Title: DNA methylation patterns in bacteria of the genus *Ensifer* during free-living growth and during nitrogen-fixing symbiosis with *Medicago* spp.

Running title: DNA methylation in the genus *Ensifer*

Authors: George C. diCenzo[1,2,#], Lisa Cangioli[2], Quentin Nicoud[3], Janis H.T. Cheng[1], Matthew J. Blow[4], Nicole Shapiro[4], Tanja Woyke[4], Emanuele G Biondi[3], Benoît Alunni[3], Alessio Mengoni[2,*], and Peter Mergaert[3,*]

Affiliations: [1] Department of Biology, Queen's University, Kingston ON, K7L 3N6, Canada

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

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

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

 # Corresponding author: George C. diCenzo, george.dicenzo@queensu.ca

* These authors contributed equally to this work.

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. Here, we explored DNA methylation in bacteria of the genus *Ensifer*, including its potential role in regulating terminal differentiation during nitrogen-fixing symbiosis with legumes. Using single-molecule real-time sequencing, six genome-wide methylated motifs were identified across four *Ensifer* strains. Five motifs were strain-specific, nearly fully methylated across the genomes in all tested conditions, and were not enriched in the promoters of symbiosis, carbon source, or cell cycle-regulated genes. These results suggest that most DNA methylation is not a major regulatory mechanism in the genus *Ensifer*. Only the GANTC motif, recognized by the CcrM methyltransferase, was methylated in all strains. In actively dividing cells, methylation of GANTC increased from the *ori* to *ter* region in each replicon, in agreement with a cell cycle-dependent regulation of CcrM. There was near full genome-wide GANTC methylation in the early stage of symbiotic differentiation, followed by a moderate decrease in methylation during later stages of differentiation. This is evidence of dysregulated and constitutive CcrM activity during terminal differentiation, which we suggest is a driving factor for endoreduplication of terminally differentiated bacteroids.

INTRODUCTION

Methylation of genomic DNA is a pervasive phenomenon found in eukaryotes (1–3), archaea (4), and bacteria (4, 5). The biological roles of DNA methylation are most extensively studied in mammals, where it contributes to normal development and disease via its impact on gene expression (6). In bacteria, DNA methylation is best known for its role in restriction-modification (R-M) systems that are thought to provide defence against phage infection and limit horizontal gene transfer through the degradation of invading non-methylated DNA (7). Several methyltransferases (MTases) of R-M systems have also been implicated in phase variation in pathogens through modulating gene expression (8). A recent study of over 200 bacterial and archaeal species identified orphan MTases not belonging to R-M systems in nearly half of the genomes (4). To date, biological functions have been attributed to very few orphan MTases, namely, the Dam MTase of the γ-Proteobacteria and the CcrM MTase of the α-Proteobacteria (9). The Dam MTase of *Escherichia coli* is notable for its role in regulation of DNA replication (10, 11) and DNA repair (12) by modulating the activity of other DNA-binding proteins. The CcrM MTase was first identified in *Caulobacter crescentus* (13), with homologs since identified in diverse α-Proteobacteria (14, 15). CcrM activity was shown to be cell cycle regulated in *C. crescentus* and *Agrobacterium tumefaciens* (13, 16), leading to methylation of its cognate DNA motif (the pentanucleotide GANTC) specifically during a short period at the end of DNA replication. This leads to a switching of GANTC sites between fully methylated (methylated on both strands) and hemi-methylated (methylated only on the template strand) as a result of DNA replication (17), which serves to modulate gene expression in a cell cycle-dependent fashion (18–20). Over- and under-expression of *ccrM* results in defects in DNA replication and cell division (14, 16, 19), while its complete loss is generally lethal.

The rhizobia are a polyphyletic group of α-Proteobacteria and β-Proteobacteria that can both live free in the soil and enter into an endosymbiotic interaction with legumes (21). This interaction begins following an exchange of signals between the free-living partners (22), and it culminates in the formation of a new organ known as a root nodule within which the cytoplasm of plant cells contain thousands of N_2-fixing bacteria called bacteroids. Bacteroid formation results in the differential expression of more than a thousand genes (23, 24) and global changes in cellular metabolism (25). In legumes of the Inverted-Repeat Lacking Clade (IRLC) and the Dalbergioid clade of the family *Papilionoideae*, bacteroid development involves an additional process of

terminal differentiation (26, 27); in other legume clades, bacteroid differentiation is less pronounced and is reversible. Terminal bacteroid development, in contrast to reversible bacteroid formation, involves cell enlargement (bacteroids are 5- to 10-fold longer than their free-living counterparts) and genome endoreduplication (resulting in up to 24 copies of the genome per cell), indicative of a cell cycle transition occurring during differentiation (26). Indeed, the correct expression of cell cycle regulators in *Ensifer* (syn. *Sinorhizobium*) *meliloti*, a symbiont of *Medicago* species of the IRLC, is essential for the formation of functional bacteroids (28, 29), while over-expression of CcrM or disruption of the master cell cycle regulator CtrA can give rise to bacteroid-like morphology in free-living cells (14, 30). Additionally, mutants in the *E. meliloti* cell cycle regulators *divJ*, *cbrA*, and *cpdR1*, encoding three negative regulators of CtrA, form non-functional nodules in which bacteroids do not differentiate properly (28, 29, 31), and genes encoding several cell cycle regulators (including CcrM) are strongly downregulated in bacteroids (24). The differentiation and cell cycle switch of bacteroids is controlled by the legume host through the production of a large family of peptides, known as Nodule-specific Cysteine-Rich (NCR) peptides (32–34).

Multiple studies have provided evidence that changes in the methylation status of the DNA of legume nodule cells contributes to symbiotic development (35–37). Conversely, it remains unknown if methylation of rhizobium DNA contributes to the regulation of N_2-fixation or bacteroid development. We are aware of only one study (38) comparing DNA methylation of a rhizobium (*Bradyrhizobium diazoefficiens* USDA110) between free-living and symbiotic states (soybean nodules), and comparing these changes with differential expression data. Intriguingly, the authors identified a DNA motif that was methylated specifically in bacteroids (38). However, no clear evidence was presented that methylation of this (or any other) motif is involved in transcriptional regulation, and the number of genes both differentially expressed and differentially methylated in bacteroids did not appear to be different than expected by chance. While these data may suggest that DNA methylation does not play a major role in regulating N_2-fixation by rhizobia, they do not address the role of DNA methylation in terminal bacteroid differentiation, as *B. diazoefficiens* undergoes reversible differentiation in soybean nodules (39).

Here, we use Pacific Biosciences Single-Molecule Real-Time (SMRT) sequencing to detect genome-wide patterns of DNA methylation in four strains belonging to the genus *Ensifer*. Our results indicate that DNA methylation is poorly conserved across the genus and suggests that

DNA methylation is not a major mechanism of regulating gene expression in these organisms aside from cell cycle control. However, analysis of bacteroid samples suggests that constitutive activation of the CcrM MTase may be a contributing factor driving terminal differentiation.

MATERIALS AND METHODS

Experimental design

The overall experimental design is summarized in **Figure S1**. Genomic DNA was isolated from four wild-type *Ensifer* strains to explore how DNA methylation varies across this genus; to allow direct comparison, the four strains were grown to mid-exponential phase in minimal medium with succinate as a carbon source. To investigate how DNA methylation patterns differ between actively dividing and non-dividing cells, genomic DNA was isolated from *E. meliloti* Rm2011 growth to either mid-exponential phase or stationary phase. Genomic DNA was isolated from *E. meliloti* Rm2011 grown to mid-exponential phase with either a glycolytic (sucrose) or gluconeogenic (succinate) carbon source to investigate whether DNA methylation may play a role in regulating carbon metabolism. Furthermore, a *E. meliloti* Rm2011 derivative lacking the pSymA and pSymB replicons (named RmP3496) was studied to gain insight into whether these replicons contribute to DNA methylation patterns in *E. meliloti* Rm2011; these strains were grown with sucrose (instead of succinate) as RmP3496 lacks the succinate transporter.

In addition to the free-living samples, *E. meliloti* bacteroid samples purified from legume nodules were collected to investigate changes in DNA methylation during bacteroid differentiation and nitrogen fixation. To do so, *E. meliloti* Rm2011 and *E. meliloti* FSM-MA bacteroids were isolated from *M. sativa* whole nodules. *E. meliloti* FSM-MA bacteroids were additionally purified from *M. truncatula* whole nodules to examine the impact of host plant on bacteroid DNA methylation patterns. *E. meliloti* Rm2011 bacteroids were only isolated from *M. sativa* nodules as, unlike FSM-MA, Rm2011 forms a poor symbiosis with *M. truncatula* (40, 41). Moreover, *E. meliloti* Rm2011 and *E. meliloti* FSM-MA bacteroids were isolated from *M. sativa* nodule sections (sectioned at the white – pink border to separate the root distal infection and differentiation zone II [white] from the root proximal nitrogen-fixing zone III [pink]) to facilitate an analysis of how DNA methylation patters differ between differentiating bacteroids and fully differentiated and nitrogen-fixing bacteroids. This was followed by isolation of *E. meliloti* FSM-MA bacteroids from whole nodules of *M. truncatula* mutant lines (*dnf1*, *dnf2*, *dnf4*, *dnf5*, *dnf7*) to investigate DNA

methylation patterns in nodule bacteria blocked as various stages of differentiation.

Bacterial strains and growth conditions

Bacterial strains used in this work are listed in **Table S1**. All strains were routinely grown on TY with 2 μM $CoCl_2$ as it was required for *E. meliloti* RmP3496 (42). The MM9 minimal medium (42) consisted of the following: 40 mM MOPS, 20 mM KOH, 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, 42 nM $CoCl_2$, 38 μM $FeCl_3$, 10 μM thiamine-HCl, and either 10 mM sucrose (MM9-sucrose) or 20 mM disodium succinate (MM9-succinate). Prior to inoculation of plants with *E. meliloti*, the strains were grown in YEB medium (43).

DNA isolation from free-living cells

Overnight cultures of all strains grown in MM9-succinate or MM9-sucrose media were diluted into 10 mL of the same medium to a starting OD_{600nm} of 0.025 (0.05 for RmP3496) and incubated overnight at 30˚C with shaking (130 rpm). The next day, cultures were diluted into 40 mL of the same medium in 100 mL flasks to the OD_{600nm} values as listed in **Table S2**. To obtain mid-exponential phase samples, cultures were harvested after 15.5-16 hours of growth at OD_{600nm} values between 0.37 and 0.69 (**Table S2**). To obtain stationary phase samples, cultures were harvested after 24 hours of growth at OD_{600nm} values of ~ 1.4. In all cases, cultures were streaked on TY plates to check for contamination and then centrifuged (8,200 *g*, 10 minutes, 4°C); the full 40 mL was centrifuged for mid-exponential phase cultures, whereas only 15 mL was centrifuged for stationary phase cultures. Most of the supernatant was discarded, and the pellet resuspended in the remaining volume, transferred to a 2 mL tube, centrifuged again (16,200 *g*, room temperature, one minute), and the supernatant discarded. Three biological replicates, each starting from a separate overnight culture, were performed. DNA was isolated using phenol:chloroform extractions and ammonium acetate precipitations as described elsewhere (44), and the DNA pellets (following RNase A treatment) were resuspended in 200 μL of 10 mM Tris-HCl, pH 8.5.

DNA isolation from bacteroids

M. sativa cv. Gabès and *M. truncatula* cv. A17 were used wild-type plants for all experiments. *M. truncatula dnf1*, *dnf2*, *dnf4*, *dnf5*, and *dnf7* mutants (45), derived from the A17 wild type, were

used for collection of bacteroids blocked at various stages of differentiation. Seeds were scarified, surface sterilized, and germinated on Kalys agar as described previously (41). Fifty mL of overnight cultures of *E. meliloti* Rm2011 or FSM-MA, grown in YEB, were centrifuged (4,000 *g*, 20 minutes, room temperature) and resuspended in ~ 1,200 mL of sterile, distilled water to obtain a cell suspension at an OD_{600nm} of ~ 0.1. Germinated seedlings were immersed for one hour in the appropriate rhizobial cell suspension, and then planted in a perlite:sand (2:1) mixture. Plants were grown in a greenhouse for five to six weeks, with occasional watering with a 1 g L^{-1} nutrient solution (PlantProd solution [N-P-K, 0-15-40; Fertil]).

For whole nodule samples of wild type plants, pink nodules were collected from 53-60 plants per replicate 34-35 days post-inoculation; in the case of *dnf* mutants (and a matched wild-type A17 sample), nodules were collected from ~ 105 plants per genotype 23-24 days post-inoculation. Nodules were collected from the roots and stored in tubes in liquid nitrogen until collection was complete, at which point they were stored at -80°C until use. For sectioned nodule samples, pink nodules were collected from 103 *M. sativa* plants for each of the microsymbionts 35 to 40 days post-inoculation. Nodules were manually sectioned at the white-pink border. Nodule sections were stored in tubes over dry ice or liquid nitrogen until collection was complete, at which point they were stored at -80°C until use. Average plant shoot dry weights for all samples are listed in **Table S3**. Bacteroids were isolated from the nodule samples using Percoll gradient centrifugation as described elsewhere (26). The recovered bacteroids were resuspended in 50-100 µL of Bacteroid Extraction Buffer (BEB; 125 mM KCl, 50 mM Na-succinate, 50 mM TES, pH 7.0), and either used immediately for microscopy, flow cytometry, and DNA isolation, or stored at -80°C until use.

Nucleic acids were initially purified from most bacteroid samples using Epicentre MasterPure™ Complete DNA and RNA Purification Kit, following the protocol for DNA isolation from cell samples; the exceptions were bacteroids purified from *dnf* mutant nodules (and the matched wild-type A17 sample), for which nucleic acids were isolated by using phenol:chloroform extractions followed by ammonium acetate DNA precipitations as described elsewhere (44). For sectioned nodule samples, pure DNA was isolated by using the manufacturer's protocol for the complete removal of RNA. For whole nodule samples, the isolated DNA was further purified by treating the nucleic acids samples with RNase A, after which pure DNA was isolated by using phenol:chloroform extractions followed by ammonium acetate DNA precipitations or alternatively

using the MasterPure™ DNA clean-up protocol for the DNA from *dnf* mutant nodules and the matched wild-type A17 sample. In all cases, the final DNA pellets were resuspended in 200 µL of 10 mM Tris-HCl, pH 8.5. Three biological replicates were performed for bacteroids isolated from most whole nodules, whereas only one replicate was performed for bacteroids isolated from sectioned nodules or *dnf* mutant nodules (and the matched wild-type A17 sample) due to low quantities of starting materials.

DNA sequencing, modification detection, and motif analysis

DNA sequencing was performed at the U.S. Department of Energy Joint Genome Institute (JGI) or in-house at the University of Florence (the stationary phase samples and *dnf* mutant nodules and the matched wild-type A17 sample) using Pacific Biosciences (PacBio) sequencing technology (46). Genomic DNA was sheared to 3 kb using a Covaris LS220 (Covaris Inc., Woburn, MA, USA) or 15 kb (for stationary phase samples and bacteroids isolated from *dnf* mutant nodules and the matched wild-type A17 sample) using g-TUBEs (Covaris Inc., Woburn, MA, USA). Sheared DNA was treated with exonuclease to remove single-stranded ends and DNA damage repair mix followed by end repair and ligation of barcoded blunt adapters using SMRTbell Template Prep Kit 2.0 (PacBio, Menlo Park, CA, USA). Libraries were purified with AMPure PB beads (Beckman Coulter, Brea, CA, USA) and three or eight libraries with different barcodes were pooled at equimolar ratios and purified with AMPure PB beads. For most samples, SMRTbell template libraries were prepared using a Sequel Binding Kit 3.0 (PacBio, Menlo Park, CA, USA), and sequenced on a Sequel instrument using a v3 or v4 sequencing primer, 1M v3 SMRT cells, and Version 3.0 sequencing chemistry with 1x360 or 1x600 sequencing movie run times. The exceptions were the *E. meliloti* Rm2011 zone II and *E. meliloti* FSM-MA zone III bacteroid samples. For these samples, SMRTbell template libraries were prepared using a Sequel II Binding Kit 2.0 (PacBio, Menlo Park, CA, USA), and then sequenced on a Sequel II instrument using the tbd-sample dependent sequencing primer, 8M v1 SMRT cells, and Version 2.0 sequencing chemistry with 1x900 sequencing movie run times.

DNA modification detection and motif analysis were performed using the PacBio SMRT Link software (PacBio, Menlo Park, CA, USA). Briefly, raw reads were filtered using SFilter to remove short reads and reads derived from sequencing adapters. Filtered reads were aligned against the appropriate reference genome (**Table S4**) using BLASR (47) and modified sites were

then identified through kinetic analysis of the aligned DNA sequence data (48); the number of mapped bases per sample is provided in **Table S4**. Modified sites were then grouped into motifs using MotifFinder. These motifs represent the recognition sequences of MTase genes active in the genome (49). Downstream analyses were performed using custom Perl and R scripts.

Flow cytometry

Flow cytometry was performed as described previously (26). Freshly prepared bacteroid samples were diluted in 200 µL of BEB, heat-treated for 10 minutes in a 70°C water bath, and then stained with the DNA-binding dye diamidino-2-phenylindole (DAPI). Cell size and ploidy level of the bacteroid samples were determined using flow cytometry with a Beckman Coulter CytoFLEX S instrument. Measurements consisted of 50,000 cells. Data analysis was performed using the CytExpert 2.2.0.97 software.

Fluorescence microscopy

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

Phylogenetic analysis

The nucleotide fasta files of representative *Ensifer* species were downloaded from the National Centre for Biotechnology Information (NCBI) Genome database. A core gene phylogeny was constructed using a previously prepared pipeline (50) reliant on the use of Roary 3.11.3 (51), Prokka 1.12-beta (52), PRANK 140110 (53), trimAl (54), and RAxML 8.2.9 (55). The phylogeny was visualized with the iTol webserver (56). Identification of nodulation (*nodABC*) and nitrogenase genes (*nifHDK*) was performed with a published pipeline (50) reliant on the use of HMMER 3.1b2 (57), and the Pfam-A 31.0 (58) and TIGERFAM 15.0 (59) databases.

RESULTS

The methylomes of the genus *Ensifer*

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

We began with a base modification analyses of four wild-type strains from three species, including three nodule-forming strains (*E. meliloti* Rm2011, *E. meliloti* FSM-MA, *E. fredii* NGR234) and one plant-associated, non-symbiotic strain (*E. adhaerens* OV14). To ensure consistency, all strains were grown to mid-exponential phase in a common minimal medium with succinate as the carbon source. A total of six methylated motifs were identified, of which five were m6A modifications and one was a m4C modification (**Table 1**). Five of the six motifs were methylated specifically in one strain. Only the GANTC motif, recognized by the highly conserved cell cycle-regulated CcrM methyltransferase (13, 14), was methylated in all four strains. To further examine the conservation, or lack thereof, of DNA modification across the genus *Ensifer*, we examined the distribution of methyltransferases in the model species *E. meliloti*. Based on gene annotations, we identified 24 genes encoding putative MTases in a previous pangenome analysis of 20 *E. meliloti* strains (**Table S5**) (60). Of these 24 genes, only one (*ccrM*) was found in all 20 strains, while four were found in two strains and 19 were found in a single strain. These results suggest that DNA methylation is unlikely to play a biologically significant role in the genus *Ensifer* aside from cell cycle control via CcrM-mediated methylation, and phage defence.

In support of most DNA methylation not having a regulatory function in the genus *Ensifer*, no global effect of carbon source (sucrose [glycolytic] vs. succinate [gluconeogenic]) was observed on the DNA methylation pattern of *E. meliloti* Rm2011 (**Figure S2**), nor were any motifs enriched in the promoter regions of genes previously shown to be differentially expressed when grown with glucose vs. succinate (61). Similarly, no global differences in DNA methylation were detected between *E. meliloti* Rm2011 and RmP3496, a Rm2011 derivative lacking the pSymA and pSymB replicons that together account for 45% of the genome content of *E. meliloti* (42) (**Figure S3**).

Cell cycle regulation by the CcrM methyltransferase

A progressive increase in the extent of methylation (herein defined as the estimated fraction of reads mapping to a motif that were methylated) of GANTC sites was observed from the *ori* to *ter* regions of the chromosomes of all four strains during mid-exponential growth (**Figures 1, S4-S6**). There was a local drop in GANTC methylation around the 1.5 Mb mark in the *E. meliloti* FSM-MA chromosome (**Figure S4**); however, this was seen in only two of three replicates and corresponded to a region of high sequencing depth (**Figure S7**), suggesting the result is a sequencing artefact. In contrast to exponential phase cultures, GANTC sites displayed near full methylation (averaging ~ 95%) across the genome during early stationary phase, while all other motifs displayed near full methylation (averaging 95-99%) across the genome regardless of growth state (**Figures 1, S4-S6**). The observed pattern of GANTC methylation indicates a progressive switch from full to hemi-methylated states as DNA replication proceeds (model provided as **Figure S8**), confirming that the CcrM methyltransferase of the family *Rhizobiaceae* is cell cycle regulated as demonstrated in *C. crescentus* (13, 17, 62). Strikingly, the GANTC methylation pattern differed across replicons within each genome. For example, in *E. meliloti* Rm2011 the extent of GANTC methylation ranged from 0.80 to 0.98 for pSymB and 0.80 to 0.96 for pSymA, while for the chromosome the range was from 0.71 to 0.99 (**Figure 1**). This result suggests that replication of each replicon is asynchronous, with replication of the secondary replicons being initiated later in the cell cycle than that of the chromosome (model provided as **Figure S8**).

A previous study identified 462 cell cycle-regulated genes in *E. meliloti* through RNA-sequencing of synchronized cell populations (63), 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 (63). 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.

We found it striking that the *E. adhaerens* OV14 genomes had 2,636 to 2,740 fewer GANTC sites than the other three strains, despite having the largest genome size. Normalized by genome length, there are 1.10 GANTC sites per kb in the *E. adhaerens* OV14 genome (**Table 1**), which is similar to the 1.12 GANTC sites per kb in *C. crescentus* NA1000. In contrast, the three

legume symbionts contained more than 1.60 GANTC sites per kb across their genomes (**Table 1**). This result prompted us to examine the frequency and distribution of GANTC sites across 157 *Ensifer* genomes. As defined previously (64), the genus *Ensifer* can be broadly sub-divided into two monophyletic clades; the "symbiotic" clade (113 strains) in which nearly all strains are legume symbionts, and the "non-symbiotic" clade (44 strains) in which nearly all strains are non-symbionts (**Figure 2A**). Consistent with previous results (18), GANTC sites occurred less frequently in all genomes (0.90 to 1.83 GANTC sites per kb) than expected in a random sequence of nucleotides (~ 3.5 GANTC sites per kb). Moreover, GANTC sites were ~ 2-fold more common in intergenic regions than in coding regions (**Figure 2B**). Strikingly, there was a strong and statistically significant difference (p-value $< 1 \times 10^{-10}$; two-sample t-test) in the frequency of GANTC sites across the genomes of strains belonging to the symbiotic clade compared to strains of the non-symbiotic clades (**Figure 2B**), with an overall average of 1.70 and 1.06 GANTC sites per kb in the symbiotic and non-symbiotic clades, respectively. The difference in the frequency of GANTC sites between the two clades could not be explained by differences in the GC content of these organisms, as both clades had an average GC content of 61.9% (**Figure S9**), suggesting that the difference reflects underlying differences in the evolution, and possibly the biology, of these two clades.

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 (38, 39). To evaluate whether DNA methylation potentially contributes to regulation of terminal differentiation, we determined the DNA methylation patterns of *E. meliloti* Rm2011 and *E. meliloti* FSM-MA bacteroids purified from whole *Medicago sativa* nodules. *E. meliloti* FSM-MA bacteroids were additionally purified from *Medicago truncatula* nodules to determine whether the host plant influences bacteroid DNA methylation. *E. meliloti* Rm2011 bacteroids were not isolated from *M. truncatula* nodules as, unlike FSM-MA, Rm2011 forms a poor symbiosis with *M. truncatula* and produces moderately differentiated bacteroids in this host (40, 41).

Moreover, we exploited the spatially distinct developmental zones that are present in indeterminate nodules (65), like those formed by *M. sativa* and *M. truncatula*. At the tip of these

nodules, a bacteria-free meristem is present, responsible for the continuous growth of the nodule. Immediately below is the infection and differentiation zone II where nodule cells become infected and bacteria differentiate into the large, polyploid bacteroids. The tip and zone II of nodules appears white. Adjacent to the white zone II is the easily recognizable pinkish zone III (due to the presence of the oxygen-carrying leghemoglobin) where mature bacteroids fix nitrogen. This nodule tissue organization provided an opportunity to examine how DNA methylation patterns may differ between differentiating and differentiated bacteroids. To this end, *E. meliloti* Rm2011 and FSM-MA bacteroids were isolated from nodules hand-sectioned at the white-pink border; bacteroids isolated from the white sections represent the infecting and differentiating bacteroids (zone II) while those isolated from the pink sections represent the mature, hence terminally differentiated, N_2-fixing bacteroids (zone III).

Fluorescence microscopy and flow cytometry confirmed that nodule sectioning resulted in the isolation of distinct bacteroid populations (**Figures S10-S13**). Nearly all of the bacteroids isolated from whole-nodule samples and zone III samples were enlarged and polyploid, and most were positive for propidium iodide (PI) staining as expected for terminally differentiated bacteroids (26). In contrast, bacteroids of the zone II samples contained a mix of cell types differing in their size, ploidy level, and PI staining. These data confirmed that the whole-nodule samples and zone III samples consisted predominately of mature N_2-fixing bacteroids, whereas the zone II samples contained a mix of cells at various stages of bacteroid differentiation.

With the exception of the GANTC sites (described below), no global difference was observed in the methylation patterns of bacteroids versus free-living cells (**Figures 3, 4, S14-S17**). Although a lower percentage of each motif was detected as methylated in the bacteroid samples compared to free-living samples, this was correlated with lower sequencing depth (**Table S4**) and thus unlikely to be biologically meaningful. Unlike *B. diazoefficiens* USDA110 (38), we did not identify motifs that were methylated specifically in the *E. meliloti* bacteroids. In addition, we found little evidence for any of the methylated motifs being enriched in the promoter regions of *E. meliloti* Rm2011 genes up-regulated or down-regulated in bacteroids relative to free-living cells, as identified in published transcriptomic data for *E. meliloti* Rm1021 (23), a near-isogenic relative of strain Rm2011 also derived from the nodule isolate SU47. These data suggest that most DNA methylation is unlikely to be a significant factor in directly regulating gene expression in *E. meliloti* bacteroids.

CcrM methyltransferase activity is dysregulated during terminal differentiation

Bacteroid development involves cell enlargement and genome endoreduplication, indicative of a cell cycle transition occurring during differentiation (26). Indeed, *ccrM* and other genes encoding cell cycle regulators are strongly downregulated in terminally differentiated *E. meliloti* bacteroids (24). We were therefore interested in examining whether GANTC methylation by the CcrM MTase was disrupted in bacteroids. Our data revealed a surprising genome-wide pattern of GANTC methylation in *E. meliloti* Rm2011 and FSM-MA bacteroids, which differed from free-living cells in either the exponential or stationary phases of growth (**Figures 3, 4, S14-S17**). The majority of GANTC sites had moderate to high levels of methylation in zone II, zone III, or whole-nodule samples, averaging 0.71 to 0.95 across each replicon (**Figures 3, 4, S14-S17, and Tables S6, S7**). Most distinctive, a progressive decrease in the extent of chromosomal methylation of the GANTC sites was observed from the *ori* to *ter* in nearly all bacteroid samples, revealing a characteristic smiling pattern, which differs from the patterns seen in exponential (frowning pattern, i.e., a progressive increase from *ori* to *ter*) and stationary (consistent methylation) phase cells. The exception was the *E. meliloti* FSM-MA zone II bacteroid sample, which displayed a consistently high level of GANTC methylation across the genome (**Figure 4C**). This pattern, which is different from those in exponential phase cells as well as mature bacteroids, could correspond to the methylation status of an early stage of bacteroid differentiation. Supporting this hypothesis, the flow cytometry data was consistent with the *E. meliloti* FSM-MA zone II bacteroids having a lower average DNA content than the *E. meliloti* Rm2011 zone II bacteroids (**Figure S18**), suggesting that the former sample represents an earlier stage of differentiation than the latter sample. In contrast to GANTC, the extent of methylation of the second m6A modified motif in each genome was consistently high, irrespective of condition or replicon (**Figures 3, 4, S14-S17, and Tables S6, S7**). Similarly, sequencing depth was consistent across the length of each replicon (**Figure S7**). These observations indicate that the changes in GANTC methylation patterns are biologically meaningful and not simply a sequencing artefact.

To further explore changes in CcrM activity during bacteroid differentiation, we took advantage of a collection of *M. truncatula* mutant plant lines (*dnf1*, *dnf2*, *dnf4*, *dnf5*, *dnf7*) whose nodules contain bacteria blocked at various stages of differentiation (45, 66–71). Microscopy and flow cytometry data was consistent with past work (71) showing that bacteroids were blocked at

the earliest to latest stages of differentiation in mutant plant lines in the order *dnf1* → *dnf5* → *dnf2* → *dnf7* → *dnf4* (**Figure 5**). Nodule bacteria of *M. truncatula dnf1* mutant plants were small with one or two haploid genome copies per cell (i.e., ploidy level = 1 or 2) (**Figures 5A, 5G**), suggesting that the cell population was dominated by actively dividing cells that had not yet begun differentiation. Indeed, the GANTC methylation pattern of these cells (**Figures 5M, S19, S20**) resembled the frowning pattern of exponentially growing free-living cells (**Figure 4A**). Although the nodule bacteria of *M. truncatula dnf5* mutant plants were also small and undifferentiated into bacteroids, the majority of cells had a ploidy level of one (**Figures 5B, 5H**), suggesting these cells had ceased replication but had not yet begun the process of endoreduplication. GANTC methylation was consistently high across the chromosome of bacteria purified from *dnf5* nodules (**Figure 5N**), similar to stationary phase free-living cells (**Figure 4B**) and indicating that terminal differentiation is preceded by full GANTC methylation.

The nodule bacteria purified from *M. truncatula dnf2* mutant nodules were a mix of undifferentiated and partially differentiated bacteroids, which were polyploid to an extent similar to bacteroids purified from wild-type A17 nodules (**Figure 5I** compared to **Figure 5L**); however, their cell size was much smaller (**Figure 5C** compared to **Figure 5F**). This was similar to differentiating bacteroids purified from *M. truncatula* and *M. sativa* zone II nodule sections, many of which had high ploidy without a corresponding increase in cell size (**Figures S11, S13**). The GANTC methylation pattern of bacteroids from *dnf2* nodules (**Figure 5O**) was also similar to that of zone II nodule sections. There was a consistently high extent of GANTC methylation across the chromosome averaging 0.870, which was less than that of bacteroids purified from *dnf5* nodules (0.956) but higher than that of bacteroids purified from wild-type A17 nodules (0.804) (**Table S8**), and without the smiling pattern. The nodule bacteria purified from *M. truncatula dnf7* and *dnf4* nodules also contained a mix of undifferentiated bacteria and fully differentiated bacteroids (**Figures 5D, 5E, 5J, 5K**), with the number of undifferentiated bacteria greater in *dnf7* nodules compared to *dnf4* nodules. The GANTC methylation pattern of bacteroids purified from *dnf7* and *dnf4* nodules was similar to that of bacteroids purified from A17 nodules (**Figures 5P-Q**). Overall, the data from bacteroids purified from section nodules and *M. truncatula dnf* mutant nodules suggest that CcrM is dysregulated during terminal bacteroid differentiation and with CcrM being constitutively active during endoreduplication.

Chromosome, pSymB, and pSymA sequencing depth are unequal in *E. meliloti* bacteroids

We noticed that in each bacteroid sample, the average extent of GANTC methylation for the chromosomes of the two strains were lower (by 0.04 to 0.13) than that of pSymA or pSymB, and unlike the chromosome, the extent of GANTC methylation was relatively constant across pSymA and pSymB (**Figures S14-S17** compared to **Figures 3 and 4**). These results suggest that, unlike in free-living cells, replication of the three replicons is not well coordinated during terminal differentiation. In agreement with this hypothesis, the mean sequencing depth across pSymA and pSymB was on average ~ 33% lower than that of the chromosome in all replicates of the *E. meliloti* whole-nodule bacteroid samples (**Table 2**). Similarly, the mean sequencing depth across pSymA and pSymB was on average ~ 23% lower than that of the chromosome for the polyploid bacterial cell populations purified from *M. truncatula dnf2*, *dnf7*, and *dnf4* mutant nodules, but not for the haploid/diploid bacterial cell populations purified from *M. truncatula dnf1* and *dnf5* mutant nodules (**Table 2**). Assuming sequencing depth is correlated with copy number, this observation suggests that *E. meliloti* bacteroids carry approximately two copies each of pSymA and pSymB per three copies of the chromosome.

DISCUSSION

In this study, we examined the genome-wide DNA methylation patterns in the free-living cells of four *Ensifer* strains, and in bacteroids of two *E. meliloti* strains, and detected a total of six methylated motifs. We were able to predict cognate MTases for most of the motifs based on genome annotations, the exception being the WNCCGATG motif of *E. adhaerens* OV14. The CGCA(N_5)GTG motif of *E. meliloti* Rm2011 is presumably methylated by Smc02296 (HsdM), a predicted m6A MTase belonging to the HsdRSM type I R-M system that is known to be functional and reduce transformation efficiency (72, 73). The RCGCCTC motif of *E. meliloti* Rm2011 is possibly methylated by Smc03763, a predicted cysteine-specific MTase located upstream of the gene *vsr* that putatively encodes a very short patch repair protein. Neither of these proteins are found in the other three strains examined here. The motifs TCGA(N_8)TCGA of *E. meliloti* FSM-MA and CAGA(N_7)GTTG of *E. fredii* NGR234 are likely methylated by SMB554_16155 and NGR_c01340, respectively, which are 88% identical at the amino acid level. Homologs of these two proteins are not found in the other two strains.

With the exception of GANTC, each methylated motif was detected as methylated only in

a single strain. Moreover, MTases are, with the exception of CcrM, not conserved among *E. meliloti* strains. The lack of conservation suggests that most DNA methylation does not have a major regulatory role in the genus *Ensifer*, aside from its role in cell cycle regulation. Supporting this conclusion, no motif was enriched in the promoter regions of symbiosis, carbon source, or cell cycle-regulated genes, and we did not detect any motifs that were methylated specifically in bacteroids. However, we cannot rule out that one or more methylated motifs may influence specific gene expression during free-living growth, differentiation, or N_2-fixation through extended motifs or proximity to other promoter elements, similar to the interplay between CcrM and GcrA during cell cycle regulation in *C. crescentus* (20, 74).

As previously published studies have provided evidence for a cell cycle transition occurring during terminal bacteroid differentiation (26), we were particularly interested in CcrM, a cell cycle-regulated MTase that is broadly conserved in the α-Proteobacteria, and its cognate DNA motif, GANTC. By identifying GANTC sites in the promoter regions of a previously determined set of 462 cell cycle-regulated genes (63), we defined a candidate CcrM regulon in *E. meliloti* consisting of 111 genes. However, further studies are required to better delineate the CcrM regulon in *E. meliloti* as the presence of a GANTC site is not diagnostic of CcrM regulation; GANTC sites were found in the promoter regions of 904 transcripts that did not display cell cycle regulation, and the promoter regions of cell cycle regulated genes were not enriched in GANTC sites relative to the whole *E. meliloti* genome. Studies in *C. crescentus* suggest that the impact of the fully or hemi-methylated status of GANTC sites on gene expression is mediated, at least in part, through modulating the activity of the transcriptional regulator GcrA (20, 74). However, not all promoter sites containing a GANTC motif are regulated by GcrA in *C. crescentus*, with the relationship dependent on an extended YGAKTCK motif and the precise position of this motif relative to other promoter elements (74). Likely, CcrM-mediated gene regulation in the genus *Ensifer* is similarly dependent upon additional sequence elements beyond the GANTC motif.

Consistent with past observations (18), GANTC sites were under-represented in the genomes of 157 *Ensifer* strains, particularly within coding regions. More surprising, however, was the strong difference in the frequency of GANTC sites between the previously defined (64) symbiotic and non-symbiotic clades in the genus *Ensifer*, with the frequency of GANTC sites being ~ 60% higher in the symbiotic clade. As methylation of the GANTC motif by CcrM is known to influence gene expression in *C. crescentus* (18), our observations suggest that CcrM has a

greater impact on modulating gene expression in the symbiotic clade compared to the non-symbiotic clade. Although further work is required to understand the biological significance of the greater frequency of GANTC sites in the symbiotic clade, it is tempting to speculate it is associated with legume symbiosis.

Our data is consistent with CcrM activity influencing terminal bacteroid differentiation. The overall moderate to high rates of GANTC methylation in all *E. meliloti* bacteroid samples, coupled with the lack of a chromosome-wide pattern in the *E. meliloti* FSM-MA zone II sample, suggests that CcrM remains constitutively active throughout most of terminal differentiation. This is supported by the results for nodule bacteria purified from *M. truncatula dnf* mutant nodules, which showed that differentiation is preceded by full GANTC methylation and that GANTC methylation remains high (but moderately lower) during endoreduplication followed by another moderate drop in GANTC methylation in late stages of differentiation. Considering that over-expression of CcrM can give rise to bacteroid-like morphology in free-living cells (14), we hypothesize that constitutive CcrM MTase activity is one (of potentially multiple) factor(s) driving polyploidization of bacteroids (**Figure S8**). CcrM activity is confined to a short window in the cell cycle since the *ccrM* gene is expressed in the late phase of genome replication (63) and the CcrM protein is degraded by the Lon protease prior to cell division (75). Thus, constitutive CcrM MTase activity in differentiating bacteroids could be obtained through an aberrant expression of the gene or alternatively a lack of proteolytic degradation of the CcrM protein. In agreement with the latter possibility, Lon protease was identified as a target of the NCR247 peptide (34). It is tempting to speculate that NCR peptides like NCR247 inhibit Lon protease activity, thereby stabilizing CcrM and trigger the bacteroid differentiation. However, the CcrM MTase does not appear to remain active in fully differentiated bacteroids, with the lower GANTC methylation near the chromosomal *ter* regions suggesting that loss of CcrM MTase activity occurs slightly prior to completion of genome endoreduplication (model provided as **Figure S8**). These hypotheses are consistent with *M. truncatula – E. meliloti* nodule zone-specific RNA-sequencing data (24), which showed that *ccrM* expression in the root distal portion of zone II is ~ 2-fold higher than in the root proximal portion of zone II, and ~ 10-fold higher than in zone III.

Our analyses also provide insight into the genome replication dynamics of *E. meliloti* during free-living growth and terminal bacteroid differentiation. Notably, flow cytometry data of *E. meliloti* bacteroids purified from zone II nodule sections and *M. truncatula dnf2* nodules suggest

that endoreduplication and cell enlargement largely occur subsequently, not concurrently. Genome replication might be a much faster process than cell growth or alternatively, endoreduplication might be required to drive cell enlargement. Moreover, our data are consistent with a loss of coordination of replication of the three replicons during terminal bacteroid differentiation, leading to unequal copy numbers with two copies of pSymA and pSymB per three copies of the chromosome in bacteroids. This relative change in replicon copy number occurs concomitantly with differentiation and polyploidization, as supported by the relative abundance of the replicons differing in nodule bacterial cells that have experienced endoreduplication (i.e., cells retrieved from *M. truncatula dnf2*, *dnf7*, and *dnf4* mutant nodules) but not in bacterial cells that had not yet undergone endoreduplication (i.e., cells retrieved from *M. truncatula dnf1* and *dnf5* mutant nodules). This differs from free-living cells, where copy number of the three replicons was approximately equal based on average sequencing depth. In contrast to our results, a previous comparison of the relative abundance of the three replicons in free-living *E. meliloti* Rm1021 versus bacteroids, no changes were detected using comparative genome hybridization with microarrays (26). We believe that the difference between our present data and the previous analysis is due to the subtlety of the differences and the lower sensitivity of the microarray hybridization method compared to high throughput sequencing.

We also observed that during free-living exponential growth, the extent of GANTC methylation at the *ori* of pSymA and pSymB is higher than at the *ori* of the chromosome, while the *ter* of the pSymA and pSymB has a slightly lower extent of GANTC methylation than the *ter* of the chromosome. As GANTC methylation occurs at a fixed stage of the cell cycle corresponding to the end of chromosome replication, our observations indicate that pSymA and pSymB replication is initiated later in the cell cycle than initiation of chromosome replication, while their replication terminates slightly before completion of chromosomal replication and the activation of CcrM (model provided as **Figure S8**). These results provide additional support for the notion of spatiotemporal regulation of DNA replication and partitioning in the multipartite *E. meliloti* genome as proposed previously (63, 76). Similarly, replication of chromosome II of *Vibrio cholerae* is delayed relative to chromosome I, leading to the replication of these two replicons terminating at approximately the same time (77). Thus, co-ordinating the timing or replication termination may be a general feature of multipartite genomes.

DATA AVAILABILITY

Most sequencing data is available through the JGI Genome Portal (genome.jgi.doe.gov/portal/) under Proposal 503835, as well as through NCBI (see **Table S4** for BioSample accessions). The data for stationary phase cultures and bacteroids isolated from *dnf* mutant nodules are available only through the NCBI (BioProject accessions PRJNA706182 and PRJNA705832; see **Table S4** for BioSample accessions). Flow cytometry FCS files are provided as **Dataset S2**. All custom scripts to perform the analyses described in this study are available through GitHub (github.com/diCenzo-Lab/003_2021_Ensifer_DNA_methylation).

FUNDING

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

ACKNOWLEDGMENTS

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

REFERENCES

1. Greenberg,M.V.C. and Bourc'his,D. (2019) The diverse roles of DNA methylation in mammalian development and disease. *Nat Rev Mol Cell Biol*, **20**, 590–607.
2. Zhang,H., Lang,Z. and Zhu,J.-K. (2018) Dynamics and function of DNA methylation in plants. *Nat Rev Mol Cell Biol*, **19**, 489–506.
3. Tang,Y., Gao,X.-D., Wang,Y., Yuan,B.-F. and Feng,Y.-Q. (2012) Widespread existence of cytosine methylation in yeast DNA measured by gas chromatography/mass spectrometry. *Anal. Chem.*, **84**, 7249–7255.
4. Blow,M.J., Clark,T.A., Daum,C.G., Deutschbauer,A.M., Fomenkov,A., Fries,R., Froula,J., Kang,D.D., Malmstrom,R.R., Morgan,R.D., *et al.* (2016) The epigenomic landscape of prokaryotes. *PLOS Genet*, **12**, e1005854.
5. Sánchez-Romero,M.A., Cota,I. and Casadesús,J. (2015) DNA methylation in bacteria: from the methyl group to the methylome. *Curr Opin Microbiol*, **25**, 9–16.
6. Gopalakrishnan,S., Van Emburgh,B.O. and Robertson,K.D. (2008) DNA methylation in development and human disease. *Mutat Res*, **647**, 30–38.
7. Vasu,K. and Nagaraja,V. (2013) Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiol Mol Biol Rev*, **77**, 53–72.
8. Atack,J.M., Tan,A., Bakaletz,L.O., Jennings,M.P. and Seib,K.L. (2018) Phasevarions of bacterial pathogens: methylomics sheds new light on old enemies. *Trends Microbiol*, **26**, 715–726.
9. Adhikari,S. and Curtis,P.D. (2016) DNA methyltransferases and epigenetic regulation in bacteria. *FEMS Microbiol Rev*, **40**, 575–591.
10. Campbell,J.L. and Kleckner,N. (1990) *E. coli oriC* and the *dnaA* gene promoter are sequestered from *dam* methyltransferase following the passage of the chromosomal replication fork. *Cell*, **62**, 967–979.
11. Kang,S., Lee,H., Han,J.S. and Hwang,D.S. (1999) Interaction of SeqA and Dam methylase on the hemimethylated origin of *Escherichia coli* chromosomal DNA replication. *J. Biol. Chem.*, **274**, 11463–11468.
12. Lahue,R., Au,K. and Modrich,P. (1989) DNA mismatch correction in a defined system. *Science*, **245**, 160–164.

13. Zweiger,G., Marczynski,G. and Shapiro,L. (1994) A *Caulobacter* DNA methyltransferase that functions only in the predivisional cell. *J Mol Biol*, **235**, 472–485.

14. Wright,R., Stephens,C. and Shapiro,L. (1997) The CcrM DNA methyltransferase is widespread in the alpha subdivision of proteobacteria, and its essential functions are conserved in *Rhizobium meliloti* and *Caulobacter crescentus*. *J Bacteriol*, **179**, 5869–5877.

15. Brilli,M., Fondi,M., Fani,R., Mengoni,A., Ferri,L., Bazzicalupo,M. and Biondi,E.G. (2010) The diversity and evolution of cell cycle regulation in alpha-proteobacteria: a comparative genomic analysis. *BMC Syst Biol*, **4**, 52.

16. Kahng,L.S. and Shapiro,L. (2001) The CcrM DNA methyltransferase of *Agrobacterium tumefaciens* Is essential, and Its activity is cell cycle regulated. *J. Bacteriol.*, **183**, 3065–3075.

17. Kozdon,J.B., Melfi,M.D., Luong,K., Clark,T.A., Boitano,M., Wang,S., Zhou,B., Gonzalez,D., Collier,J., Turner,S.W., *et al.* (2013) Global methylation state at base-pair resolution of the *Caulobacter* genome throughout the cell cycle. *Proc Natl Acad Sci USA*, **110**, E4658–E4667.

18. Gonzalez,D., Kozdon,J.B., McAdams,H.H., Shapiro,L. and Collier,J. (2014) The functions of DNA methylation by CcrM in *Caulobacter crescentus*: a global approach. *Nucleic Acids Res*, **42**, 3720–3735.

19. Gonzalez,D. and Collier,J. (2013) DNA methylation by CcrM activates the transcription of two genes required for the division of *Caulobacter crescentus*: regulation of cell division by DNA methylation. *Mol Microbiol*, **88**, 203–218.

20. Fioravanti,A., Fumeaux,C., Mohapatra,S.S., Bompard,C., Brilli,M., Frandi,A., Castric,V., Villeret,V., Viollier,P.H. and Biondi,E.G. (2013) DNA binding of the cell cycle transcriptional regulator GcrA depends on N6-adenosine methylation in *Caulobacter crescentus* and other Alphaproteobacteria. *PLOS Genet*, **9**, e1003541.

21. Wang,E.T. and Young,J.P.W. (2019) History of rhizobial taxonomy. In Wang,E.T., Tian,C.F., Chen,W.F., Young,J.P.W., Chen,W.X. (eds), *Ecology and Evolution of Rhizobia: Principles and Applications*. Springer, Singapore, pp. 23–39.

22. Oldroyd,G.E.D. (2013) Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nat Rev Microbiol*, **11**, 252–263.

23. Barnett,M.J., Toman,C.J., Fisher,R.F. and Long,S.R. (2004) A dual-genome Symbiosis Chip for coordinate study of signal exchange and development in a prokaryote–host interaction. *Proc Natl Acad Sci USA*, **101**, 16636–16641.

24. Roux,B., Rodde,N., Jardinaud,M.-F., Timmers,T., Sauviac,L., Cottret,L., Carrère,S., Sallet,E., Courcelle,E., Moreau,S., *et al.* (2014) An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. *Plant J*, **77**, 817–837.

25. diCenzo,G.C., Tesi,M., Pfau,T., Mengoni,A. and Fondi,M. (2020) Genome-scale metabolic reconstruction of the symbiosis between a leguminous plant and a nitrogen-fixing bacterium. *Nat Commun*, **11**, 2574.

26. Mergaert,P., Uchiumi,T., Alunni,B., Evanno,G., Cheron,A., Catrice,O., Mausset,A.-E., Barloy-Hubler,F., Galibert,F., Kondorosi,A., *et al.* (2006) Eukaryotic control on bacterial cell cycle and differentiation in the *Rhizobium*-legume symbiosis. *Proc Natl Acad Sci USA*, **103**, 5230–5235.

27. Czernic,P., Gully,D., Cartieaux,F., Moulin,L., Guefrachi,I., Patrel,D., Pierre,O., Fardoux,J., Chaintreuil,C., Nguyen,P., *et al.* (2015) Convergent evolution of endosymbiont differentiation in Dalbergioid and inverted repeat-lacking clade legumes mediated by nodule-specific cysteine-rich peptides. *Plant Physiol*, **169**, 1254–1265.

28. Pini,F., Frage,B., Ferri,L., De Nisco,N.J., Mohapatra,S.S., Taddei,L., Fioravanti,A., Dewitte,F., Galardini,M., Brilli,M., *et al.* (2013) The DivJ, CbrA and PleC system controls DivK phosphorylation and symbiosis in *Sinorhizobium meliloti*. *Mol Microbiol*, **90**, 54–71.

29. Kobayashi,H., De Nisco,N.J., Chien,P., Simmons,L.A. and Walker,G.C. (2009) *Sinorhizobium meliloti* CpdR1 is critical for co-ordinating cell cycle progression and the symbiotic chronic infection. *Mol Microbiol*, **73**, 586–600.

30. Pini,F., De Nisco,N.J., Ferri,L., Penterman,J., Fioravanti,A., Brilli,M., Mengoni,A., Bazzicalupo,M., Viollier,P.H., Walker,G.C., *et al.* (2015) Cell cycle control by the master regulator CtrA in *Sinorhizobium meliloti*. *PLOS Genet*, **11**, e1005232.

31. Gibson,K.E., Campbell,G.R., Lloret,J. and Walker,G.C. (2006) CbrA is a stationary-phase regulator of cell surface physiology and legume symbiosis in *Sinorhizobium meliloti*. *J Bacteriol*, **188**, 4508–4521.

32. Van de Velde,W., Zehirov,G., Szatmari,A., Debreczeny,M., Ishihara,H., Kevei,Z., Farkas,A., Mikulass,K., Nagy,A., Tiricz,H., *et al.* (2010) Plant peptides govern terminal differentiation of bacteria in symbiosis. *Science*, **327**, 1122–1126.

33. Penterman,J., Abo,R.P., De Nisco,N.J., Arnold,M.F.F., Longhi,R., Zanda,M. and Walker,G.C. (2014) Host plant peptides elicit a transcriptional response to control the *Sinorhizobium meliloti* cell cycle during symbiosis. *Proc Natl Acad Sci USA*, **111**, 3561–3566.

34. Farkas,A., Maróti,G., Dürgő,H., Györgypál,Z., Lima,R.M., Medzihradszky,K.F., Kereszt,A., Mergaert,P. and Kondorosi,E. (2014) *Medicago truncatula* symbiotic peptide NCR247 contributes to bacteroid differentiation through multiple mechanisms. *Proc Natl Acad Sci USA*, **111**, 5183–5188.

35. Satgé,C., Moreau,S., Sallet,E., Lefort,G., Auriac,M.-C., Remblière,C., Cottret,L., Gallardo,K., Noirot,C., Jardinaud,M.-F., *et al.* (2016) Reprogramming of DNA methylation is critical for nodule development in *Medicago truncatula*. *Nat Plants*, **2**, 16166.

36. Nagymihaly,M., Veluchamy,A., Györgypál,Z., Ariel,F., Jégu,T., Benhamed,M., Sz\Hucs,A., Kereszt,A., Mergaert,P. and Kondorosi,E. (2017) Ploidy-dependent changes in the epigenome of symbiotic cells correlate with specific patterns of gene expression. *Proc Natl Acad Sci USA*, **114**, 4543–4548.

37. Pecrix,Y., Staton,S.E., Sallet,E., Lelandais-Brière,C., Moreau,S., Carrère,S., Blein,T., Jardinaud,M.-F., Latrasse,D., Zouine,M., *et al.* (2018) Whole-genome landscape of *Medicago truncatula* symbiotic genes. *Nat Plants*, **4**, 1017–1025.

38. Davis-Richardson,A.G., Russell,J.T., Dias,R., McKinlay,A.J., Canepa,R., Fagen,J.R., Rusoff,K.T., Drew,J.C., Kolaczkowski,B., Emerich,D.W., *et al.* (2016) Integrating DNA methylation and gene expression data in the development of the soybean-*Bradyrhizobium* N_2-fixing symbiosis. *Front Microbiol*, **7**, 518.

39. Barrière,Q., Guefrachi,I., Gully,D., Lamouche,F., Pierre,O., Fardoux,J., Chaintreuil,C., Alunni,B., Timchenko,T., Giraud,E., *et al.* (2017) Integrated roles of BclA and DD-carboxypeptidase 1 in *Bradyrhizobium* differentiation within NCR-producing and NCR-lacking root nodules. *Sci Rep*, **7**, 9063.

40. Moreau,D., Voisin,A.-S., Salon,C. and Munier-Jolain,N. (2008) The model symbiotic association between *Medicago truncatula* cv. Jemalong and *Rhizobium meliloti* strain 2011

leads to N-stressed plants when symbiotic N2 fixation is the main N source for plant growth. *J Exp Bot*, **59**, 3509–3522.

41. Kazmierczak,T., Nagymihaly,M., Lamouche,F., Barrière,Q., Guefrachi,I., Alunni,B., Ouadghiri,M., Ibijbijen,J., Kondorosi,E., Mergaert,P., *et al.* (2017) Specific host-responsive associations between *Medicago truncatula* accessions and *Sinorhizobium strains*. *Mol Plant Microbe Interact*, **30**, 399–409.
42. diCenzo,G.C., MacLean,A.M., Milunovic,B., Golding,G.B. and Finan,T.M. (2014) Examination of prokaryotic multipartite genome evolution through experimental genome reduction. *PLOS Genet*, **10**, e1004742.
43. Krall,L., Wiedemann,U., Unsin,G., Weiss,S., Domke,N. and Baron,C. (2002) Detergent extraction identifies different VirB protein subassemblies of the type IV secretion machinery in the membranes of *Agrobacterium tumefaciens*. *Proc Natl Acad Sci USA*, **99**, 11405–11410.
44. Cowie,A., Cheng,J., Sibley,C.D., Fong,Y., Zaheer,R., Patten,C.L., Morton,R.M., Golding,G.B. and Finan,T.M. (2006) An integrated approach to functional genomics: construction of a novel reporter gene fusion library for *Sinorhizobium meliloti*. *Appl Environ Microbiol*, **72**, 7156–7167.
45. Starker,C.G., Parra-Colmenares,A.L., Smith,L., Mitra,R.M. and Long,S.R. (2006) Nitrogen fixation mutants of *Medicago truncatula* fail to support plant and bacterial symbiotic gene expression. *Plant Physiol*, **140**, 671–680.
46. Eid,J., Fehr,A., Gray,J., Luong,K., Lyle,J., Otto,G., Peluso,P., Rank,D., Baybayan,P., Bettman,B., *et al.* (2009) Real-time DNA sequencing from single polymerase molecules. *Science*, **323**, 133–138.
47. Chaisson,M.J. and Tesler,G. (2012) Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinform*, **13**, 238.
48. Flusberg,B.A., Webster,D.R., Lee,J.H., Travers,K.J., Olivares,E.C., Clark,T.A., Korlach,J. and Turner,S.W. (2010) Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods*, **7**, 461–465.
49. Clark,T.A., Murray,I.A., Morgan,R.D., Kislyuk,A.O., Spittle,K.E., Boitano,M., Fomenkov,A., Roberts,R.J. and Korlach,J. (2012) Characterization of DNA methyltransferase

specificities using single-molecule, real-time DNA sequencing. *Nucleic Acids Res*, **40**, e29–e29.

50. diCenzo,G.C., Debiec,K., Krzysztoforski,J., Uhrynowski,W., Mengoni,A., Fagorzi,C., Gorecki,A., Dziewit,L., Bajda,T., Rzepa,G., *et al.* (2018) Genomic and biotechnological characterization of the heavy-metal resistant, arsenic-oxidizing bacterium *Ensifer* sp. M14. *Genes*, **9**, 379.
51. Page,A.J., Cummins,C.A., Hunt,M., Wong,V.K., Reuter,S., Holden,M.T.G., Fookes,M., Falush,D., Keane,J.A. and Parkhill,J. (2015) Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*, **31**, 3691–3693.
52. Seemann,T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, **30**, 2068–2069.
53. Löytynoja,A. (2014) Phylogeny-aware alignment with PRANK. In Russell,D. (ed), *Methods in Molecular Biology (Methods and Protocols)*. Humana Press, Totowa, NJ, Vol. 1079, pp. 155–170.
54. Capella-Gutiérrez,S., Silla-Martínez,J.M. and Gabaldón,T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**, 1972–1973.
55. Stamatakis,A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.
56. Letunic,I. and Bork,P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res*, **44**, W242–W245.
57. Eddy,S.R. (2009) A new generation of homology search tools based on probabilistic inference. *Genome Inform*, **23**, 205–211.
58. Finn,R.D., Coggill,P., Eberhardt,R.Y., Eddy,S.R., Mistry,J., Mitchell,A.L., Potter,S.C., Punta,M., Qureshi,M., Sangrador-Vegas,A., *et al.* (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*, **44**, D279–D285.
59. Haft,D.H., Selengut,J.D., Richter,R.A., Harkins,D., Basu,M.K. and Beck,E. (2013) TIGRFAMs and genome properties in 2013. *Nucleic Acids Res.*, **41**, D387–D395.
60. diCenzo,G.C., Zamani,M., Checcucci,A., Fondi,M., Griffitts,J., Finan,T.M. and Mengoni,A. (2019) Multi-disciplinary approaches for studying rhizobium - legume symbioses. *Can J Microbiol*, **65**, 1–33.

61. diCenzo,G.C., Muhammed,Z., Østerås,M., O'Brien,S.A.P. and Finan,T.M. (2017) A key regulator of the glycolytic and gluconeogenic central metabolic pathways in *Sinorhizobium meliloti*. *Genetics*, **207**, 961–974.

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

63. De Nisco,N.J., Abo,R.P., Wu,C.M., Penterman,J. and Walker,G.C. (2014) Global analysis of cell cycle gene expression of the legume symbiont *Sinorhizobium meliloti*. *Proc Natl Acad Sci USA*, **111**, 3217–3224.

64. Fagorzi,C., Ilie,A., Decorosi,F., Cangioli,L., Viti,C., Mengoni,A. and diCenzo,G.C. (2020) Symbiotic and nonsymbiotic members of the genus *Ensifer* (syn. *Sinorhizobium*) are separated into two clades based on comparative genomics and high-throughput phenotyping. *Genome Biol Evol*, **12**, 2521–2534.

65. Vasse,J., De Billy,F., Camut,S. and Truchet,G. (1990) Correlation between ultrastructural differentiation of bacteroids and nitrogen fixation in alfalfa nodules. *J Bacteriol*, **172**, 4295–4306.

66. Wang,D., Griffitts,J., Starker,C., Fedorova,E., Limpens,E., Ivanov,S., Bisseling,T. and Long,S. (2010) A nodule-specific protein secretory pathway required for nitrogen-nixing symbiosis. *Science*, **327**, 1126–1129.

67. Bourcy,M., Brocard,L., Pislariu,C.I., Cosson,V., Mergaert,P., Tadege,M., Mysore,K.S., Udvardi,M.K., Gourion,B. and Ratet,P. (2013) *Medicago truncatula* DNF2 is a PI-PLC-XD-containing protein required for bacteroid persistence and prevention of nodule early senescence and defense-like reactions. *New Phytol*, **197**, 1250–1261.

68. Kim,M., Chen,Y., Xi,J., Waters,C., Chen,R. and Wang,D. (2015) An antimicrobial peptide essential for bacterial survival in the nitrogen-fixing symbiosis. *Proceedings of the National Academy of Sciences*, **112**, 15238–15243.

69. Domonkos,A., Horvath,B., Marsh,J.F., Halasz,G., Ayaydin,F., Oldroyd,G.E.D. and Kalo,P. (2013) The identification of novel loci required for appropriate nodule development in *Medicago truncatula*. *BMC Plant Biol*, **13**, 157.

70. Horváth,B., Domonkos,Á., Kereszt,A., Sz\Hucs,A., Ábrahám,E., Ayaydin,F., Bóka,K., Chen,Y., Chen,R., Murray,J.D., *et al.* (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.

71. Lang,C. and Long,S.R. (2015) Transcriptomic analysis of *Sinorhizobium meliloti* and *Medicago truncatula* symbiosis using nitrogen nixation–deficient nodules. *Mol Plant Microbe Interact*, **28**, 856–868.
72. Ferri,L., Gori,A., Biondi,E.G., Mengoni,A. and Bazzicalupo,M. (2010) Plasmid electroporation of *Sinorhizobium* strains: The role of the restriction gene *hsdR* in type strain Rm1021. *Plasmid*, **63**, 128–135.
73. Brumwell,S.L., MacLeod,M.R., Huang,T., Cochrane,R.R., Meaney,R.S., Zamani,M., Matysiakiewicz,O., Dan,K.N., Janakirama,P., Edgell,D.R., *et al.* (2019) Designer *Sinorhizobium meliloti* strains and multi-functional vectors enable direct inter-kingdom DNA transfer. *PLOS One*, **14**, e0206781.
74. Haakonsen,D.L., Yuan,A.H. and Laub,M.T. (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.
75. Wright,R., Stephens,C., Zweiger,G., Shapiro,L. and Alley,M.R. (1996) *Caulobacter* Lon protease has a critical role in cell-cycle control of DNA methylation. *Genes Dev*, **10**, 1532–1542.
76. Frage,B., Döhlemann,J., Robledo,M., Lucena,D., Sobetzko,P., Graumann,P.L. and Becker,A. (2016) Spatiotemporal choreography of chromosome and megaplasmids in the *Sinorhizobium meliloti* cell cycle. *Mol Microbiol*, **100**, 808–823.
77. Rasmussen,T., Jensen,R.B. and Skovgaard,O. (2007) The two chromosomes of *Vibrio cholerae* are initiated at different time points in the cell cycle. *EMBO J*, **26**, 3124–3131.

TABLE AND FIGURE LEGENDS

Table 1. Methylated motifs identified in this study.

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

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.

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 (64), 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.

Figure 3. Chromosome-wide DNA methylation of *E. meliloti* Rm2011 bacteroids. The extent of methylation of (**A**-**E**) GANTC, (**F**-**J**) CGCA(N_5)GTG, and (**K**-**O**) RCGCCTC motifs across the *E. meliloti* Rm2011 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**,**F**,**K**) Free-living cells harvested in mid-exponential phase. (**B**,**G**,**L**) Free-living cells harvested in early stationary phase. (**C**,**H**,**M**) Bacteroids isolated from *M. sativa* zone II nodule sections. (**D**,**I**,**N**) Bacteroids isolated from *M. sativa* zone III nodule sections. (**E**,**J**,**O**) Bacteroids isolated from *M. sativa* 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 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.

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**) *M. truncatula* A17 wild-type nodules. (**A**-**F**) Micrographs of *E. meliloti* FSM-MA bacteroids stained with the DNA binding dye DAPI. The scale bar represents 30 µm. (**G**-**L**) Pseudo-coloured scatterplots displaying the cell morphology (X-axis) and DNA content (Y-axis) of *E. meliloti* FSM-MA bacteroids, as determined based on flow cytometry analysis of DAPI stained cells. The red dashed ellipses indicate the position of undifferentiated bacteria as in culture (not shown) or in the *dnf1* mutant nodules (lower left ellipse) or fully mature bacteroids as in the A17 wild-type nodules (top right ellipse). (**M**-**X**) The extent of methylation of (**M**-**R**) GANTC or (**S**-**X**) TCGA(N_8)TCGA motifs across the *E. meliloti* FSM-MA chromosome, shown using a 10 kb

sliding window. 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 S19 and S20.

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

* The methylated nucleotides are indicated in boldface font.

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

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

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

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

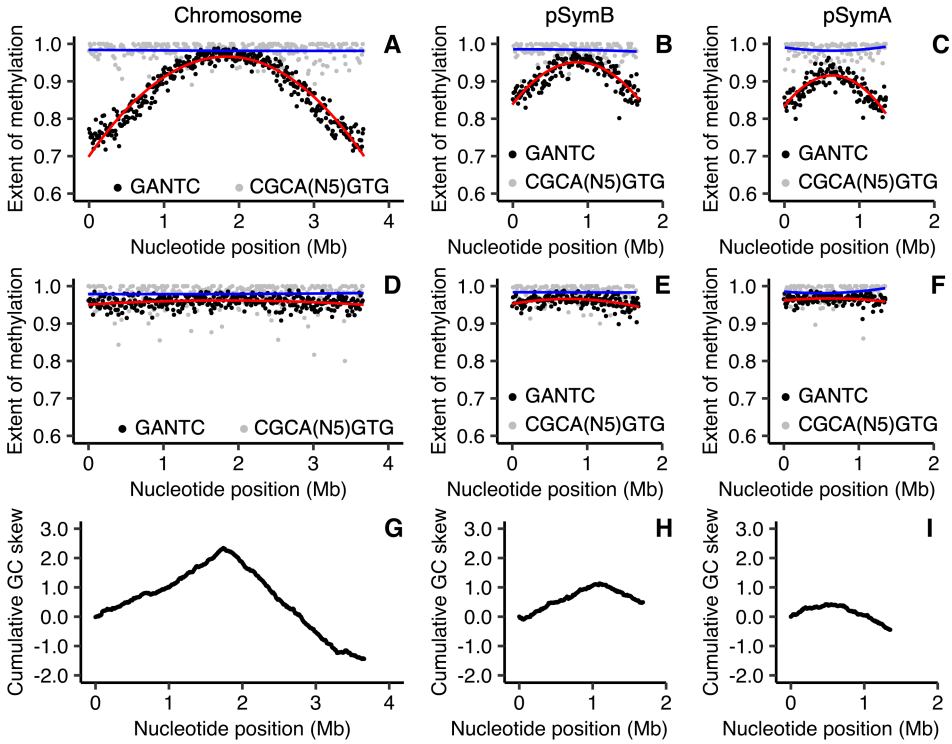

Figure 1. Genome-wide DNA methylation of *E. meliloti* Rm2011. (**A-F**) The extent of methylation is shown, using a 10 kb sliding window, of GANTC sites (black) and CGCA(N_5)GTG sites (grey) across the chromosome (**A**,**D**), pSymB (**B**,**E**), and pSymA (**C**,**F**) replicons of exponential phase (**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.

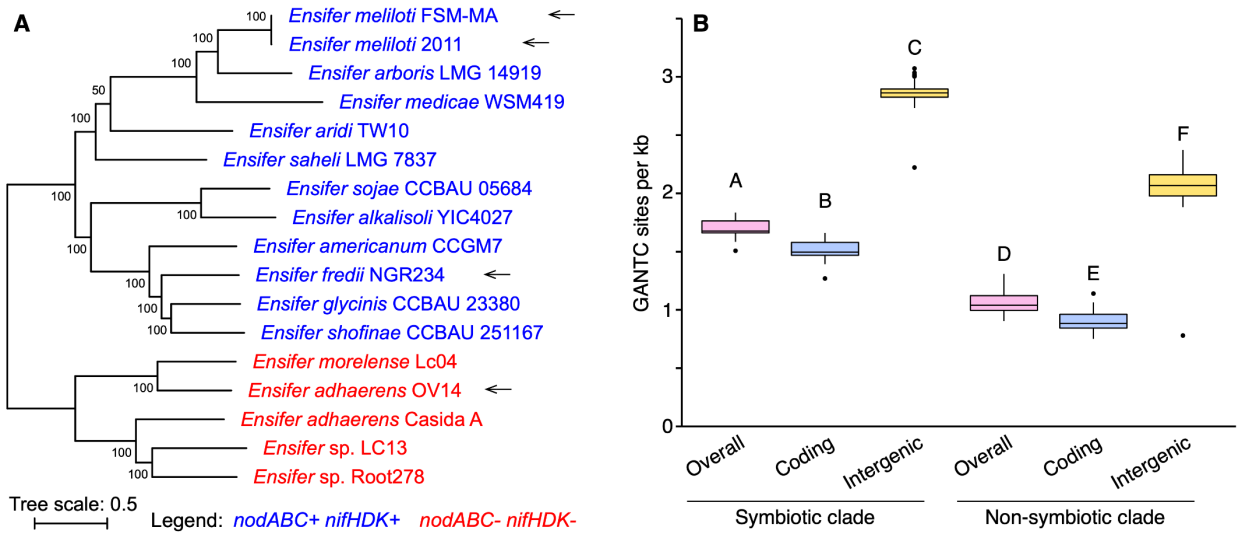

Figure 2. GANTC frequency in the genus *Ensifer*. (**A**) An unrooted maximum likelihood phylogeny of 17 representative *Ensifer* strains. The phylogeny represents the bootstrap best tree following 100 bootstrap replicates, prepared on the basis of the concatenated nucleotide alignments of 1566 core 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 (64), 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.

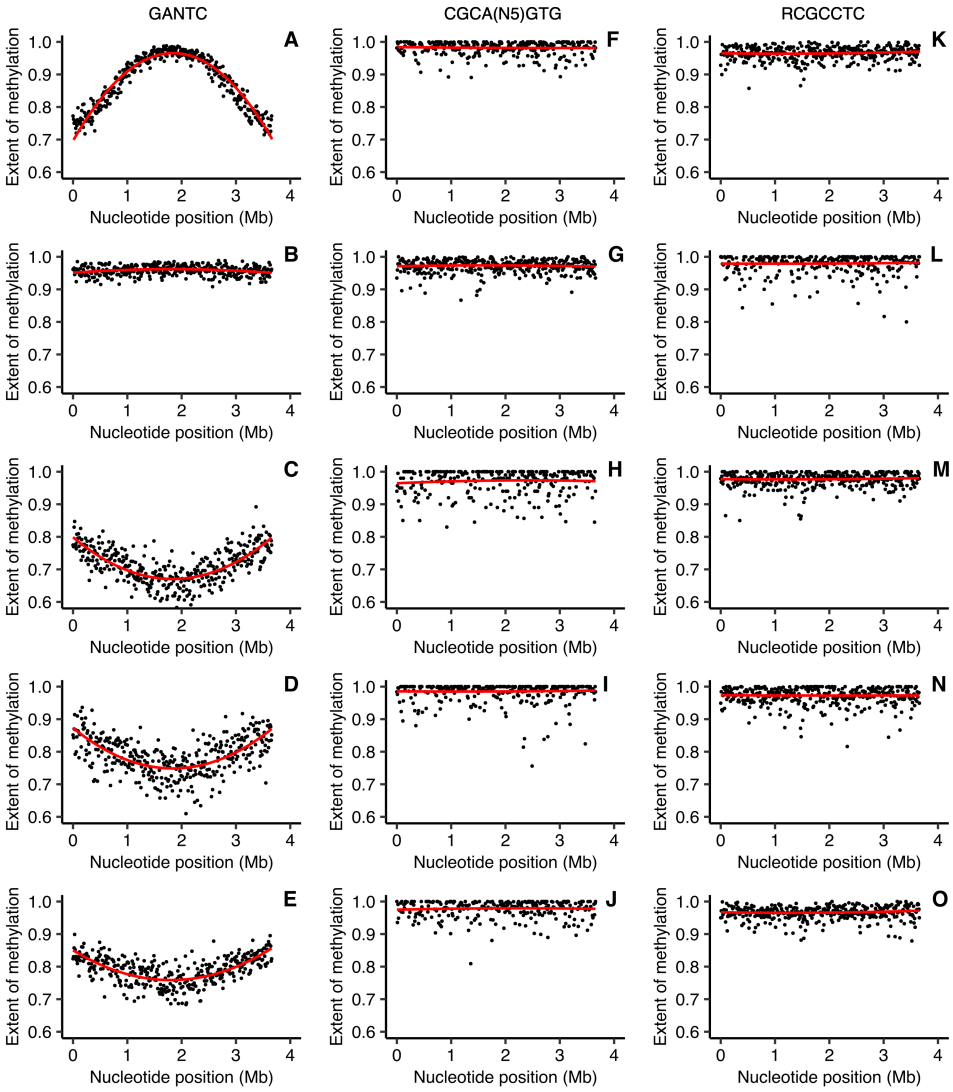

Figure 3. Chromosome-wide DNA methylation of *E. meliloti* Rm2011 bacteroids. The extent of methylation of (**A**-**E**) GANTC, (**F**-**J**) CGCA(N_5)GTG, and (**K**-**O**) RCGCCTC motifs across the *E. meliloti* Rm2011 chromosome is shown using a 10 kb sliding window. Averages from three biological replicates are shown for free-living and whole nodule samples; data represents one replicate for the zone II and zone III nodule sections. (**A**,**F**,**K**) Free-living cells harvested in mid-exponential phase. (**B**,**G**,**L**) Free-living cells harvested in early stationary phase. (**C**,**H**,**M**) Bacteroids isolated from *M. sativa* zone II nodule sections. (**D**,**I**,**N**) Bacteroids isolated from *M. sativa* zone III nodule sections. (**E**,**J**,**O**) Bacteroids isolated from *M. sativa* 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 S14 and S15.

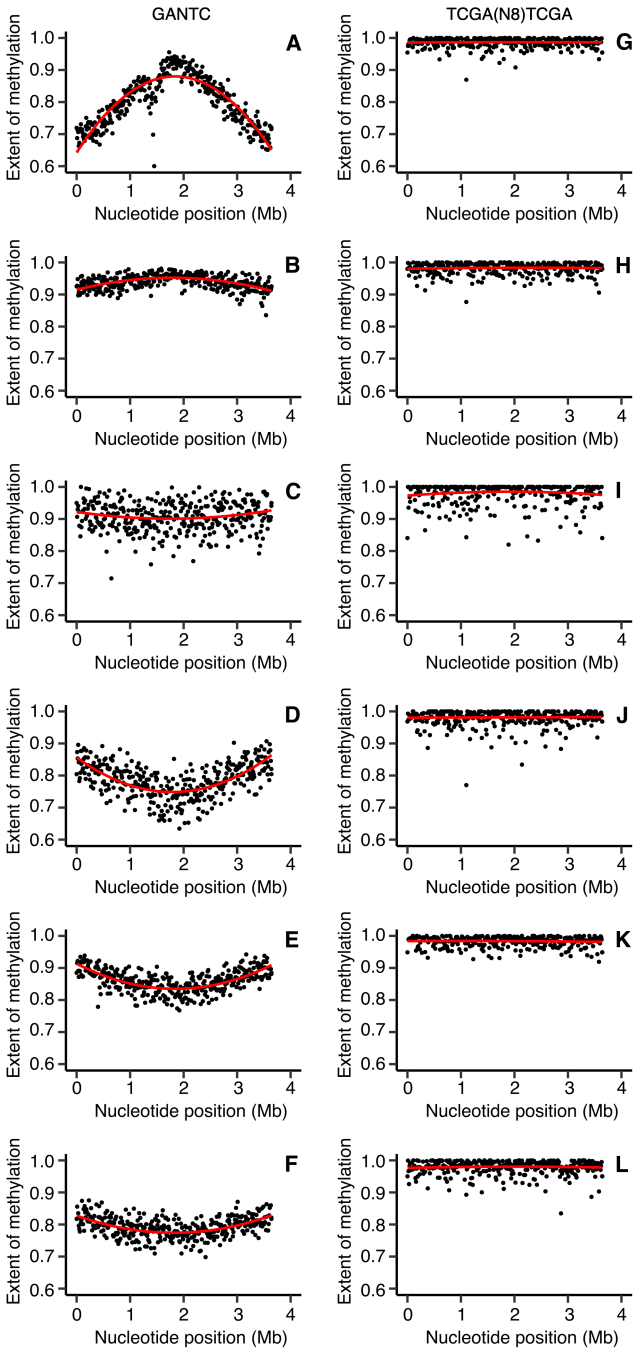

Figure 4. Chromosome-wide DNA methylation of *E. meliloti* FSM-MA bacteroids. The extent of methylation of (**A-F**) GANTC and (**G-L**) TCGA(N_8)TCGA motifs across the *E. meliloti* FSM-MA chromosome is shown using a 10 kb sliding window. Averages from three biological replicates are shown for free-living 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.

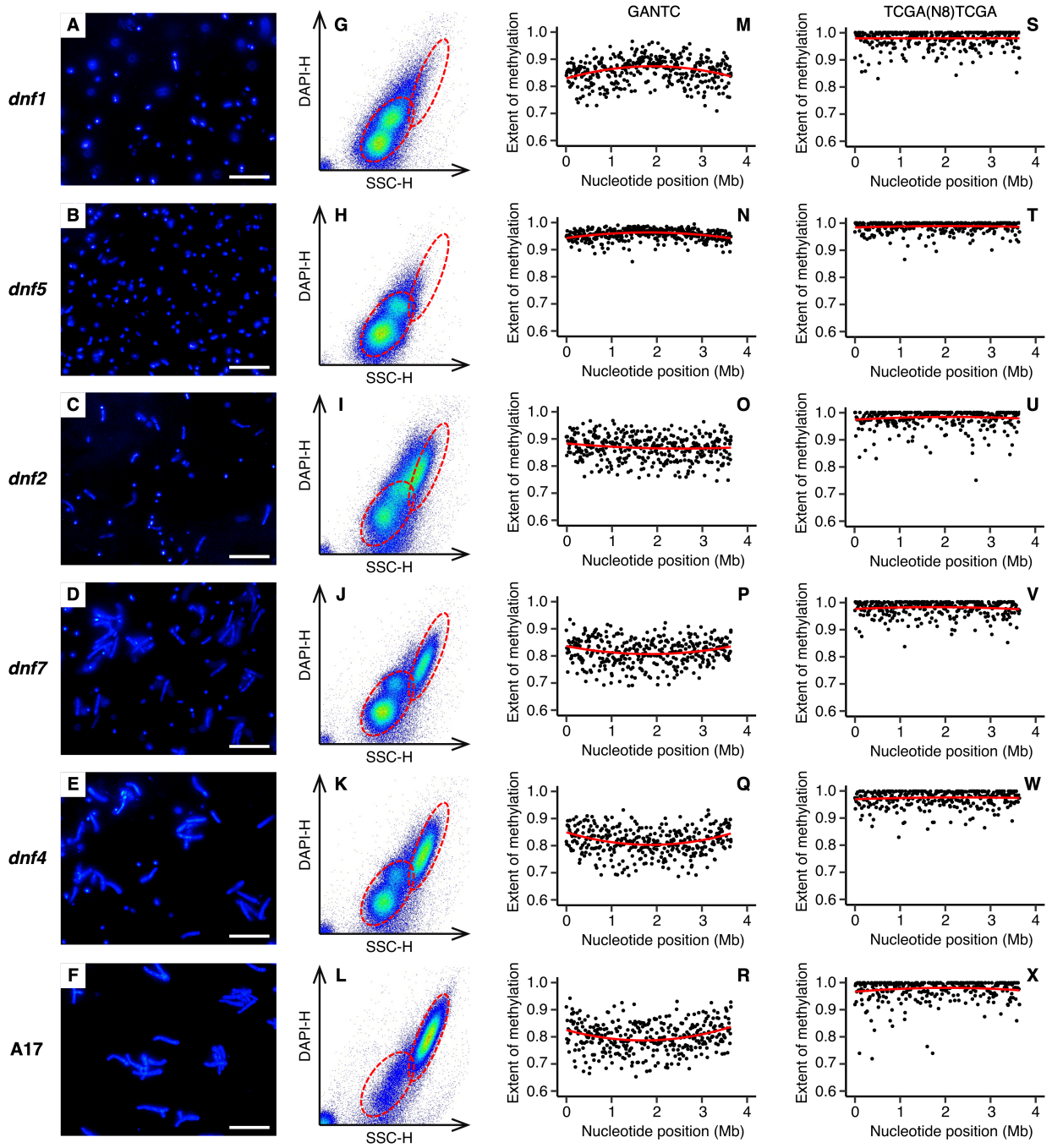

Figure 5. Bacteroid morphology and chromosomal GANTC methylation in *E. meliloti* bacteroids purified from *M. truncatula dnf* mutant nodules. Data is shown for *E. meliloti* FSM-MA bacteroids purified from (**A**,**G**,**M**,**S**) *M. truncatula dnf1* mutant nodules, (**B**,**H**,**N**,**T**) *M. truncatula 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**) *M. truncatula* A17 wild-type nodules. (**A-F**) Micrographs of *E. meliloti* FSM-MA bacteroids stained with the DNA binding dye DAPI. The scale bar represents 30 µm. (**G-L**) Pseudo-coloured scatterplots displaying the cell morphology (X-axis) and DNA content (Y-axis) of *E. meliloti* FSM-MA bacteroids, as determined based on flow cytometry analysis of DAPI stained cells. The

red dashed ellipses indicate the position of undifferentiated bacteria as in culture (not shown) or in the *dnf1* mutant nodules (lower left ellipse) or fully mature bacteroids as in the A17 wild-type nodules (top right ellipse). (**M**-**X**) The extent of methylation of (**M**-**R**) GANTC or (**S**-**X**) TCGA(N_8)TCGA motifs across the *E. meliloti* FSM-MA chromosome, shown using a 10 kb sliding window. 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 S19 and S20.