Insect Molecular Biology

Insect Molecular Biology (2013) **22**(2), 143–154 doi: 10.1111/imb.12010

Evidence of a conserved functional role for DNA methylation in termites

K. M. Glastad, B. G. Hunt and M. A. D. Goodisman

School of Biology, Georgia Institute of Technology, Atlanta, GA, USA

Abstract

Many organisms are capable of developing distinct phenotypes from the same genotype. This developmental plasticity is particularly prevalent in insects, which can produce alternate adaptive forms in response to distinct environmental cues. The ability to develop divergent phenotypes from the same genotype often relies on epigenetic information, which affects gene function and is transmitted through cell divisions. One of the most important epigenetic marks, DNA methylation, has been lost in several insect lineages, yet its taxonomic distribution and functional conservation remain uninvestigated in many taxa. In the present study, we demonstrate that the signature of high levels of DNA methylation exists in the expressed genes of two termites, *Reticulitermes flavipes* and *Coptotermes formosanus*. Further, we show that DNA methylation is preferentially targeted to genes with ubiquitous expression among morphs. Functional associations of DNA methylation are also similar to those observed in other invertebrate taxa with functional DNA methylation systems. Finally, we demonstrate an association between DNA methylation and the long-term evolutionary conservation of genes. Overall, our findings strongly suggest DNA methylation is present at particularly high levels in termites and may play similar roles to those found in other insects.

Keywords: caste, DNA methylation, epigenetics, gene expression, genome evolution, Isoptera, eusocial insect, CpG o/e.

First published online 28 December 2012.

Correspondence: Michael A. D. Goodisman, School of Biology, Georgia Institute of Technology, 310 Ferst Dr., Atlanta, GA 30332, USA. Tel.: +1 404 385 6311; fax: +1 404 894 0519; e-mail: michael.goodisman@biology.gatech.edu

Introduction

Environmental responsiveness plays a fundamental role in the success of complex life forms (West-Eberhard, 2003). Accordingly, virtually all multicellular organisms are capable of developing differently in response to environmental variation (Pfennig *et al.*, 2010). This developmental plasticity is particularly evident in insects, where many taxa display dramatic morphological differences, depending on environmental conditions (Hölldobler & Wilson, 1990; Simpson *et al.*, 2011). While the importance of environmentally responsive phenotypes is well understood (Price *et al.*, 2003; Pfennig *et al.*, 2010), our understanding of plasticity's molecular underpinnings remains quite limited.

Recent investigations indicate that epigenetic information plays an important role in regulating the development of environmentally induced phenotypic variation (Kucharski *et al.*, 2008; Burdge & Lillycrop, 2010; Schmitz & Ecker, 2012). Epigenetic modifications are not coded in the standard complement of DNA bases, yet they affect gene function and are heritable through cell divisions (Berger *et al.*, 2009). Variation in epigenetic information can lead to sustained changes in gene expression (Kota & Feil, 2010; Margueron & Reinberg, 2010), ultimately permitting variation in developmental programmes in response to environmental cues.

One important form of epigenetic modification is the methylation of cytosine bases in eukaryotic genomes. DNA methylation is present in all three domains of life (Klose & Bird, 2006; Suzuki & Bird, 2008; Zemach *et al.*, 2010), and has been linked to gene expression variation in mammals (Fraga *et al.*, 2005; Cheong *et al.*, 2006; Reik, 2007), plants (Li *et al.*, 2008; He *et al.*, 2011) and insects (Kucharski *et al.*, 2008; Lyko *et al.*, 2010; Glastad *et al.*, 2011). DNA methylation has also been linked to the outcome of alternative splicing (Lyko *et al.*, 2010; Maunakea *et al.*, 2010; Shukla *et al.*, 2011), as well as chromatin structure and modification in both plants (Zhang *et al.*, 2009; Chodavarapu *et al.*, 2010) and vertebrates (Okitsu & Hsieh, 2007; Hodges *et al.*, 2009; Jeong *et al.*, 2009). Thus DNA methylation appears to play an important role in mediating the relationship between genotype and phenotype in many taxa.

DNA methylation has recently been linked to phenotypic plasticity in eusocial insects (Kucharski *et al*., 2008; Lyko *et al*., 2010; Foret *et al*., 2012), where some of the most pronounced instances of plastic development occur (Hölldobler & Wilson, 1990; Simpson *et al*., 2011). Eusocial insects are defined by a caste system whereby individuals may develop different phenotypes, most frequently through differential environmental induction of gene expression (Evans & Wheeler, 2001; Barchuk *et al*., 2007; Smith *et al*., 2008; but see: Goodisman *et al*., 2008; Schwander *et al*., 2010). Typically, a minority of individuals belong to the reproductive castes, while the majority of individuals, belonging to the worker and soldier castes, remain effectively sterile (Hölldobler & Wilson, 1990; Wilson 1971). Importantly, castes often show extreme differences in morphology, physiology, and behaviour arising through the differential expression of genes (Robinson *et al*., 2005; Smith *et al*., 2008).

DNA methylation has been linked to caste formation in honey bees (Kucharski *et al*., 2008). Furthermore, caste-specific differences in DNA methylation are associated with alternative splicing differences between reproductive queens and sterile workers (Lyko *et al*., 2010), as well as between queen and worker-destined larvae (Foret *et al*., 2012). DNA methylation has also been identified in the genomes of several other hymenopteran eusocial insect taxa (Glastad *et al*., 2011), suggesting that it may play a wide-ranging role in developmental responsiveness. Thus, DNA methylation may be integral to maintaining the phenotypic plasticity that is central to the structure of hymenopteran eusocial insect societies.

DNA methylation has recently been identified in the genome of the termite *Coptotermes lacteus* (Lo *et al*., 2012). Termites are highly eusocial and exhibit distinct castes (Scharf *et al*., 2007; Toru & Scharf, 2011), but termites represent a completely novel origin of eusociality from the Hymenoptera; isopteran and hymenopteran insects diverged ~ 375 million years ago (Gaunt & Miles, 2002). Thus, termites provide an important evolutionary contrast to Hymenoptera for investigating the link between developmental regulation of phenotypic plasticity and DNA methylation.

In contrast to hymenopterans, which are holometabolous, termites are hemimetablolous, and therefore exhibit a developmental programme that differs substantially from ants, bees and wasps. In many termite species, larvae develop into either workers or nymphs (immature forms with rudimentary wing buds; Eggleton, 2011). Nymphs then develop into winged reproductive forms (alates). Workers, while considered a distinct caste, are composed of multiple, developmentally progressive instars and can also further develop into either soldiers, or, in some cases, a third, worker-derived, wingless reproductive form (Eggleton, 2011). Much of this developmental plasticity is informed by hormonal (endogenous) and environmental (exogenous) cues (Mao *et al*., 2005; Scharf *et al*., 2007; Toru & Scharf, 2011). Thus, development and caste are arguably more protean in termites than in the hymenopteran eusocial insects, making them an ideal system for studying the molecular basis of caste and, more generally, environmentally induced phenotypic plasticity.

The purpose of the present study was to gain a greater understanding of the nature and function of DNA methylation in termites. We investigated this issue by studying the genetic signature of DNA methylation in the eastern subterranean termite, *Reticulitermes flavipes* (Tartar *et al*., 2009; Steller *et al*., 2010) and the formosan subterranean termite, *Coptotermes formosanus* (Lo *et al*., 2012). *R. flavipes* and *C. formosanus* are members of the family Rhinotermitidae, which displays a complex caste system. For example, *Reticulitermes* and *Coptotermes* early-stage nymphs can conditionally change developmental trajectory, shifting to the wingless (worker) form (Roisin & Korb, 2011). The Rhinotermitidae also include some of the most invasive termite pests, with *R. flavipes* and *C. formosanus* being among the most devastating (Evans, 2011). In addition, these two species represent important taxa for studying molecular biology and development in termites (e.g., Park & Raina, 2004; Scharf *et al*., 2007).

In the present study, we demonstrate that a strong genetic signature of DNA methylation exists in *R. flavipes* and *C. formosanus*. Moreover, we show that DNA methylation exhibits conserved associations with patterns of gene expression, functional enrichment, and molecular evolution, as identified in other insects and non-insect invertebrates (Elango *et al*., 2009; Hunt *et al*., 2010; Sarda *et al*., 2012). Overall, our analyses suggest that DNA methylation is an important factor associated with gene function in these species, and termites will therefore represent important taxa for investigating DNA methylation outside of Holometabola.

Results and discussion

Reticulitermes flavipes *and* Coptotermes formosanus *genes show evidence of DNA methylation*

We examined the distribution of nucleotides in six *R. flavipes* and one *C. formosanus* cDNA libraries in order to assess the presence and patterns of DNA methylation in coding sequences (see methods). *R. flavipes* assembled expressed sequence tags (ESTs) were constructed from nonnormalized, cDNA libraries from three distinct, adult morphs, and two developmental stages (Steller *et al*., 2010). For several of our analyses we additionally incorporated information from another normalized *R. flavipes* library, which was constructed from termite flora-free gut tissue (Tartar *et al*., 2009). We also assembled and

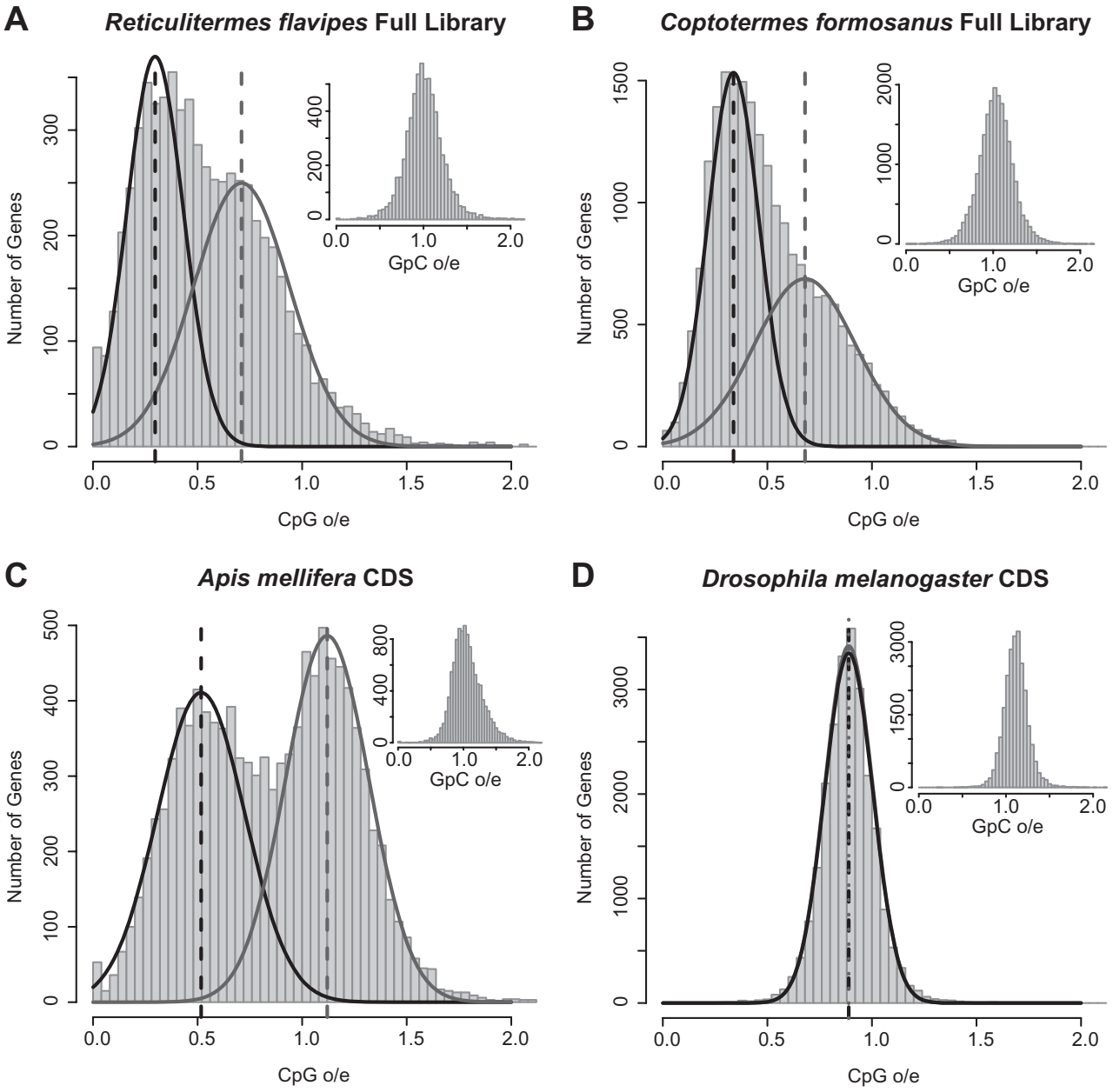

Figure 1. Distribution of normalized CpG dinucleotide content (CpG o/e) and GpC o/e (inset) for sequences from the termites (A) *Reticulitermes flavipes* and (B) *Coptotermes formosanus*, and coding sequence (CDS) from (C) the honeybee *Apis mellifera* which exhibits DNA methylation targeting to approximately half of its genes, and (D) the fruit fly *Drospohila melanogaster*, which exhibits no detected DNA methylation, and thus, no notable CpG depletion.

analysed a single, normalized library from *C. formosanus* pooled individuals derived from multiple castes (Zhang *et al*., 2012).

We calculated the normalized cytosine-phosphate-guanine (CpG) dinucleotide content from these sequence assemblies. Normalized CpG dinucleotide content [CpG observed/expected (o/e)] acts as a proxy for the presence of DNA methylation, reflecting the greatly increased tendency of 5-methylcytosine to undergo spontaneous deamination over evolutionary time (Coulondre *et al*., 1978; Shen *et al*., 1994). Thus, CpG o/e can be used to determine the targets of DNA methylation across the genome (Tweedie *et al*., 1997; Suzuki *et al*., 2007; Bock & Lengauer, 2008; Yi & Goodisman, 2009).

Strikingly, the CpG o/e distributions for both the full *R. flavipes* and *C. formosanus* assemblies exhibited evidence of heavy CpG depletion (Fig. 1A and B). Intriguingly, the CpG o/e profile observed for *R. flavipes* (and, to a lesser extent, *C. formosanus*) exhibited a mixed distribution (Fig. 1A and B). Similar mixed distributions have been

noted in several other insects with DNA methylation, such as the pea aphid *Acyrthosiphon pisum* and honey bee *Apis mellifera* (Fig. 1C; Elango *et al*., 2009; Wang & Leung, 2009; Hunt *et al*., 2010; Walsh *et al*., 2010). Such mixed distributions are characteristic of a genome in which genes can be divided into two classes; the first class is generally methylated and displays low values of CpG o/e, and the second class is generally unmethylated and displays relatively high values of CpG o/e. This can be contrasted with *Drosophila melanogaster* (Fig. 1D), for example, which has virtually no detectable DNA methylation (Suzuki & Bird, 2008; Zemach *et al*., 2010; Glastad *et al*., 2011). *D. melanogaster* CpG o/e exhibits a mean near the expected value of one owing to the absence of DNA methylation. In addition, in contrast to our termite libraries (Fig. 1A and B) and *A. mellifera* (Fig. 1C), the distribution of CpG o/e is clearly unimodal in *D. melanogaster* (Fig 1D).

We examined the distribution of GpC o/e (that is, the observed/expected ratio of 5′-GpC-3′ dinucleotides) to confirm that the patterns we observed among CpG dinucleotides reflected the action of DNA methylation and did not arise from potential artifacts related to guanine-cytosine (GC) content (Fryxell & Moon, 2005). GpC dinucleotides, unlike CpG dinucleotides, are not methylated in animal genomes and should have no mutational pressure introduced by DNA methylation; therefore, the frequency of GpC dinucleotides should be readily predictable from the frequency of G and C nucleotides in a gene (i.e., GpC o/e should equal 1). We found that mean GpC o/e in *R. flavipes* and *C. formosanus* was, indeed, quite close to 1.0 (*R. flavipes* = 1.02, *C. formosanus* = 1.03) and did not exhibit mixed distributions in either species, as was the case for CpG o/e (Fig. 1A and B, inset). This result suggests that depletion of CpG in *R. flavipes* and *C. formosanus* is indicative of the presence of DNA methylation.

The mean CpG o/e values for both termites' assemblies were very low, indicating high levels of gene methylation (Table 1; Elango *et al*., 2009). The mean (SD) CpG o/e value for our *R. flavipes* assembly was 0.54 (0.33), and the mean (SD) GpC o/e was 1.02 (0.26), which is much lower than in other investigated insect gene sets (Table 1). Indeed, the mean CpG o/e values in termites are lower than other eusocial insects including both honey bees and ants (Elango *et al*., 2009; Glastad *et al*., 2011). Both means when treating these CpG o/e distributions as a mixture of two distributions in *R. flavipes* were well below 1 as well (mean 1 = 0.30, SD = 0.14, proportion = 0.34; mean 2 = 0.68, SD = 0.32, proportion = 0.66). These results were highly similar for *C. formosanus*, which exhibited a similar level of CpG depletion (mean = 0.51, SD = 0.27), but with a less pronounced upper mean (mean 1 = 0.34, SD = 0.129, proportion = 0.5; mean 2 = 0.70, SD = 0.26, proportion = 0.5).

Table 1. Mean CpG o/e and GpC o/e for gene sets (coding sequence) from invertebrates with intragenic DNA methylation

Species	CpG o/e	GpC o/e
Reticulitermes flavipes	0.54 (0.30/0.69)	1.02
Coptotermes formosanus	0.51 (0.33/0.70)	1.03
Acyrthospihon pisum	0.91 (0.64/1.08)	1.01
Nasonia vitripennis	0.93 (0.78/1.09)	1.09
Apis mellifera	0.85 (0.50/1.10)	1.05
Ciona intestinalis	0.79 (0.62/0.92)	0.99

Parenthetical values indicate the means for the low and high peaks in the distributions.

We also investigated whether any sequences in either library contained orthologues to honey bee DNA-methyltransferases (DNMTs). While we found no sequences exhibiting significant orthology (at a threshold of e < 1e-10) to DNMTs in *R. flavipes* sequences, we detected a putative orthologue of the honey bee DNMT1 in the *C. formosanus* assembled library (e = 4e-12). This suggests that DNMT1 is present and expressed in termites, supporting a functional methylation system in this order.

Overall, these results suggest the presence of DNA methylation in *R. flavipes* and *C. formosanus* termites, potentially at particularly high levels or widely targeted across the lengths of genes, relative to other investigated insect taxa.

DNA methylation in genes expressed by different Reticulitermes flavipes *castes and stages*

We identified genes whose representation was biased among *R. flavipes* cDNA libraries in order to assess whether genes with differences in expression between castes, sexes or developmental stages tended to be methylated or unmethylated in termites. We found that contigs that exhibited any level of significant differential expression among *R. flavipes* castes or developmental stages (i.e. whose EST composition was significantly different between any two libraries) had significantly higher CpG o/e values than those with no significant bias in expression among any libraries (Wilcoxon rank sum test, $P = 6.8e\text{-}12$, W = 63571). Furthermore, as the number of significantly different pairwise comparisons between libraries increased (out of a possible six pairwise comparisons) the CpG o/e values also increased (Fig. 2; Kruskal–Wallis test, $P = 2.1e\text{-}11$, chi-squared = 49.2). These data indicate that ubiquitously expressed genes are more likely to be targeted by DNA methylation in the germline than genes exhibiting differential expression between castes or stages.

The positive relationships observed between CpG o/e and measures of expression bias in *R. flavipes* are consistent with studies of DNA methylation in other insects, which found that DNA methylation is significantly negatively associated with gene expression bias (Elango *et al*., 2009; Foret

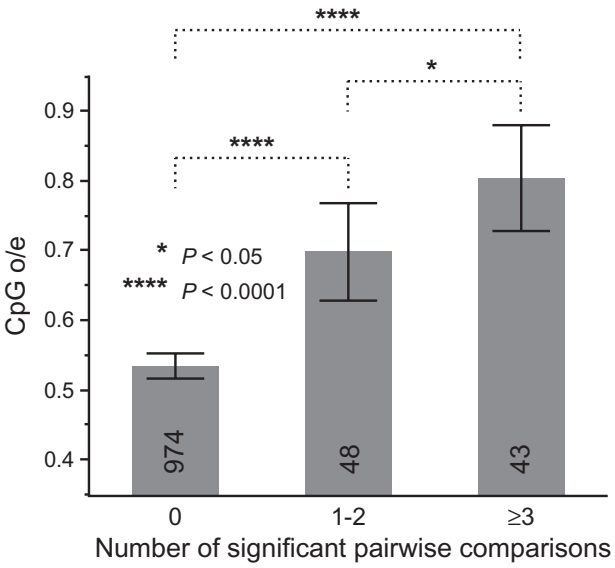

Figure 2. DNA methylation and level of expression bias in *Reticulitermes flavipes* sequences. Number of significant pairwise comparisons represents the number of times that a particular gene was significantly differentially expressed among libraries (out of six possible pairwise comparisons between alate, worker, soldier and larval libraries; Fisher's exact test after correcting for multiple comparisons). Means and 95% CIs are plotted. Significance values represent Wilcoxon signed-rank tests between CpG o/e for each category. Numbers within bars represent the number of genes falling into each class.

et al., 2009; Hunt *et al*., 2010; Xiang *et al*., 2010). Our results in *R. flavipes* confirm that the pattern whereby genes with biased expression show lower levels of DNA methylation than genes that are highly or ubiquitously expressed is common between hemimetabolous and holometabolous insect clades (Hunt *et al*., 2010).

We next investigated if distinct classes of individuals expressed genes that were targeted by different levels of DNA methylation in the germline. Thus, we compared the distributions of CpG o/e values for assembled ESTs in libraries from different *R. flavipes* castes and life stages. Surprisingly, we found that CpG o/e values varied significantly among libraries (Kruskal–Wallis test, chi-squared = 86.3, $P < 0.0001$). These results suggest that genes expressed by distinct castes exhibit differences in their overall CpG o/e profiles. Thus, some castes may be expressing a greater proportion of germline methylated genes than others. The library that was the most distinct from the others was the alate library (Fig. 3), which was considerably more biased towards low CpG o/e sequences than the other libraries. Given that our data support the view that DNA methylation is targeted to genes of low expression bias, our results may indicate that soldiers and workers (castes with the highest EST CpG o/e profile; Fig. 3) are expressing more unmethylated genes that are not expressed (or are more lowly expressed) in alates; however, caution should be taken when interpreting these results, as these EST libraries only capture a portion of expressed genes in each caste.

DNA methylation is associated with particular functional classes of genes

We next investigated if particular functional classes of genes were methylated or unmethylated in *R. flavipes* and *C. formosanus*. Specifically, we determined whether specific gene ontology (GO) functional categories of genes were enriched within the low (CpG o/e < 0.49 and 0.45 for *R. flavipes* and *C. formosanus*, respectively) and high (CpG o/e > 0.59 and 0.56 for *R. flavipes* and *C. formosanus* respectively) components of our CpG o/e distributions. Since CpG o/e corresponds to DNA methylation, this allows us to examine whether or not DNA methylation is targeted to genes with specific functional annotations, as has been found in previous studies of insects (Elango *et al*., 2009; Hunt *et al*., 2010; Sarda *et al*., 2012).

We found that several GO categories were enriched for each CpG o/e class of genes in *R. flavipes* and *C. formosanus* (Table 2), but some of the terms in each species were not significant following correction for multiple testing. Functional terms associated with low CpG o/e genes in *R. flavipes* largely involved basal metabolic processes and membrane-bound organelles, while terms enriched in high-CpG *R. flavipes* genes were associated with carbohydrate binding, cell adhesion, and ion binding and transport. In *C. formosanus*, low-CpG o/e genes were enriched for terms associated with basal catabolic and metabolic processes, similar to *R. flavipes*; however high-CpG o/e genes in *C. formosanus*, which are putatively unmethylated, were enriched with terms often associated with methylated genes in other taxa (Sarda *et al*., 2012).

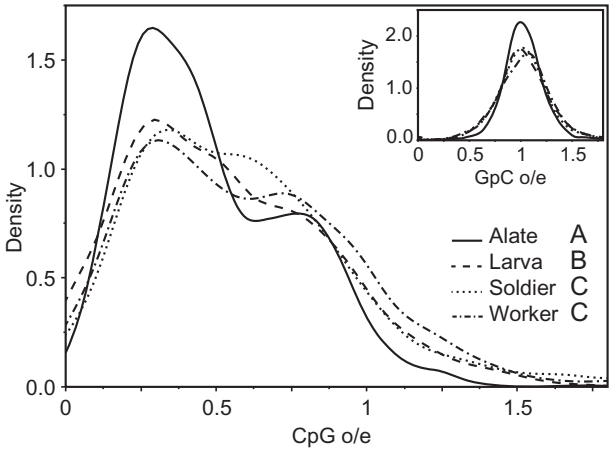

Figure 3. Variation in CpG o/e of expressed sequence tags (ESTs) between castes. Comparative density plot of CpG o/e values for assembled EST libraries of each *Reticulitermes flavipes* caste (inset: GpC o/e). Letters next to library names represent significantly differing groups as determined by Mann–Whitney *U*-tests for each pairwise comparison.

Table 2. Top five gene ontology-enriched terms within each CpG o/e class for *Reticulitermes flavipes* and *Coptotermes formosanus*

CpG o/e category	Species	Term	Category	False discovery rate	*P*
Low-CpG o/e	*R. flavipes*	Cellular metabolic process‡	P	0.0021	<0.0001
		RNA metabolic process*	P	0.0030	<0.0001
		Transferase activity	F	0.0121	<0.0001
		Generation of precursor metabolites and energy	P	0.0179	<0.0001
		Catalytic activity	F	0.0199	<0.0010
	C. formosanus	Carbohydrate metabolic process	P	0.0426	<0.0005
		Glycolysis	P	0.0500	<0.0005
		Catabolic process†	P	0.0927	0.0012
		Peptidase activity	F	0.0927	0.0013
		Macromolecule metabolic process*	P	0.4556	0.0150
High-CpG o/e	*R. flavipes*	Extracellular region	C	0.0179	<0.0001
		Ion transport†	F	0.0272	<0.0005
		Ion binding	F	0.2387	0.0067
		Carbohydrate binding	F	0.4131	0.0136
		Cell adhesion*	C	0.5087	0.0216
	C. formosanus	Ribonucleoprotein complex	C	<0.0001	<0.0001
		Ribosome	C	<0.0001	<0.0001
		Structural molecule activity	F	<0.0001	<0.0001
		Translation	P	0.0027	<0.0001
		Cellular biosynthetic process	P	0.0027	<0.0001

Gene ontology terms also enriched in either **Apis mellifera*, †*Acyrthosiphon pisum* or ‡both species methylated genes (from either Elango *et al*., 2009 or Hunt *et al*., 2010, respectively). *P* values derived using Fisher's exact test and false discovery rate values represent *P* values corrected for multiple testing using the Benjamini-Hochberg method.
P = biological process, F = molecular function, C = cellular component.

Our results suggest several categories of genes are enriched in each CpG o/e class in *R. flavipes* and *C. formosanus*. Previous studies of DNA methylation have often found that low-CpG o/e genes are associated with 'housekeeping functions' such as basal metabolic processes, and are often ubiquitously expressed both in time and space, while high-CpG o/e genes are associated with more condition-specific and complex regulation (Elango *et al*., 2009; Hunt *et al*., 2010; Glastad *et al*., 2011; Sarda *et al*., 2012). Consistent with these previous results, two of the top five *R. flavipes* low-CpG o/e enriched categories (cellular metabolic processes and DNA metabolic processes) were previously observed to be associated with methylated genes in either the honey bee (Elango *et al*., 2009) or pea aphid (Hunt *et al*., 2010), and one of the top five enriched terms among *R. flavipes* high-CpG o/e genes (ion transport) was enriched in the pea aphid (Table 2; Hunt *et al*., 2010). In *C. formosanus*, low-CpG o/e genes were similarly associated with metabolic processes, with two terms previously associated with methylated genes in other insects being present in the top five enriched categories (catabolic metabolic processes and macromolecule metabolic processes; Table 2); however, high-CpG o/e genes from *C. formosanus* displayed terms inconsistent with previous observations in insects. This may be attributable to the incomplete nature of these EST sequences and associated GO, or, less parsimoniously, a lineage-specific change in the function of unmethylated genes. Furthermore, the fact that the *R. flavipes* EST libraries were not normalized, whereas the EST libraries of *C. formosanus* were normalized, may be a factor in the taxon-specific patterns of enrichment we observe. Thus, our functional enrichment results warrant validation in future studies.

DNA methylation and evolutionary conservation in Reticulitermes flavipes *and* Coptotermes formosanus

We investigated if genes that were methylated or unmethylated in *R. flavipes* and *C. formosanus* displayed differing levels of evolutionary conservation. In order to investigate this issue, we assigned orthology to *R. flavipes* and *C. formosanus* genes based on reciprocal best hits between our full assemblies in both species and multiple organisms of increasing phylogenetic distance. That is, we determined which genes had detectable orthologues in organisms of increasing evolutionary distance, ranging from species that were relatively closely related to termites (the insect *D. melanogaster*), to those that were very distantly related (the plant *Arabidopsis thaliana*). By choosing pairs of organisms at increasing phylogenetic distances, this nested 'orthology walk' allowed us to assign a level of evolutionary conservation to a particular sequence and determine if orthologues were detectable only within insects, within multiple invertebrates, among all animals, or across eukaryotes. Using this analysis, we found that phylogenetic conservation was significantly, negatively associated with CpG o/e in both *R. flavipes* and *C. formosanus* for sequences with detectable orthologues outside of termites (Fig. 4A and

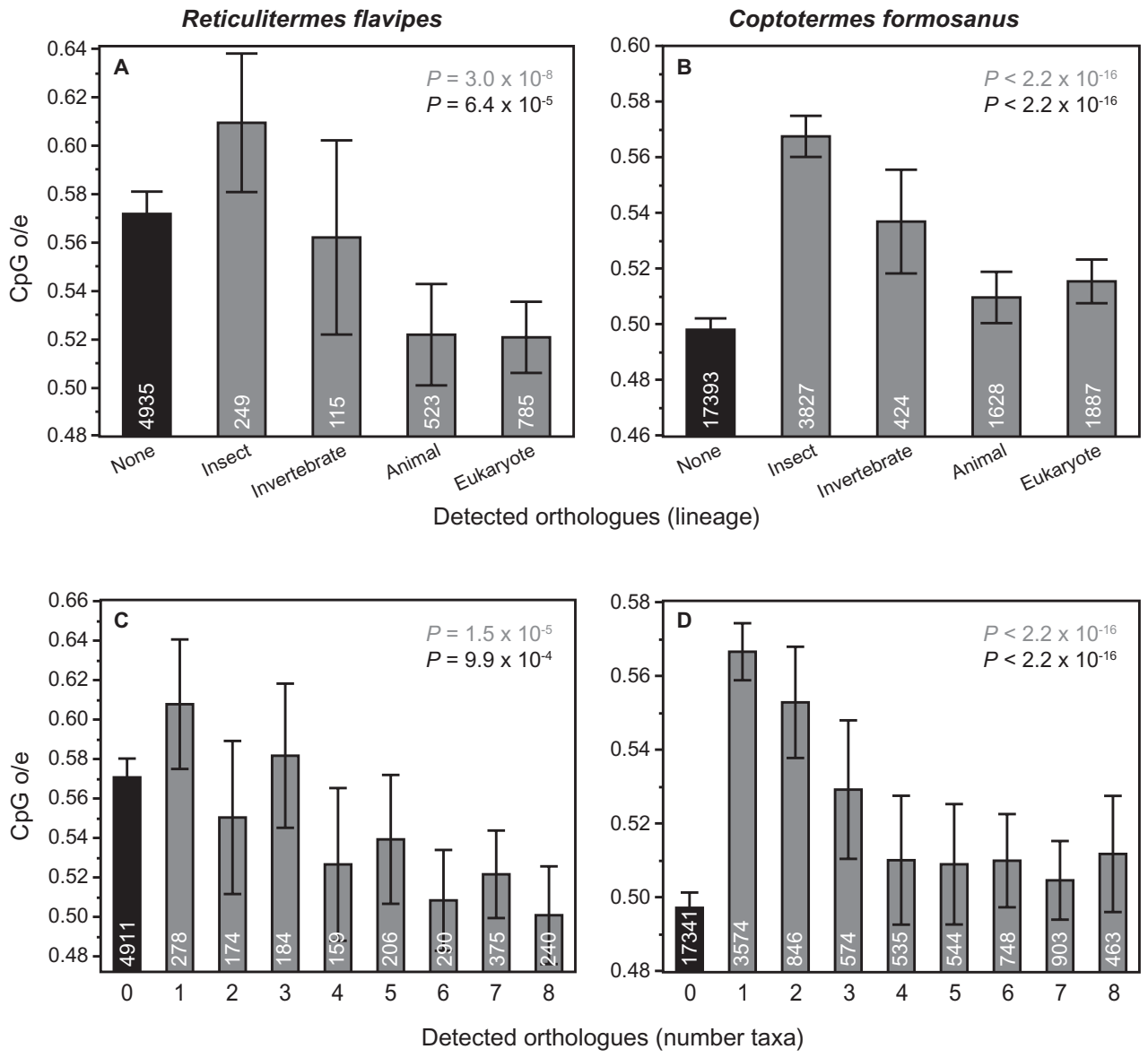

Figure 4. CpG o/e and evolutionary conservation of genes in (A,C) *Reticulitermes flavipes* and (B,D) *Coptotermes formosanus*. Evolutionary conservation was determined in two ways. First (A,B), by determining the most broad phylogenetic level an orthologue was conserved across (see main text) and second (C,D), by determining the total number of orthologues for a given gene found among four insects, two non-insect invertebrates, and two more distantly related eukaryotes. Means with 95% CIs are plotted. Grey *P* values represent Kruskal–Wallis tests for significant differences among classes using only sequences with homology in one or more nontermites; black *P* values represent Kruskal–Wallis tests for significant differences among classes with 'orphan' sequences included as a separate group. White numbers within bars represent the number of genes falling in that class.

B; Kruskal–Wallis test: $P < 3.0e\text{-}08$, chi-squared = 37.9, $P < 2.2e\text{-}16$, chi-squared: 90.9, respectively).

In a separate analysis, we assigned orthology based on reciprocal best hits to a different species set with greater representation of insect taxa (four insects, two non-insect invertebrates, and two non-invertebrate eukaryotes), assigning each sequence a value based upon the number of orthologues detected. In agreement with our other analysis phylogenetic conservation, we found that the number of species with a putative orthologue for a given gene was significantly negatively associated with CpG o/e in both species (Fig. 4C and D; Kruskal–Wallis test, *R. flavipes*: $P = 1.5e\text{-}05$, chi-squared = 34.3; *C. formosanus*: $P < 2.2e\text{-}16$, chi-squared: 94.8).

In previous studies of DNA methylation in insects, the observation has been made that genes with high levels of DNA methylation tend to be more conserved than those with lower or absent levels of DNA methylation (Hunt *et al*., 2010; Lyko *et al*., 2010; Glastad *et al*., 2011; Park *et al*., 2011). Our data, among termite genes which exhibit

Table 3. Spearman's correlations (ρ) between CpG o/e values for *Reticulitermes flavipes* or *Coptotermes formosanus* and CpG o/e values at orthologous loci in other invertebrates

Species	*R. flavipes* (ρ)	*C. formosanus* (ρ)	DNA methylation	Divergence
Acyrthospihon pisum	0.2497****	0.1811****	Yes†	378
Drosophila melanogaster	0.0349^{NS}	0.0559****	No‡	378
Apis mellifera	0.2553****	0.1852****	Yes‡	378
Tribolium castaneum	0.0589*	0.0088^{NS}	No§	378
Ixodes scapularis	0.1813****	0.1200****	?	581
Ciona intestinalis	0.1699****	0.1137****	Yes‡	782

All divergence estimates in Mya (Hedges *et al*., 2006).
$^{NS}P > 0.05$, $*P < 0.05$, $****P < 0.0001$.
†Glastad *et al*., 2011.
‡Suzuki & Bird, 2008.
§Zemach *et al*., 2010.

orthology to genes in other nontermite species, also demonstrate that those genes with higher phylogenetic conservation tend to be more methylated than those that are less conserved. This relationship held when considering species of wide phylogenetic reach (Fig. 4).

Surprisingly, however, sequences for which no orthologues were detected exhibited marked CpG depletion and a departure from the negative relationship between CpG o/e and evolutionary conservation (Fig. 4). This indicates that termite-specific genes may be more methylated than other, more conserved genes. Alternatively, since ESTs often contain noncoding components of mRNAs, these sequences may, in part, represent noncoding sequences (such as untranslated regions or even introns), which exhibit far less conservation than coding sequence. If such noncoding sequences are targeted by DNA methylation in termites, there may be less resistance to CpG depletion than in functionally constrained coding sequences. This would result in more CpG depletion in noncoding or partially noncoding EST sequences compared with coding sequences, potentially biasing the CpG content of fast-evolving noncoding sequences. However, it is presently unknown whether DNA methylation targets noncoding sequences in termites.

Patterns of DNA methylation at the gene level are conserved across insects

Strikingly, CpG o/e values in our libraries were significantly, positively correlated with the CpG o/e in orthologues in some invertebrate species we examined, but not in others (Table 3; Fig. S3). Correlations between orthologous CpG o/e values of termites and species that have empirically confirmed genomic DNA methylation (*A. mellifera* and *A. pisum*) were much higher than those between termites and insects lacking DNA methylation (*D. melanogaster* and *Tribolium castaneum*; Zemach *et al*., 2010). Interestingly, correlations for GpC o/e values between species were uniformly high and did not depend on the DNA methylation status of the species in question (Fig. S3).

Overall, these correlational data suggest that many of the same genes are methylated between disparate insect taxa (Hunt *et al*., 2010; Sarda *et al*., 2012). Since the four insect species used in our analyses diverged from Isoptera at the same time (~378 Mya; Gaunt & Miles, 2002), the difference in CpG o/e correlation probably reflect the relaxation of methylation-associated mutational pressure on CpG dinucleotides in *T. castaneum* and *D. melanogaster* relative to *A. mellifera* and *A. pisum*. The lack of variation in GpC o/e correlations in the same species supports this view (Fig. S3), as GpC dinucleotides are not targeted by DNA methylation, and thus, not subject to methylation-induced variation in content between genes. Interestingly, the correlations between CpG o/e values in our termite libraries and the gene set of the blacklegged tick, *Ixodes scapularis* (VectorBase, 2009), for which DNA methylation has not been investigated, were nearly as high as those with other invertebrates with DNA methylation. This may suggest the presence of DNA methylation in *I. scapularis*.

Conclusion

DNA methylation has recently emerged as a potentially important regulator of developmental fate in insects (Kucharski *et al*., 2008; Lyko *et al*., 2010). Such phenotypic plasticity is highly important in eusocial insects such as termites. Using assembled EST sequences in two species, we have shown that DNA methylation exists in termites at potentially higher levels than observed in any insect observed to date (Glastad *et al*., 2011). Notably, the CpG o/e values of our *C. formosanus* and, to a lesser extent, *R. flavipes* sequences were nearly as low as those observed for genes of mammals such as *Homo sapiens* (Fig. S1; mean CpG o/e *R. flavipes*: 0.54, *C. formosanus*: 0.51, *H. sapiens*: 0.48), where DNA methylation levels are much higher than in previously studied insects (Suzuki & Bird, 2008).

We have demonstrated similar functional and gene expression associations of DNA methylation to those

previously noted in the eusocial honey bee *A. mellifera* and other insects investigated to date. In addition, we have shown that patterns of DNA methylation exhibit consistency across large evolutionary distances (Table 3). Together, our results support the emerging view that DNA methylation in insects is targeted to active, ubiquitously expressed genes enriched for 'housekeeping' functions (Glastad *et al*., 2011). Moreover, our results suggest that investigations into DNA methylation and gene regulation in termites will provide an important opportunity to investigate the possibility of convergent relationships between phenotypic plasticity and epigenetic regulation.

Experimental procedures

Expressed sequence tag assembly

We obtained 12 943 raw EST sequences derived from five non-normalized *R. flavipes* libraries (reproductive, soldier, worker, early larva, and late larva) (Steller *et al*., 2010; GenBank: G0898823-G0911765), and one library containing 4578 ESTs from a normalized library derived from microorganism-free *R. flavipes* gut tissue (Tartar *et al*., 2009; GenBank: FL634956-FL640828). We also obtained 131 636 raw EST sequences derived from pooled, normalized RNA from *C. formosanus* (Zhang *et al*., 2012; GenBank: JK342531-JK474166). We filtered ESTs for polyA/T tails, low complexity, mtDNA [Accessions: NC_009498.1 (*R. flavipes*), NC_015800.1 (*C. formosanus*)], and vector contamination [National Center for Biotechnology Information (NCBI) core vector library] using SeqClean (http://compbio.dfci.harvard.edu/tgi/software/), cross_match (Huang & Madan, 1999), and RepeatMasker (Smit *et al*., 1996–2010). All EST libraries were generated through Sanger sequencing.

Each library was assembled into contiguous assembled sequence blocks (contigs) using CAP3 (Huang & Madan, 1999). Assembly criteria included 90% minimum sequence identity between overlapping regions, a 45-bp overlap minimum, and 10-bp maximum gap length within the region of overlap. Assembly was performed in three ways for our *R. flavipes* libraries: A) with the entire set of ESTs from all caste-specific (hereafter referred to as 'Rf Caste'; sequences from Steller *et al*., 2010), B) in a caste-specific manner in order to examine patterns of normalized CpG dinucleotide content in each caste's library separately ('Rf Alate', 'Rf Worker', 'Rf Soldier', 'Rf Larval'), and C), as a 'complete' assembly of all 17 521 ESTs from the six *R. flavipes* (from Tartar *et al*., 2009; Steller *et al*., 2010) libraries to serve as a full assembly ('Rf Full'). For the *C. formosanus* library, we assembled all ESTs available using the same assembly criteria as for the other assemblies ('Cf Full'). For all assemblies we only accepted sequences with a length above 250 bp after assembly, in order to ensure sufficient information content for reasonable CpG depletion estimates.

After filtering, our Rf Caste assembly yielded 1037 contigs and 3684 singletons (total library of 5038 sequences). Our caste-specific assemblies yielded the following number of unique sequences: Rf Alate = 1615 (220 contigs, 1395 singletons), Rf Worker = 1683 (323 contigs, 1360 singletons), Rf Soldier = 1544 (193 contigs, 1351 singletons), and Rf Larval = 1537 (236 contigs, 1301 singletons; early and late larval samples were pooled). The Rf Full assembly yielded 1387 contigs and 4975 singletons (6362 sequences total). The Cf Full assembly yielded 16 879 contigs and 8755 singletons (25 634 sequences total).

In order to confirm that the particularly prominent CpG depletion we observed in our termite sequences was not a product of biases in EST libraries, as compared with genome annotations from sequenced genomes, we compared CpG o/e profiles of official pea aphid gene models to a pea aphid EST set with properties similar to those of our *R. flavipes* data. We assembled a nonnormalized library of ESTs (9417 ESTs) from the pea aphid *A. pisum* (Sabater-Munoz *et al*., 2006; GenBank: CN753369-CN764460, CF546452-CF546552, CF587442-CF588411, CN582088-CN587684) in the same manner as for our *R. flavipes* libraries, to produce a library of 4269 sequences (1442 contigs and 2827 singletons) with a length > 250 bp. We then compared this with the CpG o/e distribution from the *A. pisum* Aphidbase mRNA set. We found that the mean CpG o/e value from our aphid EST assembly were highly similar to that of the Aphidbase mRNA annotations (Fig. S2; 0.97 vs. 1.00 mean CpG o/e respectively), indicating that the use of EST sequences did not bias CpG content analysis in any meaningful way.

CpG o/e analysis

Normalized CpG dinucleotide content (CpG o/e) was calculated for all assembled contigs and singletons from assembly runs in both species (Rflav Full, Rflav Caste assembly, and Cfor Full assemblies) as:

$$CpG\,\frac{o}{e} = \frac{length^2}{length} \times \frac{CpG\ count}{C\ count \times G\ count}.$$

CpG o/e is highly informative to the presence of DNA methylation at the germline, as has been demonstrated in numerous investigations (Shimizu *et al*., 1997; Elango *et al*., 2009; Yi & Goodisman, 2009; Glastad *et al*., 2011). The utility of the metric CpG o/e for assessing DNA methylation results from the fact that (1) DNA methylation is largely targeted to cytosines in 5′-CpG-3′ dinucleotides (so called 'CpG dinucleotides') in animals, and (2) 5-methylcytosines tend to undergo spontaneous deamination over evolutionary time at a much higher rate when compared with unmethylated cytosines (Coulondre *et al*., 1978; Shen *et al*., 1994). Consequently, CpG dinucelotides become depleted in genomic regions where DNA methylation is present. Thus CpG o/e reflects levels of DNA methylation in differing genomic regions. In contrast, *GpC o/e* should be unaffected by methylation-dependent mutation pressures and thus provides a null comparison (Fryxell & Moon, 2005).

We used NOCOM (Ott, 1979), a mixture modelling program in the Statistical Genetics Utility program suite, to evaluate our CpG o/e distributions for mixtures of normal distributions (namely, a two-mixture assumption).

Gene ontology and evolutionary assessment

In order to evaluate GO enrichment associated with CpG-depleted and CpG-rich genes in our dataset, we used the program BLAST2GO (Conesa *et al*., 2005) to establish orthology. This performed BLAST searches using our 'Full' assemblies (Rf and Cf Full) against the full NCBI nr protein database (e-value cutoff: 1e-6). A total of 3132 *R. flavipes* and 10 714 *C. formosanus* sequences had significant BLAST hits (1e-6). These BLAST hits

were then linked to GO classifications using default parameters, which establishes and 'maps' GO terms to each sequence based upon several databases (NCBI's gene2accession, UniPort IDs, the DBXRef table of the GO database, and Gene_Product Table of the GO database) and determines BLAST homology.

A total of 2567 *R. flavipes* and 6091 *C. formosanus* sequences were ultimately linked with one or more GO terms. We used the means and standard deviations of each component of a mixture distribution (both low and high CpG o/e) from our Full assemblies' (Rf and Cf Full) CpG o/e distributions (as imputed by the program NOCOM; Ott, 1979), to separate the assembled sequences into high- and low-CpG o/e categories. We chose cutoffs based upon the means of each species' CpG o/e distribution (0.51 and 0.54 for *C. formosanus* and *R. flavipes*, respectively), taking values $\pm$0.05 lower and higher than the means to exclude genes of ambiguous membership. BLAST2GO's inbuilt Gossip package (Blüthgen *et al*., 2005) was then used to test for enrichment of these CpG o/e classes of genes using a Fisher's exact test with correction for multiple testing using a Benjamini-Hochberg false discovery rate.

In order to assess the relationship between sequence conservation and CpG o/e we performed reciprocal best BLAST searches (tblastx, e-value <1.0e-6 (Altschul *et al*., 1990) between our Rf Full and Cf Full assemblies and the full gene sets from eight species of varying phylogenetic distance: *A. mellifera* (OGS 1.1), *D. melanogaster* (FlyBase r5.46), *Ciona intestinalis* (Ensemble KH.67), *I. scapularis* (VectorBase), *Danio rerio* (Ensemble Zv9.67), *H. sapiens* (RefSeq Acyr_2.0), *Neurospora crassa* (Assembly OR74A), and *A. thaliana* (RefSeq TAIR10). We then assigned genes a degree of nested conservation based upon whether they had BLAST hits to insects (*A. mellifera* and *D. melanogaster*), invertebrates (one insect as well as a hit to either *I. scapulari* and/or *C. intestinalis*), animals (one insect, one invertebrate and either *D. rerio* and/or *H. sapiens*), or eukaryotes (one in each previous category as well as a hit to *N. crassa* and/or *A. thaliana*; Fig. 4A and B). This provided a general picture of the breadth of conservation associated with each gene.

We also performed similar reciprocal best BLAST searches with two other insects species (*A. pisum*: RefSeq Acyr_2.0, *T. castaneum*: RefSeq Tcas_3.0) and separately analysed a dataset of all four insects, the two invertebrates used in the previous analysis, and two non-invertebrates (*A. thaliana* and *H. sapiens*), determining the number of species within which a given termite sequence was linked to an orthologue (Fig. 4C and D). We then evaluated the relationship between CpG o/e and both raw number of hits in the four-insect dataset, as well as the nested conservation metric from the cross-eukaryote dataset. We also calculated CpG o/e for the gene sets from four insect species (*A. mellifera*, *A. pisum*, *T. castaneum*, and *D. melanogaster*), two non-insect invertebrates (*I. scapularis* and *C. intestinalis*), and two non-invertebrate eukaryotes (*H. sapiens* and *A. thaliana*) to examine the correlations between the CpG o/e status in our sequences and the CpG o/e of their putative orthologues.

Expression bias and CpG o/e

In order to assess representation bias in our Rf Caste library (Steller *et al*., 2010), a custom perl script was first used to parse the CAP3 output, which provides information on which of the original raw EST sequences contributed to which contig during the assembly process. This then allowed us to determine the EST sequences that contributed to each contig, and from which caste's library they originated. The level of expression of a contig within each treatment was estimated from the number of cognate ESTs found in each library, under the assumption that it was proportional to the transcript frequencies (Romualdi *et al*., 2001; Megy *et al*., 2002). IDEG6 (Romualdi *et al*., 2003) was then used to evaluate the significance of each pairwise comparison of EST contribution to a given contig between castes (Fisher's exact test with multiple testing correction). This gave us an estimate of which contigs exhibited significantly different expression between one or more castes. Within our dataset, 91 contigs showed one or more significant pairwise comparisons between treatments. We then assigned a number from 0 to 4 (pooling 4–6 significant pairwise comparisons), representing the number of pairwise comparisons showing significant differences between libraries.

All statistical tests were performed in R (R Development Core Team, 2011) or JMP (SAS Institute Inc, Cary, NC), unless otherwise indicated.

Acknowledgements

This work was supported by the U. S. National Science Foundation (grant numbers DEB-1011349, DEB-0640690, and IOS-0821130) and the Georgia Tech-Elizabeth Smithgall Watts endowment.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403–410.

Barchuk, A.R., Cristino, A.S., Kucharski, R., Costa, L.F., Simoes, Z.L.P. and Maleszka, R. (2007) Molecular determinants of caste differentiation in the highly eusocial honeybee *Apis mellifera*. *BMC Dev Biol* **7**: 70.

Berger, S.L., Kouzarides, T., Shiekhattar, R. and Shilatifard, A. (2009) An operational definition of epigenetics. *Genes Dev* **23**: 781–783.

Blüthgen, N., Čajavec, B., Swat, M., Herzel, H. and Beule, D. (2005) Biological profiling of gene groups utilizing Gene Ontology. *Genome Inform* **16**: 106–115.

Bock, C. and Lengauer, T. (2008) Computational epigenetics. *Bioinformatics* **24**: 1–10.

Burdge, G.C. and Lillycrop, K.A. (2010) Nutrition, epigenetics, and developmental plasticity: implications for understanding human disease. *Annu Rev Nutr* **30**: 315–339.

Cheong, J., Yamada, Y., Yamashita, R., Irie, T., Kanai, A., Wakaguri, H. *et al*. (2006) Diverse DNA methylation statuses at alternative promoters of human genes in various tissues. *DNA Res* **13**: 155–167.

Chodavarapu, R.K., Feng, S., Bernatavichute, Y.V., Chen, P.-Y., Stroud, H., Yu, Y. *et al*. (2010) Relationship between nucleosome positioning and DNA methylation. *Nature* **466**: 388–392.

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M. and Robles, M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**: 3674–3676.

Coulondre, C., Miller, J.H., Farabaugh, P.J. and Gilbert, W. (1978) Molecular basis of base substitution hotspots in *Escherichia coli*. *Nature* **274**: 775–780.

Eggleton, P. (2011) An introduction to termites: biology, taxonomy and functional morphology. In *Biology of Termites: A Modern Synthesis* (Bignell, D.E., Roisin, Y. and Lo, N., eds), pp. 1–27. Springer, London.

Elango, N., Hunt, B.G., Goodisman, M.A.D. and Yi, S.V. (2009) 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.

Evans, J.D. and Wheeler, D.E. (2001) Gene expression and the evolution of insect polyphenisms. *Bioessays* **23**: 62–68.

Evans, T.A. (2011) Invasive termites. In *Biology of Termites: A Modern Synthesis* (Bignell, D.E., Roisin, Y. and Lo, N., eds), pp. 519–562. CSIRO, Canberra.

Foret, S., Kucharski, R., Pittelkow, Y., Lockett, G. and Maleszka, R. (2009) Epigenetic regulation of the honey bee transcriptome: unravelling the nature of methylated genes. *BMC Genomics* **10**: 472.

Foret, S., Kucharski, R., Pellegrini, M., Feng, S., Jacobsen, S.E., Robinson, G.E. *et al*. (2012) DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. *Proc Natl Acad Sci USA* **109**: 4968–4973.

Fraga, M.F., Ballestar, E., Paz, M.F., Ropero, S., Setien, F., Ballestart, M.L. *et al*. (2005) Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci USA* **102**: 10604–10609.

Fryxell, K.J. and Moon, W.-J. (2005) CpG mutation rates in the human genome are highly dependent on local GC content. *Mol Biol Evol* **22**: 650–658.

Gaunt, M.W. and Miles, M.A. (2002) An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. *Mol Biol Evol* **19**: 748–761.

Glastad, K.M., Hunt, B.G., Yi, S.V. and Goodisman, M.A.D. (2011) DNA methylation in insects: on the brink of the epigenomic era. *Insect Mol Biol* **20**: 553–565.

Goodisman, M.A.D., Kovacs, J.L. and Hunt, B.G. (2008) Functional genetics and genomics in ants (Hymenoptera: Formicidae): the interplay of genes and social life. *Myrmecol News* **11**: 107–117.

He, X.-J., Chen, T. and Zhu, J.-K. (2011) Regulation and function of DNA methylation in plants and animals. *Cell Res* **21**: 442–465.

Hedges, S.B., Dudley, J. and Kumar, S. (2006) TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* **22**: 2971–2972.

Hodges, E., Smith, A., Kendall, J., Xuan, Z., Ravi, K., Rooks, M. *et al*. (2009) High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing. *Genome Res* **19**: 1593–1605.

Hölldobler, B. and Wilson, E.O. (1990) *The Ants*. Belknap Press of Harvard Univ Press, Cambridge, MA.

Huang, X. and Madan, A. (1999) CAP3: a DNA sequence assembly program. *Genome Res* **9**: 868–877.

Hunt, B.G., Brisson, J.A., Yi, S.V. and Goodisman, M.A.D. (2010) Functional conservation of DNA methylation in the pea aphid and the honeybee. *Genome Biol Evol* **2**: 719–728.

Jeong, S., Liang, G., Sharma, S., Lin, J.C., Choi, S.H., Han, H. *et al*. (2009) Selective anchoring of DNA methyltransferases 3A and 3B to nucleosomes containing methylated DNA. *Mol Cell Biol* **29**: 5366–5376.

Klose, R.J. and Bird, A.P. (2006) Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci* **31**: 89–97.

Kota, S.K. and Feil, R. (2010) Epigenetic transitions in germ cell development and meiosis. *Dev Cell* **19**: 675–686.

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

Li, X.Y., Wang, X.F., He, K., Ma, Y.Q., Su, N., He, H. *et al*. (2008) High-resolution mapping of epigenetic modifications of the rice genome uncovers interplay between DNA methylation, histone methylation, and gene expression. *Plant Cell* **20**: 259–276.

Lo, N., Li, B. and Ujvari, B. (2012) DNA methylation in the termite *Coptotermes lacteus*. *Insectes Soc* **59**: 257–261.

Lyko, F., Foret, S., Kucharski, R., Wolf, S., Falckenhayn, C. and Maleszka, R. (2010) The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *Plos Biol* **8**: e1000506.

Mao, L., Henderson, G., Liu, Y. and Laine, R.A. (2005) Formosan subterranean termite (Isoptera: *Rhinotermitidae*) soldiers regulate juvenile hormone levels and caste differentiation in workers. *Ann Entomol Soc Am* **98**: 340–345.

Margueron, R. and Reinberg, D. (2010) Chromatin structure and the inheritance of epigenetic information. *Nat Rev Genet* **11**: 285–296.

Maunakea, A.K., Nagarajan, R.P., Bilenky, M., Ballinger, T.J., D'Souza, C., Fouse, S.D. *et al*. (2010) Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* **466**: 253–257.

Megy, K., Audic, S. and Claverie, J.M. (2002) Heart specific genes revealed by EST sampling. *Genome Biol* **3**: 0074.1–0074.11.

Okitsu, C.Y. and Hsieh, C.L. (2007) DNA methylation dictates histone H3K4 methylation. *Mol Cell Biol* **27**: 2746–2757.

Ott, J. (1979) Detection of rare major genes in lipid levels. *Hum Genet* **51**: 79–91.

Park, J., Peng, Z., Zeng, J., Elango, N., Park, T., Wheeler, D. *et al*. (2011) Comparative analyses of DNA methylation and sequence evolution using *Nasonia* genomes. *Mol Biol Evol* **28**: 3345–3354.

Park, Y.I. and Raina, A.K. (2004) Juvenile hormone III titers and regulation of soldier caste in *Coptotermes formosanus* (Isoptera : Rhinotermitidae). *J Insect Physiol* **50**: 561–566.

Pfennig, D.W., Wund, M.A., Snell-Rood, E.C., Cruickshank, T., Schlichting, C.D. and Moczek, A.P. (2010) Phenotypic plasticity's impacts on diversification and speciation. *Trends Ecol Evol* **25**: 459–467.

Price, T.D., Qvarnström, A. and Irwin, D.E. (2003) The role of phenotypic plasticity in driving genetic evolution. *Proc R Soc Lond B Biol Sci* **270**: 1433–1440.

R Development Core Team (2011) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna. http://www.R-project.org.

Reik, W. (2007) Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **447**: 425–432.

Robinson, G.E., Grozinger, C.M. and Whitfield, C.W. (2005) Sociogenomics: social life in molecular terms. *Nat Rev Genet* **6**: 257–270.

Roisin, Y. and Korb, J. (2011) Social organisation and the status of workers in termites. In *Biology of Termites: A Modern Synthesis* (Bignell, D.E., Roisin, Y. and Lo, N., eds), pp. 133–164. Springer, London.

Romualdi, C., Bortoluzzi, S. and Danieli, G.A. (2001) Detecting differentially expressed genes in multiple tag sampling

experiments: comparative evaluation of statistical tests. *Hum Mol Genet* **10**: 2133–2141.

Romualdi, C., Bortoluzzi, S., D'Alessi, F. and Danieli, G.A. (2003) IDEG6: a web tool for detection of differentially expressed genes in multiple tag sampling experiments. *Physiol Genomics* **12**: 159–162.

Sabater-Munoz, B., Legeai, F., Rispe, C., Bonhomme, J., Dearden, P., Dossat, C. *et al*. (2006) Large-scale gene discovery in the pea aphid *Acyrthosiphon pisum* (Hemiptera). *Genome Biol* **7**: R21.

Sarda, S., Zeng, J., Hunt, B.G. and Yi, S.V. (2012) The evolution of invertebrate gene body methylation. *Mol Biol Evol* **29**: 1907–1916.

Scharf, M.E., Buckspan, C.E., Grzymala, T.L. and Zhou, X. (2007) Regulation of polyphenic caste differentiation in the termite *Reticulitermes flavipes* by interaction of intrinsic and extrinsic factors. *J Exp Biol* **210**: 4390–4398.

Schmitz, R.J. and Ecker, J.R. (2012) Epigenetic and epigenomic variation in *Arabidopsis thaliana*. *Trends Plant Sci* **17**: 149–154.

Schwander, T., Lo, N., Beekman, M., Oldroyd, B.P. and Keller, L. (2010) Nature versus nurture in social insect caste differentiation. *Trends Ecol Evol* **25**: 275–282.

Shen, J.C., Rideout, W.M. and Jones, P.A. (1994) The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. *Nucleic Acids Res* **22**: 972–976.

Shimizu, T.S., Takahashi, K. and Tomita, M. (1997) CpG distribution patterns in methylated and non-methylated species. *Gene* **205**: 103–107.

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M. *et al*. (2011) CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479**: 74–79.

Simpson, S.J., Sword, G.A. and Lo, N. (2011) Polyphenism in insects. *Curr Biol* **21**: R738–R749.

Smit, A.F.A., Hubley, R. and Green, P. (1996–2010) RepeatMasker Open-3.3. http://www.repeatmasker.org. Accessed on 6 May 2012.

Smith, C.R., Toth, A.L., Suarez, A.V. and Robinson, G.E. (2008) Genetic and genomic analyses of the division of labour in insect societies. *Nat Rev Genet* **9**: 735–748.

Steller, M.M., Kambhampati, S. and Caragea, D. (2010) Comparative analysis of expressed sequence tags from three castes and two life stages of the termite *Reticulitermes flavipes*. *BMC Genomics* **11**: 463.

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

Suzuki, M.M., Kerr, A.R.W., De Sousa, D. and Bird, A. (2007) CpG methylation is targeted to transcription units in an invertebrate genome. *Genome Res* **17**: 625–631.

Tartar, A., Wheeler, M., Zhou, X., Coy, M., Boucias, D. and Scharf, M. (2009) Parallel metatranscriptome analyses of host and symbiont gene expression in the gut of the termite *Reticulitermes flavipes*. *Biotechnol Biofuels* **2**: 25.

Toru, M. and Scharf, M.E. (2011) Molecular basis underlying caste differentiation in termites. In *Biology of Termites: A Modern Synthesis* (Bignell, D.E., Roisin, Y. and Lo, N., eds), pp. 211–253. Springer, London.

Tweedie, S., Charlton, J., Clark, V. and Bird, A. (1997) Methylation of genomes and genes at the invertebrate-vertebrate boundary. *Mol Cell Biol* **17**: 1469–1475.

VectorBase (2009) *Ixodes scapularis* annotation, IscaW1.2. http://www.vectorbase.org. Accessed on 5 January 2012.

Walsh, T.K., Brisson, J.A., Robertson, H.M., Gordon, K., Jaubert-Possamai, S., Tagu, D. *et al*. (2010) A functional DNA methylation system in the pea aphid, *Acyrthosiphon pisum*. *Insect Mol Biol* **19**: 215–228.

Wang, Y. and Leung, F. (2009) In silico prediction of two classes of honeybee genes with CpG deficiency or CpG enrichment and sorting according to gene ontology classes. *J Mol Evol* **68**: 700–705.

West-Eberhard, M.J. (2003) *Developmental Plasticity and Evolution*. Oxford University Press, New York.

Wilson, E.O. (1971) *The Insect Societies*. Harvard University Press, Cambridge.

Xiang, H., Zhu, J.D., Chen, Q., Dai, F.Y., Li, X., Li, M.W. *et al*. (2010) Single base-resolution methylome of the silkworm reveals a sparse epigenomic map. *Nat Biotechnol* **28**: 516–U181.

Yi, S.V. and Goodisman, M.A.D. (2009) Computational approaches for understanding the evolution of DNA methylation in animals. *Epigenetics* **4**: 551–556.

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

Zhang, D., Lax, A.R., Henrissat, B., Coutinho, P., Katiya, N., Nierman, W.C. *et al*. (2012) Carbohydrate-active enzymes revealed in *Coptotermes formosanus* (Isoptera: Rhinotermitidae) transcriptome. *Insect Mol Biol* **21**: 235–245.

Zhang, X.Y., Bernatavichute, Y.V., Cokus, S., Pellegrini, M. and Jacobsen, S.E. (2009) Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in *Arabidopsis thaliana*. *Genome Biol* **10**: R62.

Supporting Information

Additional Supporting Information may be found in the online version of this article under the DOI reference: 10.1111/imb.12010

Figure S1. CpG o/e profiles (and GpC o/e – inset) for (A) *Reticulitermes flavipes*, (B) *C. formosanus*, and (C) *Homo sapiens*, as well as a (D) density graph comparing all of their distributions. These comparisons demonstrate that genic methylation levels in termites may be very similar to genic methylation levels in mammals.

Figure S2. CpG o/e (and GpC o/e – inset) of the (A) pea aphid *Acyrthosiphon pisum* official gene set and (B) assembled expressed sequence tag library similar to those in this study, demonstrating our methods of assembly do not alter the CpG o/e distributions noticeably, relative to actual coding sequence annotation.

Figure S3. Comparison of Spearman's rank correlations of gene sequence for (A) GpC o/e and (B) CpG o/e levels between select invertebrates and *Reticulitermes flavipes* and *C. formosanus*. Notably, the correlations for CpG o/e, but not GpC o/e, vary drastically, indicating the action of DNA methylation on CpG, but not GpC, dinucleotides.