Title: DNA methylation in *Ensifer* species during free-living growth and during nitrogen-fixing symbiosis with *Medicago* spp.

Running title: DNA methylation in the genus *Ensifer*

Authors: George C. diCenzo[a,b,*], Lisa Cangioli[b], Quentin Nicoud[c], Janis H.T. Cheng[a], Matthew J. Blow[d], Nicole Shapiro[d], Tanja Woyke[d], Emanuele G Biondi[c], Benoît Alunni[c], Alessio Mengoni[b,1], and Peter Mergaert[c,1]

[1] These authors contributed equally to this work.

Affiliations: [a] Department of Biology, Queen's University, Kingston ON, K7L 3N6, Canada

[b] Department of Biology, University of Florence, Florence, 50019, Italy

[c] Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell, Gif-sur-Yvette, 91198, France

[d] U.S. Department of Energy Joint Genome Institute, Berkeley CA, 94720, USA

*** Corresponding author:** George C. diCenzo
Queen's University, Department of Biology
Biosciences Complex, Room 2433
116 Barrie Street
Kingston, ON, K7P0S7, Canada
george.dicenzo@queensu.ca
+1 (613) 533-6000 x78529

Keywords: Rhizobia, Symbiotic nitrogen fixation, DNA methylation, Cell cycle regulation, CcrM

ABSTRACT

Methylation of specific DNA sequences is ubiquitous in bacteria and has known roles in immunity and regulation of cellular processes, such as the cell cycle. Using single-molecule real-time sequencing, six genome-wide methylated motifs were identified across four *Ensifer* strains, five of which were strain-specific. Only the GANTC motif, recognized by the cell cycle-regulated CcrM methyltransferase, was methylated in all strains. In actively dividing cells, methylation of GANTC motifs increased progressively from the *ori* to *ter* regions in each replicon, in agreement with a cell cycle-dependent regulation of CcrM. In contrast, there was near full genome-wide GANTC methylation in the early stage of symbiotic differentiation. This was followed by a moderate decrease in the overall extent methylation and a progressive decrease in chromosomal GANTC methylation from the *ori* to *ter* regions in later stages of differentiation. We interpret these observations as evidence of dysregulated and constitutive CcrM activity during terminal differentiation, and we hypothesize that it is a driving factor for endoreduplication of terminally differentiated bacteroids.

41 **INTRODUCTION**

42 Methylation of genomic DNA is a pervasive phenomenon found in eukaryotes (Greenberg and
43 Bourc'his, 2019; Tang et al., 2012; Zhang et al., 2018), archaea (Blow et al., 2016), and bacteria
44 (Blow et al., 2016; Sánchez-Romero et al., 2015). The biological roles of DNA methylation are
45 most extensively studied in mammals, where it contributes to normal development and disease via
46 its impact on gene expression (Gopalakrishnan et al., 2008). In bacteria, DNA methylation is best
47 known for its role in restriction-modification (R-M) systems that are thought to provide defence
48 against phage infection and limit horizontal gene transfer through the degradation of invading non-
49 methylated DNA (Vasu and Nagaraja, 2013). Several methyltransferases (MTases) of R-M
50 systems have also been implicated in phase variation in pathogens through modulating gene
51 expression (Atack et al., 2018). A recent study of over 200 bacterial and archaeal species identified
52 orphan MTases not belonging to R-M systems in nearly half of the genomes (Blow et al., 2016).
53 To date, biological functions have been attributed to very few orphan MTases, namely, the Dam
54 MTase of the γ-Proteobacteria and the CcrM MTase of the α-Proteobacteria (Adhikari and Curtis,
55 2016). The Dam MTase of *Escherichia coli* is notable for its role in regulation of DNA replication
56 (Campbell and Kleckner, 1990; Kang et al., 1999) and DNA repair (Lahue et al., 1989) by
57 modulating the activity of other DNA-binding proteins. The CcrM MTase was first identified in
58 *Caulobacter crescentus* (Zweiger et al., 1994), with homologs since identified in diverse α-
59 Proteobacteria (Brilli et al., 2010; Wright et al., 1997). CcrM activity was shown to be cell cycle
60 regulated in *C. crescentus* and *Agrobacterium tumefaciens* (Kahng and Shapiro, 2001; Zweiger et
61 al., 1994), leading to methylation of its cognate DNA motif (the pentanucleotide GANTC)
62 specifically during a short period at the end of DNA replication. This leads to a switching of
63 GANTC sites between fully methylated (methylated on both strands) and hemi-methylated

64 (methylated only on the template strand) as a result of DNA replication (Kozdon et al., 2013),
65 which serves to modulate gene expression in a cell cycle-dependent fashion (Fioravanti et al.,
66 2013; Gonzalez et al., 2014; Gonzalez and Collier, 2013). Over- and under-expression of *ccrM*
67 results in defects in DNA replication and cell division (Gonzalez and Collier, 2013; Kahng and
68 Shapiro, 2001; Wright et al., 1997), while its complete loss is lethal under some conditions.

69 The rhizobia are a polyphyletic group of α-Proteobacteria and β-Proteobacteria that can
70 both live free in the soil and enter into an endosymbiotic interaction with legumes (Wang and
71 Young, 2019). This interaction begins following an exchange of signals between the free-living
72 partners (Oldroyd, 2013), and it culminates in the formation of a new organ known as a root nodule
73 within which the cytoplasm of plant cells contain thousands of N_2-fixing bacteria called bacteroids.
74 Bacteroid formation results in the differential expression of more than a thousand genes (Barnett
75 et al., 2004; Roux et al., 2014) and global changes in cellular metabolism (diCenzo et al., 2020).
76 In legumes of the Inverted-Repeat Lacking Clade (IRLC) and the Dalbergioid clade of the family
77 *Papilionoideae*, bacteroid development involves an additional process of terminal differentiation
78 (Czernic et al., 2015; Mergaert et al., 2006); in other legume clades, bacteroid differentiation is
79 less pronounced and is reversible. Terminal bacteroid development, in contrast to reversible
80 bacteroid formation, involves cell enlargement (bacteroids are 5- to 10-fold longer than their free-
81 living counterparts) and genome endoreduplication (resulting in up to 24 copies of the genome per
82 cell), indicative of a cell cycle transition occurring during differentiation (Mergaert et al., 2006).
83 Indeed, the correct expression of cell cycle regulators in *Ensifer* (syn. *Sinorhizobium*) *meliloti*, a
84 symbiont of *Medicago* species of the IRLC, is essential for the formation of functional bacteroids
85 (Kobayashi et al., 2009; Pini et al., 2013), while over-expression of CcrM or disruption of the
86 master cell cycle regulator CtrA can give rise to bacteroid-like morphology in free-living cells

87 (Pini et al., 2015; Wright et al., 1997). Additionally, mutants in the *E. meliloti* cell cycle regulators
88 *divJ*, *cbrA*, and *cpdR1*, encoding three negative regulators of CtrA, form non-functional nodules
89 in which bacteroids do not differentiate properly (Gibson et al., 2006; Kobayashi et al., 2009, p. 1;
90 Pini et al., 2013), and genes encoding several cell cycle regulators (including CcrM) are strongly
91 downregulated in bacteroids (Roux et al., 2014). The differentiation and cell cycle switch of
92 bacteroids is controlled by the legume host through the production of a large family of peptides,
93 known as Nodule-specific Cysteine-Rich (NCR) peptides (Farkas et al., 2014; Penterman et al.,
94 2014; Van de Velde et al., 2010).

95 Multiple studies have provided evidence that changes in the methylation status of the DNA
96 of legume nodule cells contributes to symbiotic development (Nagymihaly et al., 2017; Pecrix et
97 al., 2018; Satgé et al., 2016). Conversely, it remains unknown if methylation of rhizobium DNA
98 contributes to the regulation of N_2-fixation or bacteroid development. We are aware of only one
99 study (Davis-Richardson et al., 2016) comparing DNA methylation of a rhizobium
100 (*Bradyrhizobium diazoefficiens* USDA110) between free-living and symbiotic states (soybean
101 nodules), and comparing these changes with differential expression data. Intriguingly, the authors
102 identified a DNA motif that was methylated specifically in bacteroids (Davis-Richardson et al.,
103 2016). However, no clear evidence was presented that methylation of this (or any other) motif is
104 involved in transcriptional regulation, and the number of genes both differentially expressed and
105 differentially methylated in bacteroids did not appear to be different than expected by chance.
106 While these data may suggest that DNA methylation does not play a major role in regulating N_2-
107 fixation by rhizobia, they do not address the role of DNA methylation in terminal bacteroid
108 differentiation, as *B. diazoefficiens* undergoes reversible differentiation in soybean nodules
109 (Barrière et al., 2017).

110 Here, we use Pacific Biosciences Single-Molecule Real-Time (SMRT) sequencing to
111 detect genome-wide patterns of DNA methylation in four strains belonging to the genus *Ensifer*.
112 Our results indicate that DNA methylation is poorly conserved across the genus and suggests that
113 DNA methylation is not a major mechanism of regulating gene expression in these organisms aside
114 from cell cycle control. However, analysis of bacteroid samples led us to hypothesize that
115 constitutive activation of the CcrM MTase may be a contributing factor driving terminal
116 differentiation.
117

118 **RESULTS**

119 **The methylomes of the genus *Ensifer***

120 Our experimental design, as summarized in the Materials and Methods and **Figure S1**, was
121 developed to support an investigation into multiple potential roles of DNA methylation in plant-
122 associated bacteria from the genus *Ensifer* through the use of SMRT sequencing. This was
123 accomplished by: i) including DNA samples isolated from phylogenetically diverse wild-type
124 strains, ii) examining DNA methylation in a single strain across multiple conditions (exponential
125 phase growth versus stationary phase; growth with sucrose versus growth with succinate), iii)
126 investigating the impact of a large-scale genome reduction on DNA methylation patterns, and iv)
127 isolating DNA from bacteroids purified from legume nodules.

128 We began with base modification analyses of four wild-type strains from three species,
129 including three nodule-forming strains (*E. meliloti* Rm2011, *E. meliloti* FSM-MA, *E. fredii*
130 NGR234) and one plant-associated, non-symbiotic strain (*E. adhaerens* OV14). To ensure
131 consistency, all strains were grown to mid-exponential phase in a common minimal medium with
132 succinate as the carbon source. A total of six methylated motifs were identified, of which five were

133 m6A modifications and one was a m4C modification (**Table 1**). Five of the six motifs were
134 methylated specifically in one strain. Only the GANTC motif, recognized by the highly conserved
135 cell cycle-regulated CcrM methyltransferase (Wright et al., 1997; Zweiger et al., 1994), was
136 methylated in all four strains. To further examine the conservation, or lack thereof, of DNA
137 modification across the genus *Ensifer*, we examined the distribution of methyltransferases in the
138 model species *E. meliloti*. Based on gene annotations, we identified 24 genes encoding putative
139 MTases in a previous pangenome analysis of 20 *E. meliloti* strains (**Table S1**) (diCenzo et al.,
140 2019). Of these 24 genes, only one (*ccrM*) was found in all 20 strains, while four were found in
141 two strains and 19 were found in a single strain. These results suggest that DNA methylation is
142 unlikely to play a biologically significant role in the genus *Ensifer* aside from cell cycle control
143 via CcrM-mediated methylation, and phage defence.

144 In support of most DNA methylation not having a regulatory function in the genus *Ensifer*,
145 none of the motifs methylated in *E. meliloti* Rm2011 were enriched in the promoter regions of
146 genes previously shown to be differentially expressed when grown with glucose vs. succinate
147 (diCenzo et al., 2017). Similarly, except for the GANTC motif as discussed below, no global effect
148 of carbon source (sucrose [glycolytic] vs. succinate [gluconeogenic]) was observed on the DNA
149 methylation pattern of *E. meliloti* Rm2011 (**Figure S2**). Similarly no global differences in DNA
150 methylation were detected between *E. meliloti* Rm2011 and RmP3496, a Rm2011 derivative
151 lacking the pSymA and pSymB replicons that together account for 45% of the genome content of
152 *E. meliloti* (diCenzo et al., 2014) (**Figure S3**).

153

154 **Table 1.** Methylated motifs identified in this study.

Motif*	Type†	Count¥	Frequency (motifs/kb)
E. meliloti 2011			
G**A**NTC CTN**A**G	m6A	11,169	1.67
RCG**C**CTC YGCGGAG	m4C	3,943	0.59
CGC**A**(N5)GTG GCGT(N5)C**A**C	m6A	1,085	0.16
E. meliloti FSM-MA			
G**A**NTC CTN**A**G	m6A	11,215	1.67
TCG**A**(N8)TCGA AGCT(N8)**A**GCT	m6A	2,612	0.39
E. fredii NGR234			
G**A**NTC CTN**A**G	m6A	11,111	1.61
CAG**A**(N7)GTTG GTCT(N7)C**A**AC	m6A	188	0.03
E. adhaerens OV14			
G**A**NTC CTN**A**G	m6A	8,475	1.10
WNCCG**A**TG WNGGCTAC	m6A	4,596	0.60

155 * The methylated nucleotides are indicated in boldface font.

156 † Indicates whether the modification is a N^6-methyladenoside (m6A) or N^4-methylcytosine (m4C).

157 ¥ The total times the motif appears in the genome, regardless of methylation status.

158

159

160 **Cell cycle regulation by the CcrM methyltransferase**

161 A progressive increase in the extent of methylation (herein defined as the estimated fraction of
162 reads mapping to a motif that were methylated) of GANTC sites was observed from the *ori* to *ter*
163 regions of the chromosomes of all four strains during mid-exponential growth (**Figures 1, S4-S6**).
164 There was a local drop in GANTC methylation around the 1.5 Mb mark in the *E. meliloti* FSM-
165 MA chromosome (**Figure S4**); however, this was seen in only two of three replicates and
166 corresponded to a region of high sequencing depth (**Figure S7**), suggesting the result is a

167 sequencing artefact. In contrast to exponential phase cultures, GANTC sites displayed near full
168 methylation (averaging ~ 95%) across the genome during early stationary phase, while all other
169 motifs displayed near full methylation (averaging 95-99%) across the genome regardless of growth
170 state (**Figures 1, S4-S6**). The observed pattern of GANTC methylation indicates a progressive
171 switch from full to hemi-methylated states as DNA replication proceeds (model provided as
172 **Figure S8**), confirming that the CcrM methyltransferase of the family *Rhizobiaceae* is cell cycle
173 regulated as demonstrated in *C. crescentus* (Kozdon et al., 2013; Mohapatra et al., 2014; Zweiger
174 et al., 1994). Interestingly, the genome-wide pattern of GANTC methylation displayed a smaller
175 variation in the extent of methylation from the *ori* to *ter* regions in *E. meliloti* Rm2011 when grown
176 with sucrose compared to succinate as the carbon source (**Figure S2**). While this observation could
177 suggest metabolic regulation of CcrM activity, we instead hypothesize, as displayed in **Figure**
178 **S8A**, that it is due to DNA replication being initiated later in the cell cycle when *E. meliloti* is
179 provided sucrose, as recent observations showed that central carbon metabolism influences the
180 rate of DNA polymerase processivity and timing of DNA replication initiation in *Bacillus subtilis*
181 (Nouri et al., 2018).

182 Notably, the GANTC methylation pattern differed across replicons within each genome.
183 For example, in *E. meliloti* Rm2011 the extent of GANTC methylation ranged from 0.80 to 0.98
184 for pSymB and 0.80 to 0.96 for pSymA, while for the chromosome the range was from 0.71 to
185 0.99 (**Figure 1**). This result suggests that replication of each replicon is asynchronous, with
186 replication of the secondary replicons being initiated later in the cell cycle than that of the
187 chromosome (model provided as **Figure S8**).
188

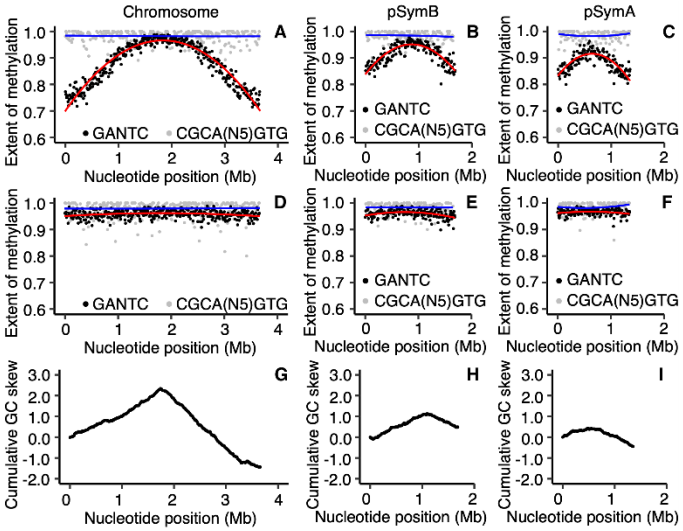


Figure 1. Genome-wide DNA methylation of *E. meliloti* Rm2011. (**A**-**F**) The extent of methylation is shown, using a 10 kb sliding window, of GANTC sites (black) and CGCA(N_5)GTG sites (grey) across the chromosome (**A**,**D**), pSymB (**B**,**E**), and pSymA (**C**,**F**) replicons of exponential phase (**A**-**C**) or early stationary phase (**D**-**F**) *E. meliloti* Rm2011. Averages from three biological replicates are shown. The red (GANTC) and blue (CGCA(N_5)GTG) lines are polynomial regression lines calculated in R using the "rlm" method and the formula "y~poly(x,2)". (**G**-**I**) Cumulative GC skews, shown using a 10 kb sliding window, across the *E. meliloti* Rm2011 chromosome (**G**), pSymB (**H**), and pSymA (**I**) replicons.

A previous study identified 462 cell cycle-regulated genes in *E. meliloti* through RNA-sequencing of synchronized cell populations (De Nisco et al., 2014), which were classified into six groups based on the timing of their expression. We identified 111 cell cycle-regulated genes, belonging to 78 transcripts, that contained at least one GANTC in the predicted promoter regions (defined as the 125 bp upstream of the transcript; **Dataset S1**), and distribution of these 111 genes across the six cell cycle gene expression groups was unbiased (De Nisco et al., 2014). As these 111 genes are both cell cycle regulated and contain a GANTC site, they represent an initial candidate CcrM regulon in *E. meliloti* although further work is required to validate the CcrM regulon.

209 We found it striking that the *E. adhaerens* OV14 genomes had 2,636 to 2,740 fewer
210 GANTC sites than the other three strains, despite having the largest genome size. Normalized by
211 genome length, there are 1.10 GANTC sites per kb in the *E. adhaerens* OV14 genome (**Table 1**),
212 which is similar to the 1.12 GANTC sites per kb in *C. crescentus* NA1000. In contrast, the three
213 legume symbionts contained more than 1.60 GANTC sites per kb across their genomes (**Table 1**).
214 This result prompted us to examine the frequency and distribution of GANTC sites across 157
215 *Ensifer* genomes. As defined previously (Fagorzi et al., 2020), the genus *Ensifer* can be broadly
216 sub-divided into two monophyletic clades; the "symbiotic" clade (113 strains) in which nearly all
217 strains are legume symbionts, and the "non-symbiotic" clade (44 strains) in which nearly all strains
218 are non-symbionts (**Figure 2A**). Consistent with previous results (Gonzalez et al., 2014), GANTC
219 sites occurred less frequently in all genomes (0.90 to 1.83 GANTC sites per kb) than expected in
220 a random sequence of nucleotides (~ 3.5 GANTC sites per kb). Moreover, GANTC sites were ~
221 2-fold more common in intergenic regions than in coding regions (**Figure 2B**). Strikingly, there
222 was a strong and statistically significant difference (p-value < 1 x 10^{-10}; two-sample t-test) in the
223 frequency of GANTC sites across the genomes of strains belonging to the symbiotic clade
224 compared to strains of the non-symbiotic clades (**Figure 2B**), with an overall average of 1.70 and
225 1.06 GANTC sites per kb in the symbiotic and non-symbiotic clades, respectively. The difference
226 in the frequency of GANTC sites between the two clades could not be explained by differences in
227 the GC content of these organisms, as both clades had an average GC content of 61.9% (**Figure**
228 **S9**), suggesting that the difference reflects underlying differences in the evolution, and possibly
229 the biology, of these two clades.
230

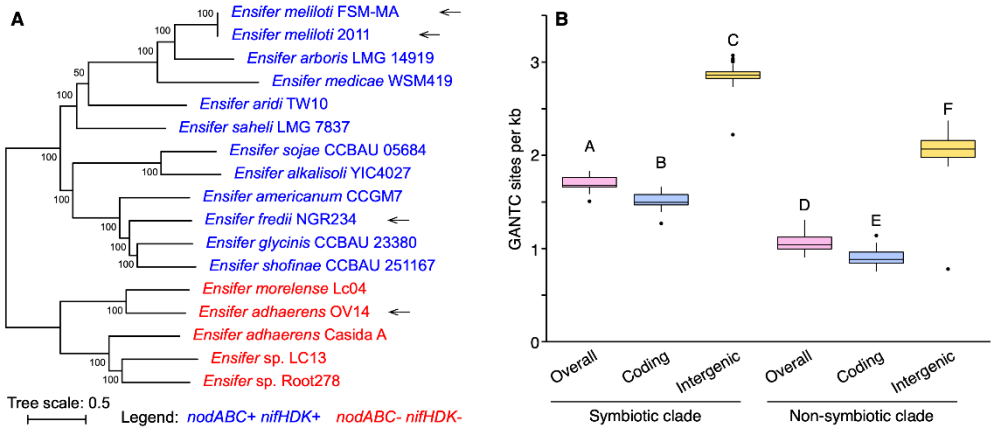


Figure 2. GANTC frequency in the genus *Ensifer*. (**A**) An unrooted maximum likelihood phylogeny of 17 representative *Ensifer* strains. The phylogeny represents the bootstrap best tree following 100 bootstrap replicates, prepared on the basis of the concatenated nucleotide alignments of 1566 core genes. Values represent the bootstrap support. N_2-fixing legume symbionts were identified by the presence of the symbiotic genes *nodABC* and *nifHDK*. They are indicated in blue, while red denotes non-symbiotic strains. The four wild-type strains used in this study are indicated with arrows. (**B**) Box plots summarizing the frequency of GANTC sites (presented as GANTC sites per kb) in 157 *Ensifer* strains is shown. The monophyletic "symbiotic" and "non-symbiotic" clades as defined previously (Fagorzi et al., 2020), are represented by 113 and 44 genomes respectively. The densities of GANTC sites across the entire genome (pink), within coding regions (blue), and within intergenic regions (yellow) are shown. Statistically different values ($p < 0.05$) are denoted by uppercase letters as determined by a one-way ANOVA followed by a Tukey's HSD post hoc test.

Contributions of DNA methylation to bacteroid differentiation

The only previously published study to examine the role of rhizobium DNA methylation during symbiosis using SMRT sequencing did so in a symbiosis where the bacteria do not undergo terminal differentiation (Barrière et al., 2017; Davis-Richardson et al., 2016). To evaluate whether DNA methylation potentially contributes to regulation of terminal differentiation, we determined

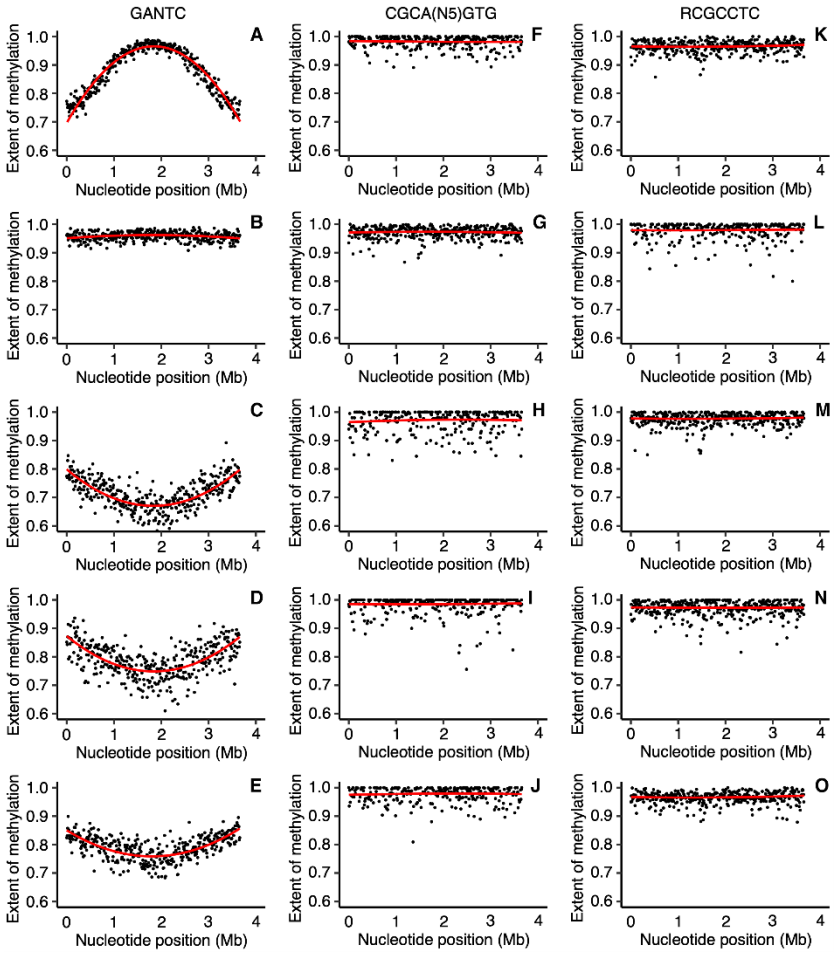
252 the DNA methylation patterns of *E. meliloti* Rm2011 and *E. meliloti* FSM-MA bacteroids purified
253 from whole *Medicago sativa* nodules. *E. meliloti* FSM-MA bacteroids were additionally purified
254 from *Medicago truncatula* nodules to determine whether the host plant influences bacteroid DNA
255 methylation. *E. meliloti* Rm2011 bacteroids were not isolated from *M. truncatula* nodules as,
256 unlike FSM-MA, Rm2011 forms a poor symbiosis with *M. truncatula* and produces moderately
257 differentiated bacteroids in this host (Kazmierczak et al., 2017; Moreau et al., 2008).

258 Moreover, we exploited the spatially distinct developmental zones that are present in
259 indeterminate nodules (Vasse et al., 1990), like those formed by *M. sativa* and *M. truncatula*. At
260 the tip of these nodules, a bacteria-free meristem is present, responsible for the continuous growth
261 of the nodule. Immediately below is the infection and differentiation zone II where nodule cells
262 become infected and bacteria differentiate into the large, polyploid bacteroids. The tip and zone II
263 of nodules appears white. Adjacent to the white zone II is the easily recognizable pinkish zone III
264 (due to the presence of the oxygen-carrying leghemoglobin) where mature bacteroids fix nitrogen.
265 This nodule tissue organization provided an opportunity to examine how DNA methylation
266 patterns may differ between differentiating and differentiated bacteroids. To this end, *E. meliloti*
267 Rm2011 and FSM-MA bacteroids were isolated from nodules hand-sectioned at the white-pink
268 border; bacteroids isolated from the white sections represent the infecting and differentiating
269 bacteroids (zone II) while those isolated from the pink sections represent the mature, hence
270 terminally differentiated, N_2-fixing bacteroids (zone III).

271 Fluorescence microscopy and flow cytometry confirmed that nodule sectioning resulted in
272 the isolation of distinct bacteroid populations (**Figures S10-S13**). Nearly all of the bacteroids
273 isolated from whole-nodule samples and zone III samples were enlarged and polyploid, and most
274 were positive for propidium iodide (PI) staining as expected for terminally differentiated

275 bacteroids (Mergaert et al., 2006). In contrast, bacteroids of the zone II samples contained a mix
276 of cell types differing in their size, ploidy level, and PI staining. These data confirmed that the
277 whole-nodule samples and zone III samples consisted predominately of mature N_2-fixing
278 bacteroids, whereas the zone II samples contained a mix of cells at various stages of bacteroid
279 differentiation.

280 With the exception of the GANTC sites (described below), no global difference was
281 observed in the methylation patterns of bacteroids versus free-living cells (**Figures 3, 4, S14-S17**).
282 Although a lower percentage of each motif was detected as methylated in the bacteroid samples
283 compared to free-living samples, this was correlated with lower sequencing depth (**Table S2**) and
284 thus unlikely to be biologically meaningful. Unlike *B. diazoefficiens* USDA110 (Davis-Richardson
285 et al., 2016), we did not identify motifs that were methylated specifically in the *E. meliloti*
286 bacteroids. In addition, we found little evidence for any of the methylated motifs being enriched
287 in the promoter regions of *E. meliloti* Rm2011 genes up-regulated or down-regulated in bacteroids
288 relative to free-living cells, as identified in published transcriptomic data for *E. meliloti* Rm1021
289 (Barnett et al., 2004), a near-isogenic relative of strain Rm2011 also derived from the nodule
290 isolate SU47. These data suggest that most DNA methylation is unlikely to be a significant factor
291 in directly regulating gene expression in *E. meliloti* bacteroids.
292



293

294 **Figure 3. Chromosome-wide DNA methylation of *E. meliloti* Rm2011 bacteroids.** The extent
295 of methylation of (**A-E**) GANTC, (**F-J**) CGCA(N_5)GTG, and (**K-O**) RCGCCTC motifs across the
296 *E. meliloti* Rm2011 chromosome is shown using a 10 kb sliding window. Averages from three
297 biological replicates are shown for free-living and whole nodule samples; data represents one
298 replicate for the zone II and zone III nodule sections. (**A**,**F**,**K**) Free-living cells harvested in mid-
299 exponential phase. (**B**,**G**,**L**) Free-living cells harvested in early stationary phase. (**C**,**H**,**M**)
300 Bacteroids isolated from *M. sativa* zone II nodule sections. (**D**,**I**,**N**) Bacteroids isolated from *M.*
301 *sativa* zone III nodule sections. (**E**,**J**,**O**) Bacteroids isolated from *M. sativa* whole nodule samples.
302 The red lines are polynomial regression lines calculated in R using the "rlm" method and the
303 formula "y~poly(x,2)". Data for pSymB and pSymA are shown in Figures S14 and S15.

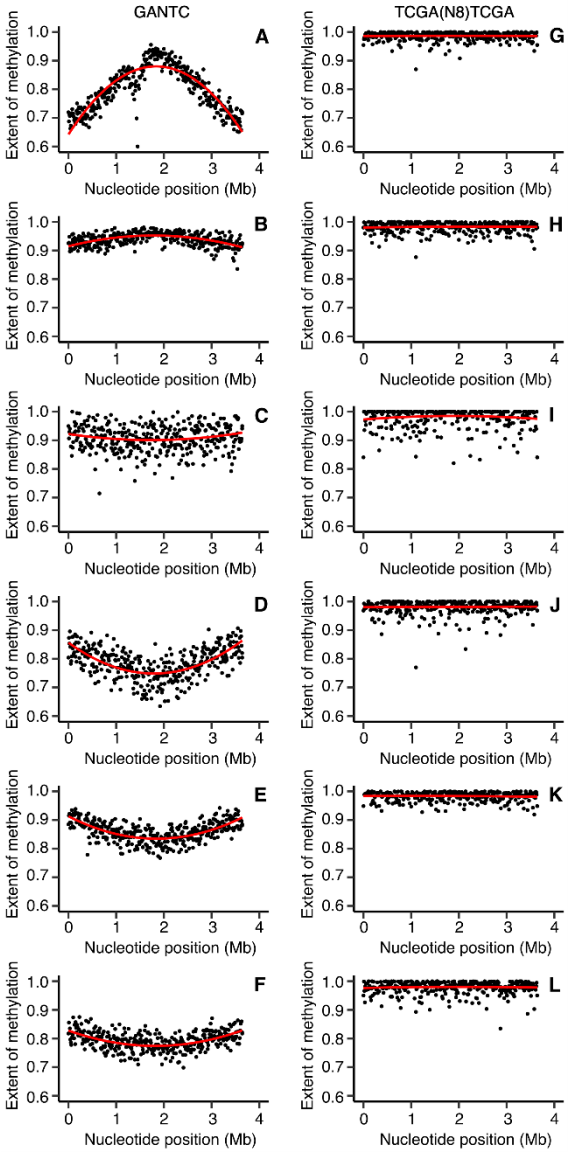


Figure 4. Chromosome-wide DNA methylation of *E. meliloti* FSM-MA bacteroids. The extent of methylation of (**A-F**) GANTC and (**G-L**) TCGA(N_8)TCGA motifs across the *E. meliloti* FSM-MA chromosome is shown using a 10 kb sliding window. Averages from three biological replicates are shown for free-living and whole nodule samples; data represents one replicate for the zone II and zone III nodule sections. (**A**,**G**) Free-living cells harvested in mid-exponential phase. (**B**,**H**) Free-living cells harvested in early stationary phase. (**C**,**I**) Bacteroids isolated from *M. sativa* zone II nodule sections. (**D,J**) Bacteroids isolated from *M. sativa* zone III nodule sections. (**E**,**K**) Bacteroids isolated from *M. sativa* whole nodule samples. (**F**,**L**) Bacteroids isolated from *M. truncatula* whole nodule samples. The red lines are polynomial regression lines calculated in R using the "rlm" method and the formula "y~poly(x,2)". Data for pSymB and pSymA are shown in Figures S16 and S17.

325

326

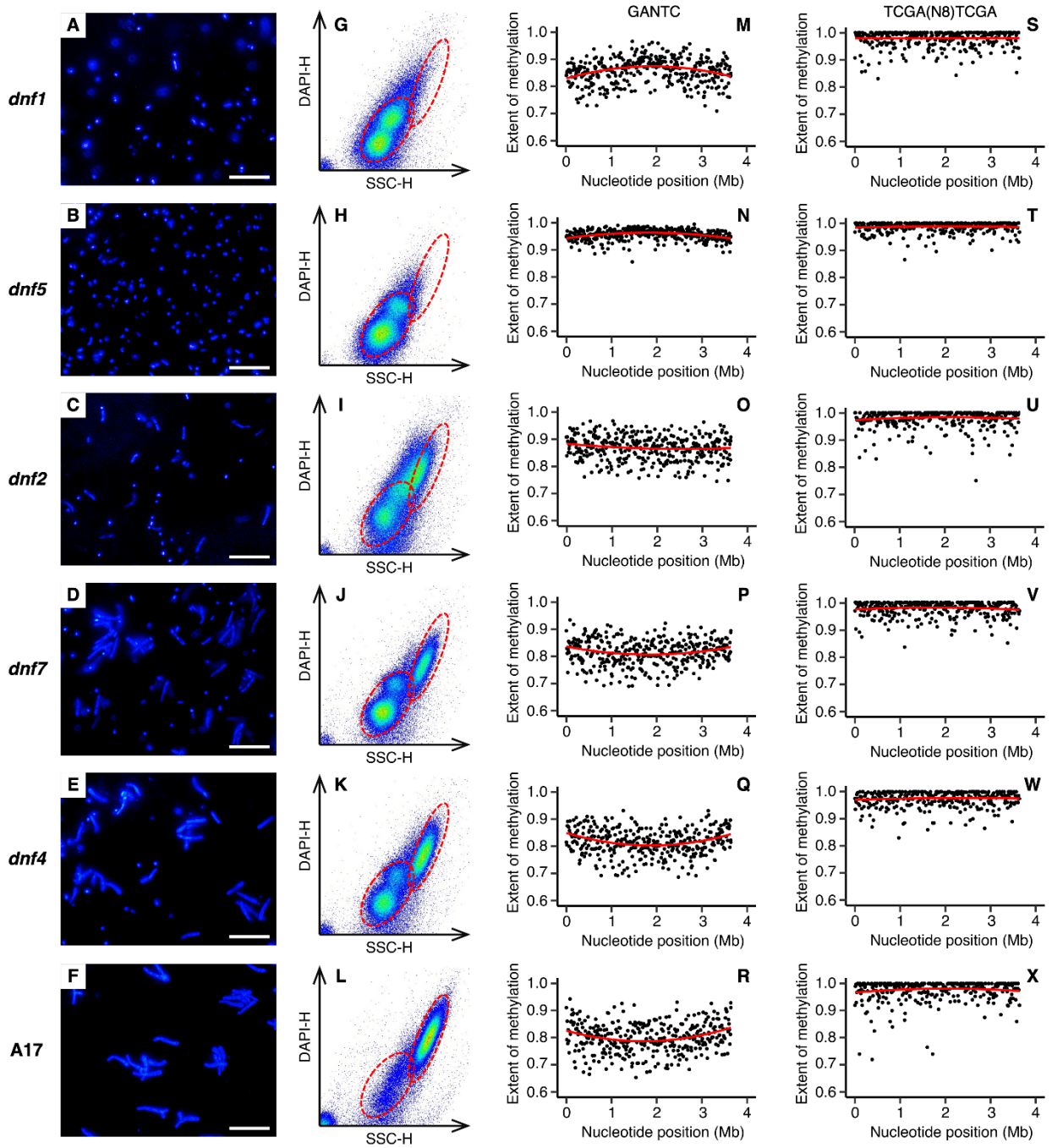
327 **CcrM methyltransferase activity is dysregulated during terminal differentiation**

328 Bacteroid development involves cell enlargement and genome endoreduplication, indicative of a
329 cell cycle transition occurring during differentiation (Mergaert et al., 2006). Indeed, expression of
330 *ccrM* and other genes encoding cell cycle regulators vary across stages of bacteroid development
331 and are strongly downregulated in mature nitrogen-fixing *E. meliloti* bacteroids (Roux et al., 2014).
332 We were therefore interested in examining whether GANTC methylation by the CcrM MTase was

333 disrupted in bacteroids. Our data revealed a surprising genome-wide pattern of GANTC
334 methylation in *E. meliloti* Rm2011 and FSM-MA bacteroids, which differed from free-living cells
335 in either the exponential or stationary phases of growth (**Figures 3, 4, S14-S17**). The majority of
336 GANTC sites had moderate to high levels of methylation in zone II, zone III, or whole-nodule
337 samples, averaging 0.71 to 0.95 across each replicon (**Figures 3, 4, S14-S17, and Tables S3, S4**).
338 Most distinctive, a progressive decrease in the extent of chromosomal methylation of the GANTC
339 sites was observed from the *ori* to *ter* in nearly all bacteroid samples, revealing a characteristic
340 smiling pattern, which differs from the patterns seen in exponential (frowning pattern, i.e., a
341 progressive increase from *ori* to *ter*) and stationary (consistent methylation) phase cells. The
342 exception was the *E. meliloti* FSM-MA zone II bacteroid sample, which displayed a consistently
343 high level of GANTC methylation across the genome (**Figure 4C**). This pattern, which is different
344 from those in exponential phase cells as well as mature bacteroids, could correspond to the
345 methylation status of an early stage of bacteroid differentiation. We did not observe the same
346 pattern in the *E. meliloti* Rm2011 zone II samples. As noted earlier (**Figures S10-S13**), the zone
347 II samples contain cells at various stages of differentiation. Given that terminal differentiation is
348 associated with an up to 24-fold increase in DNA content, small increases in the proportion of cells
349 at late stages of differentiation could mask the DNA methylation pattern of the cells at early stages
350 of differentiation. Thus, we hypothesize that the Rm2011 zone II sample captures a later stage of
351 differentiation than that captured by the FSM-MA zone II sample. Supporting this hypothesis, the
352 distribution of DNA content per cell in the flow cytometry data was flatter for *E. meliloti* Rm2011
353 zone II bacteroids compared to *E. meliloti* FSM-MA zone II bacteroids (**Figure S18**), which
354 suggests that the former sample represents a broader range of differentiation stages than the latter
355 sample. In contrast to GANTC, the extent of methylation of the second m6A modified motif in

356 each genome was consistently high, irrespective of condition or replicon (**Figures 3, 4, S14-S17,**
357 **and Tables S3, S4**). Similarly, sequencing depth was consistent across the length of each replicon
358 (**Figure S7**). These observations indicate that the changes in GANTC methylation patterns are
359 biologically meaningful and not simply a sequencing artefact.

360 To further explore changes in CcrM activity during bacteroid differentiation, we took
361 advantage of a collection of *M. truncatula* mutant plant lines (*dnf1*, *dnf2*, *dnf4*, *dnf5*, *dnf7*) whose
362 nodules contain bacteria blocked at various stages of differentiation (Bourcy et al., 2013;
363 Domonkos et al., 2013; Horváth et al., 2015; Kim et al., 2015; Lang and Long, 2015; Starker et
364 al., 2006; Wang et al., 2010). Microscopy and flow cytometry data was consistent with past work
365 (Lang and Long, 2015) showing that bacteroids were blocked at the earliest to latest stages of
366 differentiation in mutant plant lines in the order *dnf1* → *dnf5* → *dnf2* → *dnf7* → *dnf4* (**Figure 5**).
367 Nodule bacteria of *M. truncatula dnf1* mutant plants were small with one or two haploid genome
368 copies per cell (i.e., ploidy level = 1 or 2) (**Figures 5A, 5G**), suggesting that the cell population
369 was dominated by actively dividing cells that had not yet begun differentiation. Indeed, the
370 GANTC methylation pattern of these cells (**Figures 5M, S19, S20**) resembled the frowning pattern
371 of exponentially growing free-living cells (**Figure 4A**). Although the nodule bacteria of *M.*
372 *truncatula dnf5* mutant plants were also small and undifferentiated into bacteroids, the majority of
373 cells had a ploidy level of one (**Figures 5B, 5H**), suggesting these cells had ceased replication but
374 had not yet begun the process of endoreduplication. GANTC methylation was consistently high
375 across the chromosome of bacteria purified from *dnf5* nodules (**Figure 5N**), similar to stationary
376 phase free-living cells (**Figure 4B**) and indicating that terminal differentiation is preceded by full
377 GANTC methylation.
378



379

380 **Figure 5. Bacteroid morphology and chromosomal GANTC methylation in *E. meliloti***
381 **bacteroids purified from *M. truncatula dnf* mutant nodules.** Data is shown for *E. meliloti* FSM-
382 MA bacteroids purified from (**A**,**G**,**M**,**S**) *M. truncatula dnf1* mutant nodules, (**B**,**H**,**N**,**T**) *M.*
383 *truncatula dnf5* mutant nodules, (**C**,**I**,**O**,**U**) *M. truncatula dnf2* mutant nodules, (**D**,**J**,**P**,**V**) *M.*
384 *truncatula dnf7* mutant nodules, (**E**,**K**,**Q**,**W**) *M. truncatula dnf4* mutant nodules, and (**F**,**L**,**R**,**X**)
385 *M. truncatula* A17 wild-type nodules. (**A-F**) Micrographs of *E. meliloti* FSM-MA bacteroids

386 stained with the DNA binding dye DAPI. The scale bar represents 30 µm. (**G-L**) Pseudo-coloured
387 scatterplots displaying the cell morphology (X-axis) and DNA content (Y-axis) of *E. meliloti*
388 FSM-MA bacteroids, as determined based on flow cytometry analysis of DAPI stained cells. The
389 red dashed ellipses indicate the position of undifferentiated bacteria as in culture (not shown) or in
390 the *dnf1* mutant nodules (lower left ellipse) or fully mature bacteroids as in the A17 wild-type
391 nodules (top right ellipse). (**M-X**) The extent of methylation of (**M-R**) GANTC or (**S-X**)
392 TCGA(N_8)TCGA motifs across the *E. meliloti* FSM-MA chromosome, shown using a 10 kb
393 sliding window. The red lines are polynomial regression lines calculated in R using the "rlm"
394 method and the formula "y~poly(x,2)". Data for pSymB and pSymA are shown in Figures S19
395 and S20.
396
397
398 The nodule bacteria purified from *M. truncatula dnf2* mutant nodules were a mix of
399 undifferentiated and partially differentiated bacteroids, which were polyploid to an extent similar
400 to bacteroids purified from wild-type A17 nodules (**Figure 5I** compared to **Figure 5L**); however,
401 their cell size was much smaller (**Figure 5C** compared to **Figure 5F**). This was similar to
402 differentiating bacteroids purified from *M. truncatula* and *M. sativa* zone II nodule sections, many
403 of which had high ploidy without a corresponding increase in cell size (**Figures S11, S13**). The
404 GANTC methylation pattern of bacteroids from *dnf2* nodules (**Figure 5O**) was also similar to that
405 of zone II nodule sections. There was a consistently high extent of GANTC methylation across the
406 chromosome averaging 0.870, which was less than that of bacteroids purified from *dnf5* nodules
407 (0.956) but higher than that of bacteroids purified from wild-type A17 nodules (0.804) (**Table S5**),
408 and without the smiling pattern. The nodule bacteria purified from *M. truncatula dnf7* and *dnf4*
409 nodules also contained a mix of undifferentiated bacteria and fully differentiated bacteroids
410 (**Figures 5D, 5E, 5J, 5K**), with the number of undifferentiated bacteria greater in *dnf7* nodules
411 compared to *dnf4* nodules. The GANTC methylation pattern of bacteroids purified from *dnf7* and

412 *dnf4* nodules was similar to that of bacteroids purified from A17 nodules (**Figures 5P-Q**). Overall,
413 we interpret the data from bacteroids purified from section nodules and *M. truncatula dnf* mutant
414 nodules as suggesting that CcrM is dysregulated during terminal bacteroid differentiation and that
415 CcrM is constitutively active during endoreduplication.
416
417 **Chromosome, pSymB, and pSymA sequencing depth are unequal in *E. meliloti* bacteroids**
418 We noticed that in each bacteroid sample, the average extent of GANTC methylation for the
419 chromosomes of the two strains were lower (by 0.04 to 0.13) than that of pSymA or pSymB, and
420 unlike the chromosome, the extent of GANTC methylation was relatively constant across pSymA
421 and pSymB (**Figures S14-S17** compared to **Figures 3 and 4**). These results suggest that, unlike in
422 free-living cells, replication of the three replicons is not well coordinated during terminal
423 differentiation. In agreement with this hypothesis, the mean sequencing depth across pSymA and
424 pSymB was on average ~ 33% lower than that of the chromosome in all replicates of the *E. meliloti*
425 whole-nodule bacteroid samples (**Table 2**). Similarly, the mean sequencing depth across pSymA
426 and pSymB was on average ~ 23% lower than that of the chromosome for the polyploid bacterial
427 cell populations purified from *M. truncatula dnf2*, *dnf7*, and *dnf4* mutant nodules, but not for the
428 haploid/diploid bacterial cell populations purified from *M. truncatula dnf1* and *dnf5* mutant
429 nodules (**Table 2**). Assuming sequencing depth is correlated with copy number, this observation
430 suggests that *E. meliloti* bacteroids carry approximately two copies each of pSymA and pSymB
431 per three copies of the chromosome.
432

433 **Table 2.** Relative sequencing depth of each *E. meliloti* replicon.

		Relative mean sequencing depth *		
Strain	Condition	Chromosome	pSymA	pSymB
Rm2011	Mid-exponential	1.00 ± 0.00	1.00 ± 0.03	0.98 ± 0.03
Rm2011	Stationary phase	1.00 ± 0.00	1.06 ± 0.02	1.02 ± 0.03
Rm2011	*M. sativa* bacteroids	1.00 ± 0.00	0.64 ± 0.02	0.62 ± 0.02
FSM-MA	Mid-exponential	1.00 ± 0.00	1.02 ± 0.03	0.87 ± 0.05
FSM-MA	Stationary phase	1.00 ± 0.00	1.10 ± 0.00	1.01 ± 0.03
FSM-MA	*M. sativa* bacteroids	1.00 ± 0.00	0.79 ± 0.04	0.71 ± 0.02
FSM-MA	*M. truncatula* bacteroids	1.00 ± 0.00	0.66 ± 0.02	0.59 ± 0.01
FSM-MA	*dnf1* bacterial cells	1.00	0.98	0.90
FSM-MA	*dnf5* bacterial dells	1.00	0.96	0.94
FSM-MA	*dnf2* bacterial dells	1.00	0.81	0.73
FSM-MA	*dnf7* bacterial dells	1.00	0.82	0.71
FSM-MA	*dnf4* bacterial dells	1.00	0.80	0.73
FSM-MA	A17 bacteroids	1.00	0.76	0.67

434 * Sequencing depth is presented relative to the sequencing depth of the chromosome in the same
435 sample. Values are the means of triplicate samples ± standard deviation, except for the third section
436 of the table for which numbers are based on a single replicate.

437

438

439 **DISCUSSION**

440 In this study, we examined the genome-wide DNA methylation patterns in the free-living cells of
441 four *Ensifer* strains, and in bacteroids of two *E. meliloti* strains, and detected a total of six
442 methylated motifs. We were able to predict cognate MTases for most of the motifs based on
443 genome annotations, the exception being the WNCCGATG motif of *E. adhaerens* OV14. The
444 CGCA(N_5)GTG motif of *E. meliloti* Rm2011 is presumably methylated by Smc02296 (HsdM), a
445 predicted m6A MTase belonging to the HsdRSM type I R-M system that is known to be functional
446 and reduce transformation efficiency (Brumwell et al., 2019; Ferri et al., 2010). The RCGCCTC
447 motif of *E. meliloti* Rm2011 is possibly methylated by Smc03763, a predicted cysteine-specific
448 MTase located upstream of the gene *vsr* that putatively encodes a very short patch repair protein.
449 Neither of these proteins are found in the other three strains examined here. The motifs

450 TCGA(N_8)TCGA of *E. meliloti* FSM-MA and CAGA(N_7)GTTG of *E. fredii* NGR234 are likely
451 methylated by SMB554_16155 and NGR_c01340, respectively, which are 88% identical at the
452 amino acid level. Homologs of these two proteins are not found in the other two strains.

453 Except for GANTC, each methylated motif was detected as methylated only in a single
454 strain. Moreover, MTases, apart from CcrM, are not conserved among *E. meliloti* strains. The lack
455 of conservation suggests that most DNA methylation does not have a major regulatory role in the
456 genus *Ensifer*, aside from its role in cell cycle regulation. Supporting this conclusion, no motif was
457 enriched in the promoter regions of symbiosis, carbon source, or cell cycle-regulated genes, and
458 we did not detect any motifs that were methylated specifically in bacteroids. However, we cannot
459 rule out that one or more methylated motifs may influence specific gene expression during free-
460 living growth, differentiation, or N_2-fixation through extended motifs or proximity to other
461 promoter elements, similar to the interplay between CcrM and GcrA during cell cycle regulation
462 in *C. crescentus* (Fioravanti et al., 2013; Haakonsen et al., 2015).

463 As previously published studies have provided evidence for a cell cycle transition
464 occurring during terminal bacteroid differentiation (Mergaert et al., 2006), we were particularly
465 interested in CcrM, a cell cycle-regulated MTase that is broadly conserved in the α-Proteobacteria,
466 and its cognate DNA motif, GANTC. By identifying GANTC sites in the promoter regions of a
467 previously determined set of 462 cell cycle-regulated genes (De Nisco et al., 2014), we defined a
468 candidate CcrM regulon in *E. meliloti* consisting of 111 genes. However, further studies are
469 required to better delineate the CcrM regulon in *E. meliloti* as the presence of a GANTC site is not
470 diagnostic of CcrM regulation; GANTC sites were found in the promoter regions of 904 transcripts
471 that did not display cell cycle regulation, and the promoter regions of cell cycle regulated genes
472 were not enriched in GANTC sites relative to the whole *E. meliloti* genome. Studies in *C.*

473 *crescentus* suggest that the impact of the fully or hemi-methylated status of GANTC sites on gene
474 expression is mediated, at least in part, through modulating the activity of the transcriptional
475 regulator GcrA (Fioravanti et al., 2013; Haakonsen et al., 2015). However, not all promoter sites
476 containing a GANTC motif are regulated by GcrA in *C. crescentus*, with the relationship
477 dependent on an extended YGAKTCK motif and the precise position of this motif relative to other
478 promoter elements (Haakonsen et al., 2015). Likely, CcrM-mediated gene regulation in the genus
479 *Ensifer* is similarly dependent upon additional sequence elements beyond the GANTC motif.

480 Consistent with past observations (Gonzalez et al., 2014), GANTC sites were under-
481 represented in the genomes of 157 *Ensifer* strains, particularly within coding regions. More
482 surprising, however, was the strong difference in the frequency of GANTC sites between the
483 previously defined (Fagorzi et al., 2020) symbiotic and non-symbiotic clades in the genus *Ensifer*,
484 with the frequency of GANTC sites being ~ 60% higher in the symbiotic clade. As methylation of
485 the GANTC motif by CcrM is known to influence gene expression in *C. crescentus* (Gonzalez et
486 al., 2014), our observations suggest that CcrM has a greater impact on modulating gene expression
487 in the symbiotic clade compared to the non-symbiotic clade. Although further work is required to
488 understand the biological significance of the greater frequency of GANTC sites in the symbiotic
489 clade, it is tempting to speculate it is associated with legume symbiosis.

490 Our data is consistent with CcrM activity differing during terminal bacteroid differentiation
491 compared to free-living cells. The overall moderate to high rates of GANTC methylation in all *E.*
492 *meliloti* bacteroid samples, coupled with the lack of a chromosome-wide pattern in the *E. meliloti*
493 FSM-MA zone II sample, leads us to hypothesize that CcrM remains constitutively active
494 throughout most of terminal differentiation. This hypothesis is supported by the results for nodule
495 bacteria purified from *M. truncatula dnf* mutant nodules, which showed that differentiation is

496 preceded by full GANTC methylation and that GANTC methylation remains high (but moderately
497 lower) during endoreduplication followed by another moderate drop in GANTC methylation in
498 late stages of differentiation. Considering that over-expression of CcrM can give rise to bacteroid-
499 like morphology in free-living cells (Wright et al., 1997), we hypothesize that constitutive CcrM
500 MTase activity is one (of potentially multiple) factor(s) driving polyploidization of bacteroids
501 (**Figure S8**). However, further studies monitoring CcrM abundance and artificially manipulating
502 *ccrM* expression throughout bacteroid differentiation are required to conclusively determine if
503 CcrM is constitutively active during terminal differentiation and the importance of this activity to
504 the promotion of endoreduplication.

505 CcrM activity is confined to a short window in the cell cycle since the *ccrM* gene is
506 expressed in the late phase of genome replication (De Nisco et al., 2014) and the CcrM protein is
507 degraded by the Lon protease prior to cell division (Wright et al., 1996). Thus, constitutive CcrM
508 MTase activity in differentiating bacteroids could be obtained through an aberrant expression of
509 the gene or alternatively a lack of proteolytic degradation of the CcrM protein. In agreement with
510 the latter possibility, Lon protease was identified as a target of the NCR247 peptide (Farkas et al.,
511 2014). It is tempting to speculate that NCR peptides like NCR247 inhibit Lon protease activity
512 post-translationally, thereby stabilizing CcrM and triggering bacteroid differentiation. However,
513 the CcrM MTase does not appear to remain active in fully differentiated bacteroids, with the lower
514 GANTC methylation near the chromosomal *ter* regions suggesting that loss of CcrM MTase
515 activity occurs slightly prior to completion of genome endoreduplication (model provided as
516 **Figure S8**). These hypotheses are consistent with *M. truncatula – E. meliloti* nodule zone-specific
517 RNA-sequencing data (Roux et al., 2014), which showed that *ccrM* expression in the root distal
518 portion of zone II is ~ 2-fold higher than in the root proximal portion of zone II, and ~ 10-fold

519 higher than in zone III. The ~ 10-fold difference in *ccrM* expression across nodule zones suggests
520 to us that the level of *ccrM* expression during early stages of bacteroid differentiation is
521 biologically significant, a prerequisite for the constitutive CcrM activity that we hypothesize.

522 Our analyses also provide insight into the genome replication dynamics of *E. meliloti*
523 during free-living growth and terminal bacteroid differentiation. Notably, flow cytometry data of
524 *E. meliloti* bacteroids purified from zone II nodule sections and *M. truncatula dnf2* nodules suggest
525 that endoreduplication and cell enlargement largely occur subsequently, not concurrently. Genome
526 replication might be a much faster process than cell growth or alternatively, endoreduplication
527 might be required to drive cell enlargement. Moreover, our data are consistent with a loss of
528 coordination of replication of the three replicons during terminal bacteroid differentiation, leading
529 to unequal copy numbers with two copies of pSymA and pSymB per three copies of the
530 chromosome in bacteroids. This relative change in replicon copy number occurs concomitantly
531 with differentiation and polyploidization, as supported by the relative abundance of the replicons
532 differing in nodule bacterial cells that have experienced endoreduplication (i.e., cells retrieved
533 from *M. truncatula dnf2*, *dnf7*, and *dnf4* mutant nodules) but not in bacterial cells that had not yet
534 undergone endoreduplication (i.e., cells retrieved from *M. truncatula dnf1* and *dnf5* mutant
535 nodules). This differs from free-living cells, where copy number of the three replicons was
536 approximately equal based on average sequencing depth. In contrast to our results, a previous
537 comparison of the relative abundance of the three replicons in free-living *E. meliloti* Rm1021
538 versus bacteroids, no changes were detected using comparative genome hybridization with
539 microarrays (Mergaert et al., 2006). We believe that the difference between our present data and
540 the previous analysis is due to the subtlety of the differences and the lower sensitivity of the
541 microarray hybridization method compared to high throughput sequencing.

542 We also observed that during free-living exponential growth, the extent of GANTC
543 methylation at the *ori* of pSymA and pSymB is higher than at the *ori* of the chromosome, while
544 the *ter* of the pSymA and pSymB has a slightly lower extent of GANTC methylation than the *ter*
545 of the chromosome. As GANTC methylation occurs at a fixed stage of the cell cycle corresponding
546 to the end of chromosome replication, our observations indicate that pSymA and pSymB
547 replication is initiated later in the cell cycle than initiation of chromosome replication, while their
548 replication terminates slightly before completion of chromosomal replication and the activation of
549 CcrM (model provided as **Figure S8**). These results provide additional support for the notion of
550 spatiotemporal regulation of DNA replication and partitioning in the multipartite *E. meliloti*
551 genome as proposed previously (De Nisco et al., 2014; Frage et al., 2016). Similarly, replication
552 of chromosome II of *Vibrio cholerae* is delayed relative to chromosome I, leading to the replication
553 of these two replicons terminating at approximately the same time (Rasmussen et al., 2007). Thus,
554 co-ordinating the timing or replication termination may be a general feature of multipartite
555 genomes.
556

557 **MATERIALS AND METHODS**

558 **Experimental design**

559 The overall experimental design is summarized in **Figure S1**. Genomic DNA was isolated from
560 four wild-type *Ensifer* strains to explore how DNA methylation varies across this genus; to allow
561 direct comparison, the four strains were grown to mid-exponential phase in minimal medium with
562 succinate as a carbon source. To investigate how DNA methylation patterns differ between
563 actively dividing and non-dividing cells, genomic DNA was isolated from *E. meliloti* Rm2011
564 growth to either mid-exponential phase or stationary phase. Genomic DNA was isolated from *E.*

565 *meliloti* Rm2011 grown to mid-exponential phase with either a glycolytic (sucrose) or
566 gluconeogenic (succinate) carbon source to investigate whether DNA methylation may play a role
567 in regulating carbon metabolism. Furthermore, a *E. meliloti* Rm2011 derivative lacking the pSymA
568 and pSymB replicons (named RmP3496) was studied to gain insight into whether these replicons
569 contribute to DNA methylation patterns in *E. meliloti* Rm2011; these strains were grown with
570 sucrose (instead of succinate) as RmP3496 lacks the succinate transporter.

571 In addition to the free-living samples, *E. meliloti* bacteroid samples purified from legume
572 nodules were collected to investigate changes in DNA methylation during bacteroid differentiation
573 and nitrogen fixation. To do so, *E. meliloti* Rm2011 and *E. meliloti* FSM-MA bacteroids were
574 isolated from *M. sativa* whole nodules. *E. meliloti* FSM-MA bacteroids were additionally purified
575 from *M. truncatula* whole nodules to examine the impact of host plant on bacteroid DNA
576 methylation patterns. *E. meliloti* Rm2011 bacteroids were only isolated from *M. sativa* nodules as,
577 unlike FSM-MA, Rm2011 forms a poor symbiosis with *M. truncatula* (Kazmierczak et al., 2017;
578 Moreau et al., 2008). Moreover, *E. meliloti* Rm2011 and *E. meliloti* FSM-MA bacteroids were
579 isolated from *M. sativa* nodule sections (sectioned at the white – pink border to separate the root
580 distal infection and differentiation zone II [white] from the root proximal nitrogen-fixing zone III
581 [pink]) to facilitate an analysis of how DNA methylation patters differ between differentiating
582 bacteroids and fully differentiated and nitrogen-fixing bacteroids. This was followed by isolation
583 of *E. meliloti* FSM-MA bacteroids from whole nodules of *M. truncatula* mutant lines (*dnf1*, *dnf2*,
584 *dnf4*, *dnf5*, *dnf7*) to investigate DNA methylation patterns in nodule bacteria blocked as various
585 stages of differentiation.

586

587 **Bacterial strains and growth conditions**

588 Bacterial strains used in this work are listed in **Table S6**. All strains were routinely grown on TY

589 with 2 μM $CoCl_2$ as it was required for *E. meliloti* RmP3496 (diCenzo et al., 2014). The MM9
590 minimal medium (diCenzo et al., 2014) consisted of the following: 40 mM MOPS, 20 mM KOH,
591 19.2 mM NH_4Cl, 85.6 mM NaCl, 2 mM KH_2PO_4, 1 mM $MgSO_4$, 0.25 mM $CaCl_2$, 1 μg ml^{-1} biotin,
592 42 nM $CoCl_2$, 38 μM $FeCl_3$, 10 μM thiamine-HCl, and either 10 mM sucrose (MM9-sucrose) or
593 20 mM disodium succinate (MM9-succinate). Prior to inoculation of plants with *E. meliloti*, the
594 strains were grown in YEB medium (Krall et al., 2002).
595
596 **DNA isolation from free-living cells**
597 Overnight cultures of all strains grown in MM9-succinate or MM9-sucrose media were diluted
598 into 10 mL of the same medium to a starting OD_{600nm} of 0.025 (0.05 for RmP3496) and incubated
599 overnight at 30˚C with shaking (130 rpm). The next day, cultures were diluted into 40 mL of the
600 same medium in 100 mL flasks to the OD_{600nm} values as listed in **Table S7**. To obtain mid-
601 exponential phase samples, cultures were harvested after 15.5-16 hours of growth at OD_{600nm}
602 values between 0.37 and 0.69 (**Table S7**). To obtain stationary phase samples, cultures were
603 harvested after 24 hours of growth at OD_{600nm} values of ~ 1.4. In all cases, cultures were streaked
604 on TY plates to check for contamination and then centrifuged (8,200 *g*, 10 minutes, 4°C); the full
605 40 mL was centrifuged for mid-exponential phase cultures, whereas only 15 mL was centrifuged
606 for stationary phase cultures. Most of the supernatant was discarded, and the pellet resuspended in
607 the remaining volume, transferred to a 2 mL tube, centrifuged again (16,200 *g*, room temperature,
608 one minute), and the supernatant discarded. Three biological replicates, each starting from a
609 separate overnight culture, were performed. DNA was isolated using phenol:chloroform
610 extractions and ammonium acetate precipitations as described elsewhere (Cowie et al., 2006), and
611 the DNA pellets (following RNase A treatment) were resuspended in 200 μL of 10 mM Tris-HCl,
612 pH 8.5.

613 **DNA isolation from bacteroids**

614 *M. sativa* cv. Gabès and *M. truncatula* cv. A17 were used wild-type plants for all experiments. *M.*
615 *truncatula dnf1*, *dnf2*, *dnf4*, *dnf5*, and *dnf7* mutants (Starker et al., 2006), derived from the A17
616 wild type, were used for collection of bacteroids blocked at various stages of differentiation. Seeds
617 were scarified, surface sterilized, and germinated on Kalys agar as described previously
618 (Kazmierczak et al., 2017). Fifty mL of overnight cultures of *E. meliloti* Rm2011 or FSM-MA,
619 grown in YEB, were centrifuged (4,000 *g*, 20 minutes, room temperature) and resuspended in ~
620 1,200 mL of sterile, distilled water to obtain a cell suspension at an OD_{600nm} of ~ 0.1. Germinated
621 seedlings were immersed for one hour in the appropriate rhizobial cell suspension, and then planted
622 in a perlite:sand (2:1) mixture. Plants were grown in a greenhouse for five to six weeks, with
623 occasional watering with a 1 g L^{-1} nutrient solution (PlantProd solution [N-P-K, 0-15-40; Fertil]).

624 For whole nodule samples of wild type plants, pink nodules were collected from 53-60
625 plants per replicate 34-35 days post-inoculation; in the case of *dnf* mutants (and a matched wild-
626 type A17 sample), nodules were collected from ~ 105 plants per genotype 23-24 days post-
627 inoculation. Nodules were collected from the roots and stored in tubes in liquid nitrogen until
628 collection was complete, at which point they were stored at -80°C until use. For sectioned nodule
629 samples, pink nodules were collected from 103 *M. sativa* plants for each of the microsymbionts
630 35 to 40 days post-inoculation. Nodules were manually sectioned at the white-pink border. Nodule
631 sections were stored in tubes over dry ice or liquid nitrogen until collection was complete, at which
632 point they were stored at -80°C until use. Average plant shoot dry weights for all samples are listed
633 in **Table S8**. Bacteroids were isolated from the nodule samples using Percoll gradient
634 centrifugation as described elsewhere (Mergaert et al., 2006). The recovered bacteroids were
635 resuspended in 50-100 µL of Bacteroid Extraction Buffer (BEB; 125 mM KCl, 50 mM Na-

636 succinate, 50 mM TES, pH 7.0), and either used immediately for microscopy, flow cytometry, and
637 DNA isolation, or stored at -80°C until use.

638 Nucleic acids were initially purified from most bacteroid samples using Epicentre
639 MasterPure™ Complete DNA and RNA Purification Kit, following the protocol for DNA isolation
640 from cell samples; the exceptions were bacteroids purified from *dnf* mutant nodules (and the
641 matched wild-type A17 sample), for which nucleic acids were isolated by using phenol:chloroform
642 extractions followed by ammonium acetate DNA precipitations as described elsewhere (Cowie et
643 al., 2006). For sectioned nodule samples, pure DNA was isolated by using the manufacturer's
644 protocol for the complete removal of RNA. For whole nodule samples, the isolated DNA was
645 further purified by treating the nucleic acids samples with RNase A, after which pure DNA was
646 isolated by using phenol:chloroform extractions followed by ammonium acetate DNA
647 precipitations or alternatively using the MasterPure™ DNA clean-up protocol for the DNA from
648 *dnf* mutant nodules and the matched wild-type A17 sample. In all cases, the final DNA pellets
649 were resuspended in 200 µL of 10 mM Tris-HCl, pH 8.5. Three biological replicates were
650 performed for bacteroids isolated from most whole nodules, whereas only one replicate was
651 performed for bacteroids isolated from sectioned nodules or *dnf* mutant nodules (and the matched
652 wild-type A17 sample) due to low quantities of starting materials.

653

654 **DNA sequencing, modification detection, and motif analysis**

655 DNA sequencing was performed at the U.S. Department of Energy Joint Genome Institute (JGI)
656 or in-house at the University of Florence (the stationary phase samples and *dnf* mutant nodules and
657 the matched wild-type A17 sample) using Pacific Biosciences (PacBio) sequencing technology
658 (Eid et al., 2009). Genomic DNA was sheared to 3 kb using a Covaris LS220 (Covaris Inc.,

659 Woburn, MA, USA) or 15 kb (for stationary phase samples and bacteroids isolated from *dnf* mutant
660 nodules and the matched wild-type A17 sample) using g-TUBEs (Covaris Inc., Woburn, MA,
661 USA). Sheared DNA was treated with exonuclease to remove single-stranded ends and DNA
662 damage repair mix followed by end repair and ligation of barcoded blunt adapters using SMRTbell
663 Template Prep Kit 2.0 (PacBio, Menlo Park, CA, USA). Libraries were purified with AMPure PB
664 beads (Beckman Coulter, Brea, CA, USA) and three or eight libraries with different barcodes were
665 pooled at equimolar ratios and purified with AMPure PB beads. For most samples, SMRTbell
666 template libraries were prepared using a Sequel Binding Kit 3.0 (PacBio, Menlo Park, CA, USA),
667 and sequenced on a Sequel instrument using a v3 or v4 sequencing primer, 1M v3 SMRT cells,
668 and Version 3.0 sequencing chemistry with 1x360 or 1x600 sequencing movie run times. The
669 exceptions were the *E. meliloti* Rm2011 zone II and *E. meliloti* FSM-MA zone III bacteroid
670 samples. For these samples, SMRTbell template libraries were prepared using a Sequel II Binding
671 Kit 2.0 (PacBio, Menlo Park, CA, USA), and then sequenced on a Sequel II instrument using the
672 tbd-sample dependent sequencing primer, 8M v1 SMRT cells, and Version 2.0 sequencing
673 chemistry with 1x900 sequencing movie run times.

674 DNA modification detection and motif analysis were performed using the PacBio SMRT
675 Link software (PacBio, Menlo Park, CA, USA). Briefly, raw reads were filtered using SFilter to
676 remove short reads and reads derived from sequencing adapters. Filtered reads were aligned
677 against the appropriate reference genome (**Table S2**) using BLASR (Chaisson and Tesler, 2012)
678 and modified sites were then identified through kinetic analysis of the aligned DNA sequence data
679 (Flusberg et al., 2010); the number of mapped bases per sample is provided in **Table S2**. Modified
680 sites were then grouped into motifs using MotifFinder. These motifs represent the recognition
681 sequences of MTase genes active in the genome (Clark et al., 2012). Downstream analyses were

682 performed using custom Perl and R scripts.

683

684 **Flow cytometry**

685 Flow cytometry was performed as described previously (Mergaert et al., 2006). Freshly prepared
686 bacteroid samples were diluted in 200 µL of BEB, heat-treated for 10 minutes in a 70°C water
687 bath, and then stained with the DNA-binding dye diamidino-2-phenylindole (DAPI). Cell size and
688 ploidy level of the bacteroid samples were determined using flow cytometry with a Beckman
689 Coulter CytoFLEX S instrument. Measurements consisted of 50,000 cells. Data analysis was
690 performed using the CytExpert 2.2.0.97 software.

691

692 **Fluorescence microscopy**

693 One µL of each freshly prepared bacteroid sample was mixed with 1 µL of 50 µg mL^{-1} DAPI or
694 with both 1 µL of 50 µg mL^{-1} DAPI and 1 µL of 100 µg mL^{-1} propidium iodide (PI), which are
695 both DNA binding dyes. Samples were visualized at 100x magnification under oil immersion using
696 a Nikon Eclipse 80*i* fluorescence microscope with the NIS-Elements BR 4.00.01 software and a
697 Digital Sight DS-U3 camera.

698

699 **Phylogenetic analysis**

700 The nucleotide fasta files of representative *Ensifer* species were downloaded from the National
701 Centre for Biotechnology Information (NCBI) Genome database. A core gene phylogeny was
702 constructed using a previously prepared pipeline (diCenzo et al., 2018) reliant on the use of Roary
703 3.11.3 (Page et al., 2015), Prokka 1.12-beta (Seemann, 2014), PRANK 140110 (Löytynoja, 2014),
704 trimAl (Capella-Gutiérrez et al., 2009), and RAxML 8.2.9 (Stamatakis, 2014). The phylogeny was

705 visualized with the iTol webserver (Letunic and Bork, 2016). Identification of nodulation
706 (*nodABC*) and nitrogenase genes (*nifHDK*) was performed with a published pipeline (diCenzo et
707 al., 2018) reliant on the use of HMMER 3.1b2 (Eddy, 2009), and the Pfam-A 31.0 (Finn et al.,
708 2016) and TIGERFAM 15.0 (Haft et al., 2013) databases.
709

710 **DATA AVAILABILITY**

711 Most sequencing data is available through the JGI Genome Portal (genome.jgi.doe.gov/portal/)
712 under Proposal 503835, as well as through NCBI (see **Table S2** for BioSample accessions). The
713 data for stationary phase cultures and bacteroids isolated from *dnf* mutant nodules are available
714 only through the NCBI (BioProject accessions PRJNA706182 and PRJNA705832; see **Table S2**
715 for BioSample accessions). All custom scripts to perform the analyses described in this study are
716 available through GitHub (github.com/diCenzo-Lab/003_2021_Ensifer_DNA_methylation), as
717 are the flow cytometry FCS files.
718

719 **ACKNOWLEDGMENTS**

720 We are grateful to E. Mullins (Teagasc, MTA2018233) for *E. adhaerens* OV14 and M. Bourge
721 (I2BC) for help with the flow cytometry experiments.
722

723 **CONFLICT OF INTERESTS**

724 The authors declare that they have no conflict of interest.
725

726 **FUNDING**

727 Most sequencing work was performed at the U.S. Department of Energy (DOE) Joint Genome

728 Institute (JGI), through the CSP New Investigator program (proposal: 503835). The work
729 conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science
730 User Facility, is supported under Contract No. DE-AC02-05CH11231. This work was partially
731 supported by the "MICRO4Legumes" grant (Italian Ministry of Agriculture), by the grant
732 "Dipartimento di Eccellenza 2018–2022" from the Italian Ministry of Education, University and
733 Research (MIUR), and by a Discovery Grant from Natural Sciences and Engineering Research
734 Council of Canada. The present work benefited from the Imagerie-Gif core facility supported by
735 the Agence Nationale de la Recherche (ANR-11-EQPX-0029/Morphoscope, ANR-10-INBS-
736 04/FranceBioImaging, ANR-11-IDEX-0003-02/Saclay Plant Sciences). G.C.D was supported by
737 a NSERC postdoctoral fellowship and a European Molecular Biology Organization (EMBO) short
738 term fellowship for part of this work. L.C. was supported by the MICRO4Legumes grant (Italian
739 Ministry of Agriculture). Q.N. was supported by a PhD fellowship from the Paris-Saclay
740 University. J.H.T.C. was supported by a NSERC Undergraduate Summer Research Award. P.M.
741 and B.A. were supported by Saclay Plant Sciences (SPS) and grant ANR-17-CE20-0011 from the
742 Agence Nationale de la Recherche.

743

744 **REFERENCES**

745 Adhikari S, Curtis PD. 2016. DNA methyltransferases and epigenetic regulation in bacteria. *FEMS*
746 *Microbiol Rev* **40**:575–591. doi:10.1093/femsre/fuw023

747 Atack JM, Tan A, Bakaletz LO, Jennings MP, Seib KL. 2018. Phasevarions of bacterial pathogens:
748 methylomics sheds new light on old enemies. *Trends Microbiol* **26**:715–726.
749 doi:10.1016/j.tim.2018.01.008

750 Barnett MJ, Toman CJ, Fisher RF, Long SR. 2004. A dual-genome Symbiosis Chip for coordinate

751 study of signal exchange and development in a prokaryote–host interaction. *Proc Natl Acad*
752 *Sci USA* **101**:16636–16641.
753 Barrière Q, Guefrachi I, Gully D, Lamouche F, Pierre O, Fardoux J, Chaintreuil C, Alunni B,
754 Timchenko T, Giraud E, Mergaert P. 2017. Integrated roles of BclA and DD-
755 carboxypeptidase 1 in *Bradyrhizobium* differentiation within NCR-producing and NCR-
756 lacking root nodules. *Sci Rep* **7**:9063. doi:10.1038/s41598-017-08830-0
757 Blow MJ, Clark TA, Daum CG, Deutschbauer AM, Fomenkov A, Fries R, Froula J, Kang DD,
758 Malmstrom RR, Morgan RD, Pósfai J, Singh K, Visel A, Wetmore K, Zhao Z, Rubin EM,
759 Korlach J, Pennacchio LA, Roberts RJ. 2016. The epigenomic landscape of prokaryotes.
760 *PLOS Genet* **12**:e1005854. doi:10.1371/journal.pgen.1005854
761 Bourcy M, Brocard L, Pislariu CI, Cosson V, Mergaert P, Tadege M, Mysore KS, Udvardi MK,
762 Gourion B, Ratet P. 2013. *Medicago truncatula* DNF2 is a PI-PLC-XD-containing protein
763 required for bacteroid persistence and prevention of nodule early senescence and defense-
764 like reactions. *New Phytol* **197**:1250–1261. doi:10.1111/nph.12091
765 Brilli M, Fondi M, Fani R, Mengoni A, Ferri L, Bazzicalupo M, Biondi EG. 2010. The diversity
766 and evolution of cell cycle regulation in alpha-proteobacteria: a comparative genomic
767 analysis. *BMC Syst Biol* **4**:52. doi:10.1186/1752-0509-4-52
768 Brumwell SL, MacLeod MR, Huang T, Cochrane RR, Meaney RS, Zamani M, Matysiakiewicz
769 O, Dan KN, Janakirama P, Edgell DR, Charles TC, Finan TM, Karas BJ. 2019. Designer
770 *Sinorhizobium meliloti* strains and multi-functional vectors enable direct inter-kingdom
771 DNA transfer. *PLOS One* **14**:e0206781. doi:10.1371/journal.pone.0206781
772 Campbell JL, Kleckner N. 1990. *E. coli oriC* and the *dnaA* gene promoter are sequestered from
773 *dam* methyltransferase following the passage of the chromosomal replication fork. *Cell*

774 **62**:967–979. doi:10.1016/0092-8674(90)90271-F

775 Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment
776 trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**:1972–1973.
777 doi:10.1093/bioinformatics/btp348

778 Chaisson MJ, Tesler G. 2012. Mapping single molecule sequencing reads using basic local
779 alignment with successive refinement (BLASR): application and theory. *BMC Bioinform*
780 **13**:238. doi:10.1186/1471-2105-13-238

781 Clark TA, Murray IA, Morgan RD, Kislyuk AO, Spittle KE, Boitano M, Fomenkov A, Roberts
782 RJ, Korlach J. 2012. Characterization of DNA methyltransferase specificities using single-
783 molecule, real-time DNA sequencing. *Nucleic Acids Res* **40**:e29–e29.
784 doi:10.1093/nar/gkr1146

785 Cowie A, Cheng J, Sibley CD, Fong Y, Zaheer R, Patten CL, Morton RM, Golding GB, Finan
786 TM. 2006. An integrated approach to functional genomics: construction of a novel reporter
787 gene fusion library for *Sinorhizobium meliloti*. *Appl Environ Microbiol* **72**:7156–7167.
788 doi:10.1128/AEM.01397-06

789 Czernic P, Gully D, Cartieaux F, Moulin L, Guefrachi I, Patrel D, Pierre O, Fardoux J, Chaintreuil
790 C, Nguyen P, Gressent F, Da Silva C, Poulain J, Wincker P, Rofidal V, Hem S, Barrière
791 Q, Arrighi J-F, Mergaert P, Giraud E. 2015. Convergent evolution of endosymbiont
792 differentiation in Dalbergioid and inverted repeat-lacking clade legumes mediated by
793 nodule-specific cysteine-rich peptides. *Plant Physiol* **169**:1254–1265.
794 doi:10.1104/pp.15.00584

795 Davis-Richardson AG, Russell JT, Dias R, McKinlay AJ, Canepa R, Fagen JR, Rusoff KT, Drew
796 JC, Kolaczkowski B, Emerich DW, Triplett EW. 2016. Integrating DNA methylation and

797 gene expression data in the development of the soybean-*Bradyrhizobium* N_2-fixing
798 symbiosis. *Front Microbiol* **7**:518. doi:10.3389/fmicb.2016.00518
799 De Nisco NJ, Abo RP, Wu CM, Penterman J, Walker GC. 2014. Global analysis of cell cycle gene
800 expression of the legume symbiont *Sinorhizobium meliloti*. *Proc Natl Acad Sci USA*
801 **111**:3217–3224. doi:10.1073/pnas.1400421111
802 diCenzo GC, Debiec K, Krzysztoforski J, Uhrynowski W, Mengoni A, Fagorzi C, Gorecki A,
803 Dziewit L, Bajda T, Rzepa G, Drewniak L. 2018. Genomic and biotechnological
804 characterization of the heavy-metal resistant, arsenic-oxidizing bacterium *Ensifer* sp. M14.
805 *Genes* **9**:379. doi:10.3390/genes9080379
806 diCenzo GC, MacLean AM, Milunovic B, Golding GB, Finan TM. 2014. Examination of
807 prokaryotic multipartite genome evolution through experimental genome reduction. *PLOS*
808 *Genet* **10**:e1004742. doi:10.1371/journal.pgen.1004742
809 diCenzo GC, Muhammed Z, Østerås M, O'Brien SAP, Finan TM. 2017. A key regulator of the
810 glycolytic and gluconeogenic central metabolic pathways in *Sinorhizobium meliloti*.
811 *Genetics* **207**:961–974. doi:10.1534/genetics.117.300212
812 diCenzo GC, Tesi M, Pfau T, Mengoni A, Fondi M. 2020. Genome-scale metabolic reconstruction
813 of the symbiosis between a leguminous plant and a nitrogen-fixing bacterium. *Nat Commun*
814 **11**:2574. doi:10.1101/765271
815 diCenzo GC, Zamani M, Checcucci A, Fondi M, Griffitts J, Finan TM, Mengoni A. 2019. Multi-
816 disciplinary approaches for studying rhizobium - legume symbioses. *Can J Microbiol*
817 **65**:1–33. doi:10.1139/cjm-2018-0377
818 Domonkos A, Horvath B, Marsh JF, Halasz G, Ayaydin F, Oldroyd GED, Kalo P. 2013. The
819 identification of novel loci required for appropriate nodule development in *Medicago*

820 *truncatula*. *BMC Plant Biol* **13**:157. doi:10.1186/1471-2229-13-157

821 Eddy SR. 2009. A new generation of homology search tools based on probabilistic inference.
822 *Genome Inform* **23**:205–211. doi:10.1142/9781848165632_0019

823 Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo
824 A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, deWinter A, Dixon
825 J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X,
826 Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I,
827 Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson
828 M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner
829 S. 2009. Real-time DNA sequencing from single polymerase molecules. *Science* **323**:133–
830 138. doi:10.1126/science.1162986

831 Fagorzi C, Ilie A, Decorosi F, Cangioli L, Viti C, Mengoni A, diCenzo GC. 2020. Symbiotic and
832 nonsymbiotic members of the genus *Ensifer* (syn. *Sinorhizobium*) are separated into two
833 clades based on comparative genomics and high-throughput phenotyping. *Genome Biol*
834 *Evol* **12**:2521–2534. doi:10.1093/gbe/evaa221

835 Farkas A, Maróti G, Dürgő H, Györgypál Z, Lima RM, Medzihradszky KF, Kereszt A, Mergaert
836 P, Kondorosi E. 2014. *Medicago truncatula* symbiotic peptide NCR247 contributes to
837 bacteroid differentiation through multiple mechanisms. *Proc Natl Acad Sci USA*
838 **111**:5183–5188. doi:10.1073/pnas.1404169111

839 Ferri L, Gori A, Biondi EG, Mengoni A, Bazzicalupo M. 2010. Plasmid electroporation of
840 *Sinorhizobium* strains: The role of the restriction gene *hsdR* in type strain Rm1021. *Plasmid*
841 **63**:128–135. doi:10.1016/j.plasmid.2010.01.001

842 Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi

843 M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. 2016. The Pfam protein families
844 database: towards a more sustainable future. *Nucleic Acids Res* **44**:D279–D285.
845 doi:10.1093/nar/gkv1344
846 Fioravanti A, Fumeaux C, Mohapatra SS, Bompard C, Brilli M, Frandi A, Castric V, Villeret V,
847 Viollier PH, Biondi EG. 2013. DNA binding of the cell cycle transcriptional regulator
848 GcrA depends on N6-adenosine methylation in *Caulobacter crescentus* and other
849 Alphaproteobacteria. *PLOS Genet* **9**:e1003541. doi:10.1371/journal.pgen.1003541
850 Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, Korlach J, Turner SW.
851 2010. Direct detection of DNA methylation during single-molecule, real-time sequencing.
852 *Nat Methods* **7**:461–465. doi:10.1038/nmeth.1459
853 Frage B, Döhlemann J, Robledo M, Lucena D, Sobetzko P, Graumann PL, Becker A. 2016.
854 Spatiotemporal choreography of chromosome and megaplasmids in the *Sinorhizobium*
855 *meliloti* cell cycle. *Mol Microbiol* **100**:808–823. doi:10.1111/mmi.13351
856 Gibson KE, Campbell GR, Lloret J, Walker GC. 2006. CbrA is a stationary-phase regulator of cell
857 surface physiology and legume symbiosis in *Sinorhizobium meliloti*. *J Bacteriol* **188**:4508–
858 4521. doi:10.1128/JB.01923-05
859 Gonzalez D, Collier J. 2013. DNA methylation by CcrM activates the transcription of two genes
860 required for the division of *Caulobacter crescentus*: regulation of cell division by DNA
861 methylation. *Mol Microbiol* **88**:203–218. doi:10.1111/mmi.12180
862 Gonzalez D, Kozdon JB, McAdams HH, Shapiro L, Collier J. 2014. The functions of DNA
863 methylation by CcrM in *Caulobacter crescentus*: a global approach. *Nucleic Acids Res*
864 **42**:3720–3735. doi:10.1093/nar/gkt1352
865 Gopalakrishnan S, Van Emburgh BO, Robertson KD. 2008. DNA methylation in development and

human disease. *Mutat Res* **647**:30–38. doi:10.1016/j.mrfmmm.2008.08.006

Greenberg MVC, Bourc'his D. 2019. The diverse roles of DNA methylation in mammalian development and disease. *Nat Rev Mol Cell Biol* **20**:590–607. doi:10.1038/s41580-019-0159-6

Haakonsen DL, Yuan AH, Laub MT. 2015. The bacterial cell cycle regulator GcrA is a σ^{70} cofactor that drives gene expression from a subset of methylated promoters. *Genes Dev* **29**:2272–2286. doi:10.1101/gad.270660.115

Haft DH, Selengut JD, Richter RA, Harkins D, Basu MK, Beck E. 2013. TIGRFAMs and genome properties in 2013. *Nucleic Acids Res* **41**:D387–D395. doi:10.1093/nar/gks1234

Horváth B, Domonkos Á, Kereszt A, Sz\Hucs A, Ábrahám E, Ayaydin F, Bóka K, Chen Y, Chen R, Murray JD, Udvardi MK, Kondorosi E, Kaló P. 2015. Loss of the nodule-specific cysteine rich peptide, NCR169, abolishes symbiotic nitrogen fixation in the *Medicago truncatula dnf7* mutant. *Proceedings of the National Academy of Sciences* **112**:15232–15237. doi:10.1073/pnas.1500777112

Kahng LS, Shapiro L. 2001. The CcrM DNA methyltransferase of *Agrobacterium tumefaciens* Is essential, and Its activity is cell cycle regulated. *J Bacteriol* **183**:3065–3075. doi:10.1128/JB.183.10.3065-3075.2001

Kang S, Lee H, Han JS, Hwang DS. 1999. Interaction of SeqA and Dam methylase on the hemimethylated origin of *Escherichia coli* chromosomal DNA replication. *J Biol Chem* **274**:11463–11468. doi:10.1074/jbc.274.17.11463

Kazmierczak T, Nagymihaly M, Lamouche F, Barrière Q, Guefrachi I, Alunni B, Ouadghiri M, Ibijbijen J, Kondorosi E, Mergaert P, Gruber V. 2017. Specific host-responsive associations between *Medicago truncatula* accessions and *Sinorhizobium strains*. *Mol*

889 *Plant Microbe Interact* **30**:399–409. doi:10.1094/MPMI-01-17-0009-R

890 Kim M, Chen Y, Xi J, Waters C, Chen R, Wang D. 2015. An antimicrobial peptide essential for
891 bacterial survival in the nitrogen-fixing symbiosis. *Proceedings of the National Academy*
892 *of Sciences* **112**:15238–15243. doi:10.1073/pnas.1500123112

893 Kobayashi H, De Nisco NJ, Chien P, Simmons LA, Walker GC. 2009. *Sinorhizobium meliloti*
894 CpdR1 is critical for co-ordinating cell cycle progression and the symbiotic chronic
895 infection. *Mol Microbiol* **73**:586–600. doi:10.1111/j.1365-2958.2009.06794.x

896 Kozdon JB, Melfi MD, Luong K, Clark TA, Boitano M, Wang S, Zhou B, Gonzalez D, Collier J,
897 Turner SW, Korlach J, Shapiro L, McAdams HH. 2013. Global methylation state at base-
898 pair resolution of the *Caulobacter* genome throughout the cell cycle. *Proc Natl Acad Sci*
899 *USA* **110**:E4658–E4667. doi:10.1073/pnas.1319315110

900 Krall L, Wiedemann U, Unsin G, Weiss S, Domke N, Baron C. 2002. Detergent extraction
901 identifies different VirB protein subassemblies of the type IV secretion machinery in the
902 membranes of *Agrobacterium tumefaciens*. *Proc Natl Acad Sci USA* **99**:11405–11410.
903 doi:10.1073/pnas.172390699

904 Lahue R, Au K, Modrich P. 1989. DNA mismatch correction in a defined system. *Science*
905 **245**:160–164. doi:10.1126/science.2665076

906 Lang C, Long SR. 2015. Transcriptomic analysis of *Sinorhizobium meliloti* and *Medicago*
907 *truncatula* symbiosis using nitrogen nixation–deficient nodules. *Mol Plant Microbe*
908 *Interact* **28**:856–868. doi:10.1094/MPMI-12-14-0407-R

909 Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and
910 annotation of phylogenetic and other trees. *Nucleic Acids Res* **44**:W242–W245.
911 doi:10.1093/nar/gkw290

912 Löytynoja A. 2014. Phylogeny-aware alignment with PRANK. In: Russell D, editor. Methods in
913 Molecular Biology (Methods and Protocols). Totowa, NJ: Humana Press. pp. 155–170.
914 Mergaert P, Uchiumi T, Alunni B, Evanno G, Cheron A, Catrice O, Mausset A-E, Barloy-Hubler
915 F, Galibert F, Kondorosi A, Kondorosi E. 2006. Eukaryotic control on bacterial cell cycle
916 and differentiation in the *Rhizobium*-legume symbiosis. *Proc Natl Acad Sci USA*
917 **103**:5230–5235. doi:10.1073/pnas.0600912103
918 Mohapatra SS, Fioravanti A, Biondi EG. 2014. DNA methylation in *Caulobacter* and other
919 Alphaproteobacteria during cell cycle progression. *Trends Microbiol* **22**:528–535.
920 doi:10.1016/j.tim.2014.05.003
921 Moreau D, Voisin A-S, Salon C, Munier-Jolain N. 2008. The model symbiotic association between
922 *Medicago truncatula* cv. Jemalong and *Rhizobium meliloti* strain 2011 leads to N-stressed
923 plants when symbiotic N2 fixation is the main N source for plant growth. *J Exp Bot*
924 **59**:3509–3522. doi:10.1093/jxb/ern203
925 Nagymihaly M, Veluchamy A, Györgypál Z, Ariel F, Jégu T, Benhamed M, Sz\Hucs A, Kereszt
926 A, Mergaert P, Kondorosi E. 2017. Ploidy-dependent changes in the epigenome of
927 symbiotic cells correlate with specific patterns of gene expression. *Proc Natl Acad Sci USA*
928 **114**:4543–4548. doi:10.1073/pnas.1704211114
929 Nouri H, Monnier A-F, Fossum-Raunehaug S, Maciąg-Dorszyńska M, Cabin-Flaman A, Képès F,
930 Węgrzyn G, Szalewska-Pałasz A, Norris V, Skarstad K, Janniere L. 2018. Multiple links
931 connect central carbon metabolism to DNA replication initiation and elongation in *Bacillus*
932 *subtilis*. *DNA Res* **25**:641–653. doi:10.1093/dnares/dsy031
933 Oldroyd GED. 2013. Speak, friend, and enter: signalling systems that promote beneficial
934 symbiotic associations in plants. *Nat Rev Microbiol* **11**:252–263. doi:10.1038/nrmicro2990

935 Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane
936 JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis.
937 *Bioinformatics* **31**:3691–3693. doi:10.1093/bioinformatics/btv421
938 Pecrix Y, Staton SE, Sallet E, Lelandais-Brière C, Moreau S, Carrère S, Blein T, Jardinaud M-F,
939 Latrasse D, Zouine M, Zahm M, Kreplak J, Mayjonade B, Satgé C, Perez M, Cauet S,
940 Marande W, Chantry-Darmon C, Lopez-Roques C, Bouchez O, Bérard A, Debellé F,
941 Muños S, Bendahmane A, Bergès H, Niebel A, Buitink J, Frugier F, Benhamed M, Crespi
942 M, Gouzy J, Gamas P. 2018. Whole-genome landscape of *Medicago truncatula* symbiotic
943 genes. *Nat Plants* **4**:1017–1025. doi:10.1038/s41477-018-0286-7
944 Penterman J, Abo RP, De Nisco NJ, Arnold MFF, Longhi R, Zanda M, Walker GC. 2014. Host
945 plant peptides elicit a transcriptional response to control the *Sinorhizobium meliloti* cell
946 cycle during symbiosis. *Proc Natl Acad Sci USA* **111**:3561–3566.
947 doi:10.1073/pnas.1400450111
948 Pini F, De Nisco NJ, Ferri L, Penterman J, Fioravanti A, Brilli M, Mengoni A, Bazzicalupo M,
949 Viollier PH, Walker GC, Biondi EG. 2015. Cell cycle control by the master regulator CtrA
950 in *Sinorhizobium meliloti*. *PLOS Genet* **11**:e1005232.
951 Pini F, Frage B, Ferri L, De Nisco NJ, Mohapatra SS, Taddei L, Fioravanti A, Dewitte F, Galardini
952 M, Brilli M, Villeret V, Bazzicalupo M, Mengoni A, Walker GC, Becker A, Biondi EG.
953 2013. The DivJ, CbrA and PleC system controls DivK phosphorylation and symbiosis in
954 *Sinorhizobium meliloti*. *Mol Microbiol* **90**:54–71. doi:10.1111/mmi.12347
955 Rasmussen T, Jensen RB, Skovgaard O. 2007. The two chromosomes of *Vibrio cholerae* are
956 initiated at different time points in the cell cycle. *EMBO J* **26**:3124–3131.
957 doi:10.1038/sj.emboj.7601747

958 Roux B, Rodde N, Jardinaud M-F, Timmers T, Sauviac L, Cottret L, Carrère S, Sallet E, Courcelle
959 E, Moreau S, Debellé F, Capela D, de Carvalho-Niebel F, Gouzy J, Bruand C, Gamas P.
960 2014. An integrated analysis of plant and bacterial gene expression in symbiotic root
961 nodules using laser-capture microdissection coupled to RNA sequencing. *Plant J* **77**:817–
962 837. doi:10.1111/tpj.12442
963 Sánchez-Romero MA, Cota I, Casadesús J. 2015. DNA methylation in bacteria: from the methyl
964 group to the methylome. *Curr Opin Microbiol* **25**:9–16. doi:10.1016/j.mib.2015.03.004
965 Satgé C, Moreau S, Sallet E, Lefort G, Auriac M-C, Remblière C, Cottret L, Gallardo K, Noirot
966 C, Jardinaud M-F, Gamas P. 2016. Reprogramming of DNA methylation is critical for
967 nodule development in *Medicago truncatula*. *Nat Plants* **2**:16166.
968 doi:10.1038/nplants.2016.166
969 Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**:2068–2069.
970 doi:10.1093/bioinformatics/btu153
971 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
972 phylogenies. *Bioinformatics* **30**:1312–1313. doi:10.1093/bioinformatics/btu033
973 Starker CG, Parra-Colmenares AL, Smith L, Mitra RM, Long SR. 2006. Nitrogen fixation mutants
974 of *Medicago truncatula* fail to support plant and bacterial symbiotic gene expression. *Plant*
975 *Physiol* **140**:671–680. doi:10.1104/pp.105.072132
976 Tang Y, Gao X-D, Wang Y, Yuan B-F, Feng Y-Q. 2012. Widespread existence of cytosine
977 methylation in yeast DNA measured by gas chromatography/mass spectrometry. *Anal*
978 *Chem* **84**:7249–7255. doi:10.1021/ac301727c
979 Van de Velde W, Zehirov G, Szatmari A, Debreczeny M, Ishihara H, Kevei Z, Farkas A, Mikulass
980 K, Nagy A, Tiricz H, Satiat-Jeunemaître B, Alunni B, Bourge M, Kucho K, Abe M, Kereszt

981 A, Maróti G, Uchiumi T, Kondorosi E, Mergaert P. 2010. Plant peptides govern terminal
982 differentiation of bacteria in symbiosis. *Science* **327**:1122–1126.
983 doi:10.1126/science.1184057
984 Vasse J, De Billy F, Camut S, Truchet G. 1990. Correlation between ultrastructural differentiation
985 of bacteroids and nitrogen fixation in alfalfa nodules. *J Bacteriol* **172**:4295–4306.
986 doi:10.1128/jb.172.8.4295-4306.1990
987 Vasu K, Nagaraja V. 2013. Diverse functions of restriction-modification systems in addition to
988 cellular defense. *Microbiol Mol Biol Rev* **77**:53–72. doi:10.1128/MMBR.00044-12
989 Wang D, Griffitts J, Starker C, Fedorova E, Limpens E, Ivanov S, Bisseling T, Long S. 2010. A
990 nodule-specific protein secretory pathway required for nitrogen-nixing symbiosis. *Science*
991 **327**:1126–1129. doi:10.1126/science.1184096
992 Wang ET, Young JPW. 2019. History of rhizobial taxonomy In: Wang ET, Tian CF, Chen WF,
993 Young JPW, Chen WX, editors. Ecology and Evolution of Rhizobia: Principles and
994 Applications. Singapore: Springer. pp. 23–39. doi:10.1007/978-981-32-9555-1_2
995 Wright R, Stephens C, Shapiro L. 1997. The CcrM DNA methyltransferase is widespread in the
996 alpha subdivision of proteobacteria, and its essential functions are conserved in *Rhizobium*
997 *meliloti* and *Caulobacter crescentus*. *J Bacteriol* **179**:5869–5877.
998 Wright R, Stephens C, Zweiger G, Shapiro L, Alley MR. 1996. *Caulobacter* Lon protease has a
999 critical role in cell-cycle control of DNA methylation. *Genes Dev* **10**:1532–1542.
1000 doi:10.1101/gad.10.12.1532
1001 Zhang H, Lang Z, Zhu J-K. 2018. Dynamics and function of DNA methylation in plants. *Nat Rev*
1002 *Mol Cell Biol* **19**:489–506. doi:10.1038/s41580-018-0016-z

1003 Zweiger G, Marczynski G, Shapiro L. 1994. A *Caulobacter* DNA methyltransferase that functions
1004 only in the predivisional cell. *J Mol Biol* **235**:472–485. doi:10.1006/jmbi.1994.1007
1005

1006 **TABLES AND FIGURES**

1007 **Table 1.** Methylated motifs identified in this study.

Motif*	Type†	Count¥	Frequency (motifs/kb)
E. meliloti 2011			
G**A**NTC CTN**A**G	m6A	11,169	1.67
RCG**C**CTC YGCGGAG	m4C	3,943	0.59
CGC**A**(N5)GTG GCGT(N5)C**A**C	m6A	1,085	0.16
E. meliloti FSM-MA			
G**A**NTC CTN**A**G	m6A	11,215	1.67
TCG**A**(N8)TCGA AGCT(N8)**A**GCT	m6A	2,612	0.39
E. fredii NGR234			
G**A**NTC CTN**A**G	m6A	11,111	1.61
CAG**A**(N7)GTTG GTCT(N7)C**A**AC	m6A	188	0.03
E. adhaerens OV14			
G**A**NTC CTN**A**G	m6A	8,475	1.10
WNCCG**A**TG WNGGCTAC	m6A	4,596	0.60

1008 * The methylated nucleotides are indicated in boldface font.

1009 † Indicates whether the modification is a N^6-methyladenoside (m6A) or N^4-methylcytosine (m4C).

1010 ¥ The total times the motif appears in the genome, regardless of methylation status.

1011 **Table 2.** Relative sequencing depth of each *E. meliloti* replicon.

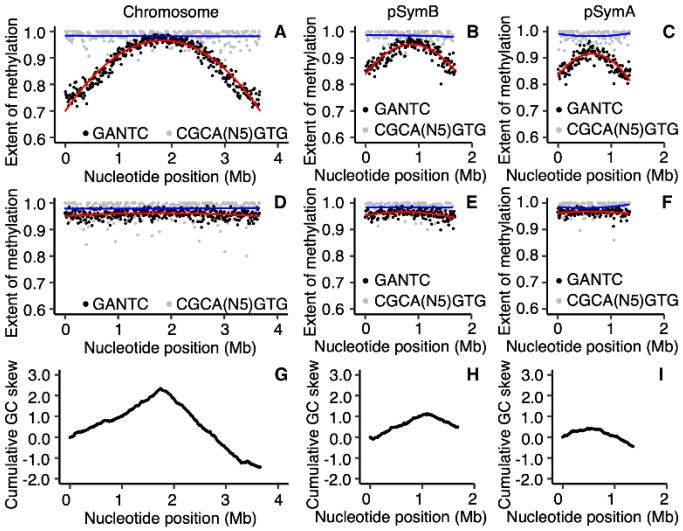
		Relative mean sequencing depth *		
Strain	Condition	Chromosome	pSymA	pSymB
Rm2011	Mid-exponential	1.00 ± 0.00	1.00 ± 0.03	0.98 ± 0.03
Rm2011	Stationary phase	1.00 ± 0.00	1.06 ± 0.02	1.02 ± 0.03
Rm2011	*M. sativa* bacteroids	1.00 ± 0.00	0.64 ± 0.02	0.62 ± 0.02
FSM-MA	Mid-exponential	1.00 ± 0.00	1.02 ± 0.03	0.87 ± 0.05
FSM-MA	Stationary phase	1.00 ± 0.00	1.10 ± 0.00	1.01 ± 0.03
FSM-MA	*M. sativa* bacteroids	1.00 ± 0.00	0.79 ± 0.04	0.71 ± 0.02
FSM-MA	*M. truncatula* bacteroids	1.00 ± 0.00	0.66 ± 0.02	0.59 ± 0.01
FSM-MA	*dnf1* bacterial cells	1.00	0.98	0.90
FSM-MA	*dnf5* bacterial dells	1.00	0.96	0.94
FSM-MA	*dnf2* bacterial dells	1.00	0.81	0.73
FSM-MA	*dnf7* bacterial dells	1.00	0.82	0.71
FSM-MA	*dnf4* bacterial dells	1.00	0.80	0.73
FSM-MA	A17 bacteroids	1.00	0.76	0.67

1012 * Sequencing depth is presented relative to the sequencing depth of the chromosome in the same

1013 sample. Values are the means of triplicate samples ± standard deviation, except for the third section

1014 of the table for which numbers are based on a single replicate.

1015



1016 **Figure 1. Genome-wide DNA methylation of *E. meliloti* Rm2011.** (**A**-**F**) The extent of
1017 methylation is shown, using a 10 kb sliding window, of GANTC sites (black) and CGCA(N_5)GTG
1018 sites (grey) across the chromosome (**A**,**D**), pSymB (**B**,**E**), and pSymA (**C**,**F**) replicons of
1019 exponential phase (**A**-**C**) or early stationary phase (**D**-**F**) *E. meliloti* Rm2011. Averages from three
1020 biological replicates are shown. The red (GANTC) and blue (CGCA(N_5)GTG) lines are
1021 polynomial regression lines calculated in R using the "rlm" method and the formula "y~poly(x,2)".
1022 (**G**-**I**) Cumulative GC skews, shown using a 10 kb sliding window, across the *E. meliloti* Rm2011
1023 chromosome (**G**), pSymB (**H**), and pSymA (**I**) replicons.

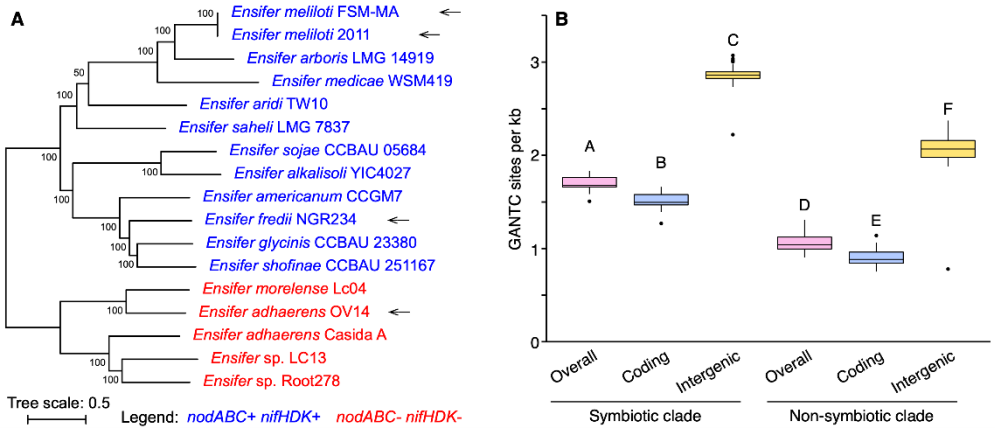
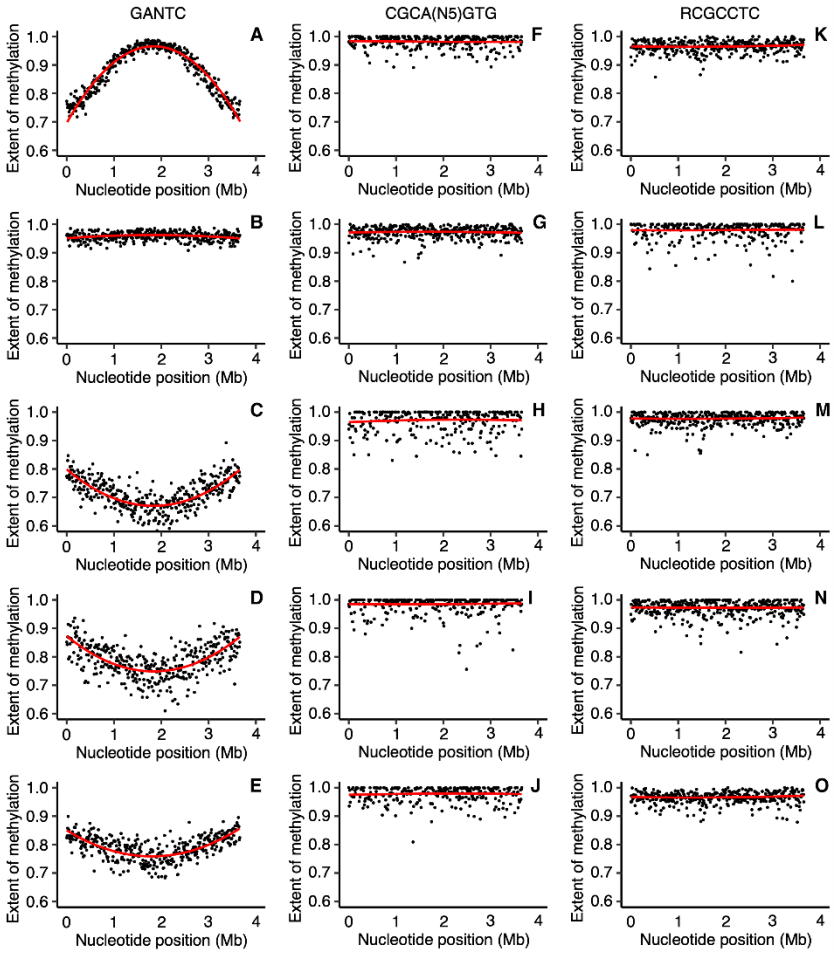
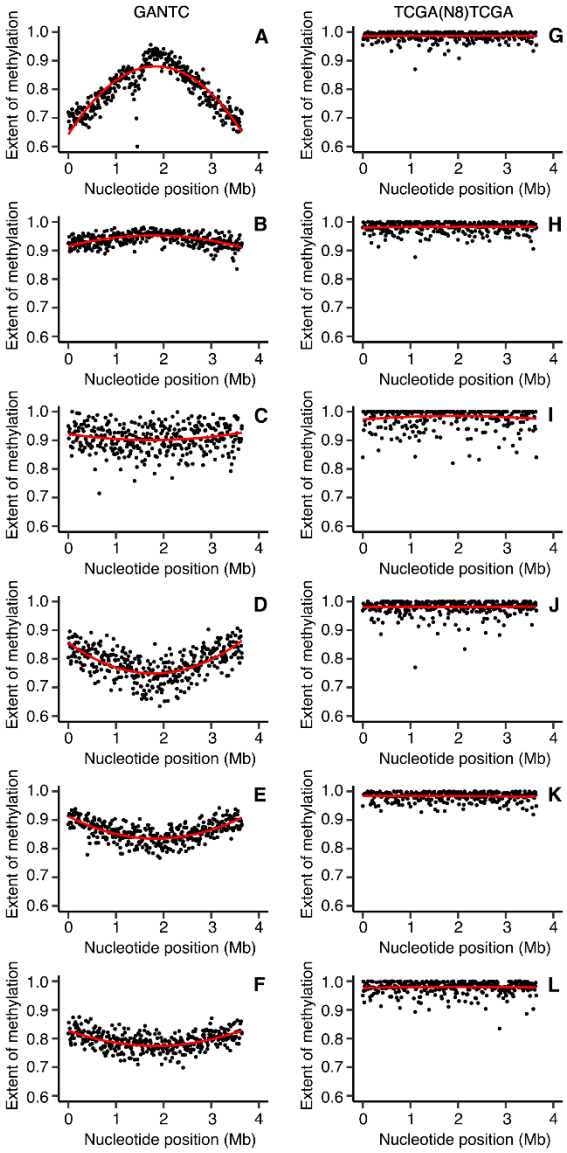


Figure 2. GANTC frequency in the genus *Ensifer*. (**A**) An unrooted maximum likelihood phylogeny of 17 representative *Ensifer* strains. The phylogeny represents the bootstrap best tree following 100 bootstrap replicates, prepared on the basis of the concatenated nucleotide alignments of 1566 core genes. Values represent the bootstrap support. N_2-fixing legume symbionts were identified by the presence of the symbiotic genes *nodABC* and *nifHDK*. They are indicated in blue, while red denotes non-symbiotic strains. The four wild-type strains used in this study are indicated with arrows. (**B**) Box plots summarizing the frequency of GANTC sites (presented as GANTC sites per kb) in 157 *Ensifer* strains is shown. The monophyletic "symbiotic" and "non-symbiotic" clades as defined previously (Fagorzi et al., 2020), are represented by 113 and 44 genomes respectively. The densities of GANTC sites across the entire genome (pink), within coding regions (blue), and within intergenic regions (yellow) are shown. Statistically different values ($p < 0.05$) are denoted by uppercase letters as determined by a one-way ANOVA followed by a Tukey's HSD post hoc test.



1038

1039 **Figure 3. Chromosome-wide DNA methylation of *E. meliloti* Rm2011 bacteroids.** The extent
1040 of methylation of (**A**-**E**) GANTC, (**F**-**J**) CGCA(N_5)GTG, and (**K**-**O**) RCGCCTC motifs across the
1041 *E. meliloti* Rm2011 chromosome is shown using a 10 kb sliding window. Averages from three
1042 biological replicates are shown for free-living and whole nodule samples; data represents one
1043 replicate for the zone II and zone III nodule sections. (**A**,**F**,**K**) Free-living cells harvested in mid-
1044 exponential phase. (**B**,**G**,**L**) Free-living cells harvested in early stationary phase. (**C**,**H**,**M**)
1045 Bacteroids isolated from *M. sativa* zone II nodule sections. (**D**,**I**,**N**) Bacteroids isolated from *M.*
1046 *sativa* zone III nodule sections. (**E**,**J**,**O**) Bacteroids isolated from *M. sativa* whole nodule samples.
1047 The red lines are polynomial regression lines calculated in R using the "rlm" method and the
1048 formula "y~poly(x,2)". Data for pSymB and pSymA are shown in Figures S14 and S15.



1049

1050 **Figure 4. Chromosome-wide DNA methylation of *E. meliloti* FSM-MA bacteroids.** The extent of
1051 methylation of (**A-F**) GANTC and (**G-L**) TCGA(N_8)TCGA motifs across the *E. meliloti* FSM-MA
1052 chromosome is shown using a 10 kb sliding window. Averages from three biological replicates are
1053 shown for free-living and whole nodule samples; data represents one replicate for the zone II and zone
1054 III nodule sections. (**A**,**G**) Free-living cells harvested in mid-exponential phase. (**B**,**H**) Free-living cells
1055 harvested in early stationary phase. (**C**,**I**) Bacteroids isolated from *M. sativa* zone II nodule sections.
1056 (**D,J**) Bacteroids isolated from *M. sativa* zone III nodule sections. (**E**,**K**) Bacteroids isolated from *M.*
1057 *sativa* whole nodule samples. (**F**,**L**) Bacteroids isolated from *M. truncatula* whole nodule samples. The
1058 red lines are polynomial regression lines calculated in R using the "rlm" method and the formula
1059 "y~poly(x,2)". Data for pSymB and pSymA are shown in Figures S16 and S17.

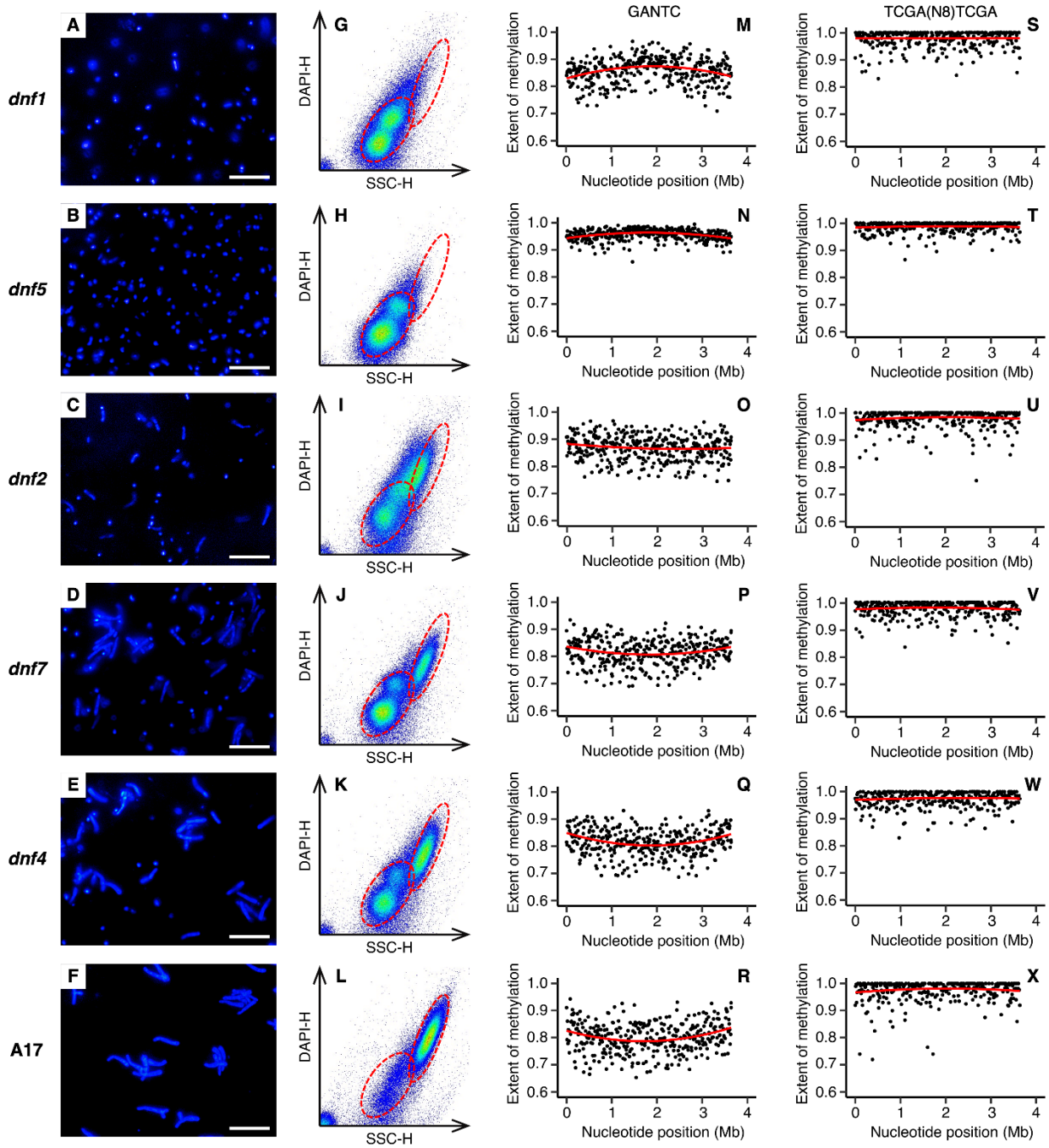


Figure 5. Bacteroid morphology and chromosomal GANTC methylation in *E. meliloti* bacteroids purified from *M. truncatula dnf* mutant nodules. Data is shown for *E. meliloti* FSM-MA bacteroids purified from (**A**,**G**,**M**,**S**) *M. truncatula dnf1* mutant nodules, (**B**,**H**,**N**,**T**) *M. truncatula dnf5* mutant nodules, (**C**,**I**,**O**,**U**) *M. truncatula dnf2* mutant nodules, (**D**,**J**,**P**,**V**) *M. truncatula dnf7* mutant nodules, (**E**,**K**,**Q**,**W**) *M. truncatula dnf4* mutant nodules, and (**F**,**L**,**R**,**X**)

1066 *M. truncatula* A17 wild-type nodules. (**A**-**F**) Micrographs of *E. meliloti* FSM-MA bacteroids
1067 stained with the DNA binding dye DAPI. The scale bar represents 30 µm. (**G**-**L**) Pseudo-coloured
1068 scatterplots displaying the cell morphology (X-axis) and DNA content (Y-axis) of *E. meliloti*
1069 FSM-MA bacteroids, as determined based on flow cytometry analysis of DAPI stained cells. The
1070 red dashed ellipses indicate the position of undifferentiated bacteria as in culture (not shown) or in
1071 the *dnf1* mutant nodules (lower left ellipse) or fully mature bacteroids as in the A17 wild-type
1072 nodules (top right ellipse). (**M**-**X**) The extent of methylation of (**M**-**R**) GANTC or (**S**-**X**)
1073 TCGA(N_8)TCGA motifs across the *E. meliloti* FSM-MA chromosome, shown using a 10 kb
1074 sliding window. The red lines are polynomial regression lines calculated in R using the "rlm"
1075 method and the formula "y~poly(x,2)". Data for pSymB and pSymA are shown in Figures S19
1076 and S20.