Does the host drive Wolbachia gene expression?

Stella Papaleo[1$], Simona Panelli[1], Ibrahim Bitar[2], Lodovico Sterzi[1], Riccardo Nodari[1,3], Francesco Comandatore[1*$]

1. Department of Biomedical and Clinical Sciences, Pediatric Clinical Research Center "Romeo and Enrica Invernizzi", University of Milan, 20157 Milan, Italy.
2. Biomedical Center, Faculty of Medicine, Charles University, Pilsen, Czechia.
3. Istituto Nazionale di Genetica Molecolare (INGM) "Romeo and Enrica Invernizzi", University of Milan, 20157 Milan, Italy.

* Corresponding author

$ The authors contributed equally

Wolbachia pipientis is an endosymbiont bacterium commonly found in association with arthropods and filarial nematodes. Several studies disclosed the main role of Wolbachia in host physiology. Reasonably, the maintenance of a strict symbiosis requires coordination between Wolbachia and host gene expressions. Experimental studies show Wolbachia gene expression varies during host development or after environmental stresses. Despite this, the mechanism behind the regulation of Wolbachia gene expression is still not clear, since their genomes have lost most promoters during genome reduction. The first published RNA-Seq study on Wolbachia showed the differential expression of a DNA methyltransferase, an enzyme able to methylate adenine or cytosine on specific patterns and often involved in the regulation of bacteria gene expression. In this work, we tested the hypothesis that DNA methylation can affect gene expression in Wolbachia. We experimentally characterised the

Wolbachia DNA methyltransferase and re-analysed RNA-Seq data from six studies on four hosts. We surprisingly found that the Wolbachia DNA methyltransferase is the only so far described able to methylate both adenine and cytosine without highly specific patterns. We also found that the nucleotide content of the genes correlates with their expressions, with a pattern compatible to be a consequence of DNA methylation. Lastly, we found a conserved eukaryotic-like promoter upstream of the Wolbachia DNA methyltransferase gene. Overall, these results may suggest a cascade mechanism in which the host induces the Wolbachia DNA methyltransferase expression, which in turn affects the bacterium gene expression.

Introduction

Wolbachia pipientis (from here Wolbachia) is probably one of the most interesting and investigated endosymbiont bacteria, colonising several species of arthropods and filarial nematodes[1,2]. In arthropods, the bacterium can manipulate the reproduction of the host[3], as well as protect it from viruses[4,5] or provide nutrients[6]. In filarial nematodes, genomic analyses and experimental evidence revealed that Wolbachia is essential for the host survival[7]. The bacterium has typically an intracellular tropism and is mainly mother-to-offspring transmitted, even if horizontal transfer among host individuals also occurs[8].

As for many other strictly host-associated bacteria, Wolbachia genome underwent an important reduction during its evolution, reaching a size that ranges from ~0.8Mb to ~1.8Mb (BV-BRC representative genomes)[9]. As observed for several endosymbiont bacterial species, genome reduction passes through the increasing of the number of mobile elements (e.g. Insertion Sequences) and the loss of unnecessary genes[10,11]. During this process, the structure of the genome and order of the genes can drastically change[11], leading to the disruption of several operons and loss of most gene promoters. Interestingly, RNA-seq studies highlighted that Wolbachia gene expression can be influenced by environmental/development hosts conditions[12–18]. This implies that Wolbachia likely possesses a mechanism for regulating gene expression, despite the low number of promoters indicating that it may not rely solely on transcription factors.

The first published Wolbachia RNA-seq study[12] compared the gene expression of Wolbachia isolated from the soma and gonad in females and males of the filarial nematode *Onchocerca ochengi*. One of the strongest differential expressions was reported for a gene annotated as a SAM-dependent DNA methyltransferase. Methyltransferases are enzymes able to methylate adenine or cytosine on specific sequence patterns of bacterial DNA; for instance, Dam methyltransferase methylates the adenine residues in the sequence GATC and Dcm

methyltransferase methylates the second cytosine of the sequences CCAGG and CCTGG [19,20]. Despite DNA methylation being a well understood mechanism of gene expression regulation in eukaryotes (i.e. epigenetics), its role in prokaryotes physiology is still poorly investigated and no studies have been published on Wolbachia. At the state of the art, there is evidence that DNA methyltransferase enzymes are involved in fundamental bacterial processes, including proofreading during DNA replication, protection from foreign DNA and regulation of gene expression. There is also evidence that DNA methylation of bacterial promoters can affect the expression of the corresponding gene by changing the RNA polymerase affinity[19,20].

In this study, we investigated the hypothesis that the regulation of Wolbachia gene expression involves DNA methylation. We experimentally determined the methylation pattern of the SAM-dependent DNA methyltransferase enzyme of the Wolbachia endosymbiont of *O. ochengi* (wOo). After determining the methylation pattern, we tested whether the frequency of this pattern into gene sequences correlates with their expression, and whether the intensity of the phenomenon is related to the expression of the SAM-dependent DNA methyltransferase gene. Finally, we identified a highly conserved putative eukaryotic-like promoter upstream of the SAM-dependent DNA methyltransferase gene on several Wolbacha genomes.

Our results show that Wolbachia harbours a SAM-dependent DNA methyltransferase enzyme with an peculiar methylation pattern. The re-analysis of six published RNA-Seq studies led us to hypothesise that this enzyme could have a key role in the regulation of gene expression in Wolbachia. Moreover, we propose that the expression of this SAM-dependent DNA methyltransferase could be regulated by Wolbachia hosts.

Material and methods

Experimental determination of the methyltransferase methylation pattern

The WOo DNA methyltransferase was artificially synthesised with codon usage adaptation and cloned into the XhoI and SalI sites of the pCOLD III expression vector (Takara Bio). The recombinant construct was transformed into *Escherichia coli* StellarTM Competent Cells (dam–/dcm–) (Takara Bio) according to the supplier's protocol (ClonTech; Protocol-at-a-Glance, PT5056-2). The *E.coli* strains were cultured in LB broth medium supplemented with ampicillin. Before inducing expression, PCR with custom primers targeting the gene of interest was performed on bacterial crude extracts to verify the outcome of transformation. MTase expression was then induced at 15°C according to the Cold Shock Expression System pColdTM DNA's protocol (Takara Bio). 24 hours after the induction, the expression of the WOo DNA methyltransferase gene was verified by SDS-PAGE. DNA was then extracted using the NucleoSpin Microbial DNA Mini kit (Macherey Nagel) and subsequently subjected to long reads sequencing using Sequel I (Pacific biosciences). Assembly of reads was done using the Microbial Assembly pipeline provided by SMRT Link v.10. With a minimum seed coverage of 30x. The assembled genome was used as a reference in the downstream analysis. Methylation status of each base of the obtained reads was then determined using the 'Base modification Analysis' pipeline included in SMRT-Portal. The pipeline will perform base modification and modified base motifs detection. The control experiment was also carried out repeating the same procedure but using a recombinant construct without the wOo DNA methyltransferase gene inserted. The 95 quantile of Mean QV and coverage values obtained from the control experiment were set as threshold for the identification of methylated bases on the transformed E. coli strain. The analyses were carried out using R.

Dataset reconstruction and data normalisation

Table S2 and gff map files of the supplementary information of Chung et al. 2020 were retrieved [21]. The table contains the gene expression quantifications computed by Chung and colleagues (2020) re-analysing seven RNA-Seq studies on Wolbachia [12–18], for a total of 128 samples (replicates) from 62 host conditions. More in details: i) four samples from two conditions on Wolbachia endosymbiont of *Onchocerca ochengi* (wOo) from Darby et al. 2012[12]; ii) five samples from five conditions on Wolbachia endosymbiont of *Dirofilaria immitis* (wDi) from Luck et al., 2014[13]; iii) six samples from two conditions on Wolbachia endosymbiont of *Drosophila melanogaster* (wMel), from Darby et al., 2014[14]; iv) 12 samples from seven conditions on wDi from Luck et al., 2015[15]; v) 55 samples from 24 conditions on wMel, from Gutzwiller et al., 2015[16]; vi) 14 samples and seven conditions on Wolbachia endosymbiont of *Brugia malayi* (wBm) from Grote et al., 2017[17]; vii) 32 samples from 15 conditions on wBm from Chung et al., 2019[18]. Preliminarily, we excluded from the dataset all the samples from Luck et al., 2015[15] because only 8% of the gene expression quantifications were greater than 0. Thus, gene expression analysis was performed on a total of 116 samples from 55 host conductions, two for wOo, five for wDi, 22 for wBm and 26 for wMel.

Gene expression values were normalised to make them comparable among studies and conditions. For each condition in each study the gene expression values were expressed as

$$\text{Log2expr} = \sum_s [\sum_i \log2(\text{expr}_{s,i} / \sum_i \text{expr}_{s,i})] / S_{num}$$

Where:

s = sample (or replicate); i = gene; $\text{expr}_{s,i}$ = the gene expression value of the gene i in the sample (replicate) s; S_{num} = total number of samples. This normalisation formula derives from the log2 fold change formula often used to compare two conditions in RNA-Seq studies:

$$\text{Log2FC} = \text{Log2expr}_{condition1} - \text{Log2expr}_{conditio21} =$$

$$\{\sum_s [\sum_i \log2(\text{expr}_{s,i} / \sum_i \text{expr}_{s,i})] / S\}_{condition1} - \{\sum_s [\sum_i \log2(\text{expr}_{s,i} / \sum_i \text{expr}_{s,i})] / S\}_{condition2}$$

Log2FC allows us to compare values from different conditions in the same experiment. To make the normalised values more comparable among the studies, the values were converted in percentiles. Indeed, the percentile may be less biassed by sample specific features, such as the number of genes.

Genomic analyses

The genome assemblies of wBm (AE017321.1), wDi (www.nematodes.org: wDi 2.2), wOo (HE660029.1) and wMel (AE017196.1) were retrieved and passed to Prodigal [22] for Open Reading Frame (ORF) calling. For each ORF, Codon Bias Index (CBI) (which measure how much a gene uses a subset of optimal codons, ranging from 0 for random using to 1 for maximum bias), the effective number of codons (Nc) and the orf length were computed using CodonW tool (Peden J.F. Analysis of codon usage. 1999. PhD Thesis, University of Nottingham, UK), while %A, %T, %G and %C were computed using an in-house Perl script.

Statistical analyses

The relationship between each genomic parameter (CBI, Nc, length, %A, %T, %G and %C, see above) and gene expression was investigated. For each of the 55 samples included in the study, the correlation with gene expression and each genomic parameter was tested using linear regression. For each regression line, the slope and the p value were obtained. To test the hypothesis that DNA methylation could affect the gene expression on the basis of thor gene nucleotide composition, the correlations between the methyltransferase gene expression and regression line slope were computed for each genomic parameter.

The presence of other genes with expression patterns comparable to that of DNA MTase gene was tested as follows. Exploiting the gene orthologous information included in the Table S6 of Chung et al., 2020 [21], we performed a Spearman co-correlation among the gene expression pattern (among the 55 samples) of the 546 single copy core genes shared

among wOo, wDi, wBm and wMel. We also compared the expression patterns of the genes by Principal Component Analysis (PCA), using R.

Analysis of the region upstream the methyltransferase gene

The 56 complete Wolbachia genome assemblies (see Table S1) available on Genome Taxonomy DataBase (GTDB) were retrieved. Genome assemblies were BLASTn searched against the nucleotide sequence of the wOo DNA MTase gene (HE660029.1, locus tag wOo_00570), and hits with coverage (i.e. hit length / slen) > 0.9 were selected. The 100 nucleotides sequence upstream the methyltransferase gene were then extracted on the basis of the BLASTn results and identical sequences were merged, using an in-house Perl script. The obtained sequences were manually visualised on http://www.vardb.org/vardb/alignments/view.html. The highly conserved region was then extracted and tested to be an eukaryotic or prokaryotic promoter for *D. melanogaster* using the Neural Network Promoter Prediction tool [23] available on the Berkeley Drosophila Genome Project website (https://www.fruitfly.org/).

Results

Experimental determination of the DNA Methyltransferase methylation pattern

The cloning and expression of the wOo methyltransferase into the Stellar E. coli dam-/dcm- strain and the PacBio sequencing, allowed to identify 166 high quality methylated bases: 151 classified as m4C (N4-methylcytosine) and 15 as m6A (N6-methyladenine) (see Table S2 and Figure S1). The investigation on the methylation pattern revealed that m4c methylations occurred in a non well conserved pattern, while the m6A methylation involved a region rich in guanine in positions -4, -8 and -19 (Figure 1).

Dataset reconstruction

For each of the six datasets included in the study, the normalised gene expression values were obtained as described in material and methods. The ORF sequences of the wOo, wDi, wBm and wMel genome assemblies were analysed to determine the CBI, Nc, length, %A, %T, %G and %C. The obtained information about gene nucleotide composition was merged to relative gene expression data (Supplemental information available on repository...).

Statistical analyses

The slope and p value of the linear regression analyses performed between the seven genetic parameters (CBI, Nc, gene length, %A, %T, %G and %C) and gene expression are reported in Table S3. Length and %G significantly correlate with gene expression in all 55 samples included in the study (100%), while %A in 53/55 (96%), CBI 42/55 (78%), %T 40/55 (73%), %C 22/55 (40%) and Nc 7/55 (13%). The obtained slopes for the regression lines range as follows: for gene length range from 8.8×10^{-5} to 3×10^{-4}, %G from 0.008 to 0.053, CBI from 0.3 to 0.7, %T from -0.01 to 0.01, %C from -0.01 to 0.02 and Nc from 0.002 to 0.008.

For each of the seven genetic parameters, the correlation between the expression of the DNA MTase gene and the regression slope was tested by linear regression. The correlations obtained were significant for all parameters except %C, which had a p-value of 0.08. For %G, %C, %A, Nc, and length, the slope increases with higher expression of the DNA MTase. Conversely, %T and CBI exhibit a decreasing trend with increasing DNA MTase expression (Figures 2-4 and Figures S2-S5).

The co-correlation analysis among the expression patterns of the single copy core genes shared among wOo, wDi, wBm and wMel, showed that no other gene has an expression pattern very similar to that of the DNA MTase gene, indeed the highest correlation value was 0.84, and the median value was 0.02 (Figure S6). In general, the absence of clusters of co-expressed genes is also evident from PCA analysis (Figure S7).

Analysis of the region upstream the methyltransferase gene

The analysis of the 100pb region upstream the DNA MTase gene on 56 complete Wolbachia genome assemblies led to the identification of a highly conserved region of about 60 pb containing three TATA box sequences (Figure 5). The Neural Network Promoter Prediction tool available on the Berkeley Drosophila Genome Project website, confirmed that this region contains a putative eukaryotic promoter (score 0.95), returning also a negative result for putative prokaryotic promoter searching.

Discussion

Wolbachia pipientis is one of the most diffused endosymbiont bacteria [8]. It evolved symbiotic relationships with arthropods and filarial nematodes, becoming often essential for the survival and reproduction of its hosts [1,24]. The close symbiotic relationship with the host has had important consequences on the bacterial genome structure, including an important reduction [25], with frequent rearrangements and the loss of non-essential regions [11], which may cause the disruption of several operons and promoters. The loss of several promoters could have limited evolutionary costs, considering that the intracellular tropism of the bacterium probably protects it from external environment stresses. However, it is reasonable that such a close symbiotic relationship requires coordination between gene expression of Wolbachia and the host. Indeed, experimental studies have shown the existence of Wolbachia genes differentially expressed in host samples from different body compartments, developmental stages and/or from hosts subjected to environmental stress [21]. Nevertheless, at the state of the art, the gene regulation mechanisms in Wolbachia are poorly understood.

The first RNA-Seq study published in Wolbachia investigated differential gene expression in the endosymbiont of the filarial nematode *Onchocerca ochengi* (wOo), comparing host gonad and soma, from males and females. One of the most differentially expressed genes was annotated as a DNA methyltransferase (DNA MTase) [12]. This class of enzymes

catalyses the transfer of a methyl group from S-adenosyl-l-methionine (SAM) to adenine or cytosine in DNA, recognizing specific sequence patterns. Methylation of bacterial gene promoters is known to influence the promoters affinity for transcription factors, leading to changes in gene expression and epigenetic modifications [19,26].

In this work, we investigated the role of Wolbachia DNA MTase as a mediator of gene expression regulation. In absence of promoters, we hypothesise that the intra-genes DNA methylation could affect gene expression. To test this hypothesis, we i) experimentally determined the sequence pattern recognised by the wOo DNA MTase; ii) studied expression data of 55 samples from six published RNA-Seq studies on Wolbachia; iii) investigated the presence of putative eukaryotic-like promoters upstream the DNA MTase gene.

In silico determination of the sequence pattern recognized by DNA MTase is challenging [27]. This is especially true in this case, considering that this is the first Wolbachia DNA MTase enzyme ever characterised. An experimental method to determine the enzyme methylation pattern is to express the DNA MTase in an *Escherichia coli* strain depleted for any DNA MTase gene, and then identify the methylated bases and motifs using SMART PacBio sequencing [27]. Surprisingly, the wOo DNA MTase was found to be able to methylate both adenine and cytosine, without highly conserved patterns (Figure 1). Our findings also suggest that the enzyme has a 10-fold greater affinity for cytosine than adenine. To our knowledge this is the first DNA MTase enzyme described with these peculiar features.

Regarding the analysis of gene expression, we started from the idea that if DNA methylation within the gene affects its expression, we should observe a correlation between gene nucleotide composition and gene expression. Furthermore, the direction (positive or negative) and the strength (the slope of the regression line) of this correlation should change with variation of DNA MTase gene expression. It is well documented in literature that gene length and codon usage could be correlated to gene expression [27,28]. In particular, highly

expressed genes tend to have lower Effective Number of Codons (Nc) and higher Codon Bias Index (CBI). For each of the 55 samples, we computed the correlation between CBI, Nc, gene length, %A, %T, %G and %C versus the relative gene expression. The result shows correlation between gene nucleotide composition and gene expression in several samples. Interestingly, we found that the slope of the obtained regression lines often correlates with the expression of the DNA MTase in that sample.

In more detail, the expression of genes is positively correlated with their %G in all samples and the correlation slope increases proportionally to the expression of DNA MTase (Figure 2). Regarding %T, the correlation with gene expression was positive (the higher the %T the higher the gene expression) when DNA methyltransferase is poorly expressed, while it became negative (the higher the %T the lower the gene expression) in the sample with high DNA MTase expression (Figure 3). Furthermore, it is interesting to note that with poorly expressed DNA MTase, gene expression correlates positively with the Codon Bias Index (CBI); the correlation becomes weaker with the increasing DNA MTase expression (Figure 4), suggesting that the DNA MTase effect on gene expression may be independent from the adaptation of codon usage to gene expression. Conventionally, gene sequences are complementary to the transcribed DNA strand, thus corresponding to the messenger RNA sequence. It means that the observed correlation between gene expression and %G and %T may be interpreted as the effects of the methylation of cytosine and adenine on the transcribed strand. The positive to negative slope trend observed for %T, is suggestive of the importance that methylation of adenine may have in the regulation of gene expression, in particular considering the AT-richness of the Wolbachia genome.

The step forward in our analysis was to investigate how the expression of the DNA MTase gene can be regulated. We analysed a collection of representative Wolbachia genomes, finding a highly conserved genomic region located upstream of the DNA MTase gene, containing TATA box regulatory sequences (Figure 5). TATA box sequences are normally found in the core promoter region of a minority of eukaryotic genes, particularly those involved

in responses to stress and changes in environmental conditions. Indeed, housekeeping genes tend to be not regulated by TATA-containing promoters, in opposition to those expressed only in specific environmental conditions that tend to be more highly regulated [29,30].

Our hypothesis suggests that host organisms may influence Wolbachia gene expression by regulating the expression of bacterial DNA MTase via TATA boxes located upstream of the gene. Consequently, the activity of the DNA MTase enzyme could potentially impact the expression of numerous other genes within the Wolbachia genome.

While this mechanism may not lead to highly precise regulation of gene expression, it could offer a level of precision adequate for the establishment of successful symbiosis with intracellular bacteria. It would be interesting to investigate whether similar mechanisms are present in other endosymbiont or intracellular organisms.

Despite the results presented in this work are far from being definitive, we hope to have opened a new way in understanding the symbiotic relationship between Wolbachia and its hosts. In our opinion, this is a magnificent example of how evolution works, recycling metabolic pieces for similar aims, sometimes reaching surprising stable *equilibria*.

Captions

Figure 1. Methylation patterns of the DNA methyltransferase gene of Wolbachia endosymbiont of *Onchocerca ochengi* (wOo)

Sequence logo of the sequences flanking (from -20 to +20 positions) the methylated positions found in the genome of Stellar *Escherichia coli* dam-/dcm- strain expressing the DNA methyltransferase gene of Wolbachia endosymbiont of *Onchocerca ochengi* (wOo). The height of the letters depends on the conservation of the position among the sequences. a) the sequence logo of the m4C methylations on cytosines; b) the m6A methylation of adenine.

Figure 2. Correlation between the expression of genes and their percentage of guanine

For each of the 55 samples included in the study, the correlation between the expression of the genes and the relative %G was investigated. The scatter plots of %G vs gene expression of three exemplificative samples selected from the 55 samples are reported: a) the correlation plot with highest slope (Hs), b) the correlation plot with the medium slope (Ms) and c) the correlation plot with the lowest slope (Ls). In each of these three plots, the regression line is coloured in dark grey and the statistics are reported on the chart. In figure d, the correlation plot between the slopes obtained for each of the 55 samples, and the relative expression of the DNA methyltransferase gene is reported. Points relative to significant correlations (p value < 0.05) are in green while non significant in grey. The regression line is in green and the statistics are reported on the chart.

Figure 3. Correlation between the expression of genes and their percentage of thymine

For each of the 55 samples included in the study, the correlation between the expression of the genes and the relative %T was investigated. The scatter plots of %T vs gene expression of three exemplificative samples selected from the 55 samples are reported: a) the correlation plot with highest slope (Hs), b) the correlation plot with the medium slope (Ms) and c) the correlation plot with the lowest slope (Ls). In each of these three plots, the regression line is coloured in dark grey and the statistics are reported on the chart. In figure d, the correlation plot between the slopes obtained for each of the 55 samples, and the relative expression of the DNA methyltransferase gene is reported. Points relative to significant correlations (p value < 0.05) are in green while non significant in grey. The regression line is in green and the statistics are reported on the chart.

Figure 4. Correlation between the expression of genes and the relative Codon Bias Index (CBI)

For each of the 55 samples included in the study, the correlation between the expression of the genes and the relative Codon Bias Index (CBI) was investigated. The scatter plots of CBI vs gene expression of three exemplificative samples selected from the 55 samples are reported: a) the correlation plot with highest slope (Hs), b) the correlation plot with the medium slope (Ms) and c) the correlation plot with the lowest slope (Ls). In each of these three plots, the regression line is coloured in dark grey and the statistics are reported on the chart. In figure d, the correlation plot between the slopes obtained for each of the 55 samples, and the relative expression of the DNA methyltransferase gene is reported. Points relative to significant correlations (p value < 0.05) are in green while non significant in grey. The regression line is in green and the statistics are reported on the chart.

Figure 5. Genomic region upstream the DNA methyltransferase gene in 56 Wolbachia genomes

The sequences upstream of the DNA methyltransferase gene on 56 high quality Wolbachia genomes are shown, after identical sequences were merged. The colour of the base's background (from white to blue) reports the level of conservation of the position. The position of the TATA-boxes are indicated by coloured rectangles on the bottom, and the relative putative translation start point (position +25) were highlighted on the alignment using the same colours of the relative TATA-boxes. The ATG starting codon of the DNA methyltransferase gene is highlighted in violet.

Funding

The study was supported by the National Institute of Virology and Bacteriology (Program EXCELES, ID project no. LX22NPO5103) funded by the European Union–Next Generation EU.

References

1. Werren, J. H., Baldo, L. & Clark, M. E. Wolbachia: master manipulators of invertebrate biology. *Nat. Rev. Microbiol.* **6**, 741–751 (2008).
2. Zug, R. & Hammerstein, P. Still a host of hosts for Wolbachia: analysis of recent data suggests that 40% of terrestrial arthropod species are infected. *PLoS One* **7**, e38544 (2012).
3. Werren, J. H. Biology of Wolbachia. *Annu. Rev. Entomol.* **42**, 587–609 (1997).
4. Hedges, L. M., Brownlie, J. C., O'Neill, S. L. & Johnson, K. N. Wolbachia and virus protection in insects. *Science* **322**, 702 (2008).
5. Teixeira, L., Ferreira, A. & Ashburner, M. The bacterial symbiont Wolbachia induces

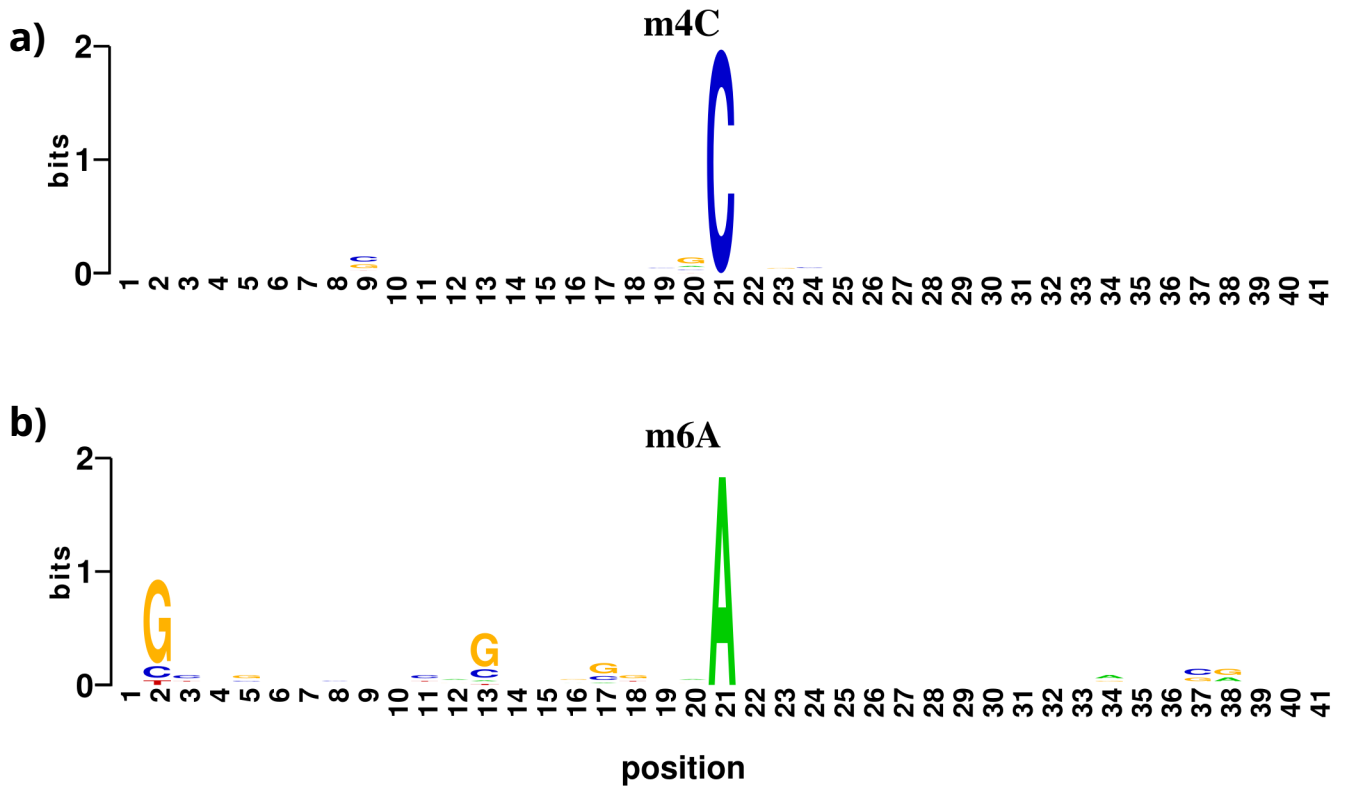
resistance to RNA viral infections in Drosophila melanogaster. *PLoS Biol.* **6**, e2 (2008).

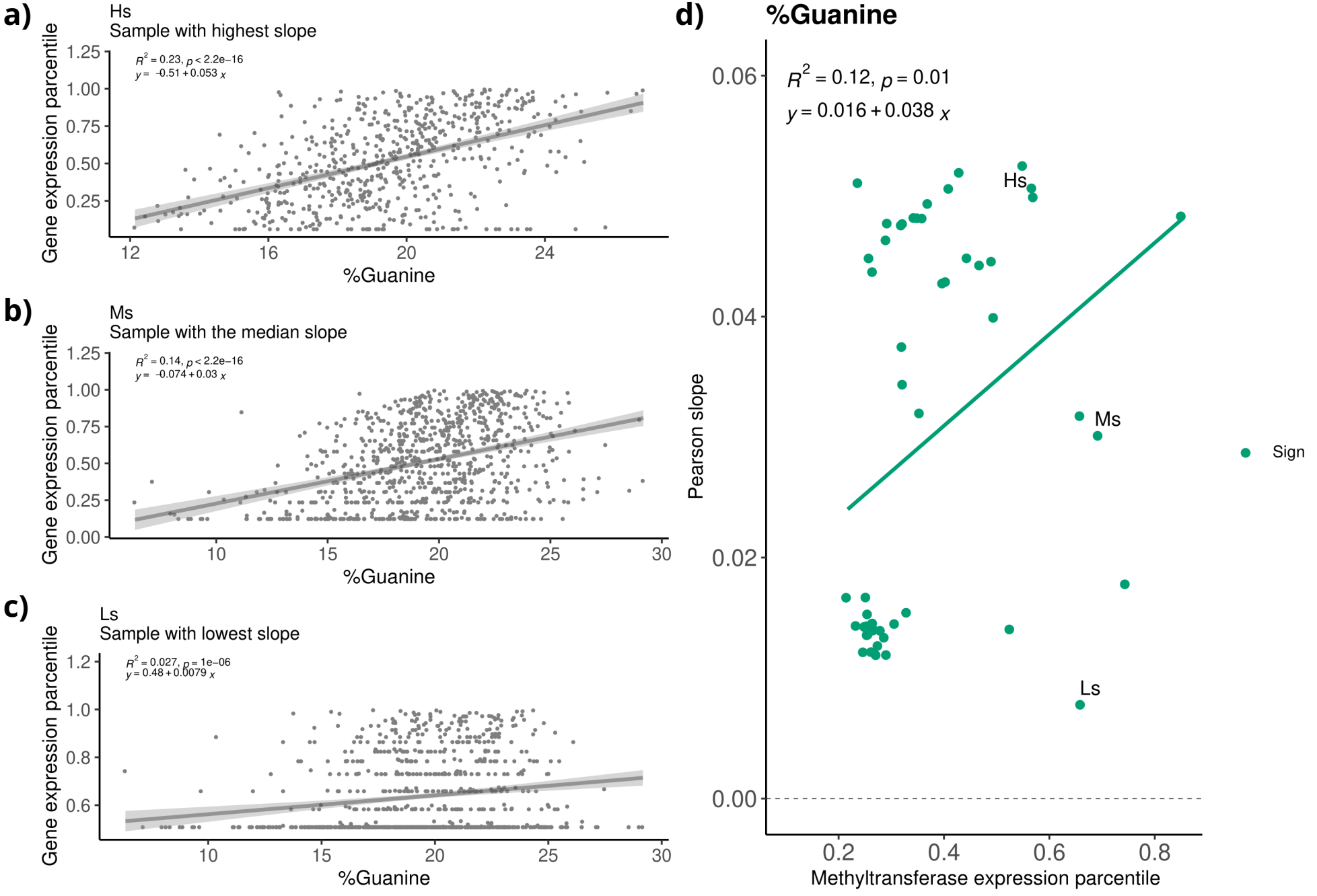
6. Newton, I. L. G. & Rice, D. W. The Jekyll and Hyde Symbiont: Could Be a Nutritional Mutualist? *J. Bacteriol.* **202**, (2020).
7. Bandi, C. *et al.* Effects of tetracycline on the filarial worms Brugia pahangi and Dirofilaria immitis and their bacterial endosymbionts Wolbachia. *Int. J. Parasitol.* **29**, 357–364 (1999).
8. Scholz, M. *et al.* Large scale genome reconstructions illuminate Wolbachia evolution. *Nat. Commun.* **11**, 5235 (2020).
9. Olson, R. D. *et al.* Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR. *Nucleic Acids Res.* **51**, D678–D689 (2023).
10. Manzano-Marín, A. & Latorre, A. Snapshots of a shrinking partner: Genome reduction in Serratia symbiotica. *Sci. Rep.* **6**, 32590 (2016).
11. Comandatore, F. *et al.* Supergroup C Wolbachia, mutualist symbionts of filarial nematodes, have a distinct genome structure. *Open Biol.* **5**, 150099 (2015).
12. Darby, A. C. *et al.* Analysis of gene expression from the Wolbachia genome of a filarial nematode supports both metabolic and defensive roles within the symbiosis. *Genome Res.* **22**, 2467–2477 (2012).
13. Luck, A. N. *et al.* Concurrent transcriptional profiling of Dirofilaria immitis and its Wolbachia endosymbiont throughout the nematode life cycle reveals coordinated gene expression. *BMC Genomics* **15**, 1041 (2014).
14. Darby, A. C. *et al.* Integrated transcriptomic and proteomic analysis of the global response of Wolbachia to doxycycline-induced stress. *ISME J.* **8**, 925–937 (2014).
15. Luck, A. N. *et al.* Tissue-specific transcriptomics and proteomics of a filarial nematode and its Wolbachia endosymbiont. *BMC Genomics* **16**, 920 (2015).
16. Gutzwiller, F. *et al.* Dynamics of Wolbachia pipientis Gene Expression Across the Drosophila melanogaster Life Cycle. *G3* **5**, 2843–2856 (2015).
17. Grote, A. *et al.* Defining Brugia malayi and Wolbachia symbiosis by stage-specific dual

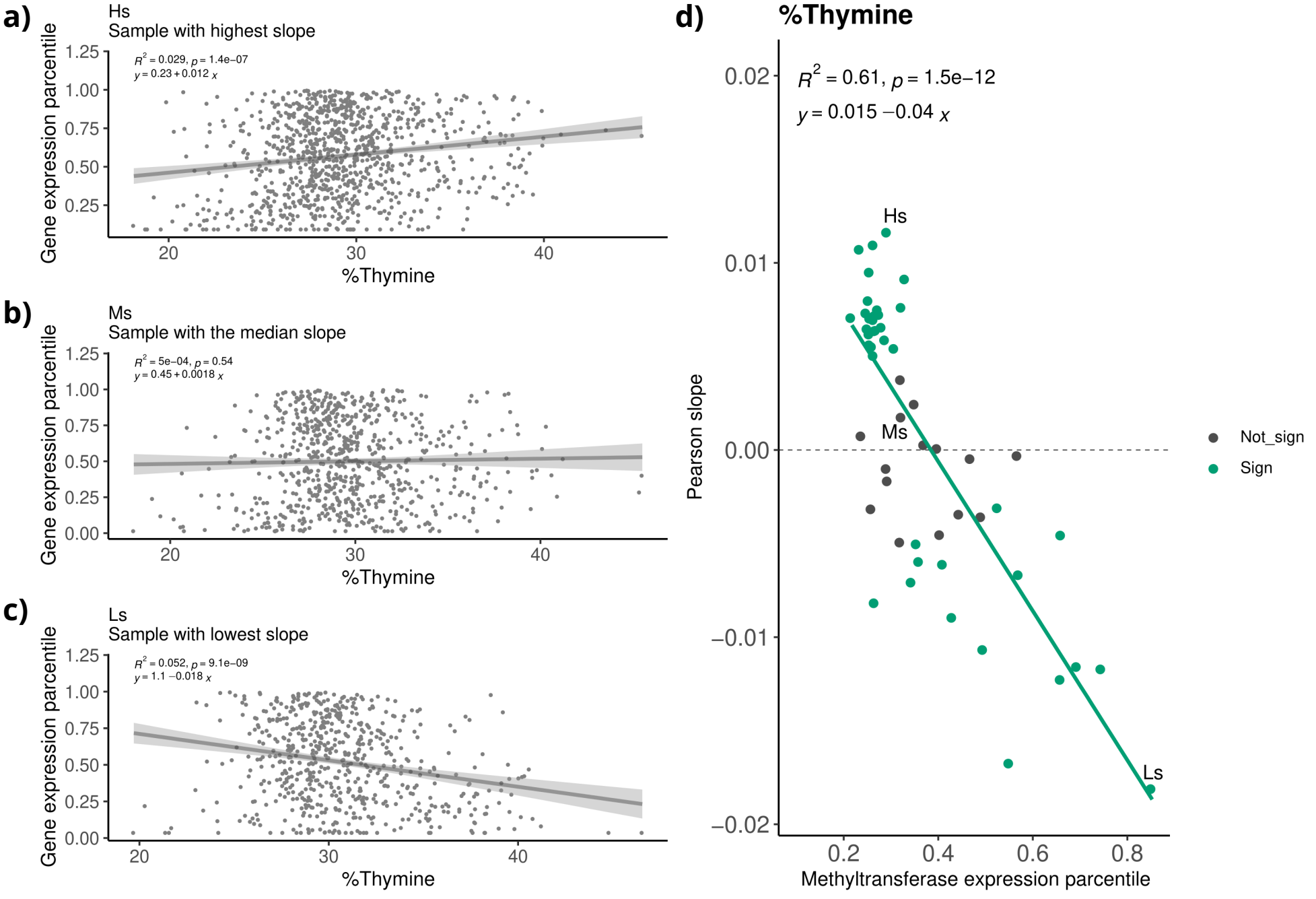
RNA-seq. *PLoS Negl. Trop. Dis.* **11**, e0005357 (2017).

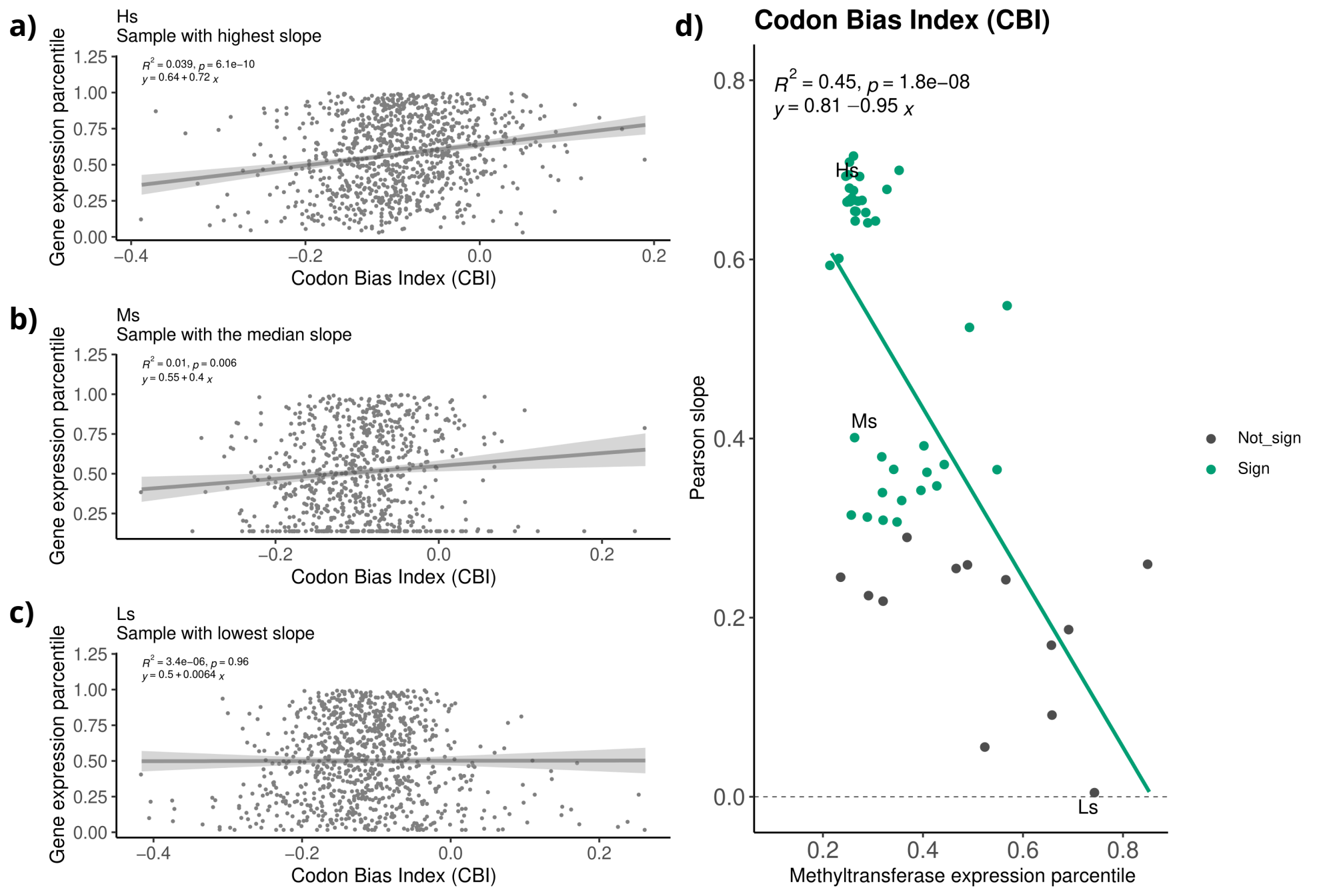
18. Chung, M. *et al.* Drug Repurposing of Bromodomain Inhibitors as Potential Novel Therapeutic Leads for Lymphatic Filariasis Guided by Multispecies Transcriptomics. *mSystems* **4**, (2019).
19. Papaleo, S. *et al.* The red thread between methylation and mutation in bacterial antibiotic resistance: How third-generation sequencing can help to unravel this relationship. *Front. Microbiol.* **13**, 957901 (2022).
20. Sánchez-Romero, M. A. & Casadesús, J. The bacterial epigenome. *Nat. Rev. Microbiol.* **18**, 7–20 (2020).
21. Chung, M. *et al.* A Meta-Analysis of Transcriptomics Reveals a Stage-Specific Transcriptional Response Shared Across Different Hosts. *G3* **10**, 3243–3260 (2020).
22. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
23. Reese, M. G. Application of a time-delay neural network to promoter annotation in the Drosophila melanogaster genome. *Comput. Chem.* **26**, 51–56 (2001).
24. Comandatore, F. *et al.* Phylogenomics and analysis of shared genes suggest a single transition to mutualism in Wolbachia of nematodes. *Genome Biol. Evol.* **5**, 1668–1674 (2013).
25. Dale, C. & Moran, N. A. Molecular Interactions between Bacterial Symbionts and Their Hosts. *Cell* **126**, 453–465 (2006).
26. Gao, Q. *et al.* Bacterial DNA methyltransferase: A key to the epigenetic world with lessons learned from proteobacteria. *Front. Microbiol.* **14**, 1129437 (2023).
27. Jensen, T. Ø. *et al.* Genome-wide systematic identification of methyltransferase recognition and modification patterns. *Nat. Commun.* **10**, 3311 (2019).
28. Quax, T. E. F., Claassens, N. J., Söll, D. & van der Oost, J. Codon Bias as a Means to Fine-Tune Gene Expression. *Mol. Cell* **59**, 149–161 (2015).
29. Basehoar, A. D., Zanton, S. J. & Pugh, B. F. Identification and distinct regulation of yeast TATA box-containing genes. *Cell* **116**, 699–709 (2004).

30. Bae, S.-H., Han, H. W. & Moon, J. Functional analysis of the molecular interactions of TATA box-containing genes and essential genes. *PLoS One* **10**, e0120848 (2015).


a)
m4C
bits
b)
m6A
bits
position


a)
Hs
Sample with highest slope
$R^2 = 0.23$, $p < 2.2e-16$
$y = -0.51 + 0.053\,x$
Gene expression parcentile
%Guanine
b)
Ms
Sample with the median slope
$R^2 = 0.14$, $p < 2.2e-16$
$y = -0.074 + 0.03\,x$
Gene expression parcentile
%Guanine
c)
Ls
Sample with lowest slope
$R^2 = 0.027$, $p = 1e-06$
$y = 0.48 + 0.0079\,x$
Gene expression parcentile
%Guanine
d)
%Guanine
$R^2 = 0.12$, $p = 0.01$
$y = 0.016 + 0.038\,x$
Hs
Ms
Ls
Pearson slope
Methyltransferase expression parcentile
Sign


a)
Hs
Sample with highest slope
$R^2 = 0.029$, $p = 1.4e-07$
$y = 0.23 + 0.012\,x$
Gene expression parcentile
%Thymine
b)
Ms
Sample with the median slope
$R^2 = 5e-04$, $p = 0.54$
$y = 0.45 + 0.0018\,x$
Gene expression parcentile
%Thymine
c)
Ls
Sample with lowest slope
$R^2 = 0.052$, $p = 9.1e-09$
$y = 1.1 - 0.018\,x$
Gene expression parcentile
%Thymine
d)
%Thymine
$R^2 = 0.61$, $p = 1.5e-12$
$y = 0.015 - 0.04\,x$
Hs
Ms
Ls
Pearson slope
Methyltransferase expression parcentile
Not_sign
Sign


a)
Hs
Sample with highest slope
$R^2 = 0.039$, $p = 6.1e-10$
$y = 0.64 + 0.72\ x$
Gene expression parcentile
Codon Bias Index (CBI)
b)
Ms
Sample with the median slope
$R^2 = 0.01$, $p = 0.006$
$y = 0.55 + 0.4\ x$
Gene expression parcentile
Codon Bias Index (CBI)
c)
Ls
Sample with lowest slope
$R^2 = 3.4e-06$, $p = 0.96$
$y = 0.5 + 0.0064\ x$
Gene expression parcentile
Codon Bias Index (CBI)
d)
Codon Bias Index (CBI)
$R^2 = 0.45$, $p = 1.8e-08$
$y = 0.81 - 0.95\ x$
Hs
Ms
Ls
Pearson slope
Methyltransferase expression parcentile
Not_sign
Sign

Eukaryotic-like promotor of Wolbachia DNA methyltransferase gene