G3: Genes|Genomes|Genetics Early Online, published on January 3, 2018 as doi:10.1534/g3.117.1112

Genome-wide characterisation of DNA methylation in an invasive Lepidopteran pest, the cotton bollworm *Helicoverpa armigera*

Christopher M. Jones[*], Ka S. Lim[†], Jason W. Chapman[‡] & Chris Bass[§]

*Vector Biology, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, L3 5QA, UK

[†]Computational and Analytical Sciences, Rothamsted Research, Harpenden, Hertfordshire, AL5 2JQ, UK

[‡]Centre for Ecology and Conservation, University of Exeter, Penryn, Cornwall, TR10 9EZ, UK

[§]College of Life and Environmental Sciences, Biosciences, University of Exeter, Penryn, Cornwall, TR10 9EZ, UK

Running title: DNA methylation in the cotton bollworm

Keywords: cytosine methylation, epigenetics, phenotypic plasticity, flight activity, gene expression

Corresponding author: Christopher Jones, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, L3 5QA, UK, +44(0)151 705 3100, chris.jones@lstmed.ac.uk

ORCID: 0000-0002-6504-6224

Abstract

The genes and genomes of insect pests are shaped by the wide array of selective forces encountered in their environments. While the molecular adaptations that evolve are beginning to be understood at the genomic and transcriptomic level they have been less well characterised at an epigenetic level. Here, we present a genome-wide map of DNA methylation, at single-nucleotide resolution for the cotton bollworm moth, *Helicoverpa armigera*; a globally invasive pest of agriculture. We show that methylation is almost identical in the larvae and adults of *H. armigera* and that, through whole genome bisulfite sequencing, at the most ~0.9% of CpG sites in this species are methylated. We find that DNA methylation occurs primarily in exons, is positively correlated with gene expression and methylated genes are enriched for cellular housekeeping roles. *H. armigera* has an exceptional capacity for long-range migration. To explore the role of methylation in influencing the migratory phenotype of *H. armigera* we performed targeted bisulfite sequencing on selected loci from sixteen genes that were differentially expressed between adult moths exhibiting distinct flight performance in behavioural assays. While most CpG sites in these genes were not methylated between flight phenotypes we identified hyper-methylation in a demethylase (*KDM4*) that targets lysine-specific histone modifications which are strongly associated with transcription and methylation. The *H. armigera* methylome provides new insights into the role of DNA methylation in a noctuid moth and is a valuable resource for further research into the epigenetic control of adaptive traits in this important pest.

Introduction

DNA methylation is an ancient epigenetic modification that pervades a wide range of organisms. Despite the conserved biochemistry of methylation its function and magnitude are highly variable across taxa with, for example, methylation levels three-orders of magnitude lower in the genomes of insects compared to those from the animal or plant kingdoms (Zemach *et al.* 2010). Furthermore, the catalogue of DNA methyltransferases (DNMTs) found in animals - the enzymes needed to maintain (DNMT1) or catalyse *de novo* methylation (DNMT3) - differ between insect species and are completely absent in some cases (e.g., *Drosophila melanogaster* (Raddatz *et al.* 2013)). Nevertheless, the presence of a functional DNA methylation system across the class Insecta with conserved patterns of methylation (Sarda *et al.* 2012; Hunt *et al.* 2013a; Bewick *et al*. 2016) suggests an important, although poorly understood, role for this epigenetic mark on the biology of insects.

Advances in whole genome sequencing coupled with bisulfite DNA treatment have led to single-nucleotide resolution maps of methylation in a range of invertebrates (Lyko *et al.* 2010; Xiang *et al.* 2010; Wurm *et al.* 2011; Wang *et al.* 2013; Wang *et al.* 2014). These studies have shown that insect methylation is primarily confined to CpG dinucleotides (cytosine followed by guanine), occurs primarily in gene bodies (exons + introns) and that hyper-methylated genes are generally associated with cellular 'housekeeping' roles whereas hypo-methylated genes are more tissue specific (Sarda *et al.* 2012). Experimental measurements of methylation mirror those inferred indirectly from the computation of the observed to expected CpG ratio (CpG O/E) which measures the depletion of CpG dinucleotides (Bird 1980). The CpG O/E is lower in methylated genes due to the mutagenic conversion of methylated cytosine to thymine (deamination) over time leaving a historical imprint of methylation. A bimodal distribution of CpG O/E has been shown in several insects indicating the presence of two classes of 'lowly' and 'highly' methylated genes and used as evidence for active genome-wide methylation.

There is a clear positive correlation between robust intragenic methylation and constitutive gene expression in insects (Xiang *et al.* 2010; Hunt *et al.* 2013b; Libbrecht *et al.* 2016). This relationship is strengthened in the context of nucleosome dynamics with spatial concordance between methylation and an additional epigenetic marker, histone post-translational modifications (hPTMs), that are thought to act in concert to regulate transcriptional activity (Hunt *et al.* 2013a; Glastad *et al.* 2015). Intragenic DNA methylation is therefore thought to regulate active transcription in insects but whether this is restricted to conserved genetic pathways, or can be extended to influence a phenotypic response, is still relatively unknown.

Phenotypic plasticity in morphological and behavioural traits represents a promising role for DNA methylation in insects, yet, evidence for this remains equivocal. In eusocial insects, such as honeybees and ants, evidence that methylation drives caste differentiation (e.g., development into a worker or queen bee) has been provided through whole-genome sequencing and *DNMT* silencing (Kucharski *et al.* 2008; Lyko *et al.* 2010), but this has recently been challenged (Wang *et al.* 2013; Libbrecht *et al.* 2016; Standage *et al.* 2016). Other promising examples of behaviours that may be impacted by methylation are those exhibited in response to shifting or deteriorating environments, such as dispersal or migration. A comparative methylome analysis of the brain tissue from the solitarious and density-dependent gregarious forms of the migratory locust (*Locusta migratoria*) showed that differentially methylated genes were associated mainly with synaptic plasticity (Wang *et al.* 2014). Furthermore, genes differentially expressed between the gregarious and solitarious phases were shown to have signs of CpG depletion (Robinson *et al.* 2011) and genes encoding methylation machinery (*DNMT1*, *DNMT2* and *methyl CpG binding domain protein 2/3*) are differentially transcribed in certain tissues of the two phases (Robinson *et al.* 2016). Finally, beyond insects, differentially methylated regions have been identified between migratory and non-migratory life stages of other organisms, such as the rainbow trout *Oncorhynchus mykiss* (Baerwald *et al.* 2016), suggesting that the

development of migratory forms in response to environmental cues may be linked to variation in methylation patterns.

The Old World bollworm (*Helicoverpa armigera*), is a globally distributed agricultural pest noctuid moth which causes considerable economic damage worldwide (Kriticos *et al.* 2015). More recently, *H. armigera* has invaded the 'New World' with evidence of multiple incursions occurring in South America and subsequent spread over the continent and into Central America (Tay *et al*. 2017). The invasiveness of *H. armigera* is accentuated by adaptive life-history strategies such as extensive polyphagy (Cunningham & Zalucki 2014), resistance to insecticides and Bt toxins (Downes *et al*. 2016), and facultative long-range migratory movements (Farrow and Daly 1987). The recent release of the *H. armigera* genome has shown that gene loss and transcriptional plasticity have facilitated polyphagy in this species (Celorio-Mancera *et al.* 2012; Pearce *et al*. 2017) and similar processes may underlie other traits, including long-distance migration (Jones *et al.* 2015). However, the role of epigenetic processes in regulating the life-history of important Lepidoptera is virtually unknown. Common to the Lepidoptera, *H. armigera* possesses DNMT1 but lacks the *'de novo'* methylase, DNMT3 although it is becoming clear that the association between the presence of specific DNMTs and methylation is not binary, and indeed, DNMT1 may compensate for the lack of DNMT3 in some cases (Bewick *et al*. 2016). A map of methylation levels in this species would complement insights from the recently published genome (Pearce *et al*. 2017).

Here, we present a detailed analysis of the methylome of *H. armigera* through whole-genome bisulfite sequencing (WGBS) and analyse the patterns of methylation in the context of published insect methylomes to date. Previously, we have shown that the flight propensity of *H. armigera* is associated with the differential expression of a suite of candidate genes associated with lipid metabolism, flight muscle function and hormonal control (Jones *et al.* 2015). We therefore extended our analyses using targeted bisulfite sequencing to investigate

potential methylation differences in a subset of these genes between insects demonstrating distinct flight performances in behavioural assays.

Materials and Methods

Mass spectrometry total DNA methylation analysis

Selected reaction monitoring mass spectrometry (SRM MS) was used to quantify global levels of 5-hydroxymethyl-2'-deoxycytidine (5HmdC) and 5-methyl-2'-deoxycytidine (5mdC). The assay measures 5HmdC and 5mdC concentrations as a percentage of 2'-deoxyguanosine (dG). The calibrated ranges for the analytes were 0-1.25% for 5HmdC and 0-12.5% for 5mdC using a fixed 40 pmol amount of dG as an internal standard. MS was performed on genomic DNA (gDNA) extracted from either the heads and thoraxes of four adult moths (2 males and 2 females) or four larvae (L3 life stage). gDNA was extracted using the EZNA Insect DNA Kit (Omega Biotek) and treated with RNaseA (Thermo Life Sciences). Three biological replicates (pools of four insects) per life stage were analysed.

Whole genome bisulfite sequencing

For the methylome analysis genomic DNA (gDNA) was extracted from the heads and thoraxes of four male and four female *H. armigera* and pooled for sequencing. Insects were collected as adults from Bt cotton fields in Qiuxian (Hebei province, China, 36.81°N, 115.16°E) and reared for one generation in the insectaries at Rothamsted Research. Adults were snap-frozen in liquid nitrogen and gDNA extracted using the DNeasy Blood and Tissue Kit (Qiagen). Methyl-MaxiSeq™ (Zymo Research) libraries were prepared from 100 ng of bisulfite treated gDNA (EZ DNA Methylation-Lightning Kit™). Bisulfite-converted DNA was amplified with a primer that contained part of an adaptor sequence plus four random nucleotides followed by two additional amplifications to add on the remaining adaptor sequence and to barcode the fragments. PCR products were purified using the DNA Clean & Concentrator-5™ (Zymo

Research). Sequencing was run on the Illumina HiSeq 2500 platform. Sequence reads were aligned to the *H. armigera* genome using the bisulfite sequencing aligner software Bismark (Krueger and Andrews 2011).

CpG methylation and gene methylation analysis

The methylation status of each cytosine was determined using a binomial distribution to compare methylated and non-methylated reads at each site possessing a minimum of two reads (Lyko *et al.* 2010). Methylated sites were determined at $p < 0.05$ after adjustment for multiple testing (Benjamini and Hochberg 1995). Methylation ratios (mCpG/CpG) were determined per gene and genomic function (exon, intron and 2kb upstream). Methylated genes or genomic functions were defined as those possessing a methylation ratio of over 10%. CpG depletion (CpG O/E) was calculated according to Bird *et al.* (1980). Hartigan's diptest was used to determine the modality of the distribution of methylation levels and CpG O/E using the diptest package in R software (Hartigan and Hartigan 1985). An enrichment analysis (Fisher's exact test) of GO terms for highly methylated genes (>50% methylation ratio) and genes containing zero methylated sites was performed against the reference gene set in Blast2GO at an FDR < 0.05 (Gotz *et al.* 2008).

The relationship between methylation and gene expression was explored using the RNA-seq dataset from Jones *et al.* (2015). Genomewide expression data were acquired from a population collected in Anyang (Henan province, 36.10°N, 114.20°E). Anyang is approximately 100 km from Qiuxian and insects were collected at a similar time of year (August 2013) and therefore it is expected that population differences between two groups of insects were minimal. The relationship between methylation and expression was explored using the methylation ratio and TMM-normalised FPKM values (fragments per kilobase of exon per million fragments mapped).

Flight mills and targeted bisulfite sequencing

To validate the genomewide bisulfite data, and to determine the strength of any association between DNA methylation and flight activity, targeted bisulfite sequencing was performed on adult moths flown on tethered flight mills. Female moths originating from northern Greece were flown on the tethered flight mills following the procedures outlined in (Jones *et al.* 2016). Insects were flown overnight and flight data collected between the hours of 1900 and 0915 (dark period 2000 to 0600). Individuals were snap-frozen in liquid nitrogen 1-2 h following the flight period for DNA extraction. Following an analysis of the flight behaviour a total of sixteen individuals (all female) representing two distinct groups of short- and long-distance fliers (eight in each phenotype) were chosen for DNA extraction and targeted bisulfite sequencing.

A selection of loci spanning sixteen candidate genes, capturing a range of methylation (exon methylation ratio 0.025-1) were chosen for the detection of CpG sites. Primers were designed with parameters that preferentially targeted regions between 100-300 bp and avoided annealing to CpGs. Details of selected gene regions and primer design are available in Table S1. Genomic DNA from sixteen individual moths was extracted from the head and thorax using the EZNA Insect DNA Kit (Omega Biotek) as described above. Samples were bisulfite converted using the EZ DNA Methylation-Lightning kit (Zymo Research) and purified (ZR-96 DNA Clean & Concentrator, Zymo Research). Bisulfite treated DNA (5 ng) was amplified, the amplicons pooled for barcoding and sequenced using a MiSeq V2 300bp Reagent Kit (Illumina).

Low quality reads and adapter sequences were trimmed and the sequencing reads realigned to the *H. armigera* genome using Bismark (Krueger and Andrews 2011). Nucleotides in primers were trimmed in the methylation calling and the methylation level quantified as the number of reads reporting a cytosine divided by the total number of reads at that site. Only CpG sites detected in at least one sample with at least 10 reads were considered for analysis. The fractional methylation ratio was calculated as the number of methylated cytosines over number

of cytosines per site (mCpG/CpG). Mean differences between the two groups of individual moths displaying contrasting performances on the flight mills (N = 8) were estimated using a Student's t-test.

Data Availability

The raw bisulfite sequencing data used to analyse the methylome is available at ArrayExpress (accession number E-MTAB-4779). Supplemental Table S1 provides information on the primers used to amplify selected loci for targeted bisulfite sequencing. Table S2 describes the enriched GO-terms in highly methylated genes and Table S3 lists enriched GO-terms in genes with no detectable methylation. Table S4 shows the top 25 differentially expressed genes associated with flight activity per methylation level. Table S5 details the selected loci for targeted bisulfite sequencing. Table S6 provides all the raw CpG data per individual site, single gene and the total exonic and intronic methylation ratio per gene. File S1 contains the Supplemental Figures S1-S4.

Results and Discussion

MS detection of global CpG methylation levels in H. armigera

The total level of methylcytosine (5mdC) and hydroxymethylcytosine (H5mdC) was measured using MS in adults and larvae. Despite missing DNMT3, methylation is observed in *H. armigera* and the percentage of 5mdC was almost identical in the two life stages (adults, 0.165% ± 0.009; L3 stage larvae, 0.164% ± 0.009) whereas H5mdC was undetectable in *H. armigera*. This suggests that DNA methylation is stable across life-stages of *H. armigera* and, in contrast to recent findings in the honeybee (Wojciechowski *et al.* 2014), there is no evidence for additional epigenetic regulation via hydroxymethylation in this species.

Whole genome bisulfite sequencing of DNA methylation in H. armigera

Sequencing of bisulfite converted gDNA from the heads and thoraxes of eight adult moths of *H. armigera* (four females and four males) yielded 529 million reads, of which, 28% mapped to the genome. The overall bisulfite conversion rate was high (>99%). Methylation in insects is almost exclusively at CpG dinucleotides rather than CHG or CHH sites (H = A, C or T) (Lyko and Maleszka 2011) so we focussed on methylation at CpG sites only. Of the estimated 19.7 million CpG sites in the *H. armigera* genome 73.5% were identified by sequencing (N = 14.5 million) with an average coverage of 28X.

The number of methylated CpGs (mCpG) detected (probability of methylated cytosine according to a binomial distribution, $p < 0.05$, minimum 2 reads per site) was 169,911 which represents 0.86% of all CpGs in the genome and 1.17% of those identified from bisulfite sequencing. Using a stricter threshold of 10 reads per site 0.43% of all cytosines detected were methylated and 0.34% of all cytosines in the genome. The distribution of mCpGs are presented for both thresholds in Figure S1. Comparisons with other genomewide bisulfite data require some caution due to differences in mCpG detection methodology, however, the absolute number of mCpGs detected in this study is similar to that predicted in *B. mori* (169,911 vs. 172,117) (Xiang *et al.* 2010). Based on the estimated number of genomic CpGs, however, the relative level of methylation is much greater in *H. armigera* (0.86% vs. 0.11%). Of the estimated 17,086 genes from the recently annotated *H. armigera* genome (Pearce *et al*. 2017) approximately 69.6% have at least one mCpG site.

Comparison of the level of CpG methylation in exons, introns and the 2kb region upstream of the gene transcription start site (putative promoter region) (Figure S2) revealed that exonic CpGs are more highly methylated (3.06%) than introns (0.57%) or the 2kb upsteam (1.78%) in line with previous findings that DNA methylation is primarily confined to exons in insects (Lyko *et al.* 2010; Xiang *et al.* 2010; Wang *et al.* 2013; Beeler *et al.* 2014; Bewick *et al*. 2016). The mean exon methylation ratio (calculated as the proportion of methylated cytosines

234 determined by the binomial distribution) are also much higher than in introns and 2kb upstream
235 regions (exon mean 0.053, intron 0.017, 2kb upstream 0.023). The distribution of the exon
236 methylation ratio follows a bimodal distribution with two overlapping clusters of lowly and
237 highly methylated genes similar to the patterns of methylation reported for *B. mori* and *A.*
238 *mellifera* (Sarda *et al.* 2012) (Figure 1A). There is a small bimodal pattern in regions 2kb
239 upstream of the TSS but this is probably insignificant given that methylation levels before the
240 TSS are generally low in other Lepidopterans (Xiang *et al*. 2010) and could be due to
241 inaccuracies in the annotation of intragenic regions of the genome (Figure 1B). Intronic
242 methylation ratios are by contrast unimodal in line with patterns in other insects (Figure 1 C).
243 The bimodal pattern of gene body methylation is a common feature between distantly related
244 invertebrates with functional methylation systems (Sarda *et al.* 2012). These results also
245 confirm that functional DNA methylation occurs in Lepidoptera despite the loss of DNMT3
246 from this order approximately 177.99-116.45 *Mya* and that either DNMT1 may compensate
247 for this loss or *de novo* methylation occurs through some other non-DNMT like protein
248 (Bewick *et al*. 2016).

249 There is negative correlation between exon methylation and the CpG O/E ratio (Figure 2A)
250 reflecting the propensity for methylated cytosines to be converted to thymines over time (Bird
251 1980). In contrast to the bimodal distribution of exonic methylation in *H. armigera* described
252 above, and the CpG O/E ratio in other insects (Lyko *et al.* 2010; Walsh *et al.* 2010; Falckenhayn
253 *et al.* 2013; Wang *et al.* 2014), we observe a single CpG O/E peak (Figure 2B; mean CpG O/E
254 per gene 0.991). This is consistent with available CpG O/E distributions from other Lepidoptera
255 (*B. mori* and *Danaus plexippus*) and the red flour beetle (*Tribolium castaneum*) (Xiang *et al.*
256 2010; Zhan *et al.* 2011). The common unimodel CpG O/E distribution in these species could
257 be due to reduced CpG depletion over evolutionary time, potentially a result of the loss of the
258 *de novo* methylation enzyme DNMT3 (Bewick *et al*. 2016). Nevertheless, when we classified

genes as methylated or non-methylated according to the level of exon methylation ratio (±10%), there is a clear segregation into low CpG O/E (methylated) and high CpG O/E (non-methylated) (Figure 2C) with significant differences between the mean CpG O/E of methylated (0.738) and non-methylated genes (1.042) ($F_{2944,14055} = 1.305$, $P < 0.0001$).

DNA methylation and gene expression in Helicoverpa armigera

The relationship between DNA methylation and gene expression was investigated using an RNA-seq dataset from adult *H. armigera* collected from a nearby population in China (Jones *et al.* 2015). There was a largely positive, although non-linear, relationship between intragenic methylation and expression (Spearman's rank, $\rho = 0.397$, $P < 0.0001$) (Figure 3A). The median expression of methylated genes was significantly greater than that of those non-methylated (±10%) (Wilcoxon Signed-Rank test, $P < 0.0001$) (Figure 3B). It was also notable that of the 1462 genes not expressed (FPKM = 0) 81.0% had zero exonic methylation; a large increase from the percentage of genes that have no detectable exonic methylation throughout the genome (43.1%).

These results demonstrate that DNA methylation is tightly associated with stably expressed genes in *H. armigera* and the function of methylation is likely to mirror that in other highly diverged insect orders (e.g. Hymenoptera and Orthoptera) (Lyko *et al.* 2010; Flores *et al.* 2012; Hunt *et al.* 2013b; Wang *et al.* 2013). Given the observation that methylation is spatially correlated with histone modifications (Glastad *et al.* 2015) future studies exploring the regulation of gene expression in *H. armigera* (and other Lepidoptera) via DNA methylation should be investigated in the context of chromatin organisation and the wider epigenetic landscape.

Functional enrichment of methylated genes in H. armigera

A functional enrichment analysis of those genes exhibiting high exon methylation ratios (>50%) showed that these genes are related to basic housekeeping roles such as ribosome structure, translation and gene expression (Table S2). Conversely, genes lacking any mCpGs were enriched for specialised functions such as cell signalling (G-protein coupled receptors), detoxification, olfaction and the insect cuticle (Table S3). This finding provides additional weight to the hypothesis that an important function of methylation in a diverse array of insects, including Lepidoptera, is the regulation of general cellular processes in ubiquitous, evolutionary conserved and stably expressed genes (Elango *et al.* 2009; Hunt *et al.* 2010; Xiang *et al.* 2010; Wurm *et al.* 2011; Sarda *et al.* 2012).

Validation of methylation in selected loci via targeted bisulfite sequencing

To validate the whole genome methylation data primers were designed to targeted selected loci in sixteen genes (see below for details). Excellent coverage was obtained with 305,680 to 507,593 reads per sample, an average CpG coverage ranging from 208X to 949X and a bisulfite conversion rate of >99%. Following quality control, a total of 322 CpG sites were detected above the required threshold (> 10 reads per site) in either exonic or 5'UTR regions (94 sites were detected in all sixteen samples). A comparison of methylation levels from the whole genome versus the average targeted bisulfite sequencing across all samples showed a strong positive relationship (R^2=0.78, P < 0.0001; Figure 4). The fact that the two methylation detection methods were strongly correlated despite that fact they were performed on different adult *H. armigera* samples suggests that the methylation status of most individual CpG sites is relatively stable across different individuals of this species.

Methylation of selected genes associated with flight behaviour

A whole-genome transcriptional analysis previously showed that the flight activity of *H. armigera* is associated with the differential expression of a suite of genes encompassing a range

306 of biological functions including fatty acid/ketone metabolism, flight muscle structure and ATP
307 synthesis/respiration (Jones *et al.* 2015). The mean exon methylation of these candidate genes
308 (n = 191) is 0.040 (range 0-0.684) with an CpG O/E of 0.95; indicating similar albeit slightly
309 lower levels compared to genome-wide methylation. A list of the twenty-five candidates with
310 exon methylation ratios over 10% is provided in Table S4 with the highest levels present in the
311 motor protein *dynein light chain roadblock-type 2* (HaOG207620), *NADH dehydrogenase*
312 (HaOG208245), the lysine-specific demethylase *KDM4* (HaOG212852) and an orthologue of
313 the *Drosophila* hypoxia-related gene, *tnz CG4365* (HaOG210853). Selected loci from sixteen
314 candidates were chosen to validate the whole genome analysis (Table S5).

315 To examine whether these genes also show signs of differential methylation in the context of
316 flight behaviour a flight mill experiment was performed on *H. armigera* collected from
317 northern Greece. Female moths showed continuous variation in flight performance with a mean
318 total distance flown during a single night of 13,619 m. Flight mill data collected from multiple
319 noctuid moth species (*H. armigera*, *Spodoptera frugiperda* and *Spodoptera exempta*) indicate
320 that insects that fly for longer distances, in general, engage in fewer flights (A. Pearson & C.M.
321 Jones unpublished data). Using this approach, we discriminated between long-distance (N = 8,
322 mean distance = 21,586 m, mean number of flights = 7.5) and short-distance fliers (N = 8, mean
323 distance = 5246 m, mean number of flights = 44.25) for comparison of methylation levels in
324 the targeted gene set (Figure S3).

325 For the majority of loci we observed few differences in the methylation levels between short-
326 and long-distance fliers with high concordance between the flight groups ($R^2 = 0.84$, $P <$
327 0.0001) (Figure S4). For example, in the ketone metabolism gene, succinyl-CoA:3-ketoacid
328 coenzyme A transferase 1 (*OXCT*), the fractional methylation ratios per CpG site are almost
329 identical across four exons (Figure 5A). This suggests that the transcriptional activity of many
330 genes associated with flight performance in *H. armigera* is not influenced by DNA methylation

(although in this preliminary study we have only looked at a comparatively small subset of previously identified candidate genes). However, there were two examples of genes (comprising a total of eight CpG sites) where methylation levels were significantly different between the flight phenotypes (Table 1).

The top three hyper-methylated sites in the long-distance fliers - with fractional methylation ratios 0.231-0.323 greater compared to short-distance fliers - were all present in *KDM4* (Table 1). In accordance with relatively high exonic methylation (~50%) and a low CpG O/E value (0.58) a large percentage of CpG sites in *KDM4* were methylated (Figure 5B). *KDM4* encodes a demethylase that removes di- and tri-methyl groups from lysines 9 and 36 in histone H3 (H3K9 and H3K36) (Klose *et al.* 2006) and therefore plays a role in reversing histone methylation which itself is associated with transcriptional activity. The co-localisation of DNA methylation and histone post-translational modifications (e.g. H3K9me3 and H3K36me3) are strongly associated with stably expressed genes (Hunt *et al.* 2013b; Glastad *et al.* 2015). For example, Glastad *et al.* (2015) show that over 90% of methylated genes also feature H3K4me3 or H3K36me3. The consequences of hyper-methylation in the *KDM4* gene itself in the context of an energetic activity such as flight that requires a strong transcriptional response is unknown. It has been shown that the loss of *KDM4* in *Drosophila* impedes the transcriptional activation of ecdysone signalling (Tsurumi *et al.* 2013); a pathway with increasingly recognised importance in adult insect behaviour (Schwedes and Carney 2012).

A functional enrichment analysis has previously shown that genes associated with the inosine monophosphate (IMP) biosynthesis pathway and purine/ATP metabolism were enriched in over-expressed genes associated with increased flight activity (Jones *et al.* 2015). Genes with these GO-terms are not, however, highly methylated (mean exonic methylation 0.035) except for *PFAS*, an enzyme that encodes phosphoribosylformylglycinamidine synthase. This gene contained the only other strongly differentially methylated site between the flight phenotypes

(Table 1; Figure 5C). This enzyme catalyses part of the pathway involved in ionosine purine biosynthesis and ATP turnover but whether the expression of this pathway induced by the demands of a highly energetic activity such as migratory flight requires mediation via a hyper-methylated site requires further investigation.

While DNA methylation in the exonic regions of insect genomes is associated with transcription this methylation largely occurs in genes with basic regulatory functions and generally not in those genes that are differentially expressed between phenotypes (Hunt *et al*. 2013a; Libbrecht *et al*. 2016; Sarda *et al*. 2012). Indeed, the function of DNA methylation in the context of expression in insects is still largely unknown and is likely to require further study using all components of the epigenome (Glasted *et al*. 2016). In this context, it is unlikely that differential methylation will contribute largely to the contrasting flight capacities exhibited by *H. armigera* in this study. Nevertheless, the differentially methylated sites described above do represent viable targets to determine the functional significance of methylation on expression and/or flight behaviour. At the single-base resolution the induction of methylation *in vivo* via the CRISPR/Cas9-based system (McDonald *et al.* 2016) represents a promising future application to determine the role of differentially methylated sites in insects. At the genomewide scale, chemical disruption of methylation via a demethylating agent has been shown to lead to subtle changes in sex allocation in the parasitic wasp *Nasonia vitripennis* (Cook *et al.* 2015). Migration is a complex syndrome consisting of a combination of several morphological, behavioural and physiological traits (Liedvogel *et al.* 2011; Chapman *et al.* 2015). It seems plausible, therefore, that the disruption of DNA methylation in migratory insects containing a functional methylation system, including *H. armigera*, could also result in subtle but significant changes in one of the many biochemical pathways that contribute to this behaviour. The knockdown of methyltransferases via CRISPR or RNA interference (RNAi; e.g. [14]) also represents a potential experimental tool.

Conclusion

The description of the single base-resolution methylome of *H. armigera* presented here provide an insight into genome-wide DNA methylation in a noctuid moth. Our findings reveal that, as reported for other insects, methylation is sparse in this species, with close to ~1% of CpG sites identified as methylated in sharp contrast to the 60-90% methylation levels observed in mammals. Methylation in *H. armigera* is predominantly exonic and significantly enriched in genes involved in basal cellular housekeeping roles. The degree of genic methylation in this species is positively correlated with gene expression, although the relationship is not linear with methylated genes exhibiting higher median expression levels than non-methylated genes, consistent with the results of other insect species. Recent studies have provided some initial evidence of a relationship between methylation and life history divergence associated with long-distance migration (Wang *et al.* 2014; Baerwald *et al.* 2016). Our preliminary exploration of the role of this epigenetic mark in the regulation of the expression of candidate genes associated with this trait in *H. armigera* suggests the transcription of only a minor subset of genes may be influenced by methylation. These genes, however, represent promising candidates for further characterisation in the context of methylation and other epigenetic marks, such as histone modifications. Finally, we envisage that the *H. armigera* methylome will be a valuable resource for further research into the epigenetic control of adaptive traits in this important insect pest (e.g. resistance to Bt toxins and insecticides (Downes *et al.* 2016)), especially now the full genome is available (Pearce *et al.* 2017).

Acknowledgements

We would like to thank the *Helicoverpa* Genome Consortium for providing access to the genome prior to its publication and George Mironidis (University of Thessaloniki) for insect samples. Rothamsted Research is a national institute of bioscience strategically funded by the

405 UK Biotechnology and Biological Sciences Research Council (BBSRC). This work was
406 supported by a BBSRC Future Leader Fellowship (grant number: BB/N012011/1) to CMJ.

References

Baerwald MR, Meek MH, Stephens MR, Nagarajan RP, Goodbla AM, Tomalty KM, Thorgaard GH, May B, Nichols KM. 2016. Migration-related phenotypic divergence is associated with epigenetic modifications in rainbow trout. *Mol Ecol* **25**(8): 1785-1800.

Beeler SM, Wong GT, Zheng JM, Bush EC, Remnant EJ, Oldroyd BP, Drewell RA. 2014. Whole-Genome DNA Methylation Profile of the Jewel Wasp (Nasonia vitripennis). *G3-Genes Genomes Genetics* **4**(3): 383-388.

Bewick AJ, Vogel KJ, Moore AJ, Schmitz RJ. 2016. Evolution of DNA methylation across Insects. *Molecular Biology and Evolution* **34**(3):654-665.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate – a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological* **57**(1): 289-300.

Bird AP. 1980. DNA methylation and the frequency of CpG in Animal DNA. *Nucleic Acids Research* **8**(7): 1499-1504.

Celorio-Mancera MdlP, Heckel DG, Vogel H. 2012. Transcriptional analysis of physiological pathways in a generalist herbivore: responses to different host plants and plant structures by the cotton bollworm, Helicoverpa armigera. *Entomologia Experimentalis Et Applicata* **144**(1): 123-133.

Chapman JW, Reynolds DR, Wilson K. 2015. Long-range seasonal migration in insects: mechanisms, evolutionary drivers and ecological consequences. *Ecology Letters* **18**(3): 287-302.

Cook N, Pannebakker BA, Tauber E, Shuker DM. 2015. DNA Methylation and Sex Allocation in the Parasitoid Wasp Nasonia vitripennis. *American Naturalist* **186**(4): 513-518.

Cunningham JP, Zalucki MP. 2014. Understanding Heliothine (Lepidoptera: Heliothinae) Pests: What is a host plant? *Journal of Economic Entomology* **107**: 881-896.

Downes S, Walsh T, Tay WT. 2016. Bt resistance in Australian insect pest species. *Current Opinion in Insect Science* **15**: 78-83.

Elango N, Hunt BG, Goodisman MAD, Yi SV. 2009. DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, Apis mellifera. *Proceedings of the National Academy of Sciences of the United States of America* **106**(27): 11206-11211.

Falckenhayn C, Boerjan B, Raddatz G, Frohme M, Schoofs L, Lyko F. 2013. Characterization of genome methylation patterns in the desert locust Schistocerca gregaria. *Journal of Experimental Biology* **216**(8): 1423-1429.

Farrow RA, Daly JC. 1987. Long range movements as an adaptive strategy in the genus *Heliothis* (Lepidoptera, Noctuidae) – a review of its occurrence and detection in four pest species. *Australian Journal of Zoology* **35**(1): 1-24.

Flores K, Wolschin F, Corneveaux JJ, Allen AN, Huentelman MJ, Amdam GV. 2012. Genome-wide association between DNA methylation and alternative splicing in an invertebrate. *Bmc Genomics* **13**.

Glastad KM, Hunt BG, Goodisman MAD. 2015. DNA Methylation and Chromatin Organization in Insects: Insights from the Ant Camponotus floridanus. *Genome Biology and Evolution* **7**(4): 931-942.

Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research* **36**(10): 3420-3435.

Hartigan JA, Hartigan PM. 1985. The dip test of unimodality. *Annals of Statistics* **13**(1): 70-84.

Hunt BG, Brisson JA, Yi SV, Goodisman MAD. 2010. Functional Conservation of DNA Methylation in the Pea Aphid and the Honeybee. *Genome Biology and Evolution* **2**: 719-728.

Hunt BG, Glastad KM, Yi SV, Goodisman MAD. 2013a. The Function of Intragenic DNA Methylation: Insights from Insect Epigenomes. *Integrative and Comparative Biology* **53**(2): 319-328.

Hunt BG, Glastad KM, Yi SV, Goodisman MAD. 2013b. Patterning and Regulatory Associations of DNA Methylation Are Mirrored by Histone Modifications in Insects. *Genome Biology and Evolution* **5**(3): 591-598.

Jones CM, Papanicolaou A, Mironidis GK, Vontas J, Yang Y, Lim KS, Oakeshott JG, Bass C, Chapman JW. 2015. Genomewide transcriptional signatures of migratory flight activity in a globally invasive insect pest. *Molecular Ecology* **24**(19): 4901-4911.

Jones HB, Lim KS, Bell JR, Hill JK, Chapman JW. 2016. Quantifying interspecific variation in dispersal ability of noctuid moths using an advanced tethered flight technique. *Ecology and evolution* **6**(1): 181-190.

Klose RJ, Yamane K, Bae Y, Zhang D, Erdjument-Bromage H, Tempst P, Wong J, Zhang Y. 2006. The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36. *Nature* **442**(7100): 312-316.

Kriticos DJ, Ota N, Hutchison WD, Beddow J, Walsh T, Tay WT, Borchert DM, Paula-Moreas SV, Czepak C, Zalucki MP. 2015. The Potential Distribution of Invading Helicoverpa armigera in North America: Is It Just a Matter of Time? *Plos One* **10**(3).

Krueger F, Andrews SR. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**(11): 1571-1572.

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

Libbrecht R, Oxley PR, Keller L, Kronauer DJC. 2016. Robust DNA Methylation in the Clonal Raider Ant Brain. *Current Biology* **26**(3): 391-395.

Liedvogel M, Akesson S, Bensch S. 2011. The genetics of migration on the move. *Trends in Ecology & Evolution* **26**(11): 561-569.

Lyko F, Foret S, Kucharski R, Wolf S, Falckenhayn C, Maleszka R. 2010. The Honey Bee Epigenomes: Differential Methylation of Brain DNA in Queens and Workers. *Plos Biology* **8**(11).

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

McDonald JI, Celik H, Rois LE, Fishberger G, Fowler T, Rees R, Kramer A, Martens A, Edwards JR, Challen GA. 2016. Reprogrammable CRISPR/Cas9-based system for inducing site-specific DNA methylation. *Biology Open*. 5: 866-874; doi: 10.1242/bio.019067

Pearce S, Clarke D, East P, Elefekih S, Gordan K *et al*., 2017 Genomic basis for the pest status of two Helicoverpa species. *BMC Biology in press*.

Raddatz G, Guzzardo PM, Olova N, Fantappie MR, Rampp M, Schaefer M, Reik W, Hannon GJ, Lyko F. 2013. Dnmt2-dependent methylomes lack defined DNA methylation patterns. *Proceedings of the National Academy of Sciences of the United States of America* **110**(21): 8627-8631.

Robinson KL, Tohidi-Esfahani D, Lo N, Simpson SJ, Sword GA. 2011. Evidence for Widespread Genomic Methylation in the Migratory Locust, Locusta migratoria (Orthoptera: Acrididae). *Plos One* **6**(12).

Robinson KL, Tohidi-Esfahani D, Ponton F, Simpson SJ, Sword GA, Lo N. 2016. Alternative migratory locust phenotypes are associated with differences in the expression of genes encoding the methylation machinery. *Insect molecular biology* **25**(2): 105-115.

Sarda S, Zeng J, Hunt BG, Yi SV. 2012. The Evolution of Invertebrate Gene Body Methylation. *Molecular Biology and Evolution* **29**(8): 1907-1916.

Schwedes CC, Carney GE. 2012. Ecdysone signaling in adult Drosophila melanogaster. *Journal of Insect Physiology* **58**(3): 293-302.

Standage DS, Berens AJ, Glastad KM, Severin AJ, Brendel VP, Toth AL. 2016. Genome, transcriptome and methylome sequencing of a primitively eusocial wasp reveal a greatly reduced DNA methylation system in a social insect. *Molecular Ecology* **25**(8): 1769-1784.

Tay WT, Walsh TK, Downes S, Anderson C, Jermiin LS, Wong TKF, Piper MC, Chang ES, Macedo IB, Czepak C, Behere GT, Silvie P, Soria MF, Frayssinet M, Gordon KHJ. 2017. Mitochondrial DNA and trade data support multiple origins of *Helicoverpa armigera* (Lepidoptera, Noctuidae) in Brazil. *Scientific Reports* **7** 45302; doi: 10.1038/srep45302Tsurumi A, Dutta P, Yan S-J, Sheng R, Li WX. 2013. Drosophila Kdm4 demethylases in histone H3 lysine 9 demethylation and ecdysteroid signaling. *Scientific Reports* **3**.

Walsh TK, Brisson JA, Robertson HM, Gordon K, Jaubert-Possamai S, Tagu D, Edwards OR. 2010. A functional DNA methylation system in the pea aphid, Acyrthosiphon pisum. *Insect Molecular Biology* **19**: 215-228.

Wang X, Wheeler D, Avery A, Rago A, Choi J-H, Colbourne JK, Clark AG, Werren JH. 2013. Function and Evolution of DNA Methylation in Nasonia vitripennis. *PLoS genetics* **9**(10).

Wang XH, Fang XD, Yang PC, Jiang XT, Jiang F, Zhao DJ, Li BL, Cui F, Wei JN, Ma CA *et al.* 2014. The locust genome provides insight into swarm formation and long-distance flight. *Nature Communications* **5**: 1-9.

Wojciechowski M, Rafalski D, Kucharski R, Misztal K, Maleszka J, Bochtler M, Maleszka R. 2014. Insights into DNA hydroxymethylation in the honeybee from in-depth analyses of TET dioxygenase. *Open Biology* **4**(8).

Wurm Y, Wang J, Riba-Grognuz O, Corona M, Nygaard S, Hunt BG, Ingram KK, Falquet L, Nipitwattanaphon M, Gotzek D *et al.* 2011. The genome of the fire ant Solenopsis invicta. *Proceedings of the National Academy of Sciences of the United States of America* **108**(14): 5679-5684.

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. *Nature Biotechnology* **28**(5): 516-U181.

Zemach A, McDaniel IE, Silva P, Zilberman D. 2010. Genome-Wide Evolutionary Analysis of Eukaryotic DNA Methylation. *Science* **328**(5980): 916-919.

Zhan S, Merlin C, Boore JL, Reppert SM. 2011. The Monarch Butterfly Genome Yields Insights into Long-Distance Migration. *Cell* **147**(5): 1171-1185.

Figures

Figure 1. The distribution of the methylation ratios per genomic function. Distribution of methylation in A) exons, B) introns and C) 2kb from transcriptional start site.

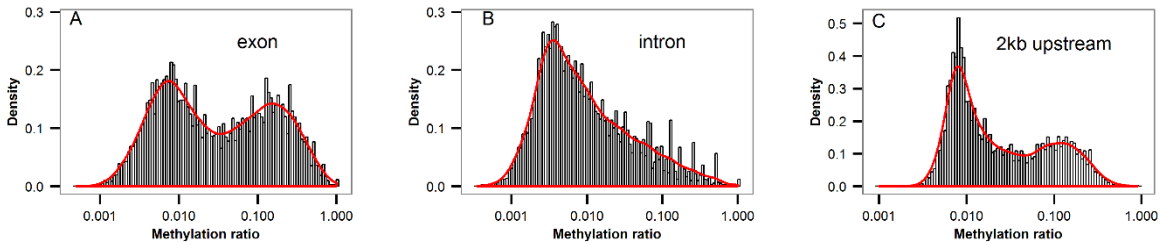

Figure 2. Patterns of methylation inferred from the CpG O/E statistic. A) Correlation between the CpG O/E and experimentally deduced methylation ratio. B) Unimodal distribution of CpG O/E. C) Distribution of methylated (red) and non-methylated (blue) genes (±10% exon methylation ratio) per CpG O/E statistic.

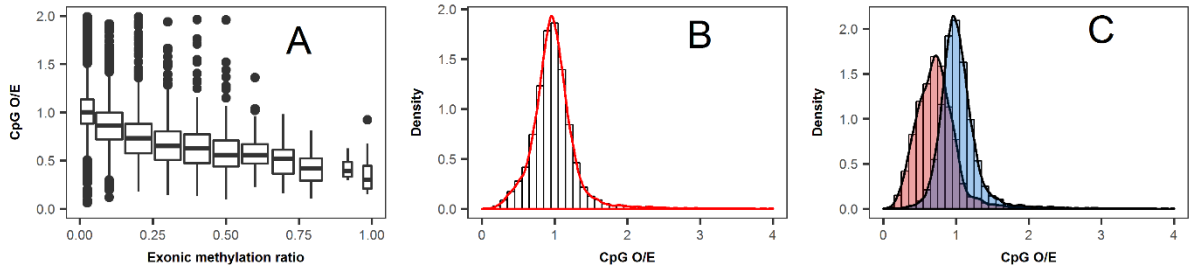

Figure 3. The relationship between methylation and gene expression in *H. armigera*. A) Distribution of RNA-seq (TMM-FPKM $\log_{10}$) expression stratified by exon methylation bins (10%). B) Density plots for expression data for individual genes per methylated (red) or non-methylated (blue) status based on ±10% exonic methylation.

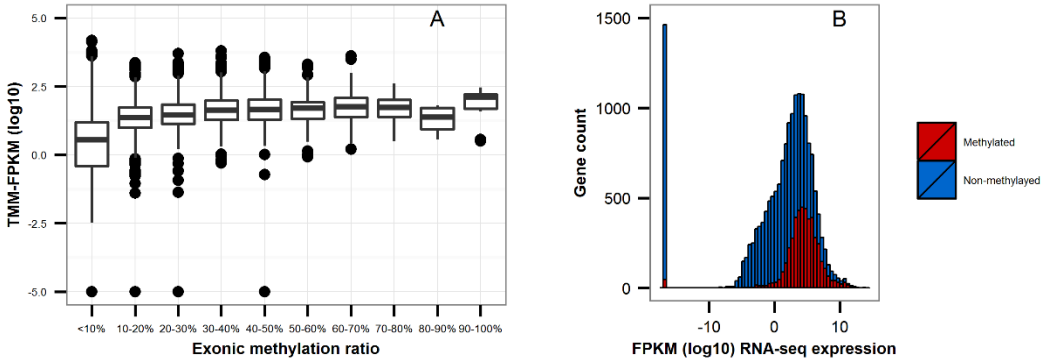

Figure 4. Correlation between methylation in selected loci analysed by whole-genome and targeted sequencing.

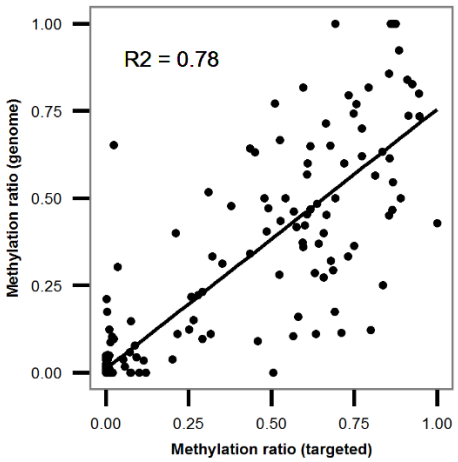

Figure 5. Methylation of selected loci in three genes as detected using targeted bisulfite sequencing. The bar graphs present the average methylation ratio at each CpG site detected for A) succinyl-CoA:3-ketoacid coenzyme A transferase 1 (*OXCT1*) B) lysine-specific histone demethylase (*KDM4*) and C) phosphoribosylformylglycinamidine synthase (*PFAS*). Methylation ratios were calculated as the average for each flight phenotype. The three CpG sites in *KDM4* and one site in *PFAS* that were significantly hyper-methylated ($p < 0.05$) in long-distance flying insects are shown with * in the top panel. Dashed lines represent exon-exon boundaries for *OXCT*.

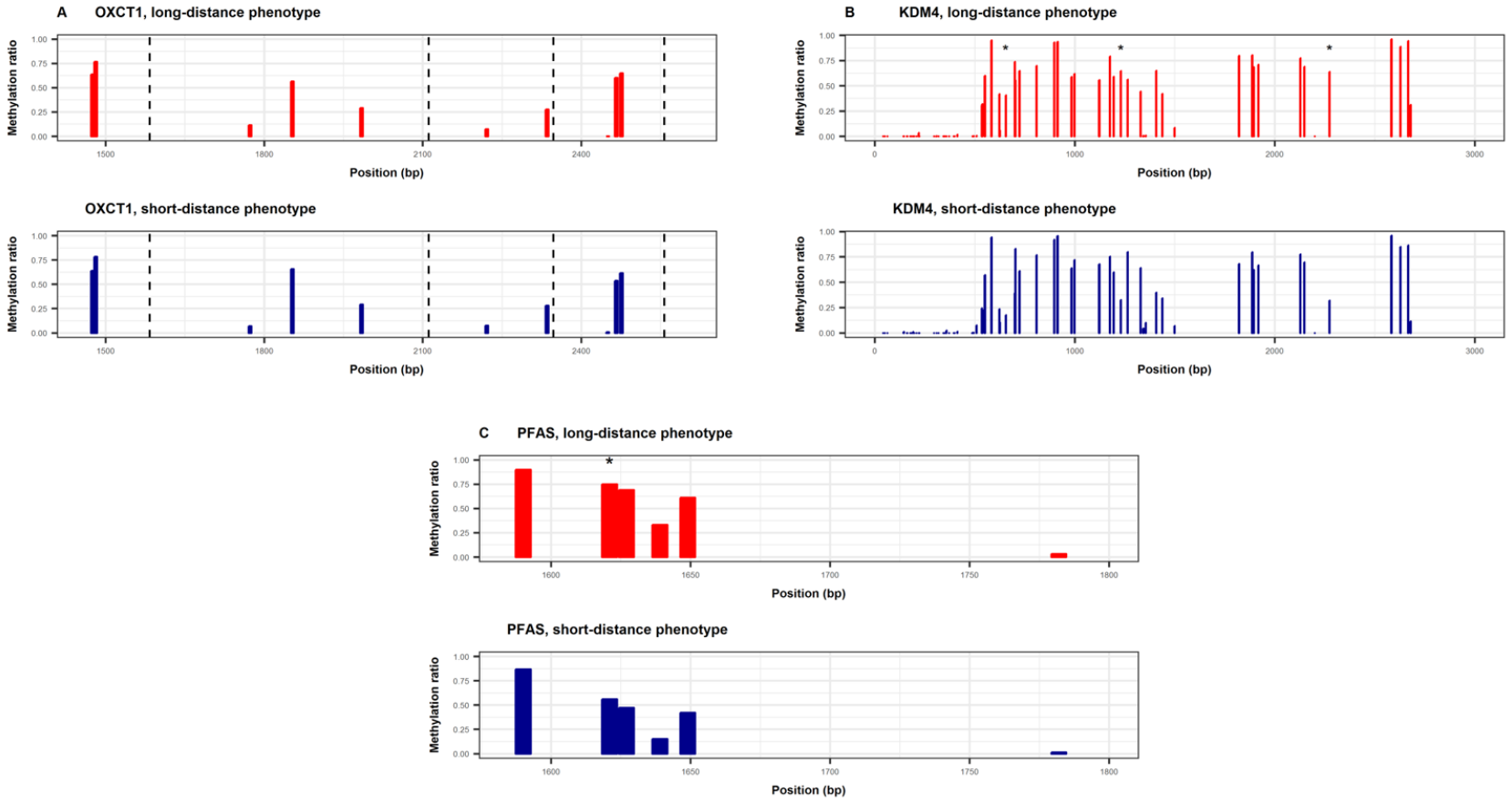

Table 1. Hyper- and hypo-methylated sites in selected loci when comparing short- and long-distance flight phenotypes.

							FRACTIONAL METHYLATION ($_{M}CG/CG_{ALL}$)	
GENE NAME	DESCRIPTION	EXON NO.	SCAFFOLD NO.	POSITION	METH DIFF[$]	P-VALUE	LONG-DISTANCE	SHORT-DISTANCE
HaOG212852	KDM4	1	480	8283	0.323	0.0007	0.648	0.325
HaOG212852	KDM4	1	480	7239	0.320	0.0312	0.639	0.319
HaOG212852	KDM4	1	480	8858	0.231	0.0474	0.408	0.177
HaOG206723	phosphoribosylformylglycinamidine synthase-like	3	211	76644	0.190	0.0298	0.750	0.560
HaOG202339	mobile element jockey-like	2	109	667671	0.005	0.0192	0.005	0.000
HaOG216422	phosphorylated CTD-interacting factor 1-like	1	86	1009788	-0.024	0.0247	0.000	0.024
HaOG206745	succinyl-CoA:3-ketoacid coenzyme A transferase 1	4	211	282391	-0.035	0.0467	0.902	0.937
HaOG202350	phosphoribosyl pyrophosphate synthetase	1	11	223889	-0.044	0.0476	0.833	0.876

[$]Methylation differences between phenotypes determined by average methylation ratio across individual samples.