Evidence that host-mediated epigenetic modifications regulate gene expression in *Wolbachia pipientis*

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

Abstract

Wolbachia pipientis is an obligate intracellular bacterium, associated with several arthropods and filarial nematodes. *Wolbachia* establishes strict symbiotic relationships with its hosts, with the consequent loss of many genes and regulatory regions. Despite this, experimental studies show that *Wolbachia* gene expression is coordinated to host needs, but the mechanism is still unknown. The first published RNA-Seq study on *Wolbachia* evidenced a strong differential expression of a DNA methyltransferase (MTase). In bacteria, this enzyme methylates either adenines or cytosines on specific motifs, contributing to the regulation of gene expression. In this work, we tested the hypothesis that the activity of MTase modulates the expression of *Wolbachia* genes. We first determined the methylation motif of the *Wolbachia* MTase by expressing it in *Escherichia coli*. Surprisingly, the experiment revealed that the *Wolbachia* MTase methylates both adenine and cytosine, without recognising highly specific motifs. Then, re-analysing data from six RNA-Seq studies, we found that the nucleotide content of *Wolbachia* genes correlates with their expressions, with a pattern compatible to be a consequence of the DNA methylation. Lastly, we identified MTase as the *Wolbachia* gene with the most conserved binding site for the Ccka/CtrA signalling transduction system, a mechanism likely involved the host-bacterium communication. Overall, these findings suggest a cascade mechanism in which the host activates the *Wolbachia* Ccka/CtrA signalling system, thus inducing the expression of the MTase gene. Then, the subsequent DNA methylation will affect the expression of several *Wolbachia* genes on the basis of their cytosine and adenine content.

Introduction

Wolbachia pipientis (from here *Wolbachia*) is an intracellular bacterium belonging to the *Alphaproteobacteria* class. It is the most prevalent endosymbiotic microbe in the animal world, colonising many arthropods species and filarial nematodes(1–4). In arthropods, the bacterium can manipulate the reproduction of the host(5), protect it from viruses(6, 7) and provide nutrients(8). In filarial nematodes, genomic analyses and experimental evidence revealed that *Wolbachia* is essential for the host survival and development (9). The bacterium is strictly intracellular, with a tropism for microtubules and with the ability to colonise endoplasmic reticulum (ER), with important effects on host cell replication(2). It is mainly mother-to-offspring transmitted, being actively transported by specialised cells of the host from the soma to germ cells (2). Nevertheless, horizontal transfer among host individuals can occur (10, 11), ehnancing the bacterium spreading (11).

As for many other strictly host-associated bacteria, the *Wolbachia* genome underwent an important reduction during its evolution, reaching a size that ranges from ~0.8 Mb to ~1.8 Mb (BV-BRC representative genomes)(12). 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 many genes(13–15). During this process, the structure of the genome and the order of genes can change dramatically(14), leading to the disruption of several operons and loss of most gene promoters. Interestingly, RNA-seq studies have highlighted that *Wolbachia* gene expression can be influenced by host environmental/developmental conditions(16–22). This suggests that *Wolbachia* likely has a mechanism for regulating gene expression, despite the low number of promoters indicating it may not depend solely on transcription factors.

The first published *Wolbachia* RNA-seq study(16) compared gene expression of *Wolbachia* colonising somatic tissues and gonads in females and males of the filarial nematode

77 *Onchocerca ochengi*. One of the strongest differential expressions was reported for a gene
78 annotated as a S-adenosyl-methyonine-dependent (SAM)-dependent DNA
79 methyltransferase (hereafter ”MTase”). In general, DNA methyltransferase enzymes are able
80 to methylate either adenine or cytosine on specific motifs of bacterial DNA. The most studied
81 MTase are the DNA adenine methyltransferase (Dam) and DNA cytosine methyltransferase
82 (Dcm) of *Escherichia coli*. The first one methylates the adenine residues in the sequence
83 GATC, while the other one targets the second cytosine within the motifs CCAGG and
84 CCTGG (23, 24). Furthermore, in *Alphaproteobacteria*, one of most characterised
85 methyltransferase is the CcrM, which methylates the adenine in the motif GAnTC, and it is
86 involved in the regulation of the cell cycle in *Caulobacter crescentus (25)*.
87
88 Despite DNA methylation has deeply been investigated in eukaryotes, as an epigenetic
89 mechanism of gene expression, its role in prokaryotes is still poorly investigated and no
90 studies have been published on *Wolbachia*.
91
92 Despite the effect of DNA methylation on gene expression has been deeply investigated in
93 eukaryotes (i.e. epigenetics), its role in prokaryotes physiology is still poorly understood and
94 no studies have been published on *Wolbachia.* At the state of the art, there is evidence that
95 DNA methyltransferase enzymes are involved in fundamental bacterial processes, including
96 proofreading during DNA replication, protection from foreign DNA and regulation of gene
97 expression. A specific type of DNA methyltransferase, called orphan DNA
98 methyltransferases, methylates gene promoters, changing their affinity for transcription
99 factors with important effects on gene expression (23, 24).
100
101 In this study, we investigated the hypothesis that DNA methylation is a key mechanism for
102 the regulation of the *Wolbachia* gene expression. First, we experimentally determined the
103 methylation motif of the MTase enzyme encoded in the genome of the *Wolbachia*
104 endosymbiont of *O. ochengi* (wOo). Then, we tested whether the frequency of this motif into

105 gene sequences correlates with their expression, and whether the intensity of the
106 phenomenon is related to the levels of expression of the MTase gene. Finally, we identified
107 regulatory regions upstream of the MTase gene in *Wolbachia* genomes. Our results suggest
108 that the host may specifically induce the expression of the *Wolbachia* MTase gene through a
109 signalling transduction system located on the bacterium's membrane. Subsequently, the
110 MTase methylates the *Wolbachia* genome, altering the expression of methylated genes
111 based on their nucleotide composition.
112

113 Material and methods

114 **Experimental determination of the wOo SAM-dependent DNA methyltransferase**
115 **methylation pattern**

116 The wOo SAM-dependent DNA methyltransferase (MTase) was artificially synthesised with
117 codon usage adaptation and cloned into the XhoI and SalI sites of the pCOLD III expression
118 vector (Takara Bio). The recombinant construct was transformed into *Escherichia coli*
119 StellarTM Competent Cells (*dam*–/*dcm*–) (Takara Bio) according to the supplier's protocol
120 (ClonTech; Protocol-at-a-Glance, PT5056-2). The *E. coli* strains were cultured in LB broth
121 medium supplemented with ampicillin. Before inducing expression, PCR with custom primers
122 targeting the gene of interest was performed on bacterial crude extracts to verify the
123 outcome of transformation. MTase expression was then induced at 15°C, according to the
124 protocol provided by the manufacturer for the Cold Shock Expression System of pColdTM
125 plasmids (Takara Bio). 24 hours after the induction, the expression of the wOo MTase gene
126 was verified by SDS-PAGE. DNA was then extracted using the NucleoSpin Microbial DNA
127 Mini kit (Macherey Nagel) and subsequently subjected to long reads sequencing using
128 Sequel I (Pacific biosciences). Assembly of reads was done using the Microbial Assembly
129 pipeline provided by SMRT Link v.10.1 with a minimum seed coverage of 30x. The
130 assembled genome was used as a reference in the downstream analysis. The methylation
131 status of each base of the obtained reads was then determined using the 'Base modification

132 Analysis' pipeline included in SMRT-Portal. The pipeline performs base modification and
133 modified base motifs detection. A control experiment was carried out transforming *E. coli*
134 Stellar TM Competent Cells with a pCOLD III vector without the wOo MTase gene inserted.
135 The 95th percentile of Mean QV and coverage values obtained from the control experiment
136 were set as minimum thresholds for the identification of methylated bases on the
137 transformed *E. coli* strain. These last analyses were performed using R.
138
139 **Dataset reconstruction and data normalisation**
140 The table S2 and gff map files of the supplementary information of Chung et al. 2020 were
141 retrieved (26). The table contains the gene expression quantifications computed by Chung
142 and colleagues (2020) re-analysing seven RNA-Seq studies on *Wolbachia* (16–22), for a
143 total of 128 samples (replicates) from 62 host conditions. More in details: i) four samples
144 from two conditions on *Wolbachia* endosymbiont of *Onchocerca ochengi* (wOo) from Darby
145 et al. 2012(16); ii) five samples from five conditions on *Wolbachia* endosymbiont of
146 *Dirofilaria immitis* (wDi) from Luck et al., 2014(17); iii) six samples from two conditions on
147 *Wolbachia* endosymbiont of *Drosophila melanogaster* (wMel), from Darby et al., 2014(18);
148 iv) 12 samples from seven conditions on wDi from Luck et al., 2015(19); v) 55 samples from
149 24 conditions on wMel, from Gutzwiller et al., 2015(20); vi) 14 samples from seven
150 conditions on *Wolbachia* endosymbiont of *Brugia malayi* (wBm) from Grote et al., 2017(21);
151 vii) 32 samples from 15 conditions on wBm from Chung et al., 2019(22). Preliminarily, we
152 excluded from the dataset all the samples from Luck et al., 2015(19) because only 8% of the
153 gene expression quantifications were greater than 0. Thus, gene expression analysis was
154 performed on a total of 116 samples from 55 host conditions, two for wOo, five for wDi, 22
155 for wBm and 26 for wMel.
156
157 Gene expression values were normalised to make them comparable among studies and
158 conditions. For each condition in each study the gene expression values were expressed as

159 $$\text{Log2expr} = \sum_s [\sum_i \log2(\text{expr}_{s,i} / \sum_i \text{expr}_{s,i})] / S_{num}$$

160 Where:

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

164 $$\text{Log2FC} = \text{Log2expr}_{condition1} - \text{Log2expr}_{conditio21} =$$

165 $$\{\sum_s [\sum_i \log2(\text{expr}_{s,i} / \sum_i \text{expr}_{s,i})] / S\}_{condition1} - \{\sum_s [\sum_i \log2(\text{expr}_{s,i} / \sum_i \text{expr}_{s,i})] / S\}_{condition2}$$

166

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

171

172 **Genes composition parameters determination**

173 The genome assemblies of wBm (AE017321.1), wDi (www.nematodes.org: wDi 2.2), wOo
174 (HE660029.1) and wMel (AE017196.1) were retrieved and passed to Prodigal (27) for Open
175 Reading Frame (ORF) calling. For each ORF, Codon Bias Index (CBI) (which measure how
176 much a gene uses a subset of optimal codons, ranging from 0 for random using, to 1 for
177 maximum bias), the effective number of codons (Nc) and the ORF length were computed
178 using CodonW tool (Peden J.F. Analysis of codon usage. 1999. PhD Thesis, University of
179 Nottingham, UK), while %A, %T, %G, %C were computed using an in-house Perl script.
180 Conventionally, ORF sequences correspond to the transcribed messenger RNA (mRNA)
181 sequence, which is complementary to the DNA template strand. Thus, the nucleotide
182 composition of the genes is derived by complementing the ORF compositions: the %A of a
183 gene is determined as the %T of the corresponding ORF, and so on. Finally, the %C/%A
184 ratio is calculated. All subsequent analyses will focus on the nucleotide composition of the
185 genes rather than on ORF compositions.

186 **Statistical analyses**

187 The relationship between each gene composition parameter (CBI, Nc, length, %A, %T, %G,
188 %C and C/A ratio, see above) and gene expression was investigated. More in detail, for
189 each of the 55 samples included in the study, linear regression was used to assess the
190 correlation between the global gene expression reported for that sample and each of the
191 gene composition parameters. For each regression line, the slope and the p value were
192 obtained. Slope values indicate the strength of change in gene expression in relation to the
193 gene composition parameter: the higher the slope, the more the gene expression varies as a
194 consequence of that gene composition parameter. In a second analysis, to investigate if
195 MTase activity affects the expression of genes on the basis of their composition parameters,
196 the correlation between the linear regression slope and the MTase gene expression was
197 computed for each gene composition parameter.

198

199 The presence of other genes with expression patterns comparable to that of MTase gene
200 was tested as follows. Exploiting the gene orthologous information included in the Table S6
201 of Chung et al., 2020 (26), we performed a Spearman co-correlation among the expressions
202 of the 546 single copy core genes shared among wOo, wDi, wBm and wMel. The expression
203 patterns were also investigated by Principal Component Analysis (PCA), using R.

204

205 **Comparison between species tree and MTase tree**

206 A dataset including 112 high quality *Wolbachia* genome assemblies spanning the host
207 genetic diversity was reconstructed, including 111 assemblies retrieved from the Genome
208 Taxonomy DataBase (GTDB) and the wDi genome used for the previous analysis (indeed,
209 the wBm, wMel and wOo assemblies were already present in the GTDB database) (see
210 Table S1). The genome assemblies were subjected to Prodigal (27) for ORF calling, and the
211 amino acid sequences were passed to OrthoFinder (28) for orthologue analysis. The
212 nucleotide sequences of the single copy core genes were then retrieved, aligned using the
213 MUSCLE tool (29), tested for recombinations using the PhiPack tool (30), trimmed using the

214 trimal tool (-gt 0.5 setting) (31), and finally concatenated. The obtained concatenate was
215 passed to RAxML8 (32) for phylogenetic analysis using the GTR+I+G model, as previously
216 determined using ModelTest-NG tool (33). The orthologue group containing the MTase
217 genes was determined and subjected to phylogenetic analysis following the same flow
218 described above (using the TVM+I+G model and excluding the step of the concatenation of
219 the genes). The obtained trees were then compared using the Cophylo R library (34).
220

221 **Analysis of the region upstream the methyltransferase gene**

222 The positions and orientation of the MTase genes on the 112 genome assemblies were
223 determined by BLASTn searches. Then, for each genome assembly, the 100 nucleotides
224 upstream of the MTase gene were extracted and screened for the presence of some of the
225 most important regulatory regions: i) CtrA binding site (motif TTAA-N7-TTAA (35)); ii)
226 Pribnow box (motif TATAAT (36)); iii) CAAT box (motif HYYRRCCAWWSR(37)); iv) GC box
227 (motif WRDRGGHRKDKYYK(37)); v) TATA box (motif TATAWAWR(37)). Perfectly matching
228 motifs and one-mismatched motifs were considered for further analyses. The presence of
229 *cck*A and *ctr*A genes in the 112 *Wolbachia* genomes was then evaluated by BlastP search,
230 using as reference the *cck*A and *ctr*A gene sequences already reported in Lyndsey 2020
231 (38). Finally, the positions of the canonical regulatory regions upstream the MTase gene,
232 and the presence/absence of *cck*A/*ctr*A genes in the 112 *Wolbachia* genomes were
233 visualised using the gplots R library.
234

235 **Analysis of the regions upstream all genes in the 112 *Wolbachia* genomes**

236 On the basis of previous results on the MTase gene, TATA boxes were searched between
237 positions -20 and -70 upstream of all the genes of the 112 *Wolbachia* genomes, and CrtA
238 binding sites between the positions 0 and -30. The presence of short AT-rich sequences, like
239 TATA box and CtrA binding site, in AT-rich genomes (such as *Wolbachia*) could be due to
240 the chance. The possible functionality of these sequences has been assessed by
241 investigating whether TATA or CtrA binding sites are enriched upstream of specific genes in

242 the *Wolbachia* genomes. More in detail, for each of the ortholog genes previously identified
243 using OrthoFinder (see above), the frequency of *Wolbachia* strains having upstream TATA
244 or CtrA binding site was investigated using R. Lastly, both for TATA box and CtrA binding
245 site, the genes present in at least 100 out of 112 *Wolbachia* genomes and having the
246 regulatory box in >80% of the genomes were retrieved and annotated using the Clusters of
247 Orthologous Groups (COG) database.
248
249 **Analysis of the genes with high and low G/T ratio**
250 One of the gene composition parameters found to be more correlated to gene expression
251 and MTase action was C/A ratio. Thus, the expression of the genes with higher or lower C/A
252 ratio is expected to be more affected by the action of the MTase enzyme. To investigate the
253 effects of this mechanism on the *Wolbachia* metabolism and physiology, genes with low or
254 high C/A ratio values were studied as follows. All the amino acid sequences from the 112
255 *Wolbachia* strains included in the study were annotated on the basis of the Clusters of
256 Orthologous Genes (COG) database. For each strain, among the COG-annotated genes,
257 the 10 with lower C/A ratio ("low C/A ratio genes") and the 10 with higher C/A ratio ("high C/A
258 ratio genes") were retrieved. The pattern of presence/absence and C/A ratio value of all
259 these retrieved genes was graphically investigated by producing a heatmap using the gplots
260 R library.
261

262 Results

263 **Experimental determination of the wOo SAM-dependent DNA methyltransferase**
264 **methylation pattern**
265 The expression of the wOo MTase into the Stellar *E. coli dam*-/*dcm*- strain followed by the
266 SMRT PacBio sequencing allowed to identify 166 high quality methylated bases: 151 on
267 cytosine (classified as m4C - N4-methylcytosine) and 15 on adenine (classified as m6A - N6-
268 methyladenine) (see Table S2 and Figure S1). Investigation of the methylation pattern

269 revealed that m4c methylations occurred in a pattern that was not well conserved, whereas
270 m6A methylation involved a guanine-rich region at positions -4, -8 and -19 (Figure 1).
271
272 **Dataset reconstruction**
273 For each of the six datasets included in the study, the normalised gene expression values
274 were obtained as described in material and methods. The ORF sequences of the wOo, wDi,
275 wBm and wMel genome assemblies were analysed to determine the CBI, Nc, length, %A,
276 %T, %G, %C and %C/%A (from here C/A ratio). The obtained information about gene
277 composition parameters was merged to relative gene expression data (Supplementary
278 information available on repository...).
279
280 **Statistical analyses**
281 The slope and p value of the linear regression analyses performed between the eight gene
282 composition parameters (CBI, Nc, gene length, %A, %T, %G, %C and C/A ratio) and the
283 gene expression data obtained for each of the 55 samples are reported in Table S3. Length
284 and %C significantly correlate with gene expression in all the samples included in the study
285 (100%), while %T in 53/55 (96%), CBI 42/55 (78%), %A 40/55 (73%), C/A ratio 31/55 (56%),
286 %G 22/55 (40%) and Nc 7/55 (13%). The obtained slopes for the regression lines range as
287 follows: for gene length from 0.009 to 0.029, %C from 0.79 to 5.3, CBI from 30.9 to 71.7, %A
288 from -1.8 to 1.1, C/A ratio from 18.5 to 90.6, %G from -1.3 to 2.2 and Nc from 0.2 to 0.8.
289 Subsequently, for each of the gene composition parameters (and for each of the 55
290 samples) the correlation between the levels of expression of the MTase gene and the
291 regression slope obtained above (indicative of the strength of the correlation between the
292 gene expression data of the sample and the gene composition parameter) was tested by
293 linear regression. The aim was to investigate if MTase activity affects the expression of
294 genes on the basis of their composition parameters. The obtained correlations were
295 significant for all parameters except %G, which had a p-value of 0.08. For %G, %C, %T, C/A
296 ratio, Nc, and length, the slope increases with higher DNA MTase expression. Conversely,

297 %A and CBI exhibit a decreasing trend with increasing of the MTase expression (Figures 2-3
298 and Figures S2-S7).
299
300 Co-correlation analysis among the expression patterns of the single copy core genes shared
301 among wOo, wDi, wBm and wMel showed that no other gene has an expression pattern
302 overlapping that of the DNA MTase gene. Indeed, the highest correlation value was 0.84 and
303 the median value was 0.02 (Figure S8). In general, the absence of clusters of co-expressed
304 genes is also evident from PCA analysis (Figure S9).
305
306 **Comparison between species tree and MTase tree**
307 Orthology analysis identified a total of 3,561 orthologous groups, including 263 single copy
308 core genes. None of the single copy core genes were recombined. After trimming of gene
309 alignments, the obtained concatenate had a length of 223,833 bp. On the other side, the
310 MTase gene alignment was 1,342 bp long. The species tree obtained from the 263 single
311 copy genes and the MTgene tree are reported in Figure S10. The obtained topologies were
312 mainly congruent, with few signs of Horizontal Gene Transfer (HGT) events.
313
314 **Analysis of the region upstream the methyltransferase gene**
315 The MTase gene was found in all the 112 genome assemblies included in the study.
316 Unfortunately, for one genome assembly (accession GCF_018454445.1, *Wolbachia*
317 endosymbiont of *Rhagoletis cerasi* - wCer5) the MTase gene was located on the extreme of
318 a contig, and thus it was not possible to extract the 100 bp upstream of the gene
319 transcription initiation site. The presence of CtrA binding site, Pribnow, CAAT, GC and TATA
320 boxes upstream the MTase gene was assessed. Among these, at least one canonical
321 sequence (i.e. perfect match) was found only for CtrA binding site, TATA and Pribnow
322 boxes. In particular, 104 out of the 111 regions upstream of the MTase gene contained
323 canonical TATA boxes (Figure S11), 100 out of 111 canonical CtrA binding sites (Figure
324 S12) and five canonical Pribnow boxes (Figure S13).

325 **Analysis of the regions upstream all genes in the 112 *Wolbachia* genomes**

326 The assessment of the presence of CtrA binding site and TATA boxes upstream of all the
327 genes of the 112 *Wolbachia* genomes included in the study (see above) led to the discovery
328 of a total of 2,210 CtrA binding sites and 12,283 TATA boxes. The CtrA binding site was
329 predominantly found upstream of the MTase gene, but it was also frequently present
330 upstream of the Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, which is
331 involved in energy production (see Table S4 and Figure S14). Instead, the TATA boxes were
332 found with high frequency upstream of 12 genes (see Table S4 and Figure S15), including
333 the gene coding for the CtrA DNA binding response regulator.

334

335 **Analysis of the genes with high and low C/A ratio**

336 A total of 57 low C/A ratio genes and 53 high C/A ratio genes were found. Among the low
337 C/A ratio genes, the most frequent COG categories were “Translation, ribosomal structure
338 and biogenesis” (J) with 13/57 (23%) genes, “Mobilome: prophages, transposons” (X) with
339 8/57 (14%), “Posttranslational modification, protein turnover, chaperones” (O) with 7/57
340 (12%) and “Transcription” (K) with 5/57 (9%) genes. Conversely, among the high C/A ratio
341 genes, the most frequent COG categories were “Energy production and conversion” (C) with
342 15/53 (28%) genes, “Inorganic ion transport and metabolism” (P) with 11/53 (21%), “Lipid
343 transport and metabolism” (I) with 5/53 (9%) genes. The presence/absence and C/A ratio
344 patterns are shown in Figure 4.

345

346 Discussion

347 *Wolbachia* is one of the most widespread endosymbiotic bacteria (10), with an important
348 impact on the survival and/or reproduction of several arthropod and filarial nematode host
349 species (2, 5, 39)(2, 3, 39). The close symbiotic relationship with the host has had important
350 consequences on the bacterial genome structure, including frequent rearrangements (14)
351 and the loss of many genes and the loss of regulatory regions (40), with the disruption of

352 several operons and promoters. The loss of promoters could have limited evolutionary costs,
353 considering that the intracellular niche of the bacterium protects it from external environment
354 stresses. However, experimental studies showed the existence of *Wolbachia* genes
355 differentially expressed in host samples from different body compartments, developmental
356 stages and/or from hosts subjected to environmental conditions (26). Nevertheless, at the
357 state of the art, the gene regulation mechanisms in *Wolbachia* are poorly understood.
358
359 One of the most differentially expressed genes between *Wolbachia* colonising the soma and
360 gonads of the filarial nematode *Onchocerca ochengi* (wOo) was a SAM-dependent DNA
361 methyltransferase (MTase) (16). Generally, methylation of bacterial gene promoters is
362 known to influence the promoters affinity for transcription factors, leading to changes in gene
363 expression and epigenetic modifications (23, 41). In this work, we investigated the possible
364 role of the *Wolbachia* MTase as a mediator of gene expression regulation. In absence of
365 promoters, we hypothesise that intragenic DNA methylation could affect gene expression.
366 Indeed, experimental studies revealed that the efficiency of bacterial RNA polymerase is
367 influenced by epigenetic modifications (42–45).
368
369 To test our hypothesis, we i) experimentally determined the sequence motif recognised by
370 the wOo DNA MTase; ii) studied the expression data of 55 samples from six published RNA-
371 Seq studies on *Wolbachia*; iii) investigated the presence of regulatory regions upstream the
372 DNA MTase gene in *Wolbachia*.
373
374 *In silico* determination of the sequence motif recognized by MTase is challenging (46). This
375 is especially true in this case, considering that this is the first *Wolbachia* DNA MTase
376 enzyme ever characterised. An experimental approach to determining the methylation motif
377 of the enzyme is to express the MTase in an *Escherichia coli* strain depleted for any MTase
378 gene, and then identify the methylated bases and motifs using SMRT PacBio sequencing
379 (46). Surprisingly, wOo MTase was found to be able to methylate both adenine and cytosine,

380 without highly conserved patterns (Figure 1), especially for cytosine. To our knowledge this
381 is the first MTase enzyme described with these peculiar features. Our findings also suggest
382 that the enzyme has a 10-fold greater affinity for cytosine than adenine. This value has been
383 determined on an *E. coli* genome containing ~50% of AT. Instead, the genome of a
384 *Wolbachia* strain usually contains ~70% of AT and this partially reduces the MTase affinity
385 bias to 4-fold.
386
387 Regarding the analysis of gene expression, we started from the idea that if DNA methylation
388 within the gene affects its expression, we should observe a correlation between gene
389 nucleotide composition and gene expression. Furthermore, the direction (positive or
390 negative) and the strength (the slope of the regression line) of this correlation should change
391 with variation of DNA MTase gene expression. It is well-documented in literature that gene
392 length and codon usage could be correlated to gene expression (46, 47). In particular, highly
393 expressed genes tend to have lower Effective Number of Codons (Nc) and higher Codon
394 Bias Index (CBI). For each of the 55 samples, we computed the correlation between CBI,
395 Nc, gene length, %A, %T, %G, %C and C/A ratio versus the relative gene expression. The
396 result shows correlation between gene nucleotide composition and gene expression in many
397 samples. Interestingly, we found that the slope of the obtained regression lines often
398 correlates with the MTase expression in that sample. In more detail, the expression of genes
399 is positively correlated with their %C in all samples and the correlation slope increases
400 proportionally to the expression of DNA MTase (Figure S2). Regarding %A, the correlation
401 with gene expression was positive (the higher the %A the higher the gene expression) in
402 samples with poorly expressed MTase, while it became negative (the higher the %A the
403 lower the gene expression) in samples with high MTase gene expression (Figure S3).
404
405 The positive to negative slope trend observed for %A suggests that methylation of adenine
406 might play an important role in the regulation of gene expression, in particular considering
407 the AT-richness of the *Wolbachia* genome. Lastly, we found that the ratio of %C and %A

408 (C/A ratio) is a good parameter to study *Wolbachia* gene expression: it significantly positively
409 correlates with gene expression in all samples except for those with very low expression of
410 MTase (Figure 2). This strongly supports the hypothesis that the methylations of cytosine
411 and adenine could be the key to understanding the regulation of gene expression in
412 *Wolbachia*. Interestingly, as shown in Figure 4, the genes with low C/A ratio are mainly
413 involved in basal metabolism pathways (such as replication and translation) while those with
414 high C/A ratio are mostly part of pathways involved in stress response (such as the
415 production of energy and ion/molecule transport). Interestingly, among the low or high C/A
416 ratio genes there are, respectively, *fts*Z and *fts*W, both involved in the bacterium cell
417 replication. This suggests that the host could modulate the bacterium replication: a possible
418 key for a successful symbiotic relationship. In this picture, *Wolbachia* could have a stress-
419 related mitochondrion-like role for the host, as previously proposed by Darby and colleagues
420 (2012). Furthermore, it is interesting to note that highly expressed genes have high Codon
421 Bias Index (CBI) (as expected from literature) only for samples with low levels of MTase
422 expression; in those with highly expressed MTase this correlation became not significant
423 (Figure 3). This suggests that the MTase effect on gene expression might have evolved
424 independently from the codon usage adaptation.
425
426 The step forward in our analysis was to investigate how the expression of the MTase gene
427 can be regulated. To this aim we analysed a collection of representative *Wolbachia*
428 genomes, screening the 100-bp region upstream of the MTase gene for the presence of
429 regulatory sequences. This analysis highlighted the presence of CtrA binding site and TATA
430 box sequences highly conserved among *Wolbachia* strains. The CtrA binding site is a
431 sequence bounded by the CtrA response regulator protein, which is part of the two
432 component regulatory system Ccka/CtrA. This system is composed of Ccka, a membrane
433 sensor histidine kinase, and CtrA, an intracellular effector protein. When triggered by an
434 external stimulus, the Ccka protein activates CtrA, which in turn binds to the CtrA binding
435 site initiating the expression of the downstream genes. While the signalling mechanisms

436 behind the activation of Ccka are not fully understood, the downstream response regulator,
437 CtrA, is a well-described master regulator with transcription factor activity, conserved in
438 *Alphaproteobacteria (48)*. The presence of a highly conserved CtrA binding site upstream
439 the MTase gene locus, strongly suggests that the CcKA/CtrA signal transduction pathway
440 could have a role in the regulation of the expression of this gene. Indeed, the host could
441 modulate the expression of the *Wolbachia* MTase gene by stimulating the Ccka receptor on
442 the membrane on the bacterial cell. Furthermore, the presence of a conserved CtrA binding
443 site upstream of only two genes suggests that this could be a highly regulated mechanism
444 subjected to a strong selective pressure.
445
446 Surprisingly, we also found highly conserved TATA boxes (i.e. eukaryotic regulatory
447 sequences) upstream of the MTase gene. In eukaryotes, TATA boxes tend to be placed
448 upstream of stress-responsive genes, whose expression must rapidly and variably be tuned
449 in response to specific environmental conditions and changing physiological needs (49, 50) .
450 The TATA box is bound by a specific protein, the TATA box binding protein (TBP), which
451 then recruits transcription factors. *Wolbachia* does not encode TBP proteins, so the
452 functionality of these TATA box sequences can only be hypothesised, e.g. assuming that
453 TBPs are supplied by the host. Interestingly, we found 12 *Wolbachia* genes having highly
454 conserved TATA boxes. These genes include *ctr*A response regulator (see above), genes
455 involved in energy production and a component of the Type IV secretion system.
456
457 To explain how the host could regulate gene expression in *Wolbachia*, we propose the
458 following model: i) specific molecules in the host cells activate the Ccka membrane histidine
459 kinase; ii) Ccka phosphorylates the intracellular CtrA response regulator; iii) the
460 phosphorylated CtrA binds to the CtrA binding site upstream of the MTase gene inducing its
461 expression in *Wolbachia*; iv) the MTase enzyme methylates adenine and cytosine of the
462 *Wolbachia* genome; v) the expression of several *Wolbachia* genes changes on the basis of

463 their nucleotide content, particularly enhancing the transcription of genes with a high ratio of
464 cytosine / adenine. The model is graphically summarised in Figure 5.
465 Moreover, the presence of a conserved TATA box upstream the *ctr*A gene could suggest a
466 possible role for the host in the regulation of *ctr*A gene expression.
467
468 While the hypothesised mechanism may not lead to highly precise regulation of gene
469 expression, it could be a rough yet functional mechanism, sufficient for the establishment of
470 a successful intracellular symbiosis. In the future, it would be interesting to investigate
471 whether similar mechanisms are present in other endosymbiont or intracellular organisms.
472
473 Despite the results presented in this work are far from being definitive, we hope to have
474 opened a new way in understanding the symbiotic relationship between *Wolbachia* and its
475 hosts. In our opinion, this is a magnificent example of how evolution works, recycling
476 metabolic pieces for similar aims, sometimes reaching surprisingly stable *equilibria*.
477
478
479
480
481
482
483

484 # Captions

485 **Figure 1. Methylation patterns of the DNA methyltransferase gene of *Wolbachia***
486 **endosymbiont of *Onchocerca ochengi* (wOo)**
487 Sequence logo of the sequences flanking (from -20 to +20 positions) the methylated
488 positions found in the genome of Stellar *Escherichia coli* dam-/dcm- strain expressing the
489 DNA methyltransferase gene of *Wolbachia* endosymbiont of *Onchocerca ochengi* (wOo).
490 The height of the letters depends on the conservation of the position among the sequences.
491 a) the sequence logo of the m4C methylations on cytosines; b) the m6A methylation of
492 adenine.
493
494 **Figure 2. Correlation between the expression of genes and their %C / %A ratio**
495 For each of the 55 samples included in the study, the correlation between the expression of
496 the genes and the relative %C/%A ("C/A ratio") was investigated. The scatter plots of C/A
497 ratio vs gene expression of three exemplificative samples selected from the 55 samples are
498 reported: a) the correlation plot with highest slope (Hs), b) the correlation plot with the
499 medium slope (Ms) and c) the correlation plot with the lowest slope (Ls). In each of these
500 three plots, colour is green whether the correlation is significant (p value < 0.05) and grey if it
501 is not, the regression line is coloured in dark grey and the statistics are reported on the chart.
502 In figure d, the correlation plot between the slopes obtained for each of the 55 samples, and
503 the relative expression of the DNA methyltransferase gene is reported. Points relative to
504 significant correlations (p value < 0.05) are in green while non significant in grey. The
505 regression line is in green and the statistics are reported on the chart.
506
507 **Figure 3. Correlation between the expression of genes and their Codon Bias Index**
508 **(CBI)**
509 For each of the 55 samples included in the study, the correlation between the expression of
510 the genes and the relative Codon Bias Index (CBI) was investigated. The scatter plots of CBI

511 vs gene expression of three exemplificative samples selected from the 55 samples are
512 reported: a) the correlation plot with highest slope (Hs), b) the correlation plot with the
513 medium slope (Ms) and c) the correlation plot with the lowest slope (Ls). In each of these
514 three plots, colour is green whether the correlation is significant (p value < 0.05) and grey if it
515 is not, the regression line is coloured in dark grey and the statistics are reported on the chart.
516 In figure d, the correlation plot between the slopes obtained for each of the 55 samples, and
517 the relative expression of the DNA methyltransferase gene is reported. Points relative to
518 significant correlations (p value < 0.05) are in green while non significant in grey. The
519 regression line is in green and the statistics are reported on the chart.
520
521 **Figure 4. COG annotated genes with highest or lowest C/A ratio**
522 The heatmap shows the pattern of the genes with high or low C/A ratio among the 112
523 *Wolbachia* strains included in the study. On the left, the Maximum Likelihood phylogenetic
524 tree obtained from concatenate of the single copy core genes. In the middle, heatmap
525 reports the pattern of the 10 lower C/A ratio COG annotated genes: rows refer to strains
526 (matching with the phylogenetic tree on the left) and columns to genes (COG annotation is
527 reported). Cell colours report the C/A ratio value following the legend on lower left. Asterisks
528 are reported for genes among those in the 10 lower C/A ratio for the relative strain. More on
529 the right, a heatmap reports the 10 higher C/A ratio COG annotated genes. Finally, on the
530 right, *Wolbachia* strains hosts and supergroups are shown.
531
532 **Figure 5. Graphical representation of the mechanism proposed for the regulation of**
533 **the *Wolbachia* gene expression**
534 The figure summarises the model here proposed for the *Wolbachia pipientis* gene
535 expression regulation. *Wolbachia* is an intracellular bacterium able to colonise the cytoplasm
536 of host cells. In the figure the bacterium, represented with violet membranes and white
537 cytoplasm, is surrounded by the host cytoplasm, coloured in green. At step 1 (on the left):
538 the host activates the Ccka sensor histidine kinase (in violet) placed on the *Wolbachia*

membrane (here we represented this activation as mediated by a molecular factor produced by the host, but other mechanisms are possible), then the Ccka phosphorilates the intracellular transcription regulator CtrA which, induces the transcription of the *Wolbachia* DNA *methyltransferase* gene (MTase) binding the CtrA binding site upstream it. At step 2 (in the middle): the MTase enzyme methylates the adenines and cytosines of the *Wolbachia* DNA and methyl groups are represented with dots coloured as the bases. At step 3 (on the right): RNA polymerases are represented by pink icons from which mRNA molecules originate. In highly expressed genes more RNA polymerases bind the gene simultaneously. The greater is the cytosines / adenines ratio in a gene the more that gene is expressed.

Acknowledgments

F.C. would like to thank his mentor, Claudio Bandi, for introducing him to the world of *Wolbachia* and for his helpful comments to this paper.

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. R. Zug, P. Hammerstein, 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).

2. J. Porter, W. Sullivan, The cellular lives of Wolbachia. *Nat. Rev. Microbiol.* **21**, 750–766 (2023).

3. J. H. Werren, Biology of Wolbachia. *Annu. Rev. Entomol.* **42**, 587–609 (1997).

4. C. Bandi, T. J. Anderson, C. Genchi, M. L. Blaxter, Phylogeny of Wolbachia in filarial nematodes. *Proc. Biol. Sci.* **265**, 2407–2413 (1998).

5. J. H. Werren, L. Baldo, M. E. Clark, Wolbachia: master manipulators of invertebrate biology. *Nat. Rev. Microbiol.* **6**, 741–751 (2008).

6. L. M. Hedges, J. C. Brownlie, S. L. O'Neill, K. N. Johnson, Wolbachia and virus protection in insects. *Science* **322**, 702 (2008).

7. L. Teixeira, A. Ferreira, M. Ashburner, The bacterial symbiont Wolbachia induces resistance to RNA viral infections in Drosophila melanogaster. *PLoS Biol.* **6**, e2 (2008).

8. I. L. G. Newton, D. W. Rice, The Jekyll and Hyde Symbiont: Could Be a Nutritional Mutualist? *J. Bacteriol.* **202** (2020).

9. C. Bandi, *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).

10. M. Scholz, *et al.*, Large scale genome reconstructions illuminate Wolbachia evolution. *Nat. Commun.* **11**, 5235 (2020).

11. S. J. A. Tolley, P. Nonacs, P. Sapountzis, Horizontal Transmission Events in Ants: What Do We Know and What Can We Learn? *Front. Microbiol.* **10**, 296 (2019).

12. R. D. Olson, *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).

13. A. Manzano-Marín, A. Latorre, Snapshots of a shrinking partner: Genome reduction in Serratia symbiotica. *Sci. Rep.* **6**, 32590 (2016).

14. F. Comandatore, *et al.*, Supergroup C Wolbachia, mutualist symbionts of filarial nematodes, have a distinct genome structure. *Open Biol.* **5**, 150099 (2015).

15. S. Mahmood, E. Nováková, J. Martinů, O. Sychra, V. Hypša, Supergroup F Wolbachia with extremely reduced genome: transition to obligate insect symbionts. *Microbiome* **11**, 22 (2023).

16. A. C. Darby, *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).

598 17. A. N. Luck, *et al.*, Concurrent transcriptional profiling of Dirofilaria immitis and its
599 Wolbachia endosymbiont throughout the nematode life cycle reveals coordinated gene
600 expression. *BMC Genomics* **15**, 1041 (2014).

601 18. A. C. Darby, *et al.*, Integrated transcriptomic and proteomic analysis of the global
602 response of Wolbachia to doxycycline-induced stress. *ISME J.* **8**, 925–937 (2014).

603 19. A. N. Luck, *et al.*, Tissue-specific transcriptomics and proteomics of a filarial nematode
604 and its Wolbachia endosymbiont. *BMC Genomics* **16**, 920 (2015).

605 20. F. Gutzwiller, *et al.*, Dynamics of Wolbachia pipientis Gene Expression Across the
606 Drosophila melanogaster Life Cycle. *G3* **5**, 2843–2856 (2015).

607 21. A. Grote, *et al.*, Defining Brugia malayi and Wolbachia symbiosis by stage-specific dual
608 RNA-seq. *PLoS Negl. Trop. Dis.* **11**, e0005357 (2017).

609 22. M. Chung, *et al.*, Drug Repurposing of Bromodomain Inhibitors as Potential Novel
610 Therapeutic Leads for Lymphatic Filariasis Guided by Multispecies Transcriptomics.
611 *mSystems* **4** (2019).

612 23. S. Papaleo, *et al.*, The red thread between methylation and mutation in bacterial
613 antibiotic resistance: How third-generation sequencing can help to unravel this
614 relationship. *Front. Microbiol.* **13**, 957901 (2022).

615 24. M. A. Sánchez-Romero, J. Casadesús, The bacterial epigenome. *Nat. Rev. Microbiol.*
616 **18**, 7–20 (2020).

617 25. S. S. Mohapatra, A. Fioravanti, E. G. Biondi, DNA methylation in Caulobacter and other
618 Alphaproteobacteria during cell cycle progression. *Trends Microbiol.* **22**, 528–535
619 (2014).

620 26. M. Chung, *et al.*, A Meta-Analysis of Transcriptomics Reveals a Stage-Specific
621 Transcriptional Response Shared Across Different Hosts. *G3* **10**, 3243–3260 (2020).

622 27. D. Hyatt, *et al.*, Prodigal: prokaryotic gene recognition and translation initiation site
623 identification. *BMC Bioinformatics* **11**, 119 (2010).

624 28. D. M. Emms, S. Kelly, OrthoFinder: phylogenetic orthology inference for comparative
625 genomics. *Genome Biol.* **20**, 238 (2019).

626 29. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high
627 throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

628 30. T. C. Bruen, H. Philippe, D. Bryant, A simple and robust statistical test for detecting the
629 presence of recombination. *Genetics* **172**, 2665–2681 (2006).

630 31. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: a tool for automated
631 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973
632 (2009).

633 32. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of
634 large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

635 33. D. Darriba, *et al.*, ModelTest-NG: A New and Scalable Tool for the Selection of DNA
636 and Protein Evolutionary Models. *Mol. Biol. Evol.* **37**, 291–294 (2020).

637 34. L. J. Revell, phytools 2.0: an updated R ecosystem for phylogenetic comparative
638 methods (and other things). *PeerJ* **12**, e16505 (2024).

639 35. M. C. Ouimet, G. T. Marczynski, Analysis of a cell-cycle promoter bound by a response
640 regulator. *J. Mol. Biol.* **302**, 761–775 (2000).

641 36. D. Pribnow, Nucleotide sequence of an RNA polymerase binding site at an early T7
642 promoter. *Proc. Natl. Acad. Sci. U. S. A.* **72**, 784–788 (1975).

643 37. P. Bucher, Weight matrix descriptions of four eukaryotic RNA polymerase II promoter
644 elements derived from 502 unrelated promoter sequences. *J. Mol. Biol.* **212**, 563–578
645 (1990).

646 38. A. R. I. Lindsey, Sensing, Signaling, and Secretion: A Review and Analysis of Systems
647 for Regulating Host Interaction in Wolbachia. *Genes* **11**, 813 (2020).

648 39. F. Comandatore, *et al.*, Phylogenomics and analysis of shared genes suggest a single
649 transition to mutualism in Wolbachia of nematodes. *Genome Biol. Evol.* **5**, 1668–1674
650 (2013).

651 40. C. Dale, N. A. Moran, Molecular Interactions between Bacterial Symbionts and Their
652 Hosts. *Cell* **126**, 453–465 (2006).

653 41. Q. Gao, *et al.*, Bacterial DNA methyltransferase: A key to the epigenetic world with
654 lessons learned from proteobacteria. *Front. Microbiol.* **14**, 1129437 (2023).

655 42. M. R. Rountree, E. U. Selker, DNA methylation inhibits elongation but not initiation of
656 transcription in Neurospora crassa. *Genes Dev.* **11**, 2383–2395 (1997).

657 43. F. Gracias, *et al.*, Homologues of epigenetic pyrimidines: 5-alkyl-, 5-hydroxyalkyl and 5-
658 acyluracil and -cytosine nucleotides: synthesis, enzymatic incorporation into DNA and
659 effect on transcription with bacterial RNA polymerase. *RSC Chem Biol* **3**, 1069–1075
660 (2022).

661 44. C. Rausch, *et al.*, Cytosine base modifications regulate DNA duplex stability and
662 metabolism. *Nucleic Acids Res.* **49**, 12870–12894 (2021).

663 45. M. Janoušková, *et al.*, 5-(Hydroxymethyl)uracil and -cytosine as potential epigenetic
664 marks enhancing or inhibiting transcription with bacterial RNA polymerase. *Chem.*
665 *Commun.* **53**, 13253–13255 (2017).

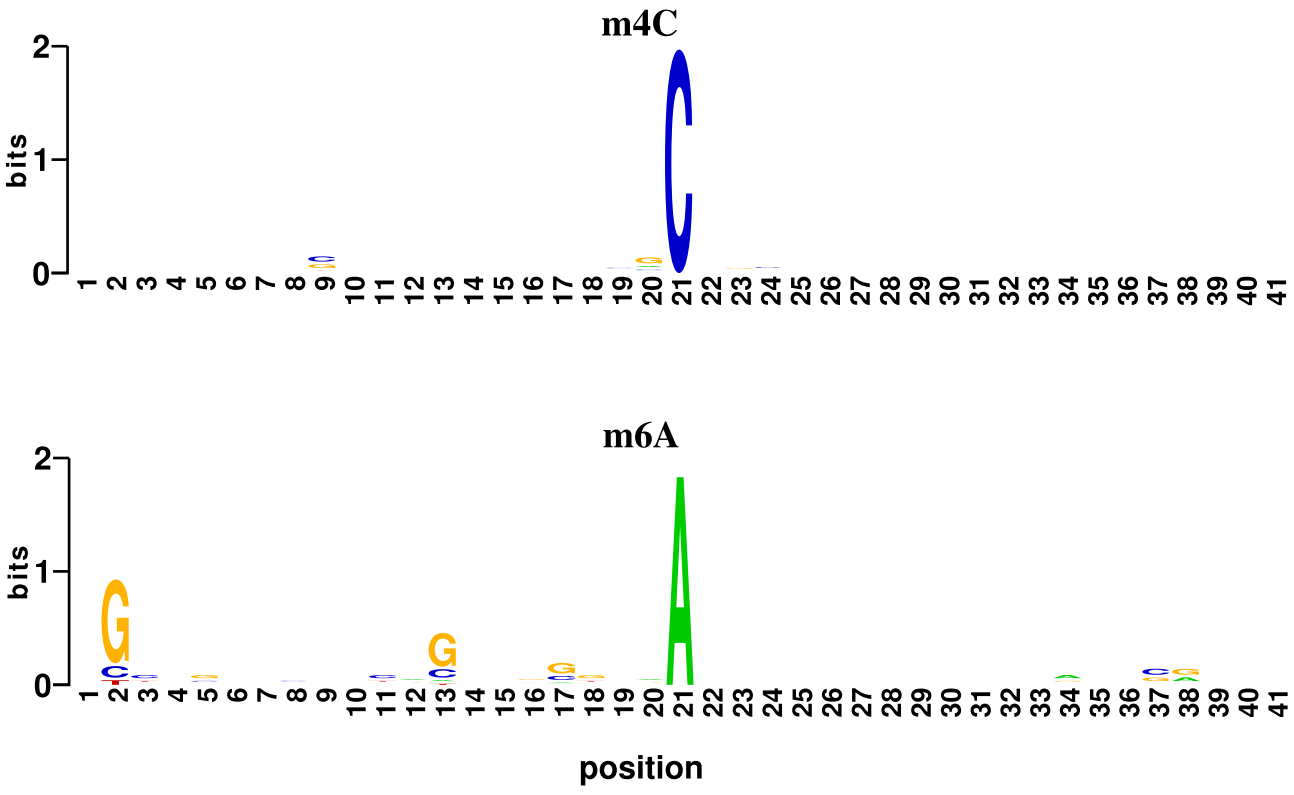
666 46. T. Ø. Jensen, *et al.*, Genome-wide systematic identification of methyltransferase
667 recognition and modification patterns. *Nat. Commun.* **10**, 3311 (2019).

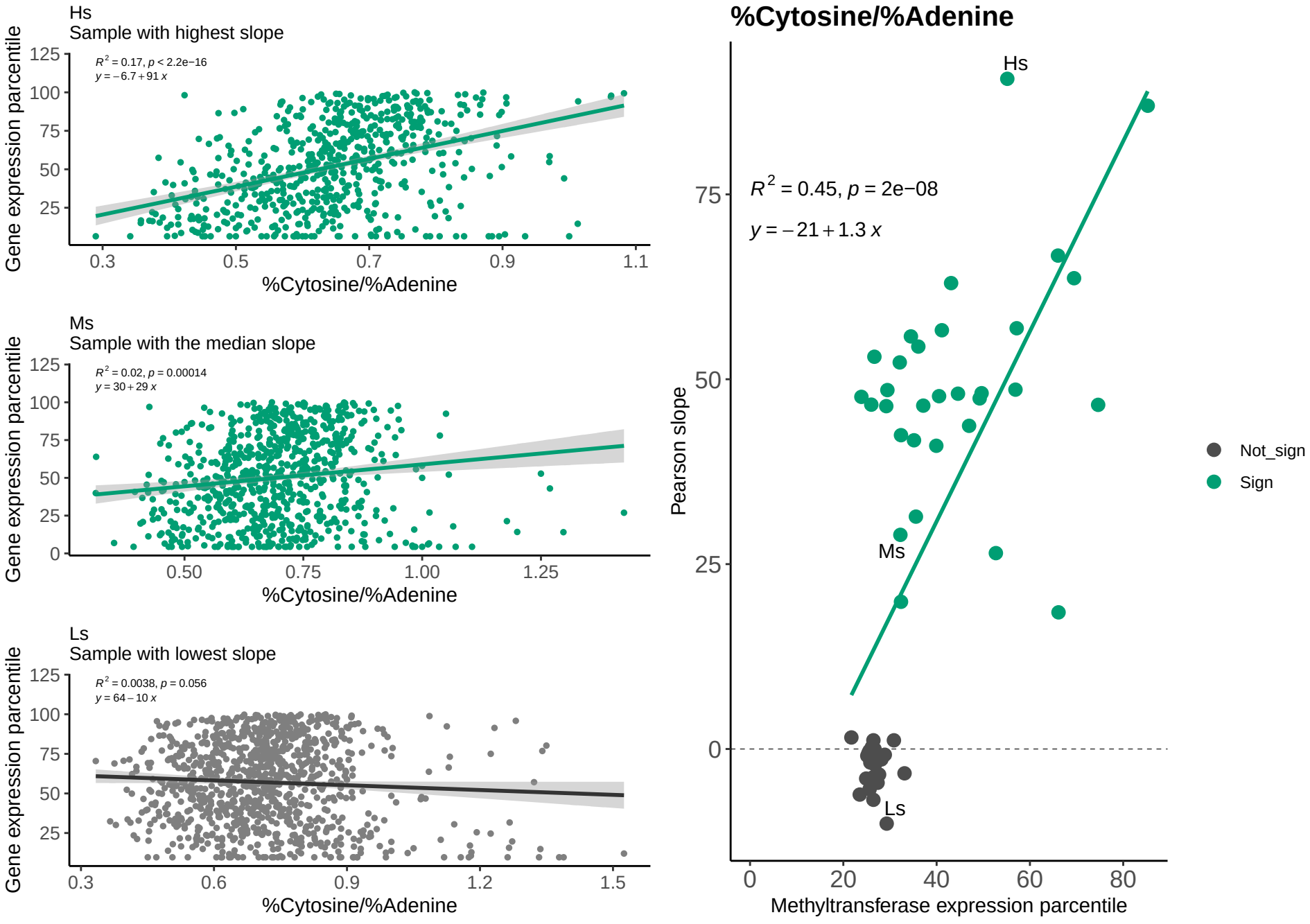
668 47. T. E. F. Quax, N. J. Claassens, D. Söll, J. van der Oost, Codon Bias as a Means to
669 Fine-Tune Gene Expression. *Mol. Cell* **59**, 149–161 (2015).

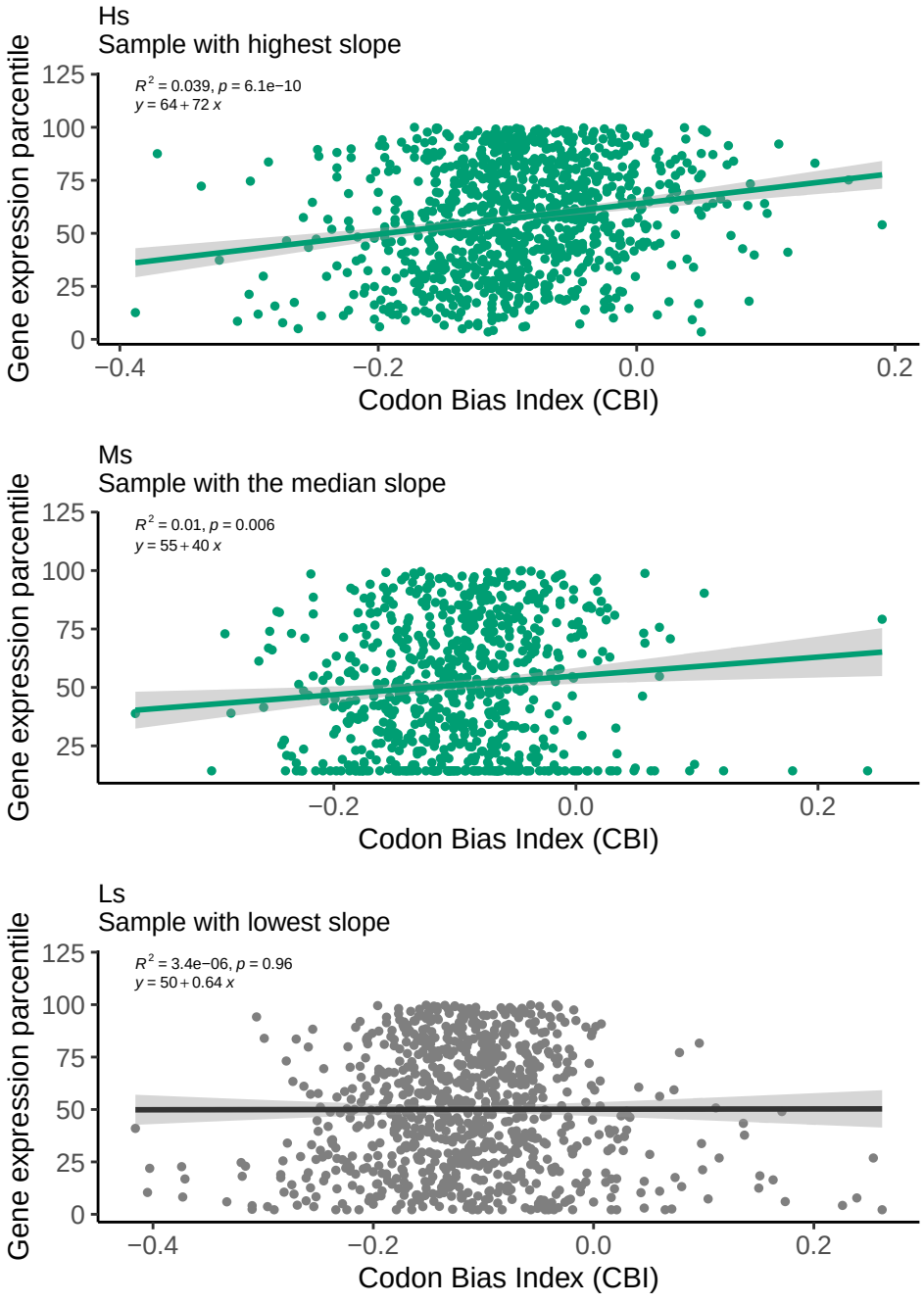
670 48. S. Narayanan, L. Kumar, S. K. Radhakrishnan, Sensory domain of the cell cycle kinase
671 CckA regulates the differential DNA binding of the master regulator CtrA in Caulobacter
672 crescentus. *Biochim. Biophys. Acta Gene Regul. Mech.* **1861**, 952–961 (2018).

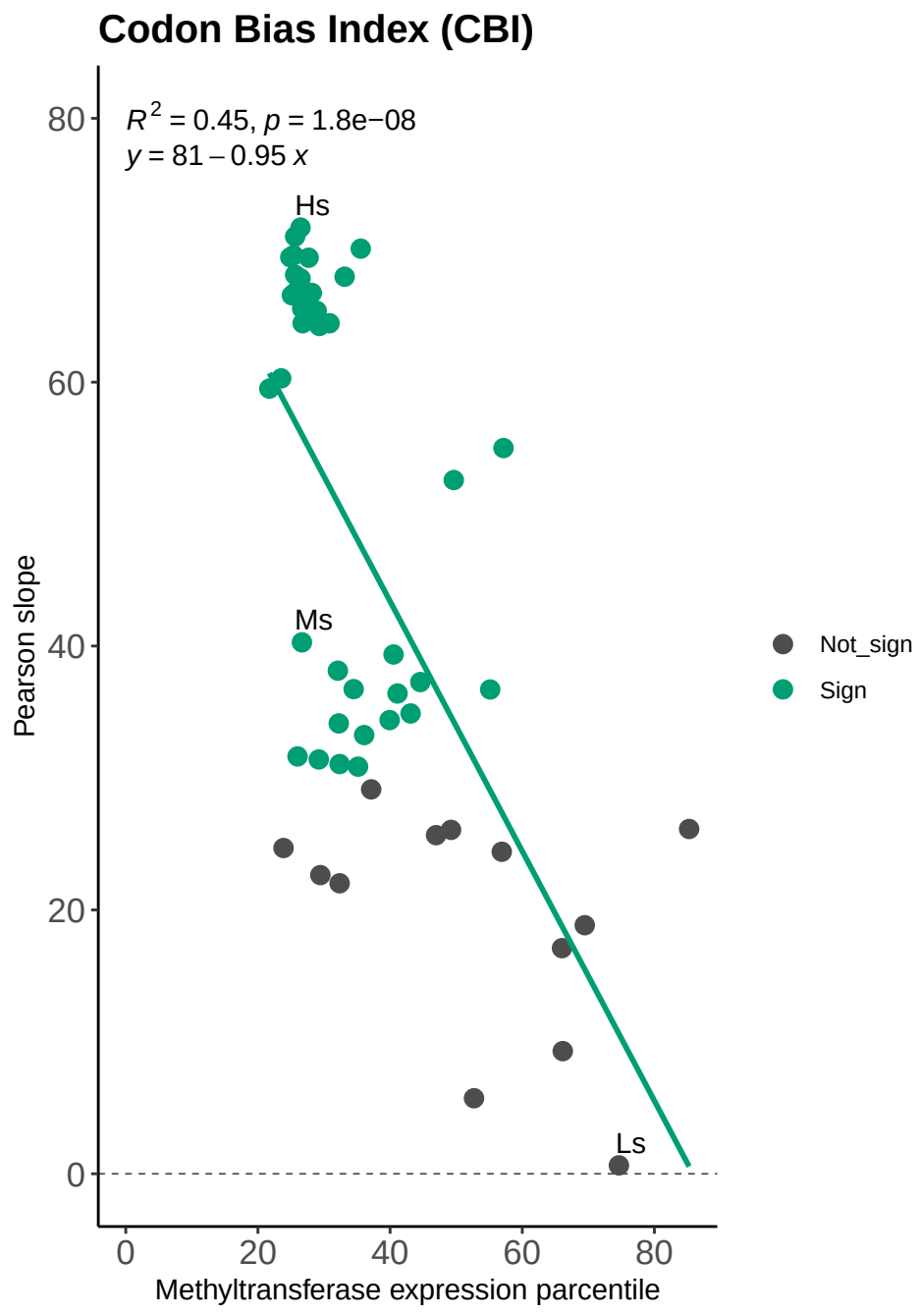
673 49. A. D. Basehoar, S. J. Zanton, B. F. Pugh, Identification and distinct regulation of yeast
674 TATA box-containing genes. *Cell* **116**, 699–709 (2004).

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


m4C
bits
m6A
bits
position

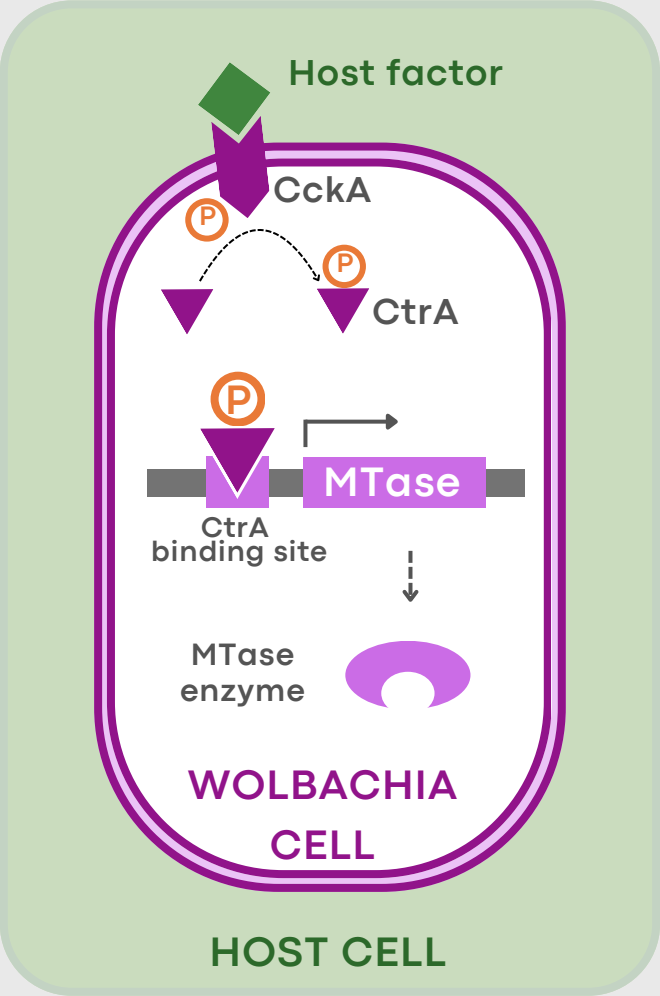

Hs
Sample with highest slope
$R^2 = 0.17, p < 2.2e-16$
$y = -6.7 + 91x$
Gene expression parcentile
%Cytosine/%Adenine
Ms
Sample with the median slope
$R^2 = 0.02, p = 0.00014$
$y = 30 + 29x$
Gene expression parcentile
%Cytosine/%Adenine
Ls
Sample with lowest slope
$R^2 = 0.0038, p = 0.056$
$y = 64 - 10x$
Gene expression parcentile
%Cytosine/%Adenine
%Cytosine/%Adenine
$R^2 = 0.45, p = 2e-08$
$y = -21 + 1.3x$
Hs
Ms
Ls
Pearson slope
Methyltransferase expression parcentile
Not_sign
Sign


Hs
Sample with highest slope
$R^2 = 0.039$, $p = 6.1e-10$
$y = 64 + 72\,x$
Gene expression parcentile
Codon Bias Index (CBI)
Ms
Sample with the median slope
$R^2 = 0.01$, $p = 0.006$
$y = 55 + 40\,x$
Gene expression parcentile
Codon Bias Index (CBI)
Ls
Sample with lowest slope
$R^2 = 3.4e-06$, $p = 0.96$
$y = 50 + 0.64\,x$
Gene expression parcentile
Codon Bias Index (CBI)

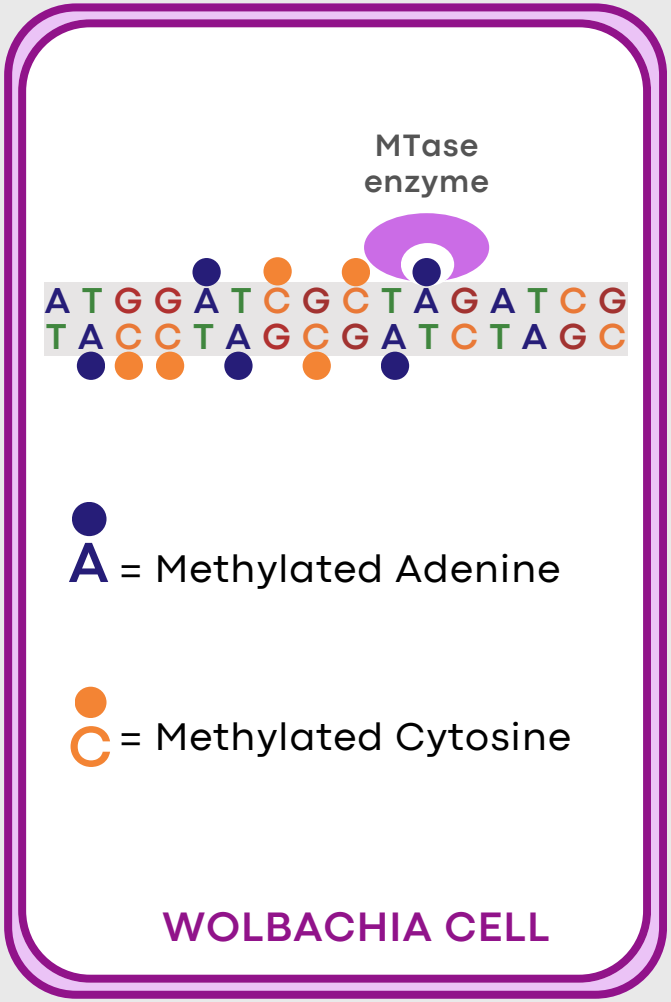
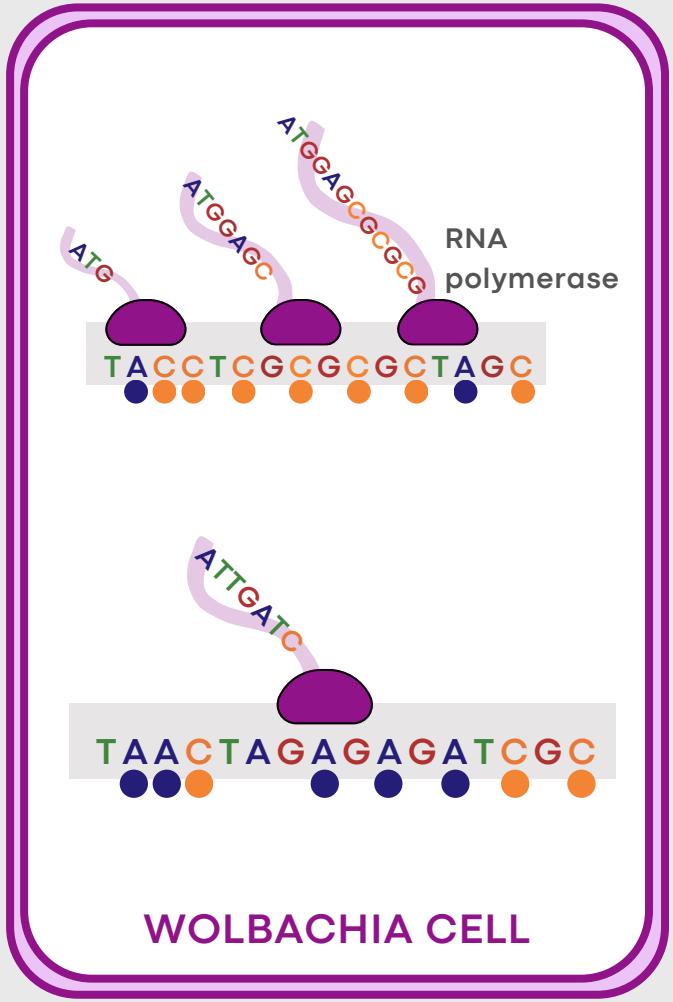

Codon Bias Index (CBI)
$R^2 = 0.45$, $p = 1.8e-08$
$y = 81 - 0.95\,x$
Hs
Ms
Ls
Pearson slope
Methyltransferase expression parcentile
Not_sign
Sign

Lowest C/A
Highest C/A
G/T ratio
Count
0
1000
2500
1
2
3
Value
M
A
B
F
C
J
I
D
T

1.MTase EXPRESSION: The host induces the MTase expression



2. DNA METHYLATION: The MTase enzyme methylates adenine and cytosines



3. GENE REGULATION: Wolbachia genes with high C/A are more expressed