Sex-specific changes in the aphid DNA methylation landscape

Thomas C. Mathers[1], Sam T. Mugford[1], Lawrence Percival-Alwyn[2, a], Yazhou Chen[1], Gemy Kaithakottil[2], David Swarbreck[2], Saskia A. Hogenhout[1, *] and Cock van Oosterhout[3, *]

[1]Department of Crop Genetics, John Innes Centre, Norwich Research Park, Norwich, United Kingdom

[2]Earlham Institute, Norwich Research Park, Norwich, United Kingdom

[2]School of Environmental Sciences, University of East Anglia, Norwich, United Kingdom

[a]Current address: The John Bingham Laboratory, NIAB, Huntingdon Road, Cambridge, United Kingdom

*Corresponding authors

E-mail: c.van-oosterhout@uea.ac.uk

E-mail: saskia.hogenhout@jic.ac.uk

Running title

Aphid DNA methylation

Keywords

Dosage compensation, epigenetic regulation, *Myzus persicae*, sex-biased gene expression, sex chromosome.

Abstract

Aphids present an ideal system to study epigenetics as they can produce diverse, but genetically identical, morphs in response to environmental stimuli. Here, using whole genome bisulphite sequencing and transcriptome sequencing of the green peach aphid (*Myzus persicae*), we present the first detailed analysis of cytosine methylation in an aphid and investigate differences in the methylation and transcriptional landscapes of male and asexual female morphs. We find that methylation primarily occurs in a CG dinucleotide (CpG) context and that exons are highly enriched for methylated CpGs, particularly at the 3' end of genes. Methylation is positively associated with gene expression, and methylated genes are more stably expressed than un-methylated genes. Male and asexual female morphs have distinct methylation profiles. Strikingly, these profiles are divergent between the sex chromosome and the autosomes; autosomal genes are hypo-methylated in males compared to asexual females, whereas genes belonging to the sex chromosome, which is haploid in males, are hyper-methylated. Overall, we find correlated changes in methylation and gene expression between males and asexual females, and this correlation is particularly strong for genes located on the sex chromosome. Our results suggest that differential methylation of sex-biased genes plays a role in *M. persicae* sexual differentiation.

Introduction

Cytosine methylation is an epigenetic mark found in many eukaryotic organisms (Bewick et al. 2016, 2017; Feng et al. 2010; Zemach and Zilberman 2010). In mammals, cytosine methylation mainly occurs in a CG dinucleotide context (CpG) (Suzuki and Bird 2008). However, in human embryonic stem cells (Guo et al. 2014), and human and mouse oocytes (Guo et al. 2014; Okae et al. 2014), cytosines are methylated in other sequence contexts (non-CpG). Plants also have high levels of non-CpG methylation that is maintained by a set of specialised CHROMOMETHYLASE enzymes not found in other eukaryotes (Bewick et al. 2017). CpG methylation is extensively detected throughout mammalian and plant genomes and often associated with suppression of gene, or transposable element, expression. In contrast, insect genomes have sparse cytosine methylation that is almost exclusively restricted to CpG sites in gene bodies (Zemach et al. 2010). Furthermore, rather than supressing gene expression, insect CpG methylation is associated with high and stable gene expression (Wang

53 et al. 2013; Patalano et al. 2015; Xiang et al. 2010; Libbrecht et al. 2016).

54 Social Hymenoptera have been used as a model system to study the function of insect DNA
55 methylation and its role caste determination (Yan et al. 2014). However, replicated
56 experimental designs have recently shown high between-sample variation (low repeatability)
57 and no evidence of statistically significant differences in CpG methylation between social
58 insect castes (Libbrecht et al. 2016). Furthermore, DNA methylation has a patchy distribution
59 across the insect phylogeny, having been lost in many species, and appears to be dispensable
60 for the evolution of sociality (Bewick et al. 2016). Development of additional model systems
61 is therefore desirable to gain a deeper understanding of the function of cytosine methylation
62 in insects.

63 Aphids have a functional DNA methylation system (Bewick et al. 2016; Hunt et al. 2010; Walsh
64 et al. 2010) and are an outgroup to holometabolous insects (Misof et al. 2014), which have
65 been the main focus of research into insect DNA methylation to date. Furthermore, aphids
66 display extraordinary phenotypic plasticity during their life cycle (Dixon 1977), in the absence
67 of confounding genetic variation, making them ideal for studying epigenetics (Srinivasan and
68 Brisson 2012). During the summer months, aphids are primarily found as alate, asexually
69 reproducing, females. These asexual females are able to produce morphologically distinct
70 morphs in response to environmental stimuli. This can include the induction of winged
71 individuals in response to crowding (Müller et al. 2001), or the production of sexually
72 reproducing forms in response to changes in temperature and day length (Blackman 1971b).
73 In the case of the production of sexually reproducing individuals, sex is determined by an XO
74 chromosomal system, where males are genetically identical to their mothers apart from the
75 random loss of one copy of the X chromosome (Wilson et al. 1997). Differences between
76 aphid morphs are known to be associated with large changes in gene expression (Jaquiéry et
77 al. 2013; Purandare et al. 2014), but whether or not changes in cytosine methylation are also
78 involved is unknown.

79 Here, we performed the first in-depth, genome-wide, analysis of aphid DNA methylation. We
80 find that asexual females and males have distinct expression and methylation profiles and
81 that changes in methylation differ between the X chromosome and autosomes. In males, the
82 autosomes are hypo-methylated relative to asexual females whilst the X chromosome is

83 hyper-methylated. Changes in gene expression and methylation between asexual females
84 and males are correlated, and this correlation is strongest for X-linked genes. Taken together
85 our findings suggest a role for DNA methylation in the regulation of aphid gene expression
86 and the establishment of sexual dimorphism.

87 Results and Discussion

88 Extensive sex-biased expression between asexual females and males

89 To identify genes with sex-biased expression in *M. persicae* clone O, we sequenced the
90 transcriptomes of asexual females and males (six biological replicates each) using RNA-seq
91 (**Supplementary Table 1**). After mapping these reads to the *M. persicae* clone O genome
92 (Mathers et al. 2017), we conducted differential expression analysis with edgeR (Robinson et
93 al. 2009). Genes were classified based on whether their expression was significantly biased
94 (edgeR; Benjamini-Hochberg (BH) corrected $p < 0.05$ and absolute fold change (FC) > 1.5)
95 towards asexual females (FAB genes) or males (MB genes). We also considered the magnitude
96 of sex bias, classifying genes as either moderately sex-biased ($1.5 \leq FC < 10$, for FAB or MB) or
97 extremely sex-biased ($FC \geq 10$, for FAB+ or MB+). In total, 3,433 genes exhibited sex-biased
98 expression (**Figure 1a**; **Supplementary Table 2**), representing 19 % of all annotated *M.*
99 *persicae* genes and 33 % of all genes with detectable expression (> 2 counts-per-million in at
100 least 3 samples, n = 10,334). MB genes outnumbered FAB genes by 18 % (1,778 vs 1,505,
101 binomial test; $p = 1.02 \times 10^{-6}$) and only a handful of FAB+ genes (15) were observed compared
102 to 135 MB+ genes (binomial test; $p = 1.28 \times 10^{-25}$; **Figure 1b**). This is consistent with patterns
103 of sex-biased gene expression in the pea aphid (Purandare et al. 2014) and other
104 invertebrates such as *Caenorhabditis* (Thomas et al. 2012; Albritton et al. 2014) and
105 *Drosophila* (Zhang et al. 2007), which also show a tendency towards an excess of male-biased
106 genes.

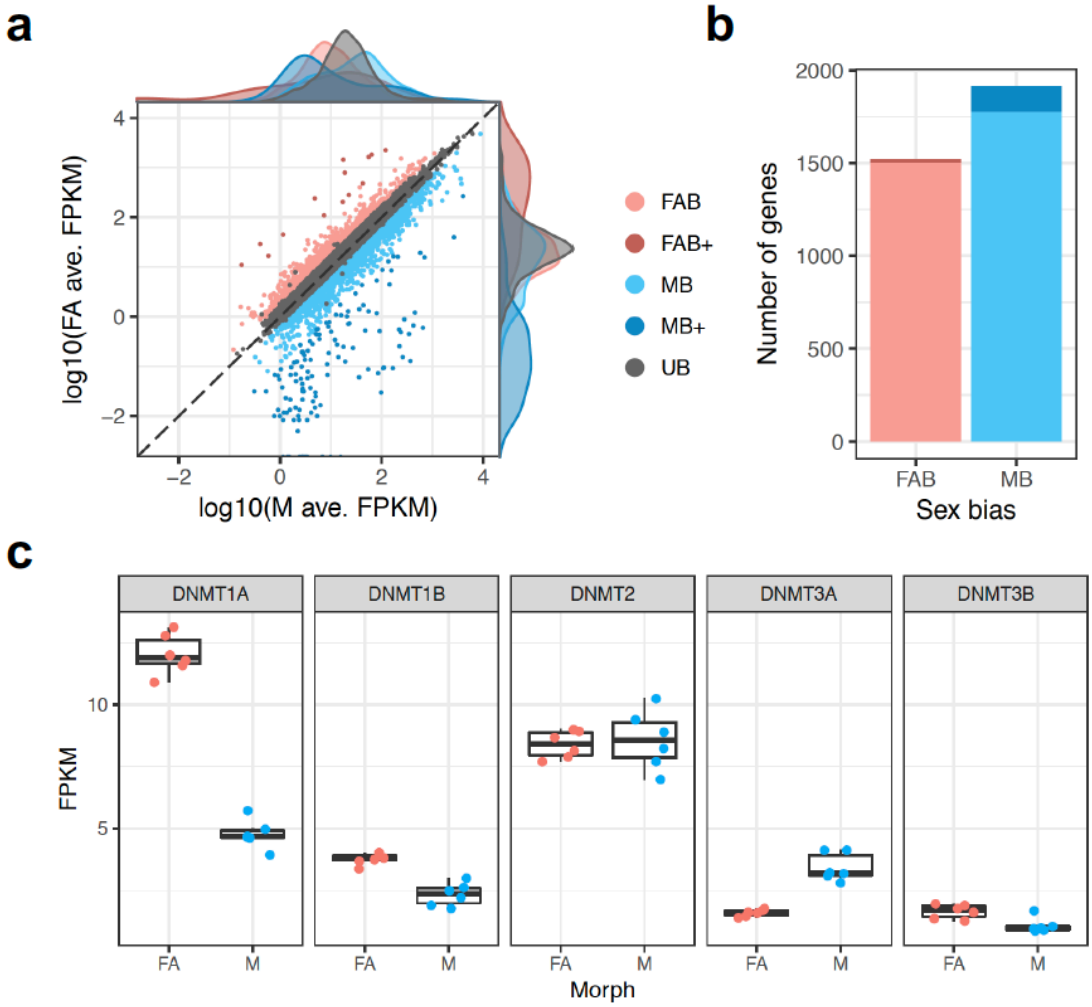


Figure 1: Differential gene expression between *M. persicae* asexual females and males. (**a**) Male (M; x-axis) and asexual female (FA; y-axis) gene expression expressed as $\log_{10}$ fragments per kilobase of transcript per million mapped reads (FPKM) averaged over 6 biological replicates for genes retained for differential expression (DE) analysis with edgeR (n = 10,334). DE genes are coloured according to the direction and magnitude of sex-bias (see main text). UB = unbiased expression (edgeR; Benjamini-Hochberg (BH) corrected $p > 0.05$ and absolute fold change (FC) > 1.5). (**b**) Male-biased (MB) genes significantly outnumber asexual female-biased (FAB) genes. (**c**) Asexual females and males differ significantly in expression at two out of five DNA methyltransferase genes (DNMT1a and DNMT3a; edgeR; BH corrected $p < 0.05$ and FC > 1.5). DNMT1b and DNMT3b are also significantly down-regulated in males (edgeR; BH corrected $p = 6.35 \times 10^{-6}$ and 0.039, respectively). However, the absolute FC of these genes falls below our cut-off of absolute FC > 1.5 (FC = 1.42 and 1.35, respectively).

120 **Methylation genes are differentially expressed**

121 Next, we used our transcriptome data to investigate expression patterns of known
122 methylation genes in *M. persicae* asexual females and males. Genome-wide patterns of DNA
123 methylation in animals are maintained by a toolkit of DNA methyltransferase genes
124 (Schübeler 2015). *De novo* DNA methylation is established by DNMT3 and DNA methylation
125 patterns are maintained by DNMT1 (Law and Jacobsen 2010). An additional homolog of
126 DNMT1 and DNMT3, DNMT2, is responsible for tRNA methylation (Goll et al. 2006) and not
127 involved in DNA methylation. Conservation of the DNA methylation toolkit varies across
128 insects (Bewick et al. 2016) with DNMT1 being associated with the presence of detectable
129 levels of DNA methylation. Aphid genomes contain a full complement of DNA methylation
130 genes with two copies of DNMT1, a single copy of DNMT2, and two copies of DNMT3
131 (Mathers et al. 2017; Nicholson et al. 2015; Walsh et al. 2010). We find that DNMT1a is down-
132 regulated in males, relative to asexual females (edgeR; BH corrected $p = 5.84 \times 10^{-40}$, abs. FC
133 = 2.25), and DNMT3a is up-regulated in males (edgeR; BH corrected $p = 3.18 \times 10^{-14}$, abs. FC =
134 2.44) (**Figure 1c**). DNMT1b and DNMT3b are also down-regulated in males (edgeR; BH
135 corrected $p = 6.35 \times 10^{-6}$ and 0.039, respectively), however the FC of these genes falls below
136 our 1.5-fold threshold. In contrast, the tRNA methyltransferase DNMT2 is uniformly
137 expressed (edgeR; BH corrected $p = 0.067$). These results suggest that changes in DNA
138 methylation may be involved in the establishment of sexual dimorphism in *M. persicae*.

139 **Genome-wide methylation patterns in *M. persicae***

140 DNA methylation has been poorly studied in insects outside of Holometabola and has only
141 been superficially described in Hemiptera as part of a broad scale comparative analysis
142 (Bewick et al. 2016). We therefore first sought to characterise genome-wide patterns of
143 methylation in *M. persicae* before going on to investigate sex-specific changes in DNA
144 methylation levels between asexual female and male morphs. To characterise genome-wide
145 DNA methylation levels at base-level resolution, we sequenced bisulphite-treated DNA
146 extracted from whole bodies of asexual females and males (three biological replicates each)
147 derived from the same clonally reproducing population (clone O), and mapped these reads to
148 the *M. persicae* clone O genome (Mathers et al. 2017) using Bismark (Krueger and Andrews
149 2011). After removal of ambiguously mapped reads and PCR duplicates, each replicate was

150 sequenced to between 24x and 37x average read depth (**Supplementary Table 3**), resulting
151 in 7,836,993 CpG sites covered by at least 5 reads in all samples.

152 *M. persicae* individuals harbour an obligate endosymbiont, *Buchnera aphidicola*, which lacks
153 a functional DNA methylation system (van Ham et al. 2003). We made use of *Buchnera*
154 derived reads in each sample to establish rates of false positive methylation calls caused by
155 incomplete cytosine conversions by mapping each sample to the *M. persicae Buchnera*
156 genome (Mathers et al. 2017) and quantifying methylation levels (**Supplementary Table 4**).
157 The average methylation level in *Buchnera* for Cs in any sequence context was 0.45% ± 0.68
158 (mean ± SD), indicating that bisulphite treatment of the aphid DNA was 99.55% efficient and
159 was consistent across samples. Based on this, we assessed methylation levels in *M. persicae*
160 for C's in a CpG, CHH and CHG context. Only Cs in a CpG context had methylation levels higher
161 than the false positive rate in *B. aphidicola*, indicating that CpG methylation is the
162 predominant form of DNA methylation in *M. persicae* (**Figure 2a**). Overall, global CpG
163 methylation levels (2.93% ± 0.32% of Cs; mean ± SD) were similar to those reported in other
164 hemipteran insects (2 – 4 %) and higher than in Hymenoptera (0.1 – 2.2 %) (Bewick et al.
165 2016). Exons were highly enriched for methylated CpGs relative to the rest of the genome (χ^2
166 $= 1.07 \times 10^8$, d.f. = 1, $p < 2.2 \times 10^{-16}$), with only 7.7% of methylated CpGs occurring in intergenic
167 regions (**Figure 2b and c**). Identification of significantly methylated CpG sites using a binomial
168 test that incorporates the false positive methylation rate (derived from *Buchnera*) showed
169 that methylation is non-randomly distributed across *M. persicae* gene bodies, with
170 methylated CpG sites biased towards the 3' end of genes despite the total number of CpG
171 sites being much higher at the 5' ends of genes (**Figure 2d**). This is likely driven by high rates
172 of methylation in 3' UTRs (**Figure 2c**). Interestingly, methylation bias towards the 3' end of
173 genes is potentially a unique feature of aphid or hemipteran CpG methylation as insects from
174 other orders show an opposite bias, with higher methylation at the 5' end of genes (Bonasio
175 et al. 2012; Hunt et al. 2013; Zemach et al. 2010).

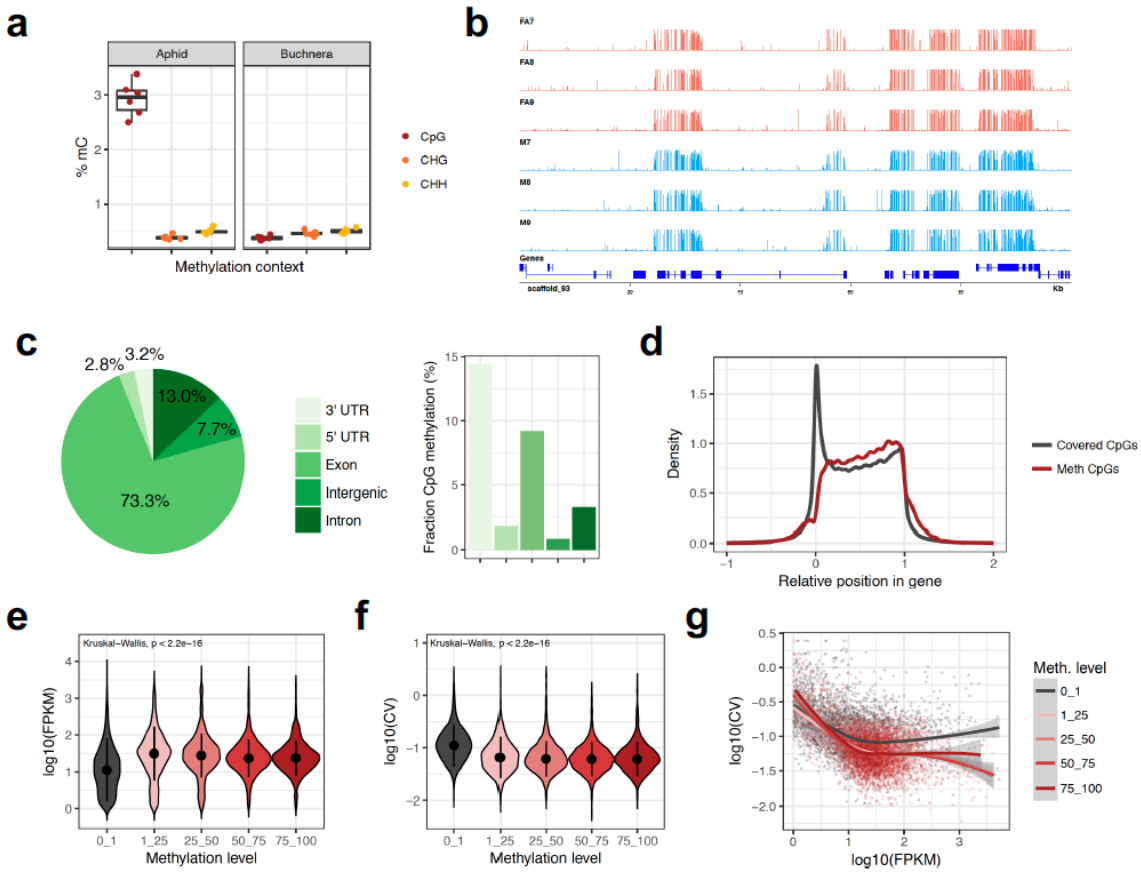


Figure 2: The *M. persicae* methylome. (**a**) *Boxplots* showing the proportion of methylated cytosines (mC) by sequence context (CpG, CHG and CHH) for *M. persicae* and its obligate endosymbiont *Buchnera aphidicola*, which lacks a functional methylation system. (**b**) Example genome browser view showing the distribution of CpG methylation in asexual females and males across the first 100Kb of scaffold_93. (**c**) The distribution of methylated CpGs across genomic features and the proportion of methylated CpGs in each feature. Methylated and un-methylated CpG counts were summed across all replicates. (**d**) The distribution of all covered CpG sites (min. 5 reads per sample) and significantly methylated CpG sites (*binomial test*, BH-corrected $p < 0.05$) across *M. persicae* gene bodies. -1 = 1Kb upstream, 0 = transcription start site, 1 = transcription stop site, 2 = 1Kb upstream. (**e**) The distribution of RNA-seq expression levels in asexual females ($\log_{10}$ FPKM) for un-methylated (0 - 1% CpG methylation) and methylated genes (FPKM = Fragments Per Kilobase of transcript per Million). Expression values were averaged across six biological replicates and methylation levels averaged across three biological replicates. Only genes with average expression levels of at least 1 FPKM in males and asexual females were included. Dots and whiskers inside the *violin plots* indicate median and interquartile range respectively. (**f**) As for (**e**) but showing the distribution of variation in expression between the six asexual female RNA-seq replicates (measured as the $\log_{10}$ transformed coefficient of variation ($\log_{10}$ CV) of FPKM) for un-methylated (0 - 1% CpG methylation) and methylated genes. (**g**) The relationship between the mean and the CV of gene expression for un-methylated and methylated genes with a trend line for each methylation level shown as a LOESS-

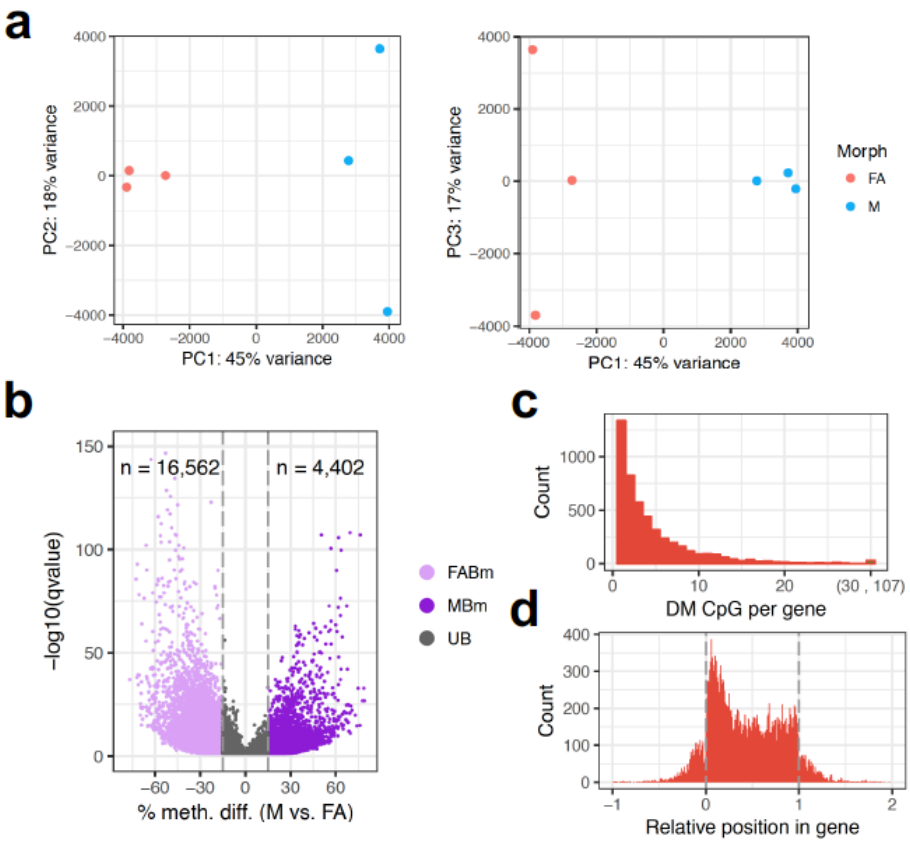
smoothed line with shaded areas indicating the 95% CI. The difference between the grey (un-methylated; 0 - 1% CpG methylation) and pink/red lines (methylated; > 1% CpG methylation) shows that methylation reduces the between-replicate variation in gene expression, particularly in highly expressed genes. The negative correlation and downwards slope of trend lines shows that higher expressed genes are better canalized, showing less between-individual variation in gene expression.

Next, we investigated the relationship between genome-wide patterns of DNA methylation and gene expression using data for asexual females (**Supplementary Table 5**). We find that the presence DNA methylation is positively associated with gene expression, with methylated genes having significantly higher expression than un-methylated genes (**Figure 2e**). We also find that methylated genes are more stably expressed than un-methylated genes (**Figure 2f**), even after accounting for the higher expression of methylated genes (**Figure 2g**). The same patterns were also observed using male methylation and gene expression data (**Supplementary Figure 1**). Taken together, these data suggest that DNA methylation in aphids may be involved in establishing and stabilising high gene expression, as has been suggested in corals (Liew et al. 2017) and holometabolous insects (Wang et al. 2013; Patalano et al. 2015; Xiang et al. 2010; Libbrecht et al. 2016).

Asexual females and males have distinct methylation profiles

To gain an overview of methylation differences between asexual female and male *M. persicae* morphs, we conducted principle component analysis based on methylation levels of 350,782 CpG sites significantly methylated (*binomial test*, BH-corrected $p < 0.05$) in at least one sample. Male and asexual female morphs clearly form distinct clusters, indicating reproducible differences in global CpG methylation (**Figure 3a**). To further characterise methylation differences between asexual females and males we conducted site-wise differential methylation (DM) analysis, identifying 20,964 DM CpG sites (> 15% methylation difference, BH corrected $p < 0.05$; **Supplementary Table 6**), 79% of which show a reduction in methylation (hypo-methylation) in males relative to asexual females and 21% the opposite (**Figure 3b**). This is significantly higher than expected by chance (see **Supplementary Figure 2**), and indicates that differences in methylation between asexual female and male morphs are unlikely to be due to random between-sample variation. Rather, alterations in CpG methylation appear to underpin differentiation between sexual morphs in aphids. These findings are striking given the absence of significant levels of sex-biased or caste-biased

228 methylation in other insect systems (Libbrecht et al. 2016; Patalano et al. 2015; Herb et al.
229 2012; Wang et al. 2015).



230

231 **Figure 3: Differential methylation between *M. persicae* asexual female and male morphs.** (**a**)
232 Principle component analysis (PCA) based on methylation levels at 350,782 CpG sites significantly
233 methylated in at least one sample. PC1 seperates the samples based on sex (45% of the variation),
234 whilst PC2 and PC3 seperate male and asexual female replicates, respectively (explaining 18 to 17 %
235 of the variation). (**b**) *Volcano plot* showing results of MethylKit (Akalin et al. 2012) site-wise tests of
236 differential methylation between asexual females (FA) and males (M). Methylation differences are
237 shown for M relative to FA. Only CpG sites showing significant differential methylation (DM) (BH
238 corrected $p < 0.05$) are shown. A minimum methylation difference threshold of 15% per site was
239 applied to define a site DM between FA and M. (**c**) The number of differentially methylated sites per
240 gene (±1Kb flanking region). (**d**) The distribution of DM CpG sites along *M. persicae* gene bodies. -1 =
241 1Kb upstream, 0 = transcription start site, 1 = transcription stop site, 2 = 1Kb upstream.

242 Overlap analysis revealed that the majority (92%) of DM CpG sites between asexual females
243 and males were located in gene bodies (± 1 Kb), with genes having between 1 and 107 DM
244 CpG sites (**Figure 3c**). These DM CpG sites were non-randomly distributed along gene bodies,
245 being biased towards the 5' end of genes (**Figure 3d**). As such, whilst overall methylation

levels are biased towards the 3' end of genes, sites with variable methylation are more likely to be at the 5' end. To directly correlate gene body methylation levels with gene expression, we also performed DM analysis at the gene level (**Supplementary Table 7**). This identified 1,344 DM genes with > 10% methylation difference (BH corrected $p < 0.05$), of which 205 showed significant male-biased methylation and 1,129 asexual female-biased methylation (**Figure 4a and b**). Considering genes with variable methylation, males have undergone a global loss of gene body methylation relative to asexual females (Wilcoxon signed-rank test, $p < 2.2 \times 10^{-22}$; **Figure 4c**).

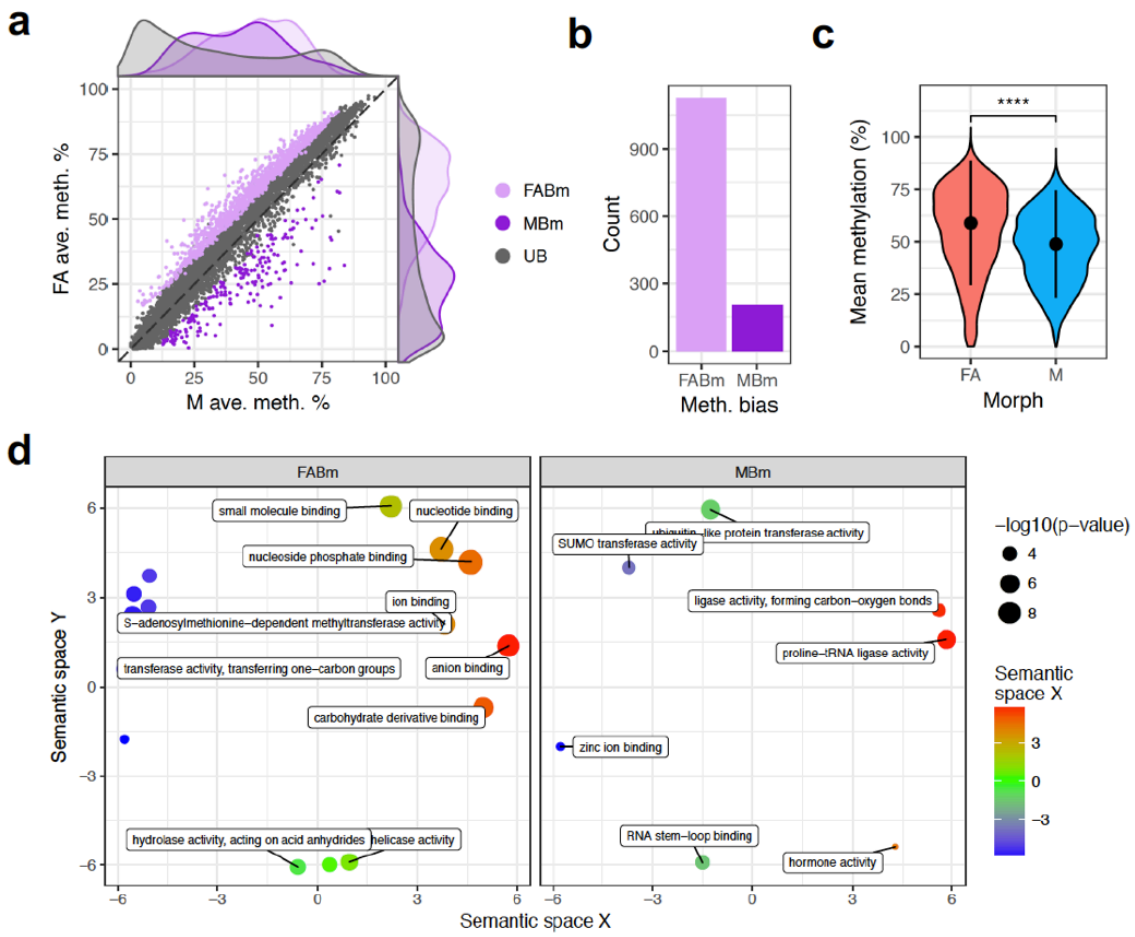


Figure 4: Genome-wide changes in gene body methylation between asexual female and male morphs. (**a**) Male (M; x-axis) and asexual female (FA; y-axis) gene-wise methylation levels averaged over 3 biological replicates for genes methylated > 1% in at least one of the two morphs (n = 6,699). Differentially methylated (DM) genes (MethylKit; > 10% methylation difference, BH corrected $p < 0.05$) are coloured according to the direction of sex-bias: MBm = male biased methylation, FABm = female-biased methylation, UB = unbiased methylation. (**b**) FABm genes outnumber MBm genes. (**c**) *Violin plot* showing the distribution of mean methylation level in FA and M for DM genes. Dots and whiskers indicate median and interquartile range respectively; **** = Wilcoxon signed-rank test $p <$

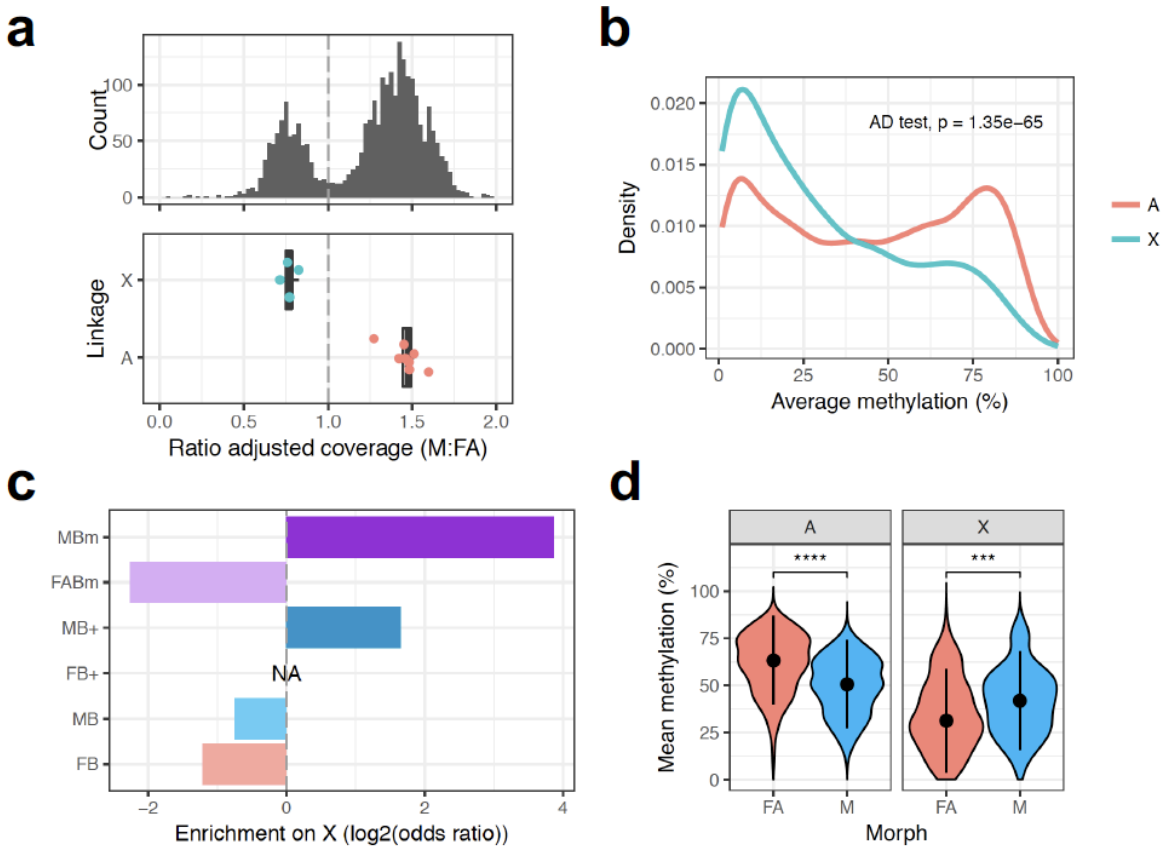
263 0.0001. (**d**) Enriched GO terms relating to molecular function plotted in semantic space for FABm
264 genes and MBm genes (for terms relating to biological process see **Supplementary Figure 3**). A full
265 list of enriched GO terms for each DM class and functional category is given in **Supplementary Table**
266 **8**).

267 Gene ontology (GO) term enrichment analysis showed that asexual female-biased
268 methylation and male-biased methylation genes were both enriched for GO terms relating to
269 core biological processes, including metabolism and regulation of gene expression (**Figure 4d**;
270 **Supplementary Figure 3**; **Supplementary Table 8**). Protein SUMOylation is enriched among
271 genes with male-biased methylation. This is interesting because protein SUMOylation is
272 essential for dosage compensation of the *C. elegans* sex chromosome (Pferdehirt and Meyer
273 2013) and plays a key role in insect development and metamorphosis (Ureña et al. 2015).
274 Changes in methylation may therefore underpin core processes involved in aphid
275 polyphenism and sex determination. Consistent with this, we also find enrichment of
276 hormone signalling amongst genes with male-biased methylation, with 3 insulin genes hyper-
277 methylated in males (2 not expressed, 1 has male-specific expression). Insulin receptors
278 determine alternative wing morphs in planthoppers (Xu et al. 2015) and have been shown to
279 interact with the core sex determination gene TRANSFORMER-*2* (Zhuo et al. 2017).

280 **The X chromosome has distinct patterns of expression and methylation**

281 We identified X-linked scaffolds in the *M. persicae* genome assembly based on the ratio of
282 male to asexual female bisulphite sequencing coverage. This approach takes advantage of the
283 hemizygous condition of the X chromosome in males, which should result in X-linked scaffolds
284 having half the read depth of autosomal scaffolds (Jaquiéry et al. 2017). As expected, we
285 observe a bimodal distribution in the ratio of male to asexual female scaffold coverage, with
286 the lower coverage peak falling at approximately half the relative coverage of the higher
287 coverage peak (**Figure 5a**; **Supplementary Table 9**). Scaffolds in this lower coverage peak are
288 putatively derived from the X chromosome. To validate the coverage results, we mapped
289 known X-linked (n=4) and autosomal (n=8) microsatellite loci to the clone O genome and
290 retrieved the male to asexual female coverage ratios of their corresponding scaffolds. The
291 coverage of these known sex-linked scaffolds also exactly matches expectations (**Figure 5a**;
292 **Supplementary Table 10**). Using a cut-off of adjusted coverage, we identified 748 X-linked
293 scaffolds and 1,852 autosomal scaffolds, totalling 68.7 and 239.7 Mb of sequence respectively

294 (**Supplementary Figure 4**). Scaffolds assigned to the X chromosome therefore account for
295 22.3% of the assembled (scaffolds ≥ 20Kb) *M. persicae* clone O genome. This is in line with
296 expectations given the most common *M. persicae* karyotype of 2n = 12 and that the X
297 chromosome is larger than the autosomes (Blackman 1971a). Using the chromosomal
298 assignment of scaffolds, we were able to assign 3,110 gene models to the X chromosome and
299 10,768 to autosomes, leaving 4,555 (24.7%) gene models on unassigned scaffolds shorter
300 than 20 Kb. The number of identified X-linked genes is not different to expectations based on
301 the assembled size of the respective chromosomal regions (binomial test, $p = 0.65$). However,
302 we find that the X chromosome is depleted in coding sequence (CDS) compared to the
303 autosomes (6.3% vs 6.5%; $\chi^2 = 5{,}821.5$, d.f. = 1, $p < 2.2 \times 10^{-16}$). This is due to the reduced CDS
304 length of X-linked genes (Wilcoxon signed-rank test, $p = 4.2 \times 10^{-4}$; **Supplementary Figure 5**).



305

306 **Figure 5: Distinct patterns of methylation and expression between the *M. persicae* X chromosome**
307 **and autosomes.** (**a**) X-linked and autosomal scaffolds (≥ 20Kb) in the *M. persicae* genome were
308 identified based on the relative coverage of BS-seq reads in males (M) compared to asexual females
309 (FA). Given the XO sex determination system of aphids, X-linked scaffolds are predicted to have half
310 autosomal coverage in males. A bimodal distribution in the ratio of M to FA coverage is clearly visible
311 (upper panel). We considered scaffolds falling in the lower coverage peak (ratio of adjusted coverage

312 < 1) as X-linked and scaffolds in the second, higher coverage peak (ratio of adjusted coverage > 1), as
313 autosomal. The assignment of scaffolds to the X chromosome or autosomes was validated by
314 comparing the M : FA ratio of coverage for scaffolds containing microsatellite markers on the X-
315 chromosome (blue dots) and autosome (red dots) (lower panel). (**b**) The distribution of gene body
316 methylation levels for X-linked and autosomal genes. (**c**) Observed / expected (odds ratio) counts of
317 DM and DE genes on the X chromosome by expression or methylation bias category. The X
318 chromosome is significantly enriched for MB+ genes (≥ 10-fold upregulation in M) and genes with
319 male-biased methylation (MBm). (**d**) The distribution of mean methylation levels in asexual females
320 (FA) and males (M) for X-linked and autosomal DM genes (MethylKit; > 10% methylation difference,
321 BH corrected $p < 0.05$). Methylation levels are significantly higher in FA than M for autosomal genes,
322 whereas M have a higher methylation than FA in X-linked genes (**d**) dots and whiskers inside the *violin*
323 *plots* indicate median and interquartile range respectively; *** = Wilcoxon signed-rank test $p < 0.001$,
324 **** = $p < 0.0001$.

325 Strikingly, the X chromosome has a distinct methylation landscape compared to the
326 autosomes (Anderson-Darling k-sample test, $p = 1.35 \times 10^{-65}$; **Figure 5b**), with fewer highly
327 methylated genes. We also find opposing dynamics of sex-biased methylation between the X
328 chromosome and the autosomes. The X chromosome is significantly enriched for genes with
329 male-biased methylation and depleted for genes with female-biased methylation (χ^2 =
330 176.65, d.f. = 2, $p < 2.2 \times 10^{-16}$; **Figure 5c**). Overall, X chromosome genes are hyper-methylated
331 in males (Wilcoxon signed-rank test, $p = 8.6 \times 10^{-4}$; **Figure 5d**) compared to the genome-wide
332 pattern of hypo-methylation (Wilcoxon signed-rank test, $p < 2.2 \times 10^{-16}$). Mirroring differences
333 in methylation between the X chromosome and the autosomes, we also find that the X
334 chromosome is enriched for genes with extreme male-biased expression (χ^2 = 42.38, d.f. = 1,
335 $p = 7.5 \times 10^{-11}$; **Figure 5c**), a phenomenon also observed in the pea aphid (Jaquiéry et al. 2013).
336 Male-biased expression of X-linked genes is therefore conserved across two distantly related
337 aphid species, and, at least in the case of *M. persicae*, this also extends to patterns of DNA
338 methylation.

339 Finally, we investigated whether changes in methylation between *M. persicae* asexual
340 females and males are associated with changes in gene expression. The relationship between
341 gene expression and gene body methylation is an open question in invertebrates and few
342 studies have directly tested for changes in expression and methylation. We find that DM
343 genes are significantly enriched for DE (χ^2 = 7.84, d.f.= 1, p = 0.005), suggesting that
344 methylation changes may be involved in the regulation of at least a subset of sex-biased

genes. In support of this, we find a weak but significant positive correlation between changes in gene expression and methylation between asexual females and males when considering genes methylated (> 1%) and expressed (> 1 FPKM) in at least one of the sexes (n = 6,699; Spearman's ρ = 0.089, $p = 2.7 \times 10^{-13}$; **Figure 6a**). Interestingly, this correlation is driven by X-linked genes which show a significantly stronger correlation between changes in expression and methylation than autosomal genes (GLM: $F_{1,6185} = 93.07$, $p < 0.0001$; **Figure 6b**). Combined with recent results demonstrating a role for chromatin accessibility in the sex-specific regulation of genes on the X chromosome and dosage compensation in the pea aphid (Richard et al. 2017), our findings suggest a key role for epigenetics in establishing patterns of X-linked gene expression in aphids.

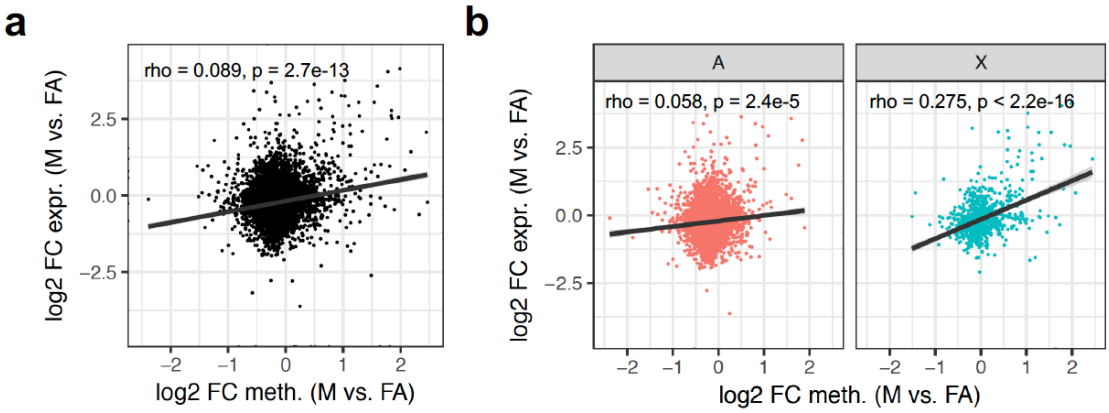


Figure 6: Correlated changes in expression and methylation between asexual females and males. (**a**) *Scatter plot* showing the relationship between fold-change (FC) in gene expression and methylation between asexual females (FA) and males (M) for genes expressed (> 1 FPKM) and methylated (> 1%) in at least one of the sexes (n = 6,699). Methylation levels of genes were estimated across the whole gene body and averaged across replicates. Positive values indicate increased expression or methylation in males, relative to asexual females; negative values indicate increased expression or methylation in asexual females, relative to males. (**b**) The correlation between gene expression changes and methylation changes between FA and M is significantly stronger for X-linked genes (X; n = 925) than autosomal genes (A; n = 5,272). Spearman's ρ was used to assess significance and strength of the relationship between change in expression and methylation for each set of genes. The trend lines indicate linear fit with shaded areas indicating 95% confidence intervals.

Conclusions

We presented the first detailed analysis of genome-wide methylation patterns in an aphid, evaluating its importance for gene expression and sexual differentiation. We found that 3,433

370 genes (19 % of the annotated genome) were differentially expressed between the males and
371 asexual females, and that there was a significant excess of male-biased genes. We also found
372 evidence suggesting that methylation plays an important role in sexual differentiation of
373 aphids, showing that DNMT1a and b are significantly downregulated in males, whereas
374 DNMT3a is upregulated in males. CpG methylation is the predominant form of DNA
375 methylation in *M. persicae* and, in contrast to other insects, exons were highly enriched for
376 methylated CpGs at the 3' end rather than the 5' end of genes. Methylation is positively
377 associated with gene expression, and in addition, methylated genes are more stably
378 expressed than un-methylated genes. Methylation was significantly reduced in males
379 compared to asexual females, yet remarkably, the X chromosome genes of males were hyper-
380 methylated. Given that differentially methylated genes were also significantly differentially
381 expressed between the sexes, we propose that changes in DNA methylation play a role in *M.*
382 *persicae* sexual differentiation. Our findings pave the way for future functional studies of DNA
383 methylation in aphids, and its potential role in the remarkable evolutionary potential of these
384 insects, and their extraordinary phenotypic plasticity.

385 # Methods

386 **Aphid rearing and sample preparation**

387 An asexual colony of *M. persicae* clone O derived from a single apterous asexual female
388 (Mathers et al. 2017) was maintained on *Brassica rapa* plants in long-day conditions (14h
389 light, 22° C day time, and 20° C night time, 48% relative humidity). Male morphs were induced
390 by transferring the colony to short-day conditions (8h light, 18° C day time, and 16° C night
391 time, 48% relative humidity) and samples were collected 2 months after transfer. Replicate
392 samples were harvested from the same populations, with each replicate consisting of 20
393 adults, with apterous asexual females collected from the long-day population, and males from
394 the short-day population. Samples were immediately frozen in liquid nitrogen prior to RNA or
395 DNA extraction. DNA (three biological replicates per morph) was extracted using the CTAB
396 protocol (Marzachi et al. 1998), with the addition of a proteinase K digestion step during the
397 initial extraction. RNA (six biological replicates per morph) was extracted using the Trizol
398 reagent according to the manufacturers' protocol (Sigma), and further purified using the
399 RNeasy kit with on-column DNAse treatment (Qiagen).

400 **Transcriptome sequencing**

401 RNA samples were sent for sequencing at the Earlham Institute (Norwich, UK) where twelve
402 non-orientated libraries were constructed using the TruSeq RNA protocol v2 (Illumina
403 #15026495 Rev.F). 1ug of total RNA was enriched for mRNA using oligo(dT) beads. The RNA
404 was then fragmented and first strand cDNA synthesised. Following end repair and adapter
405 ligation, each library was subjected to a bead based size selection using Beckman Coulter XP
406 beads (Beckman Coulter Inc., Brea, CA, USA) before performing PCR to enrich for fragments
407 containing TruSeq adapter sequences. Libraries were then pooled and sequenced on the
408 Illumina HiSeq 2000 platform (Illumina Inc., San Diego, CA, USA) (v3 chemistry; 2 x 100 bp),
409 generating between 15 and 57 million paired-end reads per sample. RNA-seq reads have been
410 deposited in the NCBI short read archive (SRA) under accession number PRJNA437622.

411 **Gene expression analysis**

412 Raw RNA-seq reads for each sample were trimmed for low quality bases and adapter
413 contamination with Trim Galore! v0.4.0 using default settings for paired end reads
414 (www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Gene-level expression
415 quantification was then performed for each sample based on the *M. persicae* clone O
416 reference genome and gene annotation (Mathers et al. 2017), using RSEM v1.2.31 (Li and
417 Dewey 2011) with STAR v2.5.2a (Dobin et al. 2013). Average expression and the coefficient of
418 variation was calculated per gene for asexual females and males separately using FPKM
419 (fragments per kilobase of transcript per million) values estimated by RSEM. We also
420 identified differentially expressed (DE) genes between asexual females and males using edgeR
421 (Robinson et al. 2009) based on gene-level expected counts estimated by RSEM. Only genes
422 with greater than 2 counts-per-million in at least three samples were retained for DE analysis
423 and we considered genes DE if they had a fold-change (FC) ≥ 1.5 and $p < 0.05$ after adjusting
424 for multiple testing using the Benjamini-Hochberg (BH) procedure (Benjamini and Hochberg
425 1995).

426 **Bisulphite sequencing**

427 Bisulphite sequencing library construction was performed using 500 ng genomic DNA per
428 sample with a BIOO Scientific NEXTflex^{TM} Bisulfite-Seq Kit (Bioo Scientific Corporation, Austin,

429 TX, USA) according to the manufacturer's instructions with the following modifications:
430 genomic DNA was sheared to 200-400 bp with a Covaris S2 sonicator (Covaris Inc., Woburn,
431 MA) using the following settings: duty cycle 10%, intensity 5, 200 cycles per burst for 120
432 seconds. The power mode was frequency sweeping, temperature 5-6°C and water level 12.
433 Libraries either received NEXTflex™ barcode #24 (GGTAGC) or #31 (CACGAT). All purified
434 libraries were QC checked with the Bioanalyzer DNA HS assay and further quantified by Qubit
435 dsDNA HS Assay Kit (Life Technologies, Carlsbad, CA, USA) before pooling as pairs. Pooled
436 libraries were further quantified by qPCR using a KAPA Library Quantification Kit - Illumina/ABI
437 Prism (Kapa Biosystems Inc., Wilmington, MA, USA) on a StepOnePlus™ Real-Time PCR System
438 (Life Technologies, Carlsbad, CA, USA). Sequencing was performed at the Earlham Institute
439 (Norwich, UK) on an Illumina HiSeq 2500 (Illumina Inc., San Diego, CA, USA) using paired-end
440 sequencing (v4 chemistry; 2 x 126 bp) with a 15% PhiX spike in, clustering to 650 K/mm^2. In
441 total, we generated between 70 and 127 million paired-end reads per sample.

442 **DNA methylation analysis**

443 Bisulphite treated reads for each sample were trimmed for low quality bases and adapter
444 contamination using Trim Galore! v0.4.0 with default settings
445 (www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Read pairs where one or both
446 reads were shorter than 75bp after trimming were discarded. We then mapped the trimmed
447 reads to the *M. persicae* clone O reference genome (Mathers et al. 2017) using Bismark
448 v0.16.1 (Krueger and Andrews 2011). Trimmed reads were also mapped to the genome of the
449 *M. persicae* strain of the obligate aphid endosymbiont *Buchnera aphidicola* (Mathers et al.
450 2017) to estimate the error rate of the C to T conversion. Reads derived from PCR duplicates
451 and that mapped to multiple locations in the genome were removed from downstream
452 analysis. The distribution of methylation across selected scaffolds was visualised using Sushi
453 (Phanstiel et al. 2014).

454 Overall levels of methylation in a CpG, CHG and CHH sequence context were estimated
455 directly from mapped reads with Bismark (Krueger and Andrews 2011). We also characterised
456 CpG methylation levels of features in the *M. persicae* clone O genome based on the reference
457 annotation (Mathers et al. 2017). Average CpG methylation levels of introns, exons, 5' UTRs,
458 3' UTRs and intergenic regions were calculated with bedtools v2.25.0 (Quinlan and Hall 2010),

459 pooling data from all replicates and counting overlapping methylated and unmethylated
460 CpGs. We also calculated per-gene methylation levels for asexual females and males
461 independently in the same way. To assess the genome-wide distribution of methylated CpGs,
462 we filtered CpG sites to those covered by at least five reads in all samples and used a binomial
463 test to identify significantly methylated sites in each sample using the C to T conversion error
464 rate (derived from mapping to *Buchnera*) as the probability of success and corrected for
465 multiple testing using the BH procedure (Benjamini and Hochberg 1995), setting the FDR at
466 5% (BH adjusted $p < 0.05$).

467 Methylation differences between asexual females and males were assessed using a principle
468 component analysis (PCA) and by identifying differentially methylated (DM) sites and genes.
469 PCA was carried out with prcomp, implemented in R v3.2.2 (R Core Team 2017), using the
470 methylation levels of CpG sites significantly methylated in a least one sample (binomial test,
471 BH adjusted $p < 0.05$). We identified DM sites and genes using logistic regression implemented
472 in MethylKit (Akalin et al. 2012) which accepts input directly from Bismark. *p* values were
473 adjusted to Q-values using the SLIM method (Wang et al. 2011) to account for multiple
474 testing. For the site-level analysis, we discarded CpG sites covered by less than 5 reads and
475 those that fell into the top 0.1% of coverage. We considered sites significantly DM if they had
476 at least a 15% methylation difference at a 5% FDR ($Q < 0.05$). At the gene level, we discarded
477 genes covered by less than 20 reads which fell into the top 1% of coverage, and called genes
478 as DM if they had at least 10% methylation difference and at a 5% FDR ($Q < 0.05$). A less
479 stringent percent methylation difference was used at the gene-level as the signal of DM may
480 be diluted over the length of the gene body. To assess the rate of false positive methylation
481 calls caused by random variation between samples we generated a null distribution of DM
482 calls at $Q < 0.05$ for a range of percentage methylation difference cut-offs using all possible,
483 non-redundant, pairs of samples where an asexual female sample is grouped with a male
484 sample ($n = 19$). Overall, these random pairings resulted in a low number of DM calls,
485 indicating our results are robust (**Supplementary Figure 2a and b**).

486 **X chromosome identification**

487 We used our whole-genome bisulphite sequencing data for males and asexual females to
488 identify X–linked scaffolds in the *M. persicae* clone O genome assembly based on the ratio of

489 male to asexual female coverage using a procedure similar to Jaquiéry et. al (2017). BAM files
490 generated by MethylKit were merged for each morph using Picard v2.1.1
491 (http://broadinstitute.github.io/picard/) to maximise the depth of coverage. We then
492 calculated per site sequence depth with SAMtools v1.3 (Li et al. 2009). The average depth of
493 the pooled asexual female and male samples was 79x and 90x, respectively. We then
494 calculated the ratio of male median depth of coverage to asexual female median depth of
495 coverage for all scaffolds longer than 20 Kb, normalising male coverage to that of asexual
496 female coverage (multiplying male median coverage by 79 / 90). This resulted in a clear
497 bimodal distribution with modes at ~0.75 and ~1.5 (**Figure 5a**). We applied a cut-off of male
498 to asexual female normalised median coverage ratio < 1 to assign scaffolds to the X
499 chromosome and > 1 to assign scaffolds to the autosomes. To validate the coverage results,
500 we mapped known X-linked (n=4) and autosomal (n=8) microsatellite loci from Sloane et. al.
501 (2001) and Wilson et. al. (2004) to the clone O genome with blastn and retrieved coverage
502 ratios for their respective scaffolds.

503 **Testing for correlation between changes in methylation and expression**

504 To investigate the relationship between changes in gene expression and methylation we
505 compared expression and methylation levels of genes in males and asexual females. Using
506 average expression (FPKM) and methylation levels, we calculated the $\log_2$ FC in expression
507 (FC_{Expr}) and methylation (FC_{Meth}), and tested for correlation using Spearman's ρ (rho). We also
508 investigated the effect of chromosomal location (X chromosome vs. autosomes) on the
509 relationship between gene expression and methylation using a general linear model (GLM).
510 The GLM was formulated with FC_{Expr} as the response variable, and FC_{Meth} as a covariate,
511 crossed with chromosome (as fixed factor). This interaction term tests whether the slopes of
512 the regression lines of the X chromosome and autosomes run parallel.

513 **Annotation of methyltransferase genes**

514 Amino acid sequences of human DNA methyltransferase genes were blasted against
515 annotated protein sequences of *Myzus persicae* Clone O (Mathers et al., 2017). The top *M.*
516 *persicae* clone O hit for each gene was then used to blast against the *M. persicae* protein set
517 in an iterative fashion until no additional genes were identified. The E value were set as 1E-
518 10.

519 **GO term enrichment analysis**

520 Go term enrichment analysis of specific gene sets was performed with BINGO (Maere et al.
521 2005) using the complete *M. persicae* clone O proteome as the reference set. Redundant
522 terms were then removed with REVIGO (Supek et al. 2011).

523 Data availability

524 Raw RNA-seq and BS-seq data generated for this study have been deposited in the NCBI short
525 read archive under accession number PRJNA437622.

526 Acknowledgements

527 This work was supported by a BBSRC Industrial Partnership Award (BB/L002108/1) to S.H.,
528 D.S. and C.v.O. and co-funded by Syngenta, the Plant Health Institute Strategy Programme
529 (BB/P012574/1) awarded to the John Innes Centre, Gatsby Charitable Foundation funding to
530 S.H., a BBSRC award (BB/N02317X/1) to C.v.O., as well as support by the Earth & Life
531 Systems Alliance (ELSA). Next-generation sequencing and library construction was delivered
532 via the Biotechnology and Biological Sciences Research Council (BBSRC) National Capability
533 in Genomics and Single Cell (BB/CCG1720/1) at Earlham Institute by members of the
534 Genomics Pipelines Group. We thank Anna Jordan at the JIC insectary for assistance with
535 aphid rearing and identification of aphid morphs.

536 Author contributions

537 T.C.M., S.T.M., Y.C., D.S., S.H. and C.v.O. conceived the study. T.C.M. performed
538 bioinformatics analysis with additional analysis performed by C.v.O, G.K. and Y.C.. S.T.M.
539 performed aphid morph phenotyping and extracted DNA and RNA. L.P.A. constructed
540 bisulphite sequencing libraries. T.C.M, C.v.O and S.H. wrote the manuscript. All authors read,
541 edited and approved the final manuscript.

542 Competing interests

543 The authors declare no competing financial interests.

544 # References

545 Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, Mason CE.
546 2012. methylKit: a comprehensive R package for the analysis of genome-wide DNA
547 methylation profiles. *Genome Biol* **13**: R87.
548 Albritton SE, Kranz AL, Rao P, Kramer M, Dieterich C, Ercan S. 2014. Sex-biased gene
549 expression and evolution of the X chromosome in nematodes. *Genetics* **197**: 865–883.
550 Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful
551 approach to multiple testing. *J R Stat Soc* **57**: 289–300.
552 Bewick AJ, Niederhuth CE, Ji L, Rohr NA, Griffin PT, Leebens-Mack J, Schmitz RJ. 2017. The
553 evolution of CHROMOMETHYLASES and gene body DNA methylation in plants. *Genome*
554 *Biol* **18**: 65.
555 Bewick AJ, Vogel KJ, Moore AJ, Schmitz RJ. 2016. Evolution of DNA methylation across
556 insects. *Mol Biol Evol*.
557 Blackman RL. 1971a. Chromosomal abnormalities in an anholocyclic biotype of *Myzus*
558 *persicae* (Sulzer). *Experientia* **27**: 704–706.
559 Blackman RL. 1971b. Variation in the photoperiodic response within natural populations of
560 *Myzus persicae* (Sulz.). *Bull Entomol Res* **60**: 533–46.
561 Bonasio R, Li Q, Lian J, Mutti NS, Jin L, Zhao H, Zhang P, Wen P, Xiang H, Ding Y, et al. 2012.
562 Genome-wide and caste-specific DNA methylomes of the ants *Camponotus floridanus*
563 and *Harpegnathos saltator*. *Curr Biol* **22**: 1755–1764.
564 Dixon AFG. 1977. Aphid ecology: life cycles, polymorphism, and population regulation. *Annu*
565 *Rev Ecol Syst* **8**: 329–353.
566 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras
567 TR. 2013. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21.
568 Feng S, Cokus SJ, Zhang X, Chen P-Y, Bostick M, Goll MG, Hetzel J, Jain J, Strauss SH, Halpern
569 ME, et al. 2010. Conservation and divergence of methylation patterning in plants and
570 animals. *Proc Natl Acad Sci U S A* **107**: 8689–94.
571 Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh C, Zhang X, Golic KG, Jacobsen SE, Bestor
572 TH. 2006. Methylation of $tRNA^{Asp}$ by the DNA methyltransferase homolog Dnmt2.
573 *Science* **311**: 395–398.
574 Guo H, Zhu P, Yan L, Li R, Hu B, Lian Y, Yan J, Ren X, Lin S, Li J, et al. 2014. The DNA

575 methylation landscape of human early embryos. *Nature* **511**: 606–610.
576 Herb BR, Wolschin F, Hansen KD, Aryee MJ, Langmead B, Irizarry R, Amdam G V, Feinberg
577 AP. 2012. Reversible switching between epigenetic states in honeybee behavioral
578 subcastes. *Nat Neurosci* **15**: 1371–1373.
579 Hunt BG, Brisson JA, Yi S V., Goodisman MAD. 2010. Functional conservation of DNA
580 methylation in the pea aphid and the honeybee. *Genome Biol Evol* **2**: 719–728.
581 Hunt BG, Glastad KM, Yi S V., Goodisman MAD. 2013. The function of intragenic DNA
582 methylation: Insights from insect epigenomes. *Integr Comp Biol* **53**: 319–328.
583 Jaquiéry J, Peccoud J, Ouisse T, Legeai F, Gouin A. 2017. Disentangling the causes for faster-X
584 evolution in aphids. *bioRxiv* 1–33.
585 Jaquiéry J, Rispe C, Roze D, Legeai F, Le Trionnaire G, Stoeckel S, Mieuzet L, Da Silva C,
586 Poulain J, Prunier-Leterme N, et al. 2013. Masculinization of the X chromosome in the
587 pea aphid. *PLoS Genet* **9**.
588 Krueger F, Andrews SR. 2011. Bismark: A flexible aligner and methylation caller for Bisulfite-
589 Seq applications. *Bioinformatics* **27**: 1571–1572.
590 Law J a, Jacobsen SE. 2010. Establishing, maintaining and modifying DNA methylation
591 patterns in plants and animals. *Nat Rev Genet* **11**: 204–220.
592 Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or
593 without a reference genome. *BMC Bioinformatics* **12**: 323.
594 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R.
595 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–
596 2079.
597 Libbrecht R, Oxley PR, Keller L, Kronauer DJC. 2016. Robust DNA methylation in the clonal
598 raider ant brain. *Curr Biol* **26**: 1–5.
599 Liew YJ, Zoccola D, Li Y, Tambutté E, Venn AA, Craig T. 2017. Epigenome-associated
600 phenotypic acclimatization to ocean acidification in a reef-building coral. *bioRxiv*.
601 Maere S, Heymans K, Kuiper M. 2005. BiNGO: A Cytoscape plugin to assess
602 overrepresentation of gene ontology categories in biological networks. *Bioinformatics*
603 **21**: 3448–3449.
604 Marzachi C, Veratti F, Bosco D. 1998. Direct PCR detection of phytoplasmas in
605 experimentally infected insects. *Ann Appl Biol* **133**: 45–54.
606 Mathers TC, Chen Y, Kaithakottil G, Legeai F, Mugford ST, Baa-Puyoulet P, Bretaudeau A,

607 Clavijo B, Colella S, Collin O, et al. 2017. Rapid transcriptional plasticity of duplicated
608 gene clusters enables a clonally reproducing aphid to colonise diverse plant species.
609 *Genome Biol* **18**: 27.
610 Misof B, Liu S, Meusemann K, Peters R. 2014. Phylogenomics resolves the timing and
611 pattern of insect evolution. *Science (80-)* **346**: 763–768.
612 Müller CB, Williams IS, Hardie J. 2001. The role of nutrition, crowding and interspecific
613 interactions in the development of winged aphids. *Ecol Entomol* **26**: 330–340.
614 Nicholson SJ, Nickerson ML, Dean M, Song Y, Hoyt PR, Rhee H, Kim C, Puterka GJ. 2015. The
615 genome of *Diuraphis noxia*, a global aphid pest of small grains. *BMC Genomics* **16**: 429.
616 Okae H, Chiba H, Hiura H, Hamada H, Sato A. 2014. Genome-wide analysis of DNA
617 methylation dynamics during early human development. *PLoS Genet* **10**: 1–12.
618 Patalano S, Vlasova A, Wyatt C, Ewels P, Camara F, Ferreira PG, Asher CL, Jurkowski TP,
619 Segonds-pichon A, Bachman M, et al. 2015. Molecular signatures of plastic phenotypes
620 in two eusocial insect species with simple societies. *PNAS*.
621 Pferdehirt RR, Meyer BJ. 2013. SUMOylation is essential for sex-specific assembly and
622 function of the *Caenorhabditis elegans* dosage compensation complex on X
623 chromosomes. *Proc Natl Acad Sci* **110**: E3810–E3819.
624 Phanstiel DH, Boyle AP, Araya CL, Snyder MP. 2014. Sushi.R: Flexible, quantitative and
625 integrative genomic visualizations for publication-quality multi-panel figures.
626 *Bioinformatics* **30**: 2808–2810.
627 Purandare SR, Bickel RD, Jaquiery J, Rispe C, Brisson J a. 2014. Accelerated evolution of
628 morph-biased genes in pea aphids. *Mol Biol Evol* **31**: 2073–2083.
629 Quinlan AR, Hall IM. 2010. BEDTools: A flexible suite of utilities for comparing genomic
630 features. *Bioinformatics* **26**: 841–842.
631 R Core Team. 2017. R: A Language and Environment for Statistical Computing.
632 Richard G, Legeai F, Prunier-Leterme N, Bretaudeau A, Tagu D, Jaquiéry J, Le Trionnaire G.
633 2017. Dosage compensation and sex-specific epigenetic landscape of the X
634 chromosome in the pea aphid. *Epigenetics Chromatin* **10**: 30.
635 Robinson MD, McCarthy DJ, Smyth GK. 2009. edgeR: A Bioconductor package for differential
636 expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.
637 Schübeler D. 2015. Function and information content of DNA methylation. *Nature* **517**: 321–
638 326.

639 Sloane M a, Sunnucks P, Wilson a C, Hales DF. 2001. Microsatellite isolation, linkage group
640 identification and determination of recombination frequency in the peach-potato
641 aphid, *Myzus persicae* (Sulzer) (Hemiptera: Aphididae). *Genet Res* **77**: 251–60.
642 Srinivasan DG, Brisson J a. 2012. Aphids: a model for polyphenism and epigenetics. *Genet*
643 *Res Int* **2012**: 431531.
644 Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. Revigo summarizes and visualizes long lists of
645 gene ontology terms. *PLoS One* **6**.
646 Suzuki MM, Bird A. 2008. DNA methylation landscapes: Provocative insights from
647 epigenomics. *Nat Rev Genet* **9**: 465–476.
648 Thomas CG, Li R, Smith HE, Woodruff GC, Oliver B, Haag ES. 2012. Simplification and
649 desexualization of gene expression in self-fertile nematodes. *Curr Biol* **22**: 2167–2172.
650 Ureña E, Pirone L, Chafino S, Pérez C, Sutherland JD, Lang V, Rodriguez MS, Lopitz-Otsoa F,
651 Blanco FJ, Barrio R, et al. 2015. Evolution of SUMO function and chain Formation in
652 insects. *Mol Biol Evol* **33**: msv242.
653 van Ham RCHJ, Kamerbeek J, Palacios C, Rausell C, Abascal F, Bastolla U, Fernández JM,
654 Jiménez L, Postigo M, Silva FJ, et al. 2003. Reductive genome evolution in *Buchnera*
655 *aphidicola*. *Proc Natl Acad Sci U S A* **100**: 581–586.
656 Walsh TK, Brisson J a., Robertson HM, Gordon K, Jaubert-Possamai S, Tagu D, Edwards OR.
657 2010. A functional DNA methylation system in the pea aphid, *Acyrthosiphon pisum*.
658 *Insect Mol Biol* **19**: 215–228.
659 Wang HQ, Tuominen LK, Tsai CJ. 2011. SLIM: A sliding linear model for estimating the
660 proportion of true null hypotheses in datasets with dependence structures.
661 *Bioinformatics* **27**: 225–231.
662 Wang X, Werren JH, Clark AG. 2015. Genetic and epigenetic architecture of sex-biased
663 expression in the jewel wasps *Nasonia vitripennis* and *giraulti*. *Proc Natl Acad Sci*
664 201510338.
665 Wang X, Wheeler D, Avery A, Rago A, Choi JH, Colbourne JK, Clark AG, Werren JH. 2013.
666 Function and Evolution of DNA Methylation in *Nasonia vitripennis*. *PLoS Genet* **9**.
667 Wilson A, Sunnucks P, Hales D. 1997. Random loss of X chromosome at male determination
668 in an aphid, *Sitobion* near *fragariae*, detected using an X-linked polymorphic
669 microsatellite marker. *Genet Res* **69**: 233–236.
670 Wilson ACC, Massonnet B, Simon J-C, Prunier-Leterme N, Dolatti L, Llewellyn KS, C FC, C RC,

671 Blackman RL, Estoup A, et al. 2004. Cross-species amplification of microsatellite loci in
672 aphids: assessment and application. *Mol Ecol …* **4**: 104–109.
673 Xiang H, Zhu J, Chen Q, Dai F, Li X, Li M, Zhang H, Zhang G, Li D, Dong Y, et al. 2010. Single
674 base–resolution methylome of the silkworm reveals a sparse epigenomic map. *Nat*
675 *Biotechnol* **28**: 516–520.
676 Xu H-J, Xue J, Lu B, Zhang X-C, Zhuo J-C, He S-F, Ma X-F, Jiang Y-Q, Fan H-W, Xu J-Y, et al.
677 2015. Two insulin receptors determine alternative wing morphs in planthoppers.
678 *Nature* **519**: 464–467.
679 Yan H, Simola DF, Bonasio R, Liebig J, Berger SL, Reinberg D. 2014. Eusocial insects as
680 emerging models for behavioural epigenetics. *Nat Rev Genet* **15**: 677–688.
681 Zemach A, McDaniel IE, Silva P, Zilberman D. 2010. Genome-wide evolutionary analysis of
682 eukaryotic DNA methylation. *Science (80-)* **328**: 916–919.
683 Zemach A, Zilberman D. 2010. Evolution of Eukaryotic DNA methylation and the pursuit of
684 safer sex. *Curr Biol* **20**: R780–R785.
685 Zhang Y, Sturgill D, Parisi M, Kumar S, Oliver B. 2007. Constraint and turnover in sex-biased
686 gene expression in the genus *Drosophila*. *Nature* **450**: 233–237.
687 Zhuo JC, Lei C, Shi JK, Xu N, Xue WH, Zhang MQ, Ren ZW, Zhang HH, Zhang CX. 2017. Tra-2
688 mediates cross-talk between sex determination and wing polyphenism in female
689 *Nilaparvata lugens*. *Genetics* **207**: 1067–1078.