellifera* (Fig. S1). The order Hymenoptera contains species that are solitary or eusocial; however, evidence for DNA methylation based on $CpG_{O/E}$ did not always co-occur with eusociality, and *vice versa*. Specifically, of the 31 Hymenopteran species investigated there was no association between social behavior and DNA methylation measured by $CpG_{O/E}$ (Pearson and $\chi^2$ P > 0.05; $\phi$-coefficient = Contingency coefficient = Cramer's V = 0.036) (Table S1). Moreover, all species examined within the order Lepidoptera are solitary, yet 8/10 showed evidence for DNA methylation based on the distribution of $CpG_{O/E}$, respectively (Table S1). Similarly, 3/4 solitary Coleopteran species were expected to have DNA methylation based on bimodality of $CpG_{O/E}$. Interestingly, six Diptera have sub-social behavior in the form of parental care (e.g., Table S1, the tsetse fly, *Glossina morsitans morsitans* [15]), yet do not have DNA methylation based on the distribution of $CpG_{O/E}$. This is contrary to the burying beetle *Nicrophorus vespilloides*, which has DNA methylation based on whole genome bisulfite sequencing (WGBS) and $CpG_{O/E}$[16], and also provide parental care through provisioning developing offspring [17] (Table S1). Overall, based on modality of $CpG_{O/E}$ DNA methylation is found in all insect orders except Diptera.

Estimates of DNA methylation from WGBS from 41 species revealed similar patterns. Levels of DNA methylation genome-wide and within coding regions were overall higher in the order Blattodea (Fig. 2). Furthermore, genome-wide and coding levels of DNA methylation were highest in the solitary cockroaches in the genus *Blattella*, *Blattella asahinai* and *Blattella germanica*, respectively. Drastic differences in levels of DNA methylation were observed between the highly eusocial species *Reticulitermes flavipes* and *Reticulitermes virginicus* (Fig. 2). Conversely, similar levels of DNA methylation were observed between closely related *Pemphigus*

species of aphids (Hemiptera) with differing social behaviors (Fig. 2). Overall, DNA methylation levels were variable in insects, with some levels reaching those observed in plants [5,6,18,19], and its contribution to complex traits still remains largely unknown despite more years of research [20–22].

**Evolution of DNMT1, DNMT2 and DNMT3 and DNA methylation.** DNMT1 is found in all orders of insects investigated in this study except Diptera (Fig. 3A; Table S2). Only DNMT2 is found in Diptera, which reflects the lack of any known DNA methylation in the genomes of all species belonging to this order (Fig. 1B; Table S1). The loss of DNA methylation most parsimoniously occurred once in all Diptera ~206.28-107.26 MYA [14], which likely coincided with the loss of DNMT1 and/or DNMT3.

Only DNMT1 and DNMT2 are found in Lepidopteran species (Fig. 3A; Table S2). This observation supports the hypothesis that DNMT3 was lost from species belonging to this order between ~177.99-116.45 MYA [14]. Also, the sister order to Lepidoptera, Trichoptera, also only possess DNMT1 and DNMT2. Therefore, the timing of this loss may be older, however only one species is represented by this order. Despite missing DNMT3 from assembled transcriptomes or genomes – as in the Lepidopteran *Bombyx mori* – evidence for DNA methylation is still observed from distributions of $CpG_{O/E}$ values (Fig. S1) and WGBS [17,23]. Similarly, 9/10 Hemiptera only possess DNMT1 and DNMT2, and DNA methylation is expected to be present given bimodality of $CpG_{O/E}$ (Table S1). DNMT1 (or 2) in Lepidoptera do not contain a PWWP domain (PF00855), which is a domain typically associated with *de novo* DNA methyltransferases in metazoa [24] that interacts with DNA and histone lysine modified nucleosomes [25,26]. PWWP domain containing proteins exist in Lepidoptera (e.g., *B. mori*), however none of these proteins contain a C-5 cytosine-specific DNA methylase domain (PF00145), which is required for DNA methylation [2,3,27]. Based on these observations two possible mechanisms include: (i) *de novo* DNA methylation does not occur in Lepidoptera via a DNMT-like protein, (ii) or DNA methylation is not reprogrammed during embryogenesis and is robustly maintained by DNMT1 during each cell replication. However, alternative mechanisms may exist, and functional tests of DNA methylation mechanisms in insects warrant further investigation.

Several duplications were observed in the DNMT1 clade, but none coincide

with presence or absence of DNA methylation or sociality. A duplication event shared by the superfamilies Apoidea (bees) and Vespoidea (wasps and ants) gave rise to what is referred to as DNMT1a and DNMT1b (Fig. 3B). This duplication is absent from species in the family Tenthredinidae and Orussidae, which suggests a timing for this event between ~343.00-283.00 MYA [28]. Relationships of superfamilies and families suggest reciprocal losses in Vespoidea: DNMT1a was lost from Vespidae (wasps), and DNMT1b from Formicidae, whereas DNMT1a and DNMT1b can both be found in Apoidea. Interestingly, mostly species-specific duplications – in-paralogs – were observed in orders Hemiptera and Lepidoptera (Fig. 3B), but several might represent allelic variation of DNMT1 for some species.

DNMT3 is the most order-poor of the DNA methyltransferases, and was only identified in species belonging to Coleoptera, Hemiptera, Hymenoptera, and Blattodea (Fig. 3A; Table S2). The association between presence of DNMT3 and DNA methylation is strong in Coleoptera with 3/7 species with a DNMT3 possessing a bimodal $CpG_{O/E}$ distribution (Table S1). A similar association is observed in Hymenoptera where 29/32 possess DNMT3 and 28 of the 29 are expected to have DNA methylation based on the distribution of $CpG_{O/E}$ (Table S1). *Lasioglossum albipes* is the one exception; this species possesses DNMT3, but is not expected to have DNA methylation based on $CpG_{O/E}$ although this contradicts a previous study [29]. As has been shown in a plant species [30,31], relaxed selection leading to non-neutral substitutions could have occurred at the DNMT3 locus in *Lasioglossum albipes*, which in turn affects the ability to methylate cytosines *de novo*. *Copidosoma floridanum* is an exception in the opposite direction and is similar to what is observed in Hemiptera and Lepidoptera; DNMT3 is not present, but DNA methylation is expected to be present.

**Sociality and DNA methylation are not evolutionary dependent.** DNA methylation has been proposed to control many aspects of sociality including behavior expressed in social interactions, caste determination, and learning and memory (reviewed in [20]). If sociality is dependent on DNA methylation we would expect a gain in DNA methylation to result in either a gain of social behavior or a predisposition to evolve social behavior. Thus, if sociality is dependent on DNA methylation we would expect a phylogenetically-corrected correlation between these

two traits. We do not find this. Sociality – including division of labor – is not evolutionary dependent on DNA methylation (Table 1). This lack of dependence holds true when DNA methylation is categorized as a continuous trait, and sociality is categorized into multiple discrete classes (Fig. 1A; Fig. 2; Table 1). Although there are differences in transition rates between social behavior and DNA methylation states, they are not dependent on one another (Table 1; Fig. S2). For example, transitions from social to solitary occur more frequently than solitary to social when DNA methylation is fixed (Fig. S2). Also, transitions from solitary to social, *in lieu* of DNA methylation, occur more frequently than transitions from social to solitary (Fig. S2). Furthermore, based on Akaike information criterion (AIC), an independent model of evolution for each trait is preferred (Table 1). These traits are not dependent when only species within Hymenoptera are considered, and similarly when all species within Diptera are removed (Table 1). This evidence supports the hypothesis that sociality and DNA methylation are not evolutionary dependent.

## CONCLUSION

We have conducted the largest phylogenetic investigation of DNA methylation associated with sociality in insects. We find that DNA methylation is too variable in insects to support a single, common role in regulating traits expressed in social interactions. DNA methylation is neither necessary nor sufficient for social evolution, and does not appear to facilitate sociality. DNA methylation can be found in both solitary and social insects, and does not always associate with presence of the *de novo* and maintenance DNA methyltransferase DNMT1 and DNMT3, respectively. DNA methylation may contribute indirectly or directly to sociality or caste formation, but it is not the quintessential mechanism by which to achieve this phenotype.

## MATERIALS AND METHODS

**Tissue collection and DNA extraction.** All samples were collected from established laboratory colonies with the following exceptions. *Cryptocercus garciai* was collected from northeast Georgia by Brian Forschler, *Polistes carolina* were collected from the campus of the University of Georgia by K. J. Vogel.

DNA extractions were performed on tissues from either freshly sacrificed

insects or ethanol preserved samples. For samples likely to contain significant gut microbial contamination (Blattodea), guts were dissected and discarded. Approximately 10 mg of material was frozen in liquid nitrogen, ground with a pestle. Samples were then extracted with the Qiagen DNEasy Mini kit following the manufacturer's instructions for animal tissues.

**Behavior classification.** Insect species were classified into one of four social behavior categories based on the presence or absence of well-defined traits [32]. *Eusocial* insects must share a common nest site, exhibit cooperative rearing of young, reproductive division of labor with sterile or less fecund workers, and have overlap of generations. Insect species that lacked any one of the previously listed traits were classified as *presocial*. Species that exhibited only parental care or cooperative brood rearing were classified as *subsocial*.

**Whole-genome bisulfite sequencing and levels of DNA methylation.** MethylC-seq libraries were prepared according to the following protocol [33]. Sequencing data for *Acyrthosiphon pisum*, *Aedes aegypti*, *Aedes albopictus*, *Anopheles gambiae*, *Apis mellifera*, *Blattella germanica*, *Camponotus floridanus*, *Copidosoma floridanum*, *Culex quinquefasciatus*, *Dinoponera quadriceps*, *Drosophila melanogaster*, *Harpegnathos saltator*, *Microplitis demolitor*, *Nasonia vitripennis*, *Nicrophorus vespilloides*, *Polistes canadensis*, *Solenopsis invicta*, *Tribolium castaneum* was aligned to their respective genome assembly using the methylpy pipeline [34] (Table S1). *Blattella asahinai* was aligned to its sister species *Blattella germanica*. *Dinoponera quadriceps* was removed from Fig. 2 and associated analyses because of possible genomic sequence contamination with other insect species. In brief, reads were trimmed of sequencing adapters using Cutadapt [35], and then mapped to both a converted forward strand (cytosines to thymines) and converted reverse strand (guanines to adenines) using bowtie [36]. Reads that mapped to multiple locations, and clonal reads were removed. Weighted methylation was calculated for each sequence context (CG, CHG and CHH) by dividing the total number of aligned methylated reads by the total number of methylated plus un-methylated reads. For the remaining species, genome-wide levels of DNA methylation were estimated using *FAST$^{m}$C* and the *animal* model [37].

CG DNA methylation in insects is enriched in coding sequences (see [16,38,39]),

and a strong correlation between genome-wide and within coding sequence levels of CG DNA methylation is observed (Fig. S3). Therefore, using genome-wide levels of CG DNA methylation estimated from *FAST$^m$C*, we were able to extrapolate levels of CG DNA methylation within coding sequence for species without sequenced genomes (Table S1).

**Measurement of $CpG_{O/E}$ and tests for bimodality.** $CpG_{O/E}$ is a metric of CpG dinucleotides normalized by G and C nucleotide content (GC content) and length (bp) of a specific region of interest (e.g., a transcript or protein coding gene) [13,40]. Due to spontaneous deamination of methylated cytosines, genes that are hypermethylated are expected to have a lower $CpG_{O/E}$ value than hypomethylated genes [41]. Thus, in a mixture of genes that are methylated and low to un-methylated, a bimodal distribution of $CpG_{O/E}$ values is expected. Conversely, a unimodal distribution is suggestive of a set of genes that are mostly low to un-methylated. The $CpG_{O/E}$ value for each gene within 125 total transcriptomes or gene annotations was defined as:

$$CpG_{O/E} = \left(\frac{l^2}{l}\right) \times \left(\frac{P_{CpG}}{P_C \times P_G}\right) \quad (1)$$

where $P_{CpG}$, $P_C$, and $P_G$ are the frequencies of CpG dinucleotides, C nucleotides, and G nucleotides, respectively, estimated from each gene of length ($l$) in bp. Only exonic sequences of a gene were considered when estimating $CpG_{O/E}$.

The modality of $CpG_{O/E}$ distributions was tested using several methods implemented in R v3.2.4, which included Ashman's D (modes v0.7.0), Silverman's test for multimodality (silvermantest v1.0), and Gaussian mixture modeling (mclust v5.2). For Ashman's D test, two normal distributions were fitted to the observed $CpG_{O/E}$ distribution and an Ashman's D ≥2 was used to distinguish uni- from bimodality. For Silverman's test a null hypothesis of 1 mode ($k = 1$) was tested using 10,000 bootstrap replicates, or until convergence was achieved. A p-value ≤0.05 was used as a cutoff to reject the null hypothesis. Finally, for Gaussian mixture modeling using mclust, two modes were modeled for each $CpG_{O/E}$ distribution, and the subsequent means and 95% confidence interval (CI) of the means were compared

with overlapping or non-overlapping CI's signifying unimodality or bimodality, respectively. Several methods for testing modality were performed because conflicting results were observed for $CpG_{O/E}$ distribution of species with sequenced DNA methylomes for the previously mentioned tests (Table S1). Based on published methylomes, Gaussian mixture modeling using mclust produced the most robust estimate of DNA methylation.

**Phylogenetic comparative methods.** The R package phytools [42] was used for all phylogenetic comparative methods. Two tests of correlated evolution were conducted. (i) Pagel's [43,44] method for detecting correlated evolution of two binary traits. Briefly, Pagel's method uses a continuous-time Markov model to simultaneously estimate transition rates in pairs of binary characters on a phylogeny. These rates are then used to test whether an independent or dependent model of evolution is preferred using the likelihood ratio test (LRT). DNA methylation was categorized into un-methylated or methylated based on the distribution of $CpG_{O/E}$, unimodal or bimodal, respectively. Sociality was categorized into social (eusocial, subsocial, and presocial) or solitary binary traits. Similarly, the *division of labor* trait was categorized into species with (caste$^{+}$) or without (caste$^{-}$) castes. (ii) Phylogenetic generalized least squares (PGLS) was used to correlate continuously categorized estimates of DNA methylation generated through WGBS and alignment to reference assemblies (methylpy [34]), or non-referenced based methods (*FAST$^{m}$C*, [37]) to discretely coded social traits (eusocial, parasocial, subsocial, presocial, and solitary). BEAST2 [43] was used to estimate a multilocus coalescent tree, which was used to control for relatedness of species (non-independence) for both comparative tests. For (i) the phylogenetic tree was estimated from a subset of previously identified orthologous protein coding loci (58 of the >1400 described in [14]). These loci were chosen because each contained no more than 5% missing species for 123 insects, and outgroups *Catajapyx aquilonaris* (Diplura) and *Daphnia pulex* (Crustacea) based on best BLASTp hit to a set of core species (*Acyrthosiphon pisum*, *Anopheles gambiae*, *Acromyrmex echinatior*, *Bombyx mori*, *Drosophila melanogaster*, *Apis mellifera*, *Ixodes scapularis*, *Nasonia vitripennis*, *Pediculus humanus*, *Tribolium castaneum*, *Zootermopsis nevadensis*, *Daphnia pulex*). Loci were aligned using PASTA [45], and Gblocks was used to identify sections of

conserved protein coding sequence [46]. For (ii) a smaller number of loci were used due to the obscurity of species and available data on Genbank. Loci used for both trees can be found in Table S3. For both trees, the program Tracer (http://tree.bio.ed.ac.uk/software/tracer/) was used to assess stationarity and effective sample size (ESS; ≥100) of the Markov Chain Monte Carlo (MCMC) chains.

**DNA methyltransferase (DNMT) phylogeny and evolution.** DNA methyltransferases (DNMT1, 2, and 3) were curated from 125 insect species through homology searches using BLASTp and previously identified DNMT1, 2, and 3 proteins in *Apis mellifera*, *Bombyx mori* (DNMT1 only), *Drosophila melanogaster* (DNMT2 only), and *Nicrophorus vespilloides*. A series of alignment and phylogenetic estimation steps were conducted eliminate partial and poor sequences, which can affect alignment and subsequently the topology of phylogeny. Protein sequences were aligned using the program PASTA [45], and back-translated using the CDS sequence to generate an in-frame codon alignment. RAxML with 1000 rapid bootstrap replicates and the GTR+G model of nucleotide substitution was used to generate the phylogeny [47]. The phylogeny was rooted to the DNMT2 clade in the program FigTree (http://tree.bio.ed.ac.uk/software/figtree/) and exported for stylization. An identical method was used to generate the Hymenoptera-specific DNMT1 gene tree. Nine iterations of alignment and phylogeny construction were performed removing sequences on long branches, which is indicative of low sequence homology. For species with assembled transcriptomes or gene annotations, additional DNA methyltransferases were identified by Interproscan [48], and filtering those sequences with a DNA methylase domain (PF00145). These DNA methylase domain-containing sequences were then subjected to the previously mentioned homology searches using BLASTp.

## ACKNOWLEDGEMENTS

We thank Lauren A. Eserman and Ben L. S. Furman for phylogenetic comparative advice. We thank Patrick Abbot, Gaelen Burke, Brian Forschler, Sarah E. Sander, Coby Schal, Kathrin Stanger-Hall, and Michael Strand for tissue and/or insect specimens. We also thank Patrick T. Griffin, Libby McKinney, and Nick A. Rohr for library preparation, and the Georgia Genomics Facility (GGF) for sequencing.

Computational resources were provided by the Georgia Advanced Computing Resource Center (GACRC). This work was supported by funding from the Office of the Vice President for Research at the University of Georgia to RJS, and by NIH award F32GM109750 to KJV.

## AUTHOR CONTRIBUTIONS

The study was conceived by AJB and RJS. All authors were involved in the experimental design. KJV and AJM curated samples. KJV and AJM classified social behavior for insects used in this study. AJB performed all analyses, with contribution from KJV to the phylogenetic analyses. AJB wrote the manuscript with contributions made by all authors.

## FIGURES AND TABLES

**Figure 1. DNA methylation in insects is not always associated with social behavior.** (A) Relationships of 123 insect species, and the outgroups *Catajapyx aquilonaris* (Dipluran) and *Daphnia pulex* (Crustacea) investigated with traits coded in binary. The tree was constructed from 58 nuclear protein coding loci, and was used in Pagel's test for evolutionary dependence [14,44]. All nodes except one had posterior probability of 1.00, which is indicated on the phylogeny. (B) A chronogram of insect order relationships with traits coded in binary. The chronogram was modified from [14]. For (A) and (B) traits are represented as shaded boxes above each species or order. Half-filled boxes indicate the trait is variable within the corresponding order.

**Figure 2. WGBS reveals extensive variation of DNA methylation in insects.** (A) Genomic levels of DNA methylation in insects ranges from zero (all Diptera examined) to ~14% (*Blattella asahinai*). Higher ranges are observed for coding regions; zero to ~42% (*Blattella germanica*). Overall, levels are highest in Blattodea, which do not always associate with social species. (B) A species tree constructed from nuclear and mitochondrial loci, which was used in Phylogenetic Generalized Least Squares (PGLS) analysis. Results from this analysis revealed that there is no correlation between social behavior and DNA methylation (Table 1). Values at nodes are posterior probabilities <0.95; all blank nodes have ≥0.95 posterior probability.

**Figure 3. Evolution of DNMT1, 2 and 3 across insects.** (A) Relationships of DNMT1, 2, and 3 in insects, Diplura and Ixodida (Arachnida). DNMT2 can be found in all insect orders investigated, while DNMT1 and 3 are more order-poor. The insect order for each sequence is provided in square brackets following the GenBank or genome annotation accession number. (B) A duplication event shared by the superfamilies Apoidea (bees) and Vespoidea (wasps and ants) gave rise to what is referred to as DNMT1a and b. Reciprocal losses occurred in Formicidae and Vespoidea, while Apoidea maintained copies of both paralogs. Other insect taxa may have experienced lineage-specific duplication events. Species names for each sequence are provided in the square brackets. *Daphnia magna* (Crustacea) was used as the outgroup. Bootstrap support values are provided at selected nodes.

**Table 1. Sociality and DNA methylation are not evolutionary dependent.** Upper half: Output from Pagel's [44] method for evolutionary dependence (implemented in phytools [42]) using the phylogeny and traits from Fig. 1A. The dependent model of trait evolution is not preferred based on the likelihood ratio test (LRT), and the independent model is preferred based on Akaike information criterion (AIC). Lower half: Output from Phylogenetic Generalized Least Squares (PGLS) implemented in phytools [42]) using the phylogeny and traits from Fig. 2. A model of zero (0) phylogenetic signal ($\lambda$) and evolution under Brownian motion was preferred over a model where $\lambda$ was estimated by maximum likelihood. For both tests, P-values represent the significance of correlations between trait *x* and *y*. P-values for PGLS are given for the preferred model.

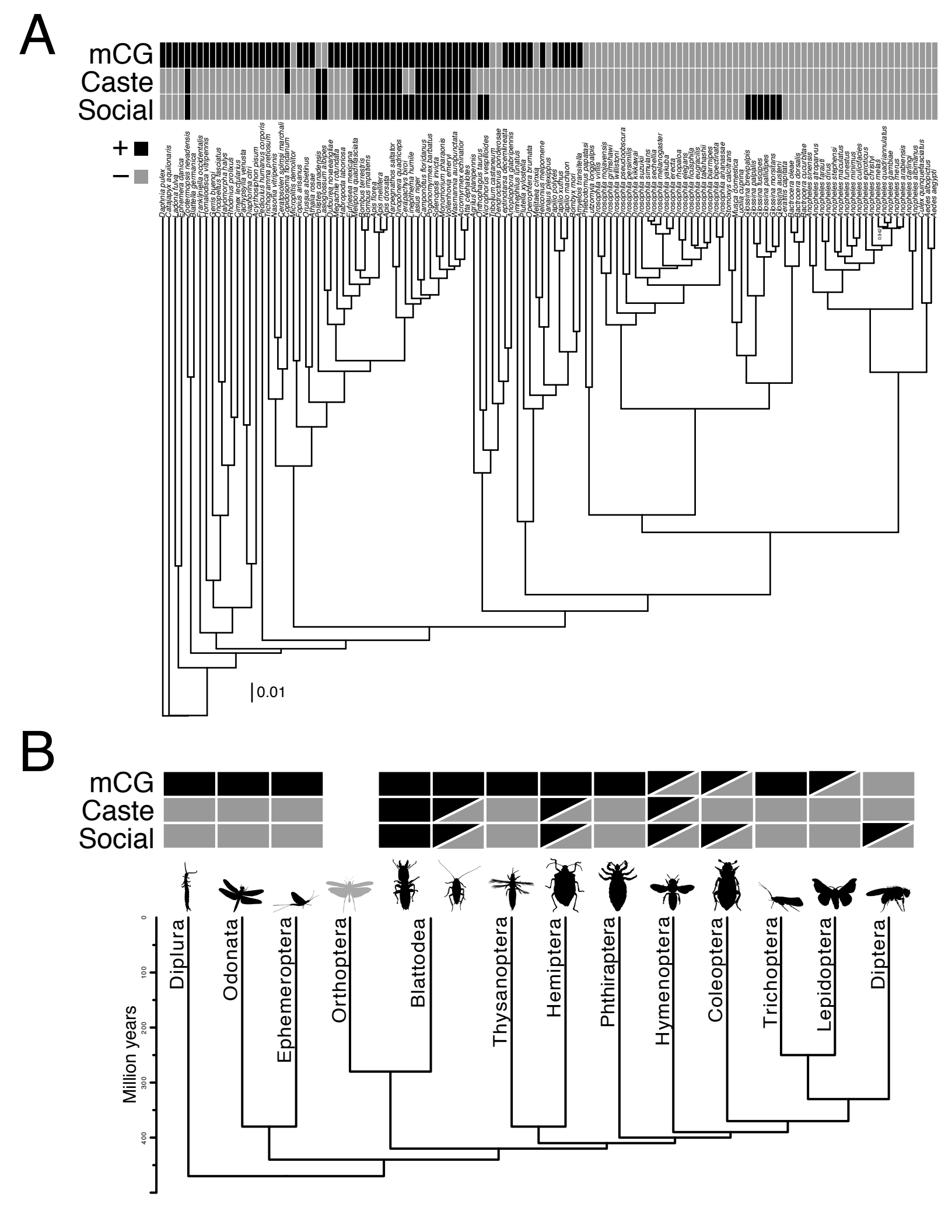
## REFERENCES

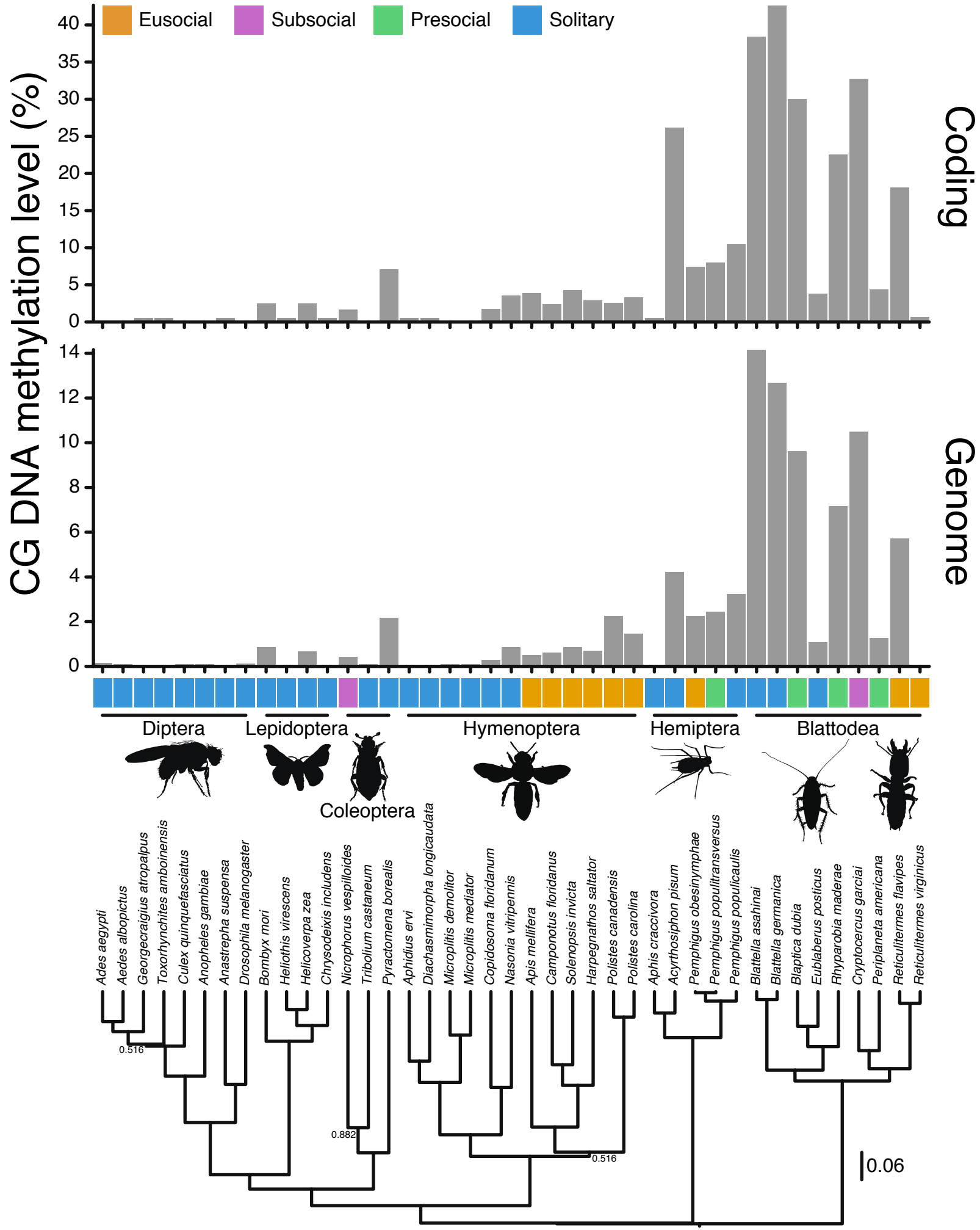

1. Kim, J. K., Samaranayake, M. & Pradhan, S. Epigenetic mechanisms in mammals. *Cell. Mol. Life Sci.* **66**, 596–612 (2008).
2. Goll, M. G. & Bestor, T. H. Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* **74**, 481–514 (2005).
3. Cheng, X. & Blumenthal, R. M. Mammalian DNA methyltransferases: a structural perspective. *Structure* **16**, 341–350 (2008).
4. Goll, M. G. Methylation of $tRNA^{Asp}$ by the DNA methyltransferase homolog Dnmt2. *Science* **311**, 395–398 (2006).
5. Zemach, A., McDaniel, I. E., Silva, P. & Zilberman, D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* **328**, 916–919 (2010).
6. Feng, S. *et al.* Conservation and divergence of methylation patterning in plants and animals. *Proc. Natl. Acad. Sci. USA.* **107**, 8689–8694 (2010).
7. Huff, J. T. & Zilberman, D. Dnmt1-independent CG methylation contributes to nucleosome positioning in diverse eukaryotes. *Cell* **156**, 1286–1297 (2014).
8. Cortijo, S. et al. Mapping the epigenetic basis of complex traits. *Science* **343**, 1145–1148 (2014).
9. Yan, H. *et al.* DNA methylation in social insects: how epigenetics can control behavior and longevity. *Annu. Rev. Entomol.* **60**, 435–452 (2015).
10. Yan, H. *et al.* Eusocial insects as emerging models for behavioural epigenetics. *Nat. Rev. Genet.* **15**, 677–688 (2014).
11. Patalano, S. *et al.* Molecular signatures of plastic phenotypes in two eusocial insect species with simple societies. *Proc. Natl. Acad. Sci. USA.* **112**, 13970–13975 (2015).
12. Bonasio, R. *et al.* Genome-wide and caste-specific DNA methylomes of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Curr. Biol.* **22**, 1755–1764 (2012).
13. Elango, N., Hunt, B. G., Goodisman, M. A. D. & Yi, S. V. DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*. *Proc. Natl. Acad. Sci. USA.* **106**, 11206–11211 (2009).
14. Misof, B. *et al.* Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**, 763–767 (2014).

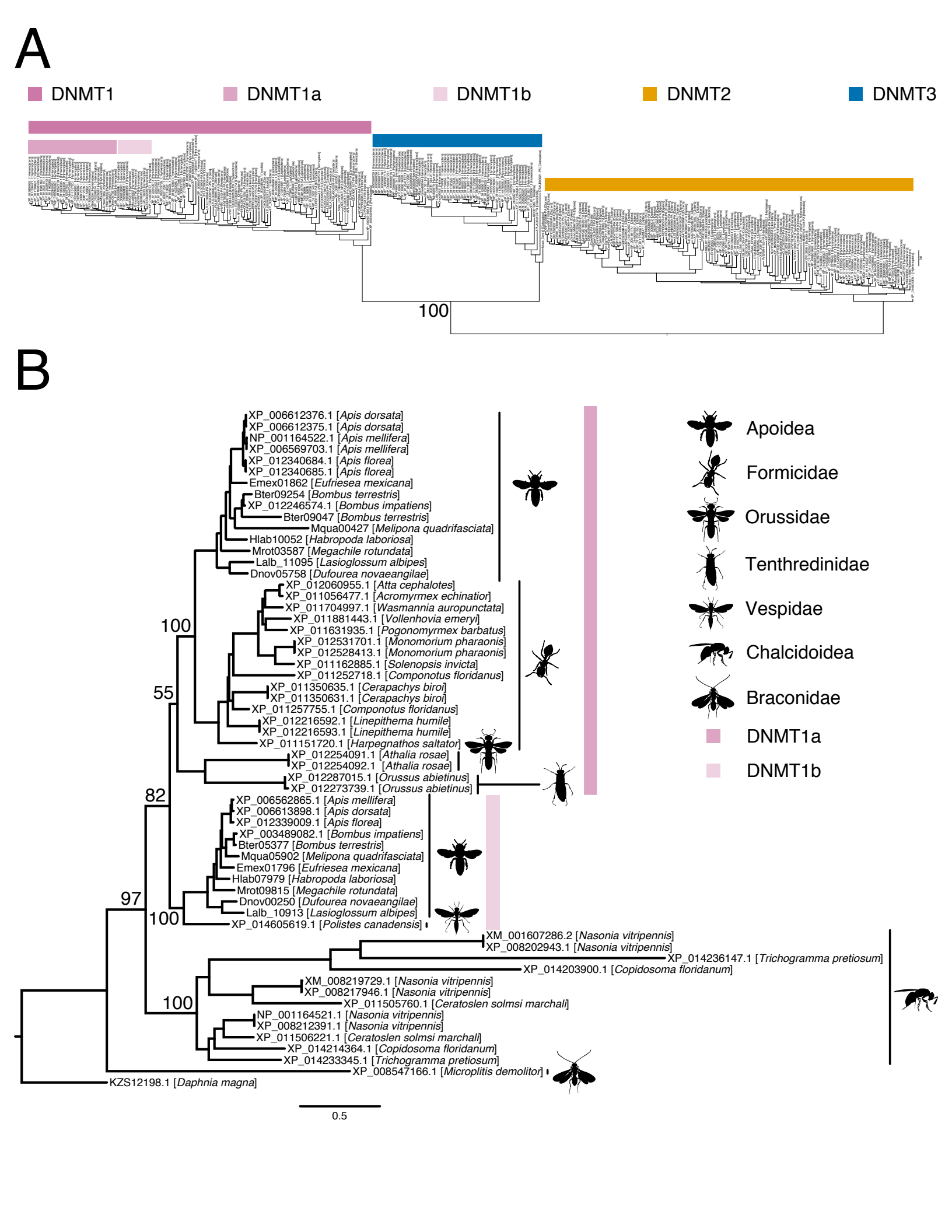
15. International *Glossina* Genome Initiative. Genome sequence of the tsetse fly (*Glossina morsitans*): vector of African trypanosomiasis. *Science* **344**, 380–386 (2014).
16. Cunningham, C. B. *et al.* The genome and methylome of a beetle with complex social behavior, *Nicrophorus vespilloides* (Coleoptera: Silphidae). *Genome Biol. Evol.* **7**, 3383–3396 (2015).
17. Walling, C. A., Stamper, C. E., Smiseth, P. T. & Moore, A. J. The quantitative genetics of sex differences in parenting. *Proc. Natl. Acad. Sci. USA.* **105**, 18430–18435 (2008).
18. Niederhuth, C. E. & Bewick, A. J. *et al.* Widespread natural variation of DNA methylation within angiosperms. (2016). doi:10.1101/045880
19. Takuno, S., Shohei, T., Jin-Hua, R. & Gaut, B. S. Evolutionary patterns of genic DNA methylation vary across land plants. *Nat. Plants* **2**, 15222 (2016).
20. Li-Byarlay, H. & Hongmei, L.-B. The function of DNA methylation marks in social insects. *Front. Ecol. Evol.* **4**, (2016).
21. Cokus, S. J. *et al.* Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* **452**, 215–219 (2008).
22. Lister, R. *et al.* Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133**, 523–536 (2008).
23. Xiang, H. *et al.* Single base–resolution methylome of the silkworm reveals a sparse epigenomic map. *Nat. Biotechnol*. **28**, 516–520 (2010).
24. Wang, Y. *et al.* Functional CpG methylation system in a social insect. *Science* **314**, 645–647 (2006).
25. Qiu, C., Sawada, K., Zhang, X. & Cheng, X. The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. *Nat. Struct. Biol.* **9**, 217–224 (2002).
26. Qin, S., Su, Q. & Jinrong, M. Structure and function of the nucleosome-binding PWWP domain. *Trends Biochem. Sci.* **39**, 536–547 (2014).
27. Bestor, T. H. & Verdine, G. L. DNA methyltransferases. *Curr. Opin. Cell Biol.* **6**, 380–389 (1994).
28. Ronquist, F. *et al.* A total-evidence approach to dating with fossils, applied to the early radiation of the hymenoptera. *Syst. Biol.* **61**, 973–999 (2012).
29. Kocher, S. D. *et al.* The draft genome of a socially polymorphic halictid bee, *Lasioglossum albipes*. *Genome Biol.* **14**, R142 (2013).

30. Bewick, A. J. *et al.* The evolution of CHROMOMETHYLASES and gene body DNA methylation in plants. (2016). doi:10.1101/054924
31. Bewick, A. J., Ji, L., Niederhuth, C. E. & Willing, E. *et al.* On the origin and evolutionary consequences of gene body DNA methylation. (2016). doi:10.1101/045542
32. Wilson, E. O. The Insect Societies. (Belknap Press, 1971).
33. Urich, M. A., Nery, J. R., Lister, R., Schmitz, R. J. & Ecker, J. R. MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. *Nat. Protoc.* **10**, 475–483 (2015).
34. Schultz, M. D. *et al.* Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* **523**, 212–216 (2015).
35. Martin, M. & Marcel, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10 (2011).
36. Langmead, B. Trapnell, C. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, doi: 10.1186/gb-2009-10-3-r25 (2009).
37. Bewick, A. J. *et al.* *FAST$^{m}$C*: a suite of predictive models for non-reference-based estimations of DNA methylation. *G3* **6**, 447–452 (2015).
38. Lyko, F. et al. The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS Biol.* **8**, e1000506 (2010).
39. Beeler, S. M. et al. Whole-genome DNA methylation profile of the jewel wasp (*Nasonia vitripennis*). *G3* **4**, 383–388 (2014).
40. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
41. Bird, A. P. DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res.* **8**, 1499–1504 (1980).
42. Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2011).
43. Bouckaert, R. *et al.* BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **10**, e1003537 (2014).
44. Pagel, M. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proceedings of the Royal Society B: Biological Sciences* **255**, 37–45 (1994).

45. Mirarab, S. *et al.* PASTA: ultra-large multiple sequence alignment for nucleotide and amino-acid sequences. *J. Comput. Biol.* **22**, 377–386 (2015).
46. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).
47. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
48. Jones, P. *et al*. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).


A
mCG
Caste
Social
+
−
Daphnia pulex
Cataiapyx aquilonaris
Ladona fulva
Ephemera danica
Ephemerella grandensis
Blatella germanica
Frankliniella occidentalis
Homalodisca vitripennis
Gerris buenoi
Oncopeltus fasciatus
Halyomorpha halys
Rhodnius prolixus
Cimex lectularius
Bactrocera dorsalis
Diaphorina citri
Acyrthosiphon pisum
Pediculus humanus corporis
Tribolium castaneum
Nasonia vitripennis
Ceratosolen solmsi marchali
Copidosoma floridanum
Microplitis demolitor
Fopius arisanus
Orussus abietinus
Athalia rosae
Athalia cordata
Lasioglossum albipes
Dufourea novaeangliae
Megachile rotundata
Euglossa hemichlora
Eufriesea mexicana
Melipona quadrifasciata
Bombus terrestris
Apis florea
Apis mellifera
Apis dorsata
Tarpeghatodes saltator
Drosophila melanogaster
Ceratapachys biroi
Linepithema humile
Camponotus floridanus
Solenopsis invicta
Pogonomyrmex barbatus
Monomorium pharaonis
Wasmannia auropunctata
Acromyrmex echinatior
Atta cephalotes
Ophthalmopsis
Ornithophagus taurus
Nicrophorus vespilloides
Tribolium castaneum
Leptinotarsa decemlineata
Anoplophora glabripennis
Innophilus lunatus
Quelleria xylostella
Bombyx mori
Melitaea cinxia
Heliconius melpomene
Danaus plexippus
Papilio xuthus
Papilio machaon
Bombyx mori
Phlebotomus papatasi
Lutzomyia longipalpis
Drosophila virilis
Drosophila mauritiana
Drosophila sechellia
Drosophila simulans
Drosophila yakuba
Drosophila erecta
Drosophila ananassae
Musca domestica
Glossina pallidipes
Glossina morsitans
Glossina austeni
Ceratitis capitata
Bactrocera dorsalis
Anopheles sinensis
Anopheles darling
Anopheles dirus
Anopheles stephensi
Anopheles funestus
Anopheles minimus
Anopheles christy
Anopheles melas
Anopheles gambiae
Anopheles arabiensis
Anopheles darling
Anopheles albimanus
Culex quinquefasciatus
Aedes aegypti
0.01
B
mCG
Caste
Social
Million years
0
100
200
300
400
Diplura
Odonata
Ephemeroptera
Orthoptera
Blattodea
Thysanoptera
Hemiptera
Phthiraptera
Hymenoptera
Coleoptera
Trichoptera
Lepidoptera
Diptera


Eusocial
Subsocial
Presocial
Solitary
CG DNA methylation level (%)
Coding
Genome
Diptera
Lepidoptera
Coleoptera
Hymenoptera
Hemiptera
Blattodea
Ades aegypti
Aedes albopictus
Georgecraigius atropalpus
Toxorhynchites amboinensis
Culex quinquefasciatus
Anopheles gambiae
Anastrepha suspensa
Drosophila melanogaster
Bombyx mori
Heliothis virescens
Helicoverpa zea
Chrysodeixis includens
Nicrophorus vespilloides
Tribolium castaneum
Pyractomena borealis
Aphidius ervi
Diachasmimorpha longicaudata
Microplitis demolitor
Microplitis mediator
Copidosoma floridanum
Nasonia vitripennis
Apis mellifera
Camponotus floridanus
Solenopsis invicta
Harpegnathos saltator
Polistes canadensis
Polistes carolina
Aphis craccivora
Acyrthosiphon pisum
Pemphigus obesinymphae
Pemphigus populitransversus
Pemphigus populicaulis
Blattella asahinai
Blattella germanica
Blaptica dubia
Eublaberus posticus
Rhyparobia maderae
Cryptocercus garciai
Periplaneta americana
Reticulitermes flavipes
Reticulitermes virginicus
0.516
0.882
0.516
0.06


A
DNMT1
DNMT1a
DNMT1b
DNMT2
DNMT3
100
B
XP_006612376.1 [Apis dorsata]
XP_006612375.1 [Apis dorsata]
NP_001164522.1 [Apis mellifera]
XP_006569703.1 [Apis mellifera]
XP_012340684.1 [Apis florea]
XP_012340685.1 [Apis florea]
Emex01862 [Eufriesea mexicana]
Bter09254 [Bombus terrestris]
XP_012246574.1 [Bombus impatiens]
Bter09047 [Bombus terrestris]
Mqua00427 [Melipona quadrifasciata]
Hlab10052 [Habropoda laboriosa]
Mrot03587 [Megachile rotundata]
Lalb_11095 [Lasioglossum albipes]
Dnov05758 [Dufourea novaeangilae]
XP_012060955.1 [Atta cephalotes]
XP_011056477.1 [Acromyrmex echinatior]
XP_011704997.1 [Wasmannia auropunctata]
XP_011881443.1 [Vollenhovia emeryi]
XP_011631935.1 [Pogonomyrmex barbatus]
XP_012531701.1 [Monomorium pharaonis]
XP_012528413.1 [Monomorium pharaonis]
XP_011162885.1 [Solenopsis invicta]
XP_011252718.1 [Componotus floridanus]
XP_011350635.1 [Cerapachys biroi]
XP_011350631.1 [Cerapachys biroi]
XP_011257755.1 [Componotus floridanus]
XP_012216592.1 [Linepithema humile]
XP_012216593.1 [Linepithema humile]
XP_011151720.1 [Harpegnathos saltator]
XP_012254091.1 [Athalia rosae]
XP_012254092.1 [Athalia rosae]
XP_012287015.1 [Orussus abietinus]
XP_012273739.1 [Orussus abietinus]
XP_006562865.1 [Apis mellifera]
XP_006613898.1 [Apis dorsata]
XP_012339009.1 [Apis florea]
XP_003489082.1 [Bombus impatiens]
Bter05377 [Bombus terrestris]
Mqua05902 [Melipona quadrifasciata]
Emex01796 [Eufriesea mexicana]
Hlab07979 [Habropoda laboriosa]
Mrot09815 [Megachile rotundata]
Dnov00250 [Dufourea novaeangilae]
Lalb_10913 [Lasioglossum albipes]
XP_014605619.1 [Polistes canadensis]
XM_001607286.2 [Nasonia vitripennis]
XP_008202943.1 [Nasonia vitripennis]
XP_014236147.1 [Trichogramma pretiosum]
XP_014203900.1 [Copidosoma floridanum]
XM_008219729.1 [Nasonia vitripennis]
XP_008217946.1 [Nasonia vitripennis]
XP_011505760.1 [Ceratoslen solmsi marchali]
NP_001164521.1 [Nasonia vitripennis]
XP_008212391.1 [Nasonia vitripennis]
XP_011506221.1 [Ceratoslen solmsi marchali]
XP_014214364.1 [Copidosoma floridanum]
XP_014233345.1 [Trichogramma pretiosum]
XP_008547166.1 [Microplitis demolitor]
KZS12198.1 [Daphnia magna]
100
55
82
97
100
100
0.5
Apoidea
Formicidae
Orussidae
Tenthredinidae
Vespidae
Chalcidoidea
Braconidae
DNMT1a
DNMT1b

| Taxa | Trait *x* | Trait *y* | Model | | | | LRT | P-value |
|---|---|---|---|---|---|---|---|---|
| | | | Independent | | Dependent | | | |
| | | | lnL | AIC | lnL | AIC | | |
| Insecta* | DNA methylation | Sociality | –59.482 | 126.964 | –58.359 | 132.718 | 2.246 | 0.691 |
| Insecta* | DNA methylation | Caste | –53.482 | 114.965 | –52.328 | 120.656 | 2.309 | 0.679 |
| Hymenoptera | DNA methylation | Sociality | –23.227 | 54.454 | –19.371 | 54.742 | 7.712 | 0.103 |
| Hymenoptera | DNA methylation | Caste | –27.385 | 62.769 | –25.578 | 67.156 | 3.613 | 0.461 |
| Diptera– | DNA methylation | Sociality | –46.632 | 101.263 | –44.112 | 104.224 | 5.039 | 0.283 |
| Diptera– | DNA methylation | Caste | –45.920 | 99.840 | –45.688 | 107.376 | 0.464 | 0.977 |
| | | | $\lambda_0$ | | $\lambda_{ML}$ | | | |
| Insecta | DNA methylation[C] | Eusocial | 33.263 | –56.525 | 40.592 | –69.184 | –7.329 | 0.447 |
| | | Subsocial | | | | | | 0.131 |
| | | Presocial | | | | | | 0.583 |
| | | Solitary | | | | | | 0.836 |
| Insecta | DNA methylation[G] | Eusocial | 84.007 | –158.013 | 88.921 | –165.843 | –4.914 | 0.440 |
| | | Subsocial | | | | | | 0.098 |
| | | Presocial | | | | | | 0.544 |
| | | Solitary | | | | | | 0.906 |

,*Catajapyx aquilonaris* (Dipluran) and *Daphnia pulex* (Crustacea) included; Likelihood ratio test (LRT); Akaike information criterion (AIC); Maximum likelihood (ML); [C], coding; [G], genome