

## RESEARCH ARTICLE

# Characterization of genome methylation patterns in the desert locust *Schistocerca gregaria*

Cassandra Falckenhayn<sup>1,2</sup>, Bart Boerjan<sup>3</sup>, Günter Raddatz<sup>1</sup>, Marcus Frohme<sup>2</sup>, Liliane Schoofs<sup>3</sup> and Frank Lyko<sup>1,\*</sup>

<sup>1</sup>Division of Epigenetics, DKFZ-ZMBH Alliance, German Cancer Research Center, Heidelberg 69120, Germany,

<sup>2</sup>Molecular Biotechnology and Functional Genome Analysis, Technical University of Applied Sciences, Wildau 15745, Germany and

<sup>3</sup>Functional Genomics and Proteomics Group, Department of Biology, KU Leuven, B-3000 Leuven, Belgium

\*Author for correspondence (f.lyko@dkfz.de)

### SUMMARY

**DNA methylation is a widely conserved epigenetic modification. The analysis of genome-scale DNA methylation patterns in various organisms suggests that major features of animal methylomes are widely conserved. However, based on the variation of DNA methyltransferase genes in invertebrates, it has also been proposed that DNA methylation could provide a molecular mechanism for ecological adaptation. We have now analyzed the methylome of the desert locust, *Schistocerca gregaria*, which represents an organism with a high degree of phenotypic plasticity. Using genome-scale bisulfite sequencing, we show here that the *S. gregaria* methylome is characterized by CpG- and exon-specific methylation and thus shares two major features with other animal methylomes. In contrast to other invertebrates, however, overall methylation levels were substantially higher and a significant fraction of transposons was methylated. Additionally, genic sequences were densely methylated in a pronounced bimodal pattern, suggesting a role for DNA methylation in the regulation of locust gene expression. Our results thus uncover a unique pattern of genome methylation in locusts and provide an important foundation for investigating the role of DNA methylation in locust phase polyphenism.**

Supplementary material available online at <http://jeb.biologists.org/cgi/content/full/216/8/1423/DC1>

Key words: DNA methylation, *Schistocerca gregaria*, locust, methylome.

Received 2 October 2012; Accepted 13 December 2012

### INTRODUCTION

As an epigenetic mechanism, DNA methylation plays a key role in the interpretation of genetic information (Bird, 2007). Examples include methylation-associated gene silencing and the regulation of developmental gene expression (Mohn and Schübeler, 2009). DNA methylation is catalyzed by a family of DNA methyltransferases (Goll and Bestor, 2005). Establishment and maintenance of DNA methylation patterns are widely assumed to be mediated by distinct enzymes, Dnmt3 and Dnmt1. Dnmt3 enzymes methylate unmethylated DNA and have traditionally been designated as ‘*de novo*’ methyltransferases that establish methylation patterns (Okano et al., 1998). In contrast, Dnmt1 enzymes show a strong preference for hemimethylated substrates (Bestor and Ingram, 1983; Gruenbaum et al., 1982) and have therefore been termed ‘maintenance methyltransferases’. These differential functional specificities of Dnmt1 and Dnmt3 are consistent with some, but not all, characteristics of DNA methylation patterns in DNA methyltransferase-deficient models (Jones and Liang, 2009). Consequently, it has been suggested that the functional specificities of Dnmt1 and Dnmt3 may be broader than previously thought (Jones and Liang, 2009).

Interestingly, the number of Dnmt1 and Dnmt3 genes can vary between organisms, and various organisms have been shown to express different sets of Dnmt enzymes (Goll and Bestor, 2005). Mammalian genomes, for example, encode one Dnmt1 gene and three paralogs of Dnmt3. This contrasts with the genome of the

parasitic wasp *Nasonia vitripennis*, which encodes three paralogs of Dnmt1 and a single Dnmt3 homologue. Such variations in the complement of DNA methyltransferases have been interpreted to reflect multiple versions of a toolkit for phenotypic adaptation (Lyko and Maleszka, 2011). During evolution, specific parts of this toolkit could have been contracted or expanded to facilitate specific requirements for genome regulation (Lyko and Maleszka, 2011).

Cells and organisms are known to adapt to changes in environmental conditions and DNA methylation could serve as an interface between environmental cues and the genome. Environment-induced epigenetic changes usually appear to be rather small in mammals (Tobi et al., 2012) and are often limited to specific genetic variations (Feinberg and Irizarry, 2010; Rakyan et al., 2003; Waterland and Jirtle, 2003). This has reinvigorated the use of non-mammalian models to further study the effects of environmental changes on the epigenome. Prominent examples include temperature-dependent sex-ratio shifts in fish (Navarro-Martin et al., 2011) and nutrition-dependent caste specification in honeybees (Kucharski et al., 2008).

In several insect species, a high degree of phenotypic plasticity has been associated with alterations in genomic DNA methylation patterns (Glastad et al., 2011; Lyko and Maleszka, 2011; Walsh et al., 2010). Consequently, an increasing number of studies has used genome-scale sequencing approaches to comprehensively analyze the methylomes of various insect species. Examples include the flour beetle (*Tribolium castaneum*), the silkworm (*Bombyx mori*), the

honey bee (*Apis mellifera*) and two ant species (*Camponotus floridanus* and *Harpegnathos saltator*) (Bonasio et al., 2012; Feng et al., 2010; Lyko et al., 2010; Xiang et al., 2010; Zemach et al., 2010). Overall cytosine methylation levels of the sequenced insect methylomes were very low (0.1–0.2%), but methylated cytosine residues were significantly enriched in CpG dinucleotide sequence context and in exons. Also, repetitive elements appeared to be depleted for DNA methylation. Together, these observations defined several conserved features of DNA methylation in invertebrates (Feng et al., 2010; Zemach et al., 2010).

Among the various known insect models, locusts represent a particularly attractive example for analyzing the effects of environmental changes on the organismal phenotype (Tanaka et al., 2012). Locusts live either solitarily or gregariously in hopper bands and swarms. Both phases differ in a number of morphological, physiological and behavioural features, depending on the maternally experienced population density (Rogers et al., 2003). This phase polyphenism could potentially be regulated by epigenetic mechanisms, and a recent biochemical analysis of *Schistocerca gregaria* genomic DNA suggested robust levels of DNA methylation in the locust genome (Boerjan et al., 2011). Similar observations were also made in another locust species, *Locusta migratoria* (Robinson et al., 2011), but details about the distribution of methylation marks in the locust genome have largely remained unknown. We have now used next-generation shotgun bisulfite sequencing of genomic DNA from *S. gregaria* to characterize the methylation patterns of neural tissues from adult gregarious animals. Our results show that the *S. gregaria* methylome is characterized by comparably high levels of DNA methylation in a CpG-specific sequence context. Further analysis revealed that a subset of transposons and a major fraction of genic sequences were densely and homogeneously methylated. These findings clearly distinguish *S. gregaria* from other insects with known methylation patterns and provide the foundation for understanding the role of DNA methylation in locust phase polyphenism.

## MATERIALS AND METHODS

### Animals

*Schistocerca gregaria* Forsskal 1775 were reared under crowded conditions to induce features typical for the gregarious phase. All animals were raised under a controlled temperature of 32°C, a photoperiod of 13 h:11 h light:dark and a relative humidity between 40 and 60%. Animals were fed daily with fresh cabbage and dried oats. The crowded adult locusts were kept in cages of 38×38×38 cm with 100–200 individuals per cage. For mRNA and methylation analysis, twenty animals (10 males and 10 females) were killed at the age of 20 days.

### Sample preparation

Nucleic acids were extracted from 20 brains and 20 metathoracic ganglia of adult *S. gregaria*. Tissues were dissected with great care under a binocular microscope in insect saline buffer (150 mmol l<sup>-1</sup> NaCl, 10 mmol l<sup>-1</sup> KCl, 4 mmol l<sup>-1</sup> CaCl<sub>2</sub>, 2 mmol l<sup>-1</sup> MgCl<sub>2</sub>, 10 mmol l<sup>-1</sup> Hepes, pH 7.0) and collected in tubes containing MagNA Lyser Green Beads (Roche, Mannheim, Germany). All samples were snap-frozen in liquid nitrogen and stored at -80°C until homogenisation using the MagNA Lyser (Roche). Total RNA was extracted using the RNeasy lipid tissue Mini Kit (Qiagen, Hilden, Germany). Messenger RNA was purified from total RNA using the Illustra Quickprep Micro mRNA purification kit (GE Healthcare, Buckinghamshire, UK). Genomic DNA was isolated using the DNeasy Blood and Tissue Kit (Qiagen).

### mRNA sequencing

The Dnmt1 catalytic domain fragment was amplified from brain cDNA using the primers CCTTGCCAGGGTTTTAGTGGA-ATGAAT (forward) and TGCATTTCCGATCTGTGCGATGTTT (reverse). PCR products were subsequently gel purified, cloned and sequenced using standard methods. The sequence of the *S. gregaria* Dnmt1 catalytic domain fragment has been deposited in GenBank (accession number JX857690).

Deep sequencing libraries were prepared from brain mRNA by using the TruSeq(R) RNA Sample Prep Kit v2 and the TruSeq(R) PE Cluster Kit v3 (Illumina, Munich, Germany). Paired-end sequencing was performed on an Illumina HiSeq system with read lengths of 105 bp and an average insert size of 250 bp. Two hundred and forty-three million read pairs were obtained on a single lane and trimmed by removing stretches of bases with a quality score <30 from the ends. Preprocessed data were assembled around the Dnmt1 catalytic fragment using the oases transcriptome assembly pipeline (Schulz et al., 2012), version 0.2.08. To account for the low expression level of Dnmt1, we used hash length (*k*) values of 19 (minimum) and 21 (maximum) and a coverage cut-off of 2.0.

### Methylation sequencing

Library preparation for genome-scale bisulfite sequencing was performed as described previously (Lyko et al., 2010). Paired-end sequencing with read lengths of 105 nucleotides was performed on an Illumina HiSeq platform. Sequencing data have been deposited in the GEO database under the accession number GSE41214. Reads were trimmed to a maximal length of 80 bp and stretches of bases with a quality score <30 at read ends were removed. Reads were mapped against the assembled contigs of the locust expressed sequence tag (EST) project (<http://titan.biotech.uiuc.edu/locust>) using BSMAP 2.0 (Xi and Li, 2009). Only reads mapping uniquely and with both read pairs having the correct distance were used in further analyses. Methylation rates were determined using a Python script distributed with the BSMAP package. To reduce the effects of sequencing errors, methylated Cs were only called when covered by more than three reads.

454 bisulfite sequencing of selected loci was performed as described previously (Grönniger et al., 2010). Briefly, DNA was treated with bisulfite using the Qiagen EpiTect Bisulfite Kit (Qiagen), according to the manufacturer's instructions. Bisulfite-treated DNA was then amplified using sequence-specific primers and 454 linker sequences (primer sequences are provided in supplementary material Table S1). PCR products were subsequently purified using the QIAquick Gel Extraction Kit (Qiagen), pooled and sequenced on a Roche 454 GS Junior system.

### Gene expression analysis

Messenger RNA isolated from brains of adult *S. gregaria* was reverse-transcribed using SuperScript III (Invitrogen, Darmstadt, Germany). qPCR analyses were performed on a LightCycler 480 Real Time PCR System (Roche) using the Absolute QPCR SYBR Green Mix (Thermo Scientific, Epsom, UK). The relative expression level was determined by the mean crossing point (Cp) value of three technical replicates in reference to  $\beta$ -tubulin. Primer sequences are provided in supplementary material Table S1.

## RESULTS

### Characterization of the *S. gregaria* methylation machinery

It is generally assumed that establishment and maintenance of DNA methylation patterns requires the coordinated activities of Dnmt3 *de novo* methyltransferases and Dnmt1 maintenance

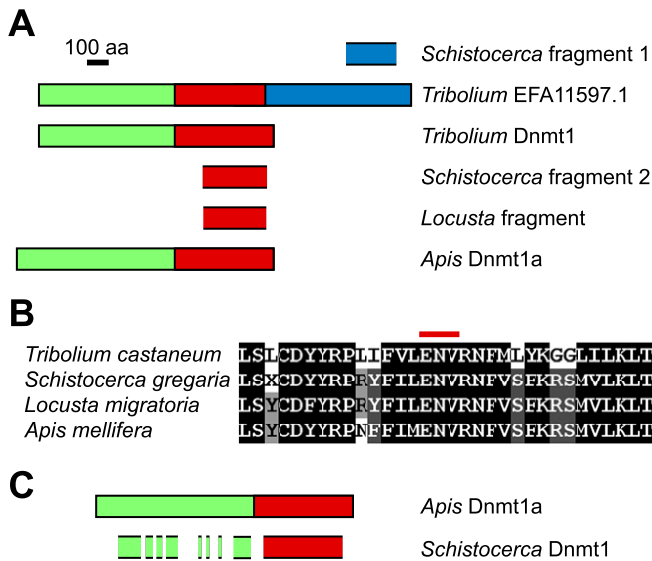


Fig. 1. Sequence analysis of the *Schistocerca gregaria* Dnmt1 homologue. (A) Schematic illustration of Dnmt1-related sequences from various insects. The Dnmt1 catalytic domain is indicated in red, the N-terminal, regulatory domain is marked in green and the cullin domain from the *Tribolium* EFA11597.1 sequence is shown in blue. The published (Boerjan et al., 2011) *S. gregaria* Dnmt1 sequence (fragment 1) shows sequence homology to the C-terminal cullin domain of *Tribolium* EFA11597.1, but not to any of the known Dnmt1 homologues. The sequence amplified in our study (fragment 2) shows substantial sequence similarity to the C-terminal catalytic domain of known Dnmt1 enzymes. (B) Fragment 2 encodes the highly conserved motif VI of canonical DNA methyltransferases (red bar). (C) Transcriptome-scale shotgun cDNA sequencing reveals additional sequence fragments with significant homologies to *Apis* Dnmt1a.

methyltransferases. Although it has been suggested that locust genomes may not encode a Dnmt3 homologue, putative Dnmt1 homologues have been described both in *S. gregaria* and in *L. migratoria* (Boerjan et al., 2011; Robinson et al., 2011).

A closer analysis of the published *S. gregaria* Dnmt1 fragment (Boerjan et al., 2011) revealed that this sequence does not contain any of the conserved catalytic motifs of Dnmt1 enzymes. Instead, the fragment showed sequence similarity to a C-terminal cullin domain from a unique Dnmt1-like sequence of *Tribolium castaneum* (Fig. 1A). Further analysis of the *Tribolium* genome sequencing data strongly suggested that the predicted Dnmt1-like gene with a cullin domain represents an artifact from the genome assembly process (data not shown). Furthermore, the *Tribolium* genome also encodes an independent Dnmt1 homologue without a cullin domain and with 45% sequence identity and 64% sequence similarity to the Dnmt1a protein from the honeybee (Fig. 1A). To confirm the presence of a Dnmt1 homologue in *S. gregaria*, we designed PCR primers against

the highly conserved catalytic DNA methyltransferase motifs IV and X and amplified putative Dnmt1-related sequences from total adult *S. gregaria* mRNA. Subsequent cloning and sequencing of PCR amplicons identified a novel sequence with clear homology to the Dnmt1 catalytic domain (Fig. 1A). This predicted protein sequence also showed very high similarity with the predicted Dnmt1 fragment from *Locusta migratoria* and with the *Apis* Dnmt1a protein (Fig. 1B).

The presence of a Dnmt1 homologue in *S. gregaria* was further confirmed by a transcriptome-scale shotgun cDNA sequencing approach and by a BLAST search for sequence fragments with significant homologies to *Apis* Dnmt1a. This identified 75 hits with an E-value <0.1 and allowed the identification and assembly of additional fragments with significant sequence homology to *Apis* Dnmt1 enzymes (Fig. 1C). In contrast, a BLAST search with the *Apis* Dnmt3 sequence failed to identify any hits with an E-value <0.1 in the *S. gregaria* transcriptome sequencing data. Together, these data suggest that DNA methylation in locusts may be established and maintained by Dnmt1, which renders the characterization of methylation patterns in this organism particularly interesting.

### General characteristics of the *S. gregaria* methylome

A previous study suggested that the global cytosine methylation level of *S. gregaria* is distinctly higher than that of other insects (Boerjan et al., 2011). To confirm this finding and to characterize the locust DNA methylation pattern, we used genome-scale shotgun bisulfite sequencing of genomic DNA from brains and metathoracic ganglia (MTG) of gregarious adult animals. These tissues were selected because they are known to play an important role in processing the environmental cues that trigger locust phase polyphenism (Burrows et al., 2011; Rogers et al., 2003). For each library, we used a single round of sequencing to obtain 180 million sequence reads, corresponding to 32 Gb of DNA sequence. We used this sequence information to calculate the total cytosine methylation level and found it to be 1.3% for brain and 1.4% for MTG. These levels are very similar to the previously reported liquid chromatography–mass spectrometry results (Boerjan et al., 2011) and suggest that DNA methylation in *S. gregaria* is substantially more prevalent than in other known insect genomes. We also found the vast majority (>90%) of 5-methylcytosine residues in a CpG dinucleotide context (Table 1), which is consistent with the sequence context of DNA methylation in all other known animal methylomes.

The genome of *S. gregaria* remains to be sequenced. As such, we could not use our methylation sequencing data for the generation of genome-wide methylation maps. However, a database of 12,748 assembled EST contigs encompassing 11.1 Mb of DNA sequence is available and was used as a reference for the mapping of our methylation data. Total cytosine methylation levels of the reference sequence were found to be 3.2% for brain and 3.1% for MTG (Table 1). These values are distinctly higher than the methylation

Table 1. DNA methylation in the *Schistocerca gregaria* brain and meta-thoracic ganglion (MTG)

	% mC	No. of mC	No. of mCG	No. of mCH	% mCG
Brain (all)	1.3	12,235,552	11,257,932	977,620	92
MTG (all)	1.4	13,508,372	12,110,466	1,397,906	90
Brain (mapped)	3.2	24,663	24,033	630	97
MTG (mapped)	3.1	14,999	14,554	445	97

Methylation calling was independently performed for all cytosines from all reads (all) and for all cytosines that could be mapped to the EST contig reference database (mapped). % mC, the fraction of methylated cytosines (relative to all reference cytosines); mCH, non-converted cytosine residues in non-CpG dinucleotide context; % mCG, the fraction of methylated cytosines in CpG dinucleotide context. Methylation calling for all cytosines was restricted to nucleotides 1–30 to minimize the effects of sequencing errors.



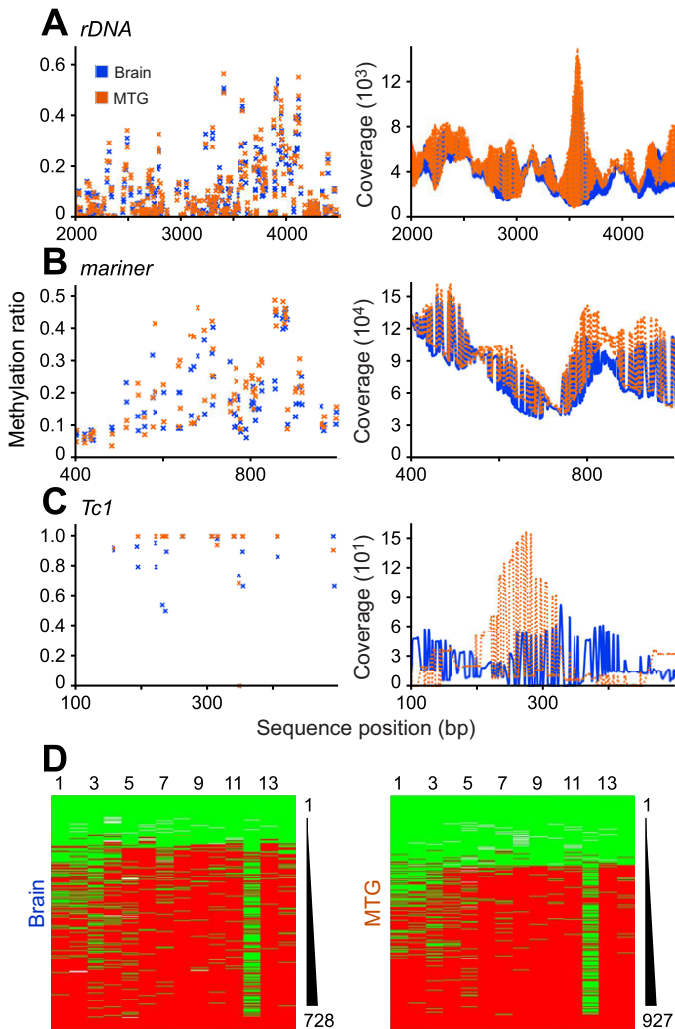


Fig. 2. Methylation analysis of repetitive sequences. (A–C) Genome-scale shotgun bisulfite sequencing analysis of *rDNA*, a *mariner*-like retroelement and a *Tc1*-like transposon. Plots show methylation ratios or coverages for specific CpG dinucleotides from the brain and the meta-thoracic ganglion (MTG). (D) 454 bisulfite sequencing analysis of the *Tc1* sequence. Sequencing results are shown as heatmaps for brain and MTG. Each row represents one sequence read; columns represent individual CpG dinucleotides. Red boxes indicate methylated CpG dinucleotides, green boxes indicate unmethylated CpG dinucleotides and sequencing gaps are shown in white. Lower right numbers indicate the sequencing coverage (reads).

levels calculated for the complete sequencing libraries (Table 1), which can be interpreted to reflect a preferential methylation of exons, consistent with the methylation patterns described in other insects (Bonasio et al., 2012; Lyko et al., 2010; Xiang et al., 2010). Further sequence analysis also allowed us to determine the sequence context of 24,663 methylated cytosine residues in brain and 14,999 methylated cytosine residues in MTG (Table 1). More than 95% of the mapped 5-methylcytosine residues were found in the CpG dinucleotide context (Table 1), which is again consistent with the methylation patterns described in other insects (Bonasio et al., 2012; Lyko et al., 2010; Xiang et al., 2010).

#### Methylation of rDNA repeats and transposons

Based on a phylogenetic methylome analysis from various species, it has previously been suggested that repetitive sequences are only

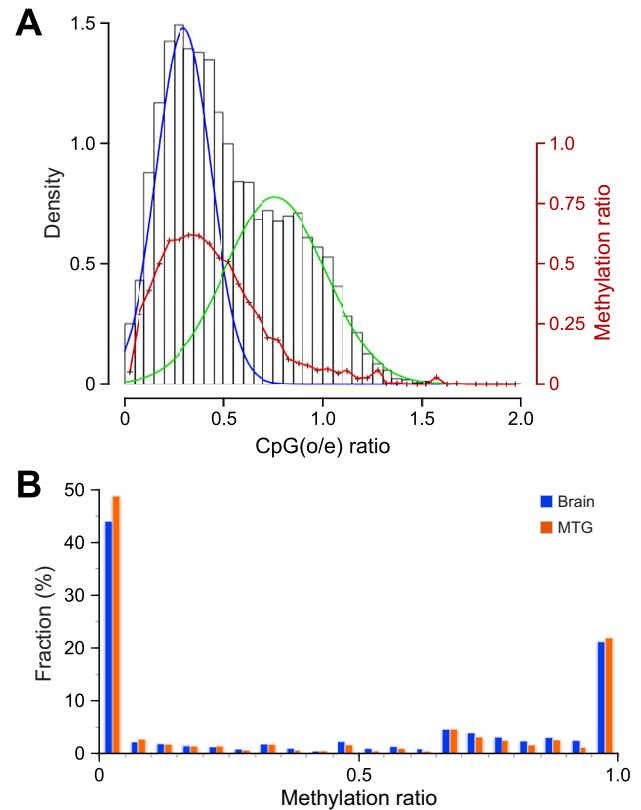


Fig. 3. Distribution of gene methylation in *S. gregaria*. (A) CpG density of *S. gregaria* genes. CpG(observed/expected) [CpG(o/e)] values were determined for all covered gene contigs and then distributed into bins with increasing CpG(o/e) values. The blue line illustrates the distribution of gene contigs with a low CpG(o/e) value, the green line illustrates the distribution of gene contigs with a high CpG(o/e) value. The red line indicates the average methylation level of the sequences contained in the bin. (B) Average methylation levels were determined for all gene contigs and then distributed into bins with increasing methylation ratios. Data are shown from the brain and the meta-thoracic ganglion (MTG).

methylated in plants and in vertebrates (Zemach and Zilberman, 2010). In addition, it has also been shown that rDNA repeats are unmethylated in silkworms and in honeybees (Lyko et al., 2010; Xiang et al., 2010). To further characterize the methylation of repetitive sequences in *S. gregaria*, we first investigated the methylation status of rDNA repeats. These sequences showed an average coverage of  $>1000\times$  in our data sets (Fig. 2A). Interestingly, we detected robust levels of rDNA methylation in brain and in MTG (Fig. 2A), which suggested that repetitive sequences are methylated in *S. gregaria*.

In subsequent steps, we also analyzed the methylation status of transposons. For example, we identified numerous copies of a *mariner*-like element, which we had sequenced with average coverages of roughly  $10,000\times$  (Fig. 2B). Individual CpGs from this transposon showed methylation ratios of up to 0.5, both in brain and in MTG (Fig. 2B), with a high degree of CpG specificity ( $>50$ -fold, data not shown). Methylation patterns were highly similar in brain and MTG samples (Fig. 2B), which provided additional confirmation for the methylation of repetitive elements.

Lastly, our data analysis also revealed a particularly high methylation level for the transposase sequence of a *Tc1*-like transposon (Fig. 2C). For this example, the sequencing coverage appeared to be rather low and we therefore sought to validate the

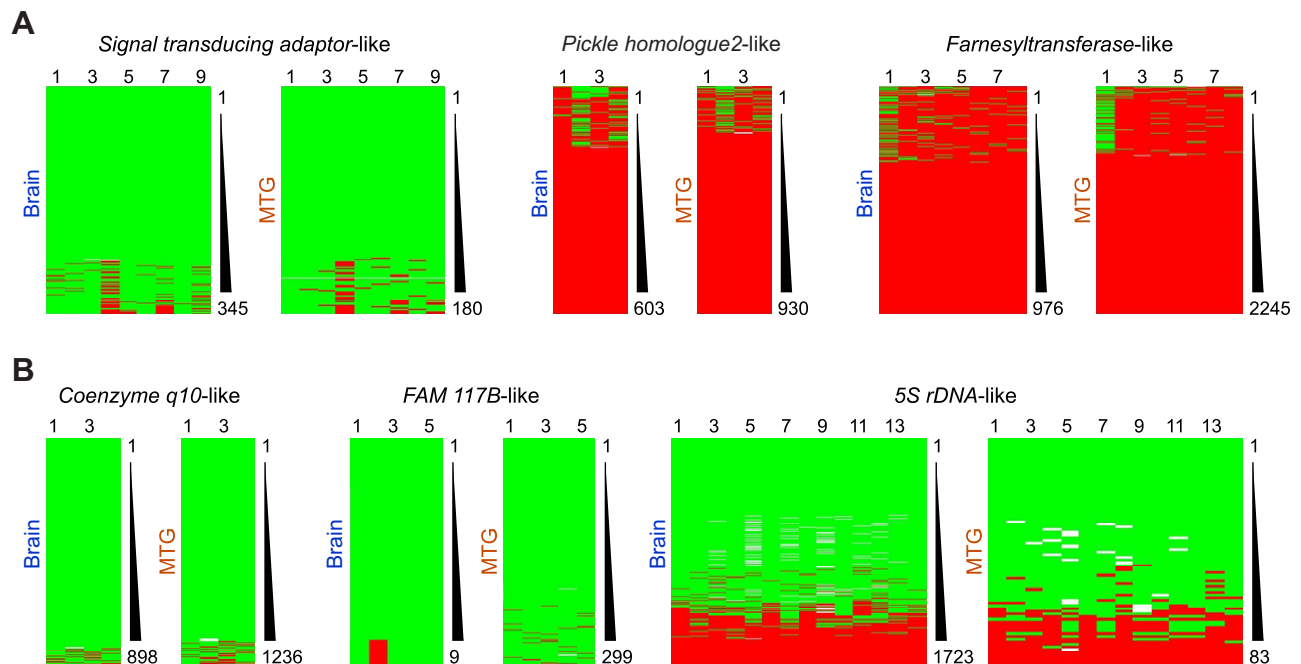


Fig. 4. Methylation pattern analysis of genic sequences. 454 bisulfite sequencing analysis of three arbitrarily selected contigs with low CpG(o/e) values (A) and three arbitrarily selected contigs with high CpG(o/e) values (B). Sequencing results are shown as heatmaps for brain and for meta-thoracic ganglion (MTG). Each row represents one sequence read; columns represent individual CpG dinucleotides. Red boxes indicate methylated CpG dinucleotides, green boxes indicate unmethylated CpG dinucleotides and sequencing gaps are shown in white. Lower right numbers indicate the sequencing coverage (reads).

methylation results by an independent sequencing approach. Using 454 bisulfite sequencing analysis of the *Tc1*-like transposase sequence, we obtained 728 reads from brain and 927 reads from MTG. Data analysis showed that this sequence was strongly methylated in the majority of reads, both in brain and in MTG (Fig. 2D). The results obtained with 454 bisulfite sequencing were in excellent agreement with the Illumina sequencing data and further illustrate the methylation of transposons in the *S. gregaria* genome.

#### Methylation of genic sequences

In animals, methylated genes are often characterized by a lower than expected density of CpG dinucleotides, which reflects the propensity of methylated Cs to be converted to thymines (Shen et al., 1994; Yi and Goodisman, 2009). We therefore determined the CpG observed/expected (o/e) ratios for all *S. gregaria* contigs. This revealed a characteristic bimodal distribution, with the two groups separated at 0.51 and a pronounced peak at low CpG(o/e) ratios (Fig. 3A, blue and green lines). We then determined the average methylation ratio for contigs with increasing CpG(o/e)

values and found that contigs with low CpG(o/e) ratios had substantially higher average methylation levels than contigs with high CpG(o/e) ratios (Fig. 3A, red line). This finding is consistent with earlier observations in honeybees (Lyko et al., 2010) and suggests that genes with low CpG(o/e) ratios are also preferentially methylated in locusts.

As an initial step towards a more detailed characterization of *S. gregaria* gene methylation patterns we further analyzed the average methylation levels of individual contigs. The results again showed a highly bimodal distribution (Fig. 3B), where approximately 45% of the contigs were completely unmethylated (average methylation ratio <0.05) and approximately 20% were completely methylated (methylation ratio >0.95). An additional 20% of the contigs showed intermediate to high methylation levels (0.65–0.95). Together, these data strongly suggest that the exons of almost half of the *S. gregaria* genes are densely methylated, which represents a major difference to the methylation patterns of other insects.

To validate these findings with an independent method and to further characterize *S. gregaria* gene methylation patterns we analyzed specific regions from six selected contigs by 454 bisulfite

Table 2. DNA methylation and mRNA expression levels for selected gene contigs

Contig	Gene name	Methylation level (%)	Expression level
LC03020A2E07.f1	n.a.	0	0.06
LC01044B2E09.f1	FAM 117B	1	n.d.
LC.4226.C1.Contig4348	Coenzyme q10 homologB	2	1.42
LC.2847.C1.Contig3007	Signal transducing adaptor	4	1.63
LC.3819.C1.Contig3946	n.a.	86	0.85
LC01054B2C09.f2	Pickle homologue 2	92	0.33
LC03023A2G08.f1	Farnesyltransferase	95	0.01

DNA methylation levels were determined by 454 bisulfite sequencing. mRNA expression levels were determined by qRT-PCR and are indicated relative to  $\beta$ -tubulin (=1). n.a., not available; n.d., not determined.

Table 3. Overview of DNA methylation in insects

Species	Dnmt1	Dnmt2	Dnmt3	% mC	Reference
<i>Apis mellifera</i>	Yes	Yes	Yes	0.1	Lyko et al., 2010
<i>Camponotus floridanus</i>	Yes	Yes	Yes	0.1	Bonasio et al., 2012
<i>Harpegnathos saltator</i>	Yes	Yes	Yes	0.1	Bonasio et al., 2012
<i>Bombyx mori</i>	Yes	Yes	No	0.1	Xiang et al., 2010
<i>Drosophila melanogaster</i>	No	Yes	No	n.d.	Zemach et al., 2010
<i>Tribolium castaneum</i>	Yes	Yes	No	n.d.	Zemach et al., 2010
<i>Schistocerca gregaria</i>	Yes	Yes	No <sup>a</sup>	1.3	Present study

Only species with published genome-scale bisulfite sequencing data are included, and the presence or absence of Dnmt1, Dnmt2 and Dnmt3 homologues is indicated. n.d., not detected.

<sup>a</sup>Sequencing of the complete genome will be required to confirm the absence of Dnmt3 in *S. gregaria*.

sequencing. The results showed that out of three arbitrarily selected contigs with low CpG(o/e) ratios, two showed high levels of methylation, while a third appeared unmethylated (Fig. 4). A similar analysis of three arbitrarily selected contigs with high CpG(o/e) ratios showed no methylation for two contigs (Fig. 4). For the 5S rDNA contig, the majority of reads (reflecting multiple copies of the locus) also showed little or no methylation (Fig. 4). These methylation patterns were highly similar between the two tissues analyzed (brain and MTG). Remarkably, similarities also extended to highly specific features, such as the methylation patterns of individual CpG dinucleotides (Fig. 4). Overall, methylation patterns also appeared to be highly homogeneous within individual genes, as most sequence reads were either completely methylated or completely unmethylated. These results provide important confirmation for the data obtained by Illumina sequencing and convincingly demonstrate the methylation of genic sequences in *S. gregaria*.

In total, 454 bisulfite sequencing analysis provided quantitative methylation data for seven contigs that presumably represent single-copy genes (Table 2). For these sequences, brain methylation levels were established and again showed a bimodal distribution, with methylation levels ranging from 0 to 95% (Table 2). In subsequent experiments we also determined the mRNA expression levels for six of these contigs using quantitative RT-PCR. The results showed that expression levels varied greatly among unmethylated and methylated gene contigs (Table 2). These data suggest that gene methylation in *S. gregaria* is not associated with highly or lowly expressed genes and indicate a complex relationship between DNA methylation and gene expression in locusts.

## DISCUSSION

Our study provides detailed insight into the genome methylation pattern of the desert locust, *S. gregaria*, and also defines a novel type of invertebrate methylome. Up until now, all known insect methylomes were characterized by sparse methylation patterns and by a global cytosine methylation level of approximately 0.1% (Table 3). While locust methylation shares CpG specificity with other insect methylomes, the global methylation level appears to be substantially higher (1.3%) and therefore more similar to vertebrate genomes.

The presently available data suggest that the locust DNA methylation system consists of two Dnmt1 paralogs (Boerjan et al., 2011; Robinson et al., 2011). A similar unconventional DNA methylation system has also been described in silkworms (Xiang et al., 2010) and is defined by the lack of a canonical *de novo* methyltransferase (Dnmt3) enzyme. Together, these findings suggest that Dnmt1 enzymes can function as *de novo* and maintenance methyltransferases *in vivo*. It is important to note that the final

confirmation of the *S. gregaria* methylation system will require the availability of a complete genome sequence.

The normalized CpG content [CpG(o/e) ratio] of DNA sequences represents a widely used tool for understanding phylogenetic aspects of DNA methylation in animals (Yi and Goodisman, 2009). Compared with other insect species (Glastad et al., 2011), the expressed genome of *S. gregaria* showed a distinct bimodal pattern with a pronounced peak for genes with low CpG(o/e) ratios. This suggests that the conversion rates of CpGs to TpGs are higher in the germline of *S. gregaria* than in other insects. Possible explanations for this observation include a longer involvement of DNA methylation in the evolution of locusts and a more extensive methylation of coding sequences. Indeed, locusts are generally considered as evolutionarily ancient insects and also appear to have a substantially higher genome methylation level than other insect species examined so far.

It has been assumed that the depletion of DNA methylation in transposons and repetitive sequences represents a conserved feature of invertebrate methylomes (Zemach et al., 2010). This conservation has been explained by the relationship of extant invertebrates to asexually reproducing unicellular organisms that have lost transposon methylation (Zemach and Zilberman, 2010). Our results clearly demonstrate that transposons are methylated in *S. gregaria*. Transposons are highly abundant in locusts (Jiang et al., 2012) and their control might therefore require additional mechanisms, such as DNA methylation. In agreement with this notion, we also found rDNA repeats to be methylated, and rDNA methylation has been shown to mark inactive rDNA copies in vertebrates (Stancheva et al., 1997). Finally, it should also be noted that a recent study in ants has identified sparse methylation patterns in ant transposons that appeared to mark active transposon populations (Bonasio et al., 2012). Further work will be required to establish the functional role of transposon methylation in locusts and other invertebrates.

Consistent with published data from other insect methylomes (Bonasio et al., 2012; Lyko et al., 2010; Xiang et al., 2010), our data suggest a robust enrichment of DNA methylation marks in exons. In honeybees, the differential methylation of exons has been associated with differential gene splicing (Lyko et al., 2010). However, honeybees and other insect methylomes are characterized by sparse methylation and low overall methylation levels of exons (Bonasio et al., 2012; Lyko et al., 2010; Xiang et al., 2010). In locusts, exons are densely methylated with high methylation levels, which may indicate a distinct functional role. Taken together, our data clearly distinguish the *S. gregaria* methylome from other known insect and invertebrate methylomes. Future studies will be required to characterize the biological function of gene methylation in locusts and to determine the role of DNA methylation in locust phase polyphenism.

## ACKNOWLEDGEMENTS

We thank André Leischwitz, Michaela Schanne and Stephan Wolf for Illumina sequencing services, and Tanja Musch for 454 sequencing support. We also thank Arnold De Loof for his critical evaluation and constructive feedback on the project and manuscript.

## AUTHOR CONTRIBUTIONS

F.L. conceived the study; B.B., G.R., M.F., L.S. and F.L. designed research; C.F. and G.R. performed research; C.F., B.B. and G.R. analyzed data; F.L. drafted and revised the article.

## COMPETING INTERESTS

No competing interests declared.

## FUNDING

This work was supported by a grant from the Deutsche Forschungsgemeinschaft to F.L. (FOR1082). B.B. and L.S. would like to acknowledge the FWO (Fonds voor Wetenschappelijk Onderzoek – Vlaanderen) and the Research Foundation of the University of Leuven for their support.

## REFERENCES

- Bestor, T. H. and Ingram, V. M. (1983). Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. *Proc. Natl. Acad. Sci. USA* **80**, 5559-5563.
- Bird, A. (2007). Perceptions of epigenetics. *Nature* **447**, 396-398.
- Boerjan, B., Sas, F., Ernst, U. R., Tobback, J., Lemièrre, F., Vandegehuchte, M. B., Janssen, C. R., Badisco, L., Marchal, E., Verlinden, H. et al. (2011). Locust phase polyphenism: does epigenetic precede endocrine regulation? *Gen. Comp. Endocrinol.* **173**, 120-128.
- Bonasio, R., Li, Q., Lian, J., Mutti, N. S., Jin, L., Zhao, H., Zhang, P., Wen, P., Xiang, H., Ding, Y. et al. (2012). Genome-wide and caste-specific DNA methylomes of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Curr. Biol.* **22**, 1755-1764.
- Burrows, M., Rogers, S. M. and Ott, S. R. (2011). Epigenetic remodelling of brain, body and behaviour during phase change in locusts. *Neural Syst. Circuits* **1**, 11.
- Feinberg, A. P. and Irizarry, R. A. (2010). Evolution in health and medicine Sackler colloquium: stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proc. Natl. Acad. Sci. USA* **107** Suppl. 1, 1757-1764.
- Feng, S., Cokus, S. J., Zhang, X., Chen, P. Y., Bostick, M., Goll, M. G., Hetzel, J., Jain, J., Strauss, S. H., Halpern, M. E. et al. (2010). Conservation and divergence of methylation patterning in plants and animals. *Proc. Natl. Acad. Sci. USA* **107**, 8689-8694.
- Glastad, K. M., Hunt, B. G., Yi, S. V. and Goodisman, M. A. (2011). DNA methylation in insects: on the brink of the epigenomic era. *Insect Mol. Biol.* **20**, 553-565.
- Goll, M. G. and Bestor, T. H. (2005). Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* **74**, 481-514.
- Grönninger, E., Weber, B., Heil, O., Peters, N., Stäb, F., Wenck, H., Korn, B., Winnefeld, M. and Lyko, F. (2010). Aging and chronic sun exposure cause distinct epigenetic changes in human skin. *PLoS Genet.* **6**, e1000971.
- Gruenbaum, Y., Cedar, H. and Razin, A. (1982). Substrate and sequence specificity of a eukaryotic DNA methylase. *Nature* **295**, 620-622.
- Jiang, F., Yang, M., Guo, W., Wang, X. and Kang, L. (2012). Large-scale transcriptome analysis of retroelements in the migratory locust, *Locusta migratoria*. *PLoS ONE* **7**, e40532.
- Jones, P. A. and Liang, G. (2009). Rethinking how DNA methylation patterns are maintained. *Nat. Rev. Genet.* **10**, 805-811.
- Kucharski, R., Maleszka, J., Foret, S. and Maleszka, R. (2008). Nutritional control of reproductive status in honeybees via DNA methylation. *Science* **319**, 1827-1830.
- Lyko, F. and Maleszka, R. (2011). Insects as innovative models for functional studies of DNA methylation. *Trends Genet.* **27**, 127-131.
- 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.
- Mohn, F. and Schübeler, D. (2009). Genetics and epigenetics: stability and plasticity during cellular differentiation. *Trends Genet.* **25**, 129-136.
- Navarro-Martín, L., Viñas, J., Ribas, L., Díaz, N., Gutiérrez, A., Di Croce, L. and Piferrer, F. (2011). DNA methylation of the gonadal aromatase (*cyp19a*) promoter is involved in temperature-dependent sex ratio shifts in the European sea bass. *PLoS Genet.* **7**, e1002447.
- Okano, M., Xie, S. and Li, E. (1998). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat. Genet.* **19**, 219-220.
- Rakyan, V. K., Chong, S., Champ, M. E., Cuthbert, P. C., Morgan, H. D., Luu, K. V. and Whitelaw, E. (2003). Transgenerational inheritance of epigenetic states at the murine *Axin(Fu)* allele occurs after maternal and paternal transmission. *Proc. Natl. Acad. Sci. USA* **100**, 2538-2543.
- Robinson, K. L., Tohidi-Esfahani, D., Lo, N., Simpson, S. J. and Sword, G. A. (2011). Evidence for widespread genomic methylation in the migratory locust, *Locusta migratoria* (Orthoptera: Acrididae). *PLoS ONE* **6**, e28167.
- Rogers, S. M., Matheson, T., Despland, E., Dodgson, T., Burrows, M. and Simpson, S. J. (2003). Mechanosensory-induced behavioural gregarization in the desert locust *Schistocerca gregaria*. *J. Exp. Biol.* **206**, 3991-4002.
- Schulz, M. H., Zerbino, D. R., Vingron, M. and Birney, E. (2012). Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **28**, 1086-1092.
- Shen, J. C., Rideout, W. M., III and Jones, P. A. (1994). The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. *Nucleic Acids Res.* **22**, 972-976.
- Stancheva, I., Lucchini, R., Koller, T. and Sogo, J. M. (1997). Chromatin structure and methylation of rat rRNA genes studied by formaldehyde fixation and psoralen cross-linking. *Nucleic Acids Res.* **25**, 1727-1735.
- Tanaka, S., Harano, K. and Nishide, Y. (2012). Re-examination of the roles of environmental factors in the control of body-color polyphenism in solitary nymphs of the desert locust *Schistocerca gregaria* with special reference to substrate color and humidity. *J. Insect Physiol.* **58**, 89-101.
- Tobi, E. W., Slagboom, P. E., van Dongen, J., Kremer, D., Stein, A. D., Putter, H., Heijmans, B. T. and Lumey, L. H. (2012). Prenatal famine and genetic variation are independently and additively associated with DNA methylation at regulatory loci within IGF2/H19. *PLoS ONE* **7**, e37933.
- Walsh, T. K., Brisson, J. A., Robertson, H. M., Gordon, K., Jaubert-Possamai, S., Tagu, D. and Edwards, O. R. (2010). A functional DNA methylation system in the pea aphid, *Acyrtosiphon pisum*. *Insect Mol. Biol.* **19** Suppl. 2, 215-228.
- Waterland, R. A. and Jirtle, R. L. (2003). Transposable elements: targets for early nutritional effects on epigenetic gene regulation. *Mol. Cell. Biol.* **23**, 5293-5300.
- Xi, Y. and Li, W. (2009). BSMAP: whole genome bisulfite sequence MAPPING program. *BMC Bioinformatics* **10**, 232.
- Xiang, H., Zhu, J., Chen, Q., Dai, F., Li, X., Li, M., Zhang, H., Zhang, G., Li, D., Dong, Y. et al. (2010). Single base-resolution methylome of the silkworm reveals a sparse epigenomic map. *Nat. Biotechnol.* **28**, 516-520.
- Yi, S. V. and Goodisman, M. A. (2009). Computational approaches for understanding the evolution of DNA methylation in animals. *Epigenetics* **4**, 551-556.
- Zemach, A. and Zilberman, D. (2010). Evolution of eukaryotic DNA methylation and the pursuit of safer sex. *Curr. Biol.* **20**, R780-R785.
- Zemach, A., McDaniel, I. E., Silva, P. and Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* **328**, 916-919.