12102–12115 *Nucleic Acids Research, 2020, Vol. 48, No. 21* *Published online 9 November 2020*
doi: 10.1093/nar/gkaa966

4-Methylcytosine DNA modification is critical for global epigenetic regulation and virulence in the human pathogen *Leptospira interrogans*

Robert A. Gaultney[1], Antony T. Vincent[2], Céline Lorioux[1], Jean-Yves Coppée[3], Odile Sismeiro[3], Hugo Varet[3,4], Rachel Legendre[3,4], Charlotte A. Cockram[5], Frédéric J. Veyrier[2] and Mathieu Picardeau[1,*]

[1]Unité Biologie des Spirochètes, Institut Pasteur, Paris, France, [2]Bacterial Symbionts Evolution, INRS-Centre Armand-Frappier, Laval, Quebec, Canada, [3]Transcriptome and Epigenome Platform, Biomics, Center for Technological Resources and Research (C2RT), Institut Pasteur, Paris, France, [4]Bioinformatics and Biostatistics Hub, Department of Computational Biology, USR 3756 CNRS, Institut Pasteur, Paris, France and [5]Spatial Regulation of Genomes Unit, Institut Pasteur, Paris, France

Received August 21, 2020; Revised October 01, 2020; Editorial Decision October 07, 2020; Accepted October 13, 2020

ABSTRACT

In bacteria, DNA methylation can be facilitated by 'orphan' DNA methyltransferases lacking cognate restriction endonucleases, but whether and how these enzymes control key cellular processes are poorly understood. The effects of a specific modification, 4-methylcytosine (4mC), are even less clear, as this epigenetic marker is unique to bacteria and archaea, whereas the bulk of epigenetic research is currently performed on eukaryotes. Here, we characterize a 4mC methyltransferase from the understudied pathogen *Leptospira* spp. Inactivating this enzyme resulted in complete abrogation of CTAG motif methylation, leading to genome-wide dysregulation of gene expression. Mutants exhibited growth defects, decreased adhesion to host cells, higher susceptibility to LPS-targeting antibiotics, and, importantly, were no longer virulent in an acute infection model. Further investigation resulted in the discovery of at least one gene, that of an ECF sigma factor, whose transcription was altered in the methylase mutant and, subsequently, by mutation of the CTAG motifs in the promoter of the gene. The genes that comprise the regulon of this sigma factor were, accordingly, dysregulated in the methylase mutant and in a strain overexpressing the sigma factor. Our results highlight the importance of 4mC in *Leptospira* physiology, and suggest the same of other understudied species.

INTRODUCTION

DNA methylation plays an important role in the regulation of eukaryotic genomes, thus participating in many biological processes (1), but the extent to which similar processes exist in pathogenic prokaryotes remains comparatively understudied. Only a small number of bacterial species have been conclusively shown to require DNA methyltransferases (MTases) to maintain wild-type virulence (2). Three different forms of DNA methylation exist in bacterial genomes: N6-methyladenine (6mA)—by far the most prevalent form—, N4-methylcytosine (4mC), which is restricted to prokaryotes and archaea; and 5-methylcytosine (5mC), which is the dominant form in eukaryotes (3). Bacterial MTases are usually associated with restriction endonucleases, thereby forming restriction-modification (R-M) systems. In these systems, restriction enzymes (REases) target foreign DNA for cleavage, while DNA methylation of the restriction site protects the host genome from degradation (4). MTases unassociated with a cognate REase, known as 'orphan' MTases can also be found in bacterial genomes (5). The 6mA MTase Dam, which regulates DNA replication timing and gene expression in *Escherichia coli* (6), is the most studied bacterial orphan MTase. Recently, genome analyses have revealed that orphan MTases are distributed in half of all prokaryote genomes but their functions remain poorly understood (7).

The advent of single-molecule, real-time (SMRT) sequencing, first commercialized in 2011, has changed the view of epigenetic gene regulation in prokaryotes. This technology enables the analysis of the frequency and distribution of methylated residues in bacterial genomes (3,8–13). The results obtained thus far from the application of SMRT

*To whom correspondence should be addressed. Tel: +33 1 45 68 83 68; Fax: +33 1 40 61 30 01; Email: mpicard@pasteur.fr
Present address : Robert A. Gaultney, Institute of Pathology, University of Bern, Bern, Switzerland.

sequencing to a growing number of prokaryotes suggest that orphan MTases might play an important role apart from self-DNA protection. However, the influence of DNA methylation in virulence has, as of yet, only been investigated in a relatively small number of studies. Alterations in the levels of Dam attenuate the virulence of enterobacteria, including *Salmonella* spp., *Yersinia pseudotuberculosis* and *Vibrio cholerae* (14). More recently, it has been shown that inactivation of 4mC and 6mA MTases in *Helicobacter pylori* has an impact on functions important for virulence and competence (13,15) and a 5mC MTase mutant strain of *C. difficile* exhibited defects in sporulation (16).

In the present study, we functionally characterized the role of a 4mC orphan MTase in the pathogen *Leptospira interrogans*, the spirochete responsible for leptospirosis, an emerging zoonotic disease negatively affecting both domesticated animals and humans. Leptospirosis is responsible for over one million severe cases and 60 000 fatalities per year worldwide (17). Despite recent progress (18), the genus *Leptospira* remains understudied, and little is known about the ability of the pathogen to adapt to the host and cause disease. Furthermore, the role played by epigenetic factors in gene regulation has not been characterized at all in this genus. *Leptospira* is a diverse genus divided into pathogenic (P subclades) and saprophytic (S) species (19). While there are major observable differences between these two clades, the factors driving this diversity remain elusive. Other spirochetes include well-known pathogens such as the agents of syphilis (*Treponema pallidum*) and Lyme disease (*Borrelia burgdorferi*).

Here, we show that one orphan MTase is not ubiquitous in *Leptospira* but only present in the P clade (pathogenic species) of the genus. The sequence motif targeted by this MTase was verified via SMRT sequencing of a transposon insertion mutant (20). Transcriptomic and phenotype comparisons of wild-type (WT), mutant, and complemented strains demonstrated that the MTase acts as a global epigenetic regulator in *L. interrogans* modulating various phenotypes, including growth, motility, and virulence. This could further result in the dysregulation of a downstream alternative sigma factor whose proper control is also necessary for bacterial virulence. We also demonstrate that the sigma factor gene's transcription is also directly controlled by methylation of CTAG motifs located in its promoter and 5′ coding region. Overall, we show that gene regulation in *L. interrogans* is complicated by the presence of epigenetic modifications essential for the bacterial life cycle.

MATERIALS AND METHODS

Bacterial strains, constructs and culture conditions

All *in vitro* and *in vivo* experiments were performed using *L. interrogans* serovar Manilae strain UP-MMC-NIID LP. *Leptospira* were cultured under normal laboratory conditions at 30°C in Ellinghausen, McCullough, Johnson and Harris (EMJH) (21) medium unless otherwise noted. For growth curves, stationary phase bacteria were diluted 200-fold into a 30 ml EMJH in 50 ml flask with agitation at 100 rpms. Aliquots were removed at the listed timepoints and the absorbance at 420nm was measured daily with an Infinite M Nano spectrophotometer and i-control 2.0 software (Tecan) until strains reached stationary phase. For experiments requiring plating, concentrated EMJH was mixed with 1.2% (solid) or 0.6% (semisolid) final volume agar noble. For semisolid plating, a small divot was gouged into the agar surface into which 5μL of mid-log phase *Leptospira* in EMJH was pipetted. Expansion diameters of the resulting colonies were measured at 5- and 7-days post-inoculation. For temperature shift experiments, cells grown to mid-log phase were diluted 4-fold into fresh medium and incubated overnight at either 30 or 37°C. Cells were then processed as required for subsequent analyses.

Plasmid constructs

All plasmid constructs used in this study are based on the leptospiral replicative vector pMaOri (22) and pRAT724 (23). A list of all strains created and used in the course of these studies can be found in Supplementary Table S4, and the primers used in Supplementary Table S5. Four vectors possessing different leptospiral promoters (from the genes *cysK*, *flaB*, *groES* and *lipL32*) in front of FLAG-tagged Loa22 were created by gene synthesis (GeneArt, Sigma). Cloning into other vectors was performed using restriction endonuclease cloning or the Gibson assembly master mix kit (NEB) according to manufacturer instructions. Plasmid transformation was performed into *E. coli* Pi1 with plating on LB agar with supplemented spectinomycin and thymidine. To generate mutants in the *lomA* coding region and in the *sigX* promoter and coding region, we employed the Q5 site-directed mutagenesis kit (NEB) using the nebasechanger online tool to design primers. Plasmid was purified from selected colonies and sent for Sanger sequencing to verify proper insertion of target DNA. To move completed constructs into *L. interrogans*, the verified plasmids were transformed into *E. coli* β2163, which was then co-incubated on 0.1 micron filters with mid-log *L. interrogans* on EMJH plates supplemented with diaminopimelic acid (DAP) as previously described (24). To select for conjugated *Leptospira* clones, bacteria on filters were re-suspended in EMJH, and plated as described above. Individual colonies were selected and screened by PCR for proper insertion of desired plasmid constructs.

Translocation and cell binding experiments

Canine kidney epithelial MDCK-1 cells (Sigma) (25) were grown on Falcon cell culture inserts with a 3.0-micron pore size (Corning #353096) in EMEM supplemented with 10% FBS and L-glutamine (complete EMEM) until completely confluent. Mid-log phase *Leptospira* were counted and re-suspended at a concentration of 40 M/ml in a 2:1 mixture of EMJH:Complete EMEM. The cell culture medium was replaced with the same mixture, and a total of 10^6 spirochetes were added to each cell culture insert and incubated for 6 h at 37°C. After co-incubation, inserts were removed and transferred to a new plate and the liquid in the bottom chamber was aliquoted and stored at −20°C. The inserts were subsequently washed, top and bottom, three times with the EMJH:EMEM mixture for 5 min each. Cells were disassociated from the culture insert membrane by scraping with a pipette tip and stored in a small volume of PBS at

−20°C. To quantify the number of bacteria, we performed qPCR of the *flaB* gene (Supplementary Table S5) against a standard curve. For the cell-associated fraction, the bacterial abundance was normalized against the number of canine GAPDH genes detected (26).

Bacterial stress assays

We performed survival assays using the alamarBlue method previously described. (27) Briefly, 100 μl of the *Leptospira* strains in EMJH at a concentration of 5×10^5/ml were added to the wells containing 2-fold serial dilutions of polymyxin B (20μM maximum concentration), colistin (20 μM), H_2O_2 (4 mM), NaCl (500 mM), lysozyme (50 mg/ml), penicillin G (4 μg/ml) and Triton X-100 (0.1%) in EMJH. Bacteria were incubated in the plate for 24 h at 30°C, after which point 20 μl of alamarBlue reagent was added into each well. The plates were incubated another 48 h and the viability was calculated as per manufacturer's instructions using wells containing no bacteria as controls.

Chromosomal conformation capture

Leptospira were grown to mid-log phase in liquid EMJH as above. Protein−DNA interactions were chemically crosslinked with fresh formaldehyde (Sigma Aldrich; 3% final concentration) for 30 min at room temperature. Crosslinking was quenched by the addition of glycine (Sigma Aldrich; 360 mM final concentration) for 20 min at room temperature. The cross-linked cells were then centrifuged at 3200 rcf for 20 min at 4°C, then washed two times with 40 ml of EMJH. After the second wash, the pellets were moved to an Eppendorf tube and spun at 7000 rcf for 3 min at 4°C, and all remaining supernatant was removed. Pellets were aliquoted to 20 mg, flash frozen on dry ice, then stored at −80°C for future use.

3C experiments were performed as described previously with minor modifications (28). Briefly, cell pellets were resuspended in 1 ml of 1× TE + complete protease inhibitor cocktail (EDTA-free, Sigma Aldrich) and cells were then subject to mechanical disruption using the Precellys Evolution tissue homogenizer. Proteins that were not crosslinked to DNA during formaldehyde fixation were then degraded by the addition of SDS (Thermo Fisher, 0.5% final concentration) for 10 min at room temperature. DNA was then digested with HpaII (New England Biolabs, 1000 U) in 1× NEB1 buffer and 1% Triton-X-100 (ThermoFisher). The remaining lysate was centrifuged (16 000 × g, 20 min, room temperature) to pellet the insoluble fraction containing the protein−DNA complexes of interest. After removing the supernatant, the pellet was resuspended in dH_2O, ligation was performed with T4 DNA ligase (Thermo Fisher) for 4 h at 16°C, followed by incubation overnight at 65°C in presence of proteinase K (EuroBio) to reverse protein-DNA crosslinks. DNA was then purified by phenol−chloroform extraction and ethanol precipitation, pellets were resuspended in 1× TE buffer + 1 mg/ml RNase A and incubated for 30 min at 37°C. 5 μg of DNA was then sonicated using the Covaris S220 Focused Ultrasonicator to yield an average fragment size of 300 bp (140 W peak incidence power, 10% duty factor, 200 cycles per burst, 7°C). The DNA was purified and processed for sequencing according to manufacturer instructions (TruSeq™ Nano DNA High Throughput Library Prep Kit, illumina). All data was processed as previously (28,29) and the 3C fastq files have been deposited on the SRA database (BioProject ID PRJNA631225). Codes and functions used to generate the figures from the raw data are available online (https://github.com/koszullab/E_coli_analysis). Briefly, reads were aligned independently by Bowtie2 (30) using a very-sensitive and iterative approach. Each read was then assigned to a restriction fragment, with uninformative events, such as self-circularized and uncut fragments discarded by filtering. The filtered fragments were then binned into 5 kb segments and the corresponding contact maps generated and normalized using the sequential component normalization (SCN) procedure (29).

Immunoblotting

Leptospira strains were grown to mid-log phase in EMJH medium, harvested by centrifugation, re-suspended in 1× SDS-page sample loading buffer, and boiled for 5 min. Samples were run on an SDS-PAGE gel and transferred to a PVDF membrane. Mouse monoclonal anti-FLAG antibody (Sigma) was used to quantify the amount of protein of interest. As a loading control, we used polyclonal anti-FlaB antibodies from rabbit. The appropriate horse radish peroxidase-conjugated secondary antibody was used for each preparation. Membranes were blotted with SuperSignal west nano substrate and scanned with a Li-Cor C-Digit blot scanner.

Hamster infections

Protocols for animal experiments conformed to the guidelines of the Animal Care and Use Committees of the Institut Pasteur (Comité d'éthique d'expérimentation animale CETEA # 2016-0019), agreed by the French Ministry of Agriculture. Bacteria grown to mid-log phase were enumerated using a Petroff-Hausser counting chamber and diluted into EMJH medium to a final concentration of 2 million bacteria/ml. Bacterial viability in all cases was confirmed to be >95% by observing motility under the microscope. The bacterial suspension was injected intraperitoneally into 3–4-week-old Syrian Golden Hamsters (Janvier Labs). Hamster health was monitored daily and deaths were recorded. When necessary, hamsters were humanely euthanized by carbon dioxide inhalation if the animal was deemed to be excessively suffering. Organs were removed from selected animals for further culturing of *Leptospira*, or to determine tissue load at 4 DPI, and all strains were verified to have the same plasmid (when applicable) as the original infectious strain.

SMRT sequencing

Genomic DNA of *Leptospira* strains was extracted with the Genomic tip 100 g kit (Qiagen) from a 35 ml culture according to manufacturer protocols and the complete genome sequence was obtained using SMRT (Pacific Biosciences) technology. PacBio sequencing was performed at

the Génome Québec Innovation Centre (McGill University, Montreal, Canada) using a Pacific BioScience RS II system. The resulting reads of each strain were *de novo* assembled with Unicycler version 0.4.7 (31). The sequences of the replicons were then linearized with various tools of EMBOSS version 6.6.0.0 (32) and polished with the Resequencing protocol of SMRTLink version 6.0.0.47841 (https://www.pacb.com). The sequencing reads of the different data sets were subsequently mapped on the genome sequences of the WT strain and analyzed at the methylation level with the Base Modification and Motif Analysis protocol, also through SMRTLink with default parameters (95% confidence interval bounds, *P*-value cutoff = 0.001, minimum methylation fraction = 0.3 and minimum Qmod score = 30) (Supplementary Table S6). The CTAG motifs were identified in the genomic sequences of the WT strain using the fuzznuc tool from the EMBOSS suite. The unmethylated motifs were determined by subtracting the methylated motifs identified by SMRTLink from the set of all of the CTAG motifs found by fuzznuc. The reads were deposited in SRA under accession numbers SRR10257437, SRR10257436 and SRR10257435 for the *L. interrogans* wild-type, *lomA* mutant and complemented strains, respectively.

Identification of the target sites of MTases

RM systems and MTases were inferred based on the specificity sequences of homologs in the reference database REBASE (33).

RNA-Seq analysis

Exponential-phase cultures of *L. interrogans* were pelleted by centrifugation at 4000g for 20 min at ambient temperature, after which the medium was removed and cells were resuspended in 1 ml TRIZOL reagent (Thermo) and stored at −20°C until all samples were ready for phase-separation purification. Once all samples were at the same stage, the RNA was extracted according to manufacturer protocol, and excess DNA was removed using the Turbo DNA free kit (Thermo). RNA concentration and quality were assessed using the 2100 Bioanalyzer system with RNA 6000 nano reagents (Agilent).

From 0.5 μg of total RNA, we depleted ribosomal RNAs using the Ribo-Zero rRNA Removal Kit (Bacteria) from Illumina. On these rRNA-depleted RNAs, the sequencing libraries were constructed using the TruSeq Stranded mRNA Sample preparation kit following the manufacturer's instructions (Illumina). The directional libraries were controlled on Bioanalyzer DNA1000 Chips (Agilent Technologies) and concentrations measured with the Qubit® dsDNA HS Assay Kit (ThermoFisher). Sequences of 65 bases were generated on the Illumina Hiseq 2500 sequencer.

Bioinformatics analysis was performed using the RNA-Seq pipeline from Sequana (34). Reads were cleaned of adapter sequences and low-quality sequences using cutadapt version 1.11 (35). Only sequences at least 25 nt in length were considered for further analysis. Bowtie version 1.2.2 (36), with default parameters, was used for alignment on the reference genome (*L. interrogans* serovar Manilae strain UP-MMC-NIID LP from MicroScope Platform; https://mage.genoscope.cns.fr/microscope/home/index.php). Genes were counted using featureCounts version 1.4.6-p3 (37) from Subreads package (parameters: -t CDS -g locus_tag -s 1).

Count data were analyzed using R version 3.5.1 (38) and the Bioconductor package DESeq2 version 1.20.0 (39). The normalization and dispersion estimation were performed with DESeq2 using the default parameters and statistical tests for differential expression were performed applying the independent filtering algorithm. A generalized linear model was set in order to test for the differential expression between WT, complemented strain, and *lomA*. For each pairwise comparison, raw *P*-values were adjusted for multiple testing according to the Benjamini and Hochberg (BH) procedure (40) and genes with an adjusted *P*-value lower than 0.05 were considered differentially expressed (Supplementary Table S7).

The genes that are differentially expressed were grouped based on COG (Clusters of Orthologous Groups). For each COG, we used the Fisher statistical test to test for over- or under-representation in differentially expressed genes. Raw *P*-values were adjusted for multiple testing using the Benjamini–Hochberg procedure and COG sets with a false discovery rate (FDR) lower than 5% were considered enriched.

The dataset is available under GEO accession number GSE138917.

Quantitative reverse transcriptase PCR (RT-qPCR)

RNA samples were prepared as for transcriptomic analyses. After the successful purification of RNA and elimination of contaminating DNA, RNA was converted into cDNA using the iScript Advance cDNA synthesis kit (BioRad) according to manufacturer's direction using 2500 ng of RNA per sample. For RT-qPCR *flaB* was used as the reference gene (See Supplementary Table S5). This gene was verified to be unchanged in the transcriptomic data as well as empirically in the lab. Samples were prepared with SsoFast EvaGreen Supermix (Biorad) and thermal cycling and quantitation was performed using a C1000 thermal cycler with attached CFX96 real-time system (Biorad). Results were obtained and analyzed using the CFX manager 3.1 software.

Fluorescence measurements

Bacteria were prepared as previously described (23) with some minor modifications. Briefly, bacteria were grown to mid-log phase as described above, and an aliquot of 2 ml was removed and centrifuged at 6500g for 10 min at room temperature. Bacteria were then washed once with room temperature PBS (Sigma, D1408), and re-pelleted as above. The bacteria were then re-suspended in 400 μl PBS and the absorbance of each tube was measured at 420 nm (as above), after which the bacterial concentrations in each solution were normalized, and 100 μl aliquots were plated into an appropriate 96-well plate for fluorescence analysis with the same model of spectrophotometer used for absorbance analyses with the preset program for GFP detection in cells (top-reading) and presets for optimal z-position

and gain. All data presented are of individual isolates of *Leptospira* for each construct created.

Statistics

Statistical analyses were performed using the suite of tools included in the Prism software used to create figures (Graphpad) unless otherwise specified in the methods. Student's *t*-tests or ANOVA with post-hoc multiple comparisons were used as appropriate using default parameters. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$. Error bars shown on graphs indicate standard error of the mean (SEM).

RESULTS

Methyltransferases in *L. interrogans*

SMRT sequencing analysis of the *Leptospira interrogans* serovar Manilae strain UP-MMC-NIID genome (41) indicated that mainly four motifs are methylated in this species (Figure 1A). There is one MTase (LIMLP_04455) associated with a Type I R-M system and one (LIMLP_04495) with a Type II system. Three other MTases not pertaining to R-M systems (orphans) are also present in the genome (Supplementary Table S1). Bioinformatic predictions (REBASE) suggest that these three putative orphan MTases are active to respectively methylate the different motifs detected with SMRT. One (LIMLP_06480) is ubiquitous in *Leptospira* and is predicted to recognize 5′-GTAC-3′ (methylated bases are underlined). The two other MTases LIMLP_11030 and LIMLP_03780 are only found in the pathogenic species, and predicted to methylate the motifs 5′-CTAG-3′ and 5′-GTATAC-3′, respectively (Supplementary Table S1). Of note, the motif 5′- CAGATC -3′ remains without assigned MTase but is probably methylated by a MTase associated with a corresponding REase given the 100% methylation of target sites (Figure 1A).

Previously, we created a library of random transposon mutants (20) in the aforementioned strain, which is highly pathogenic and used to investigate leptospiral virulence in animal models (19,27,41–45). A complete loss of methylation at the CTAG sites was observed in a LIMLP_11030 transposon mutant by SMRT sequencing, and was restored upon addition of the MTase *in trans* via plasmid (Figure 1A). In contrast, inactivation of LIMLP_11030 did not result in loss of methylation at the other methylated motifs, further indicating that the LIMLP_11030 gene product specifically recognizes the CTAG motif and no other MTase serves a redundant function in this bacterium (Figure 1A). In the wild-type *L. interrogans* 80% of the 12682 CTAG motifs in the genome are methylated, supporting the hypothesis that the MTase is an orphan, as canonical R-M system MTase typically methylate nearly 100% of target sequences to prevent self-DNA digestion. Interestingly, the MTase shares a 45% sequence identity with a MTase, MthZIM, from the thermophilic archaea *Methanobacter thermoautotriphicus* (46) (formerly *M. thermoformicicum* (47)). In this species, the *mthZIM* gene is localized with its corresponding REase, MthZIR, an enzyme absent from *L. interrogans*. There is no information about the roles of MthZIM homologs in other prokaryotic species other than the assumed protection against targeted DNA cleavage. In light of these data, we propose naming the *L. interrogans* gene and corresponding product Leptospiral Orphan Methylase A (*lomA*).

LomA is only present in the P clade (pathogenic *Leptospira* species)

LomA is highly conserved in all pathogenic (subclade P1) and intermediate (subclade P2) species (Figure 1B and Supplementary Table S1), with a lowest protein sequence similarity of 68% between the most dissimilar alleles. Neither LomA, nor a homologue, exist in any of the known S clade (saprophytic) species. This is supported by the SMRT sequencing of two saprophytic species, which had no modification on CTAG motifs (Figure 1C).

The number of CTAG motifs is highly variable among the members of subclade P1. The three species most commonly implicated in severe human infections, *L. interrogans*, *L. kirschneri* and *L. noguchii*, have a large number of CTAGs in their genomes relative to other members of the P1 clade (Figure 1B). All of these data, taken together, led us to speculate that LomA may be important for the virulence of the pathogenic species of *L. interrogans*.

The motif recognized by LomA, CTAG, appears to have a relatively even distribution across the genome with no particular over or under-represented regions, complicating prediction of the role of the MTase in absence of other data (Figure 2A). Of note, some regions seem to harbor average numbers of CTAG motifs, but decreased proportions of methylated sites. This can be observed in Figure 2A (pink and purple lines such as those seen at approx. 220 and 3260 kb—enhanced zoom in Supplementary Figure S1). Alternatively, some regions have a noticeable decrease in their total CTAG density (yellow line, Figure 2A).

Inactivation of LomA attenuates virulence

Leptospira lacking LomA, when grown under ideal laboratory conditions, exhibited a marked delay in exit from lag phase but the mutant attained a final cell density similar to that of WT (Figure 3A). The *lomA* mutant appears motile in liquid medium, but when exponential phase *lomA* was plated on semisolid EMJH agar plates, the zone of bacterial expansion was decreased relative to WT, suggesting a possible defect in bacterial motility in viscous medium as well (Figure 3B). *lomA* also displayed a ∼2-fold decrease in adherence to polarized MDCK cells grown on a porous membrane but was equally able to transverse the same cells as the WT strain (Figure 3C). Bacteria lacking the MTase are no longer able to kill hamsters (Figure 3D), however, they were observed by darkfield microscopy in tissue slurries prepared from hamster kidneys up to 14 days post-infection. At an earlier timepoint, 4 days post-infection, the tissue load of *lomA* in both kidneys and livers was drastically decreased when compared to WT (Figure 3E). Additional *in vitro* experiments such as resistance assays to oxidative and osmotic stresses and detergents did not show any altered phenotype in the *lomA* mutant but a significant decrease in cell viability of >50% was observed at 2.5 μg/ml polymyxin B in the *lomA* mutant (Figure 3F). Similar results were seen for colistin, another LPS-binding antibiotic (Supplementary Figure S2). Importantly, supplying *lomA in trans* restored the

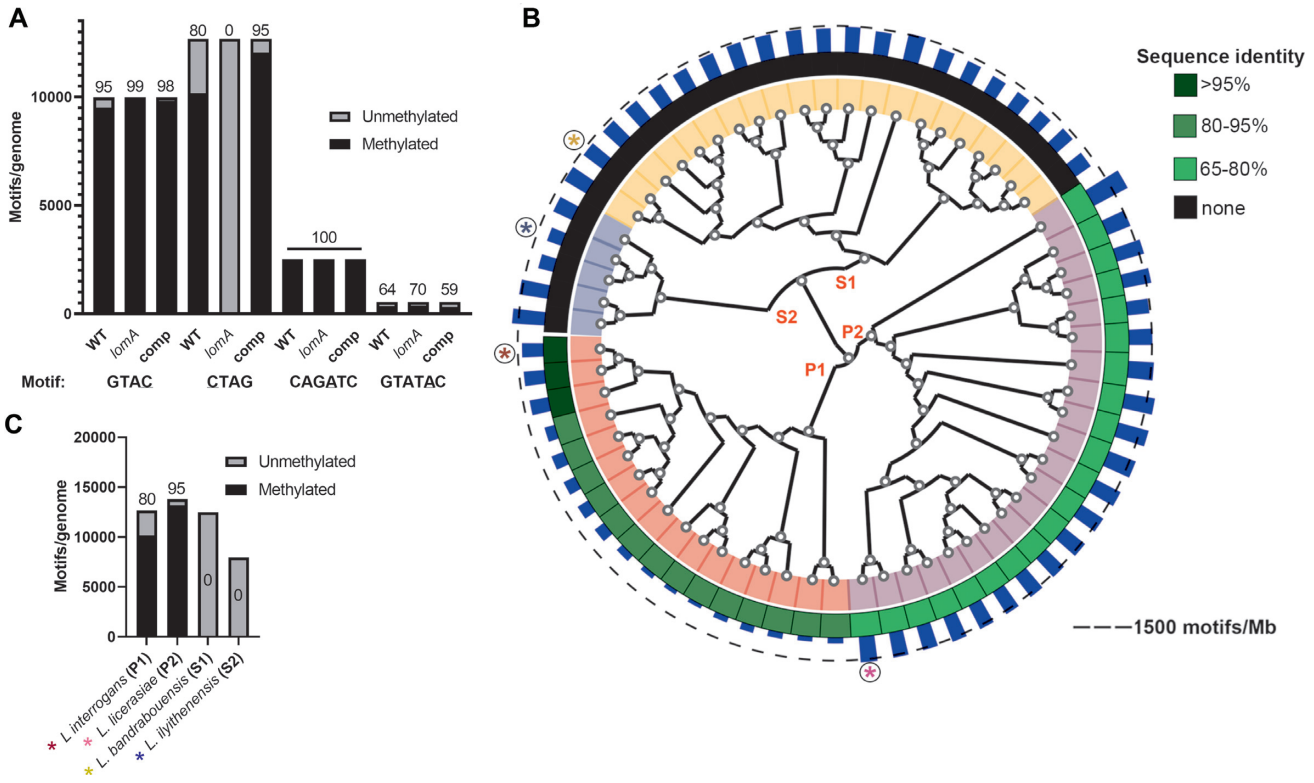

Figure 1. *L. interrogans* contains several DNA methylases and LomA is conserved among pathogenic *Leptospira*. (**A**) Methylation patterns of the WT, *lomA* mutant and complemented strains for the identified methylated motifs: GTAC, CTAG, CAGATC, and GTATAC. (**B**) Distribution of LIMLP_11030/LomA homologs in the genus *Leptospira* (the subclades P1 (light red inner circle) and P2 (violet) contain the infectious species and S1 (yellow) and S2 (light blue), saprophytes) and the number of CTAG sites in the genome of each representative species relative to the whole genome size. The dashed line corresponds to the average of 1500 motifs per Mb. The asterisks in circles correspond to the species whose genomes were analyzed for 4mCTAG content (**C**), including the species used in this study (*L. interrogans* -maroon asterisk). In (A) and (C), the numbers over each bar correspond to the percentage of methylated sites.

WT phenotype in all assays used, confirming the direct role that LomA plays in these observed phenomena (Figure 3).

To support the hypothesis that the phenotype of the *lomA* mutant is due to the absence of 4mC methylation, rather than to another non-canonical role of LomA, we re-introduced a *lomA* allele with a single-residue substitution mutation into the *lomA* strain. 4mC MTases share a relatively conserved TSPPY S-adenosyl methionine-binding N-terminal motif (48) that we converted into TAPPY in the mutant allele *lomA*S30A. As expected, the catabolically inactive LomA allele, when introduced *in trans*, was unable to complement the growth, semisolid agar expansion, and, most importantly, virulence phenotypes observed in the mutant (Supplementary Figure S3a–c). This mutant also enabled us to verify the putative toxicity of the MTase in *L. biflexa*—a species not adapted to 4mCTAG DNA methylation. Wild-type LomA is unable to transconjugate into the saprophyte *L. biflexa*, but when the *lomA*S30A mutant allele was provided on a conjugative plasmid, we were able to obtain colonies (Supplementary Figure S3d). Methylation of CTAG motifs in this unhabituated organism is evidently toxic. All together, these data support the hypothesis that LomA plays an important role in the regulation of bacterial transcription through its ability to methylate DNA.

Global alteration of the transcriptome in the *lomA* mutant

To study the functional significance of methylation at CTAG sites in pathogenic *Leptospira* and to understand the mechanisms underlying the various observed phenotypes, we used RNA-Seq to compare the transcriptomes of wild-type *L. interrogans* with that of the *lomA* mutant and complemented strains under ideal culture conditions at mid-log phase (Supplementary Table S2 and Supplementary Figure S4). Real-time qPCR was performed on all genes with an absolute change of 2-fold or greater to validate the RNA-Seq data (Supplementary Figure S5). The vast majority of genes examined by RT-qPCR were in accordance with the RNA-Seq results (Supplementary Table S2).

To investigate the correlation between differential methylation and differential gene expression, only genes that were significantly changed in both the WT and complemented strains vs. *lomA* were identified as positive to eliminate any confounding variables. We, thus, found 193 genes that were significantly and differentially expressed in the *lomA* mutant, including 20 genes with more than two-fold differences (adjusted P-value < 0.05) (Supplementary Table S2, Figure 2a and Supplementary Figure S4). To determine if a bias existed toward genes with products involved in similar biological functions, deregulated genes were categorized into COGs and the resulting frequencies were compared to the

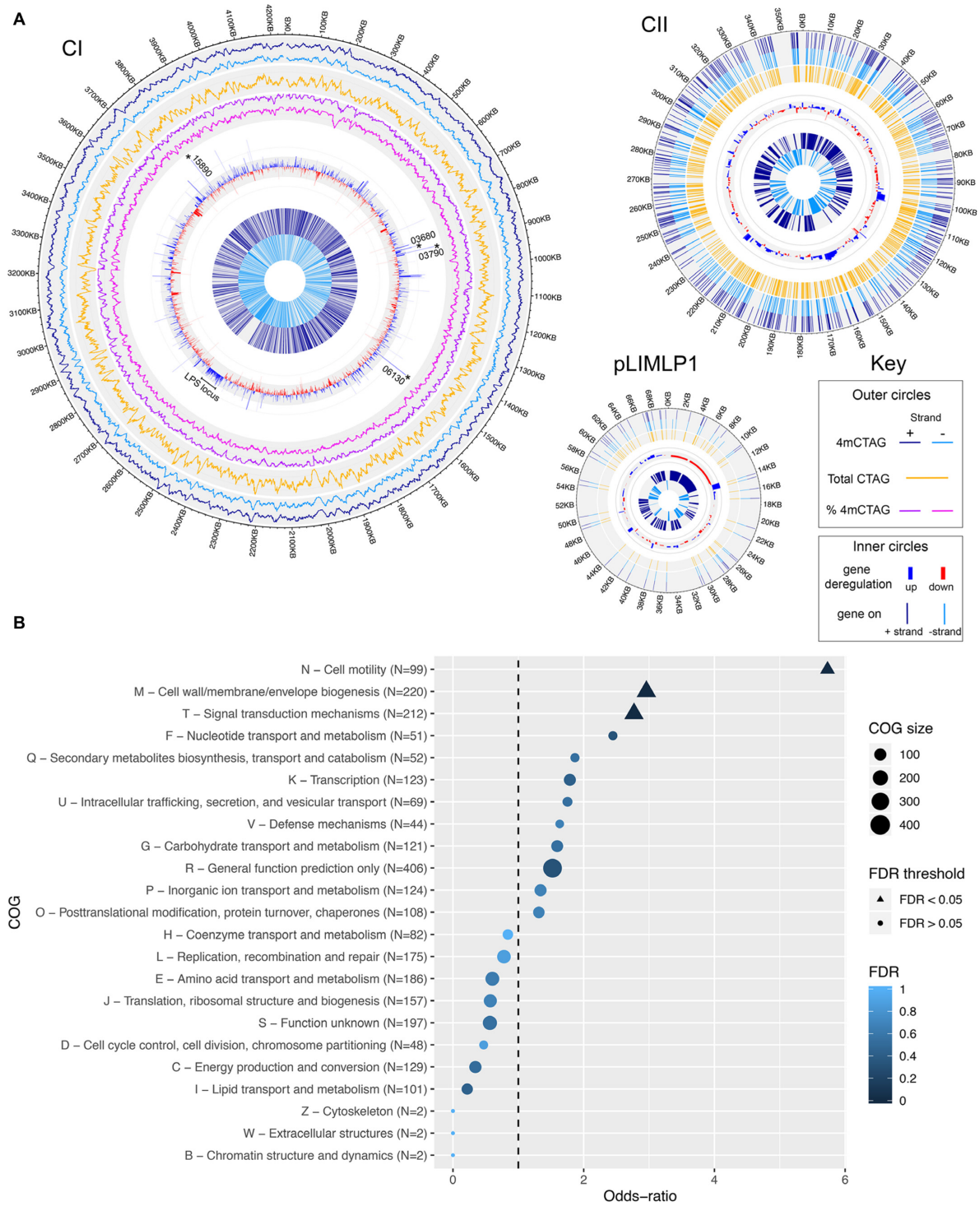

Figure 2. Comparison of the distribution of CTAG motifs, 4mC methylation and differential gene expression in *L. interrogans*. (**A**) Schematic representation of chromosomes I and II, as well as plasmid pLIMLP1 of *L. interrogans* serovar Manilae strain UP-MMC-NIID LP. The two chromosomes and large plasmid are represented, and a window size of 10kb and a step size of 100bp were used for chromosome I, consequently, the methylated CTAG site density, total CTAG density, and relative abundance of 4mCTAG are shown (see key for details). For the smaller chromosome II and pLIMLP1, the individual methylated and total CTAG sites are visible, but no relative abundance could be calculated. The blue and red bars indicate, respectively, the up and down-regulated genes in the *lomA* mutant the interior grey shaded area corresponds to a $\log_2$ fold change of 1. For legibility, the cutoff value used was $\log_2 FC = 2$, higher dysregulation values, therefore present as this value. The most highly upregulated genes, along with an ECF sigma factor, are indicated by asterisks and their locus tags. The innermost circles indicate the presence of an annotated coding sequence on the ± strands (see key, lower box). (**B**) Classification of deregulated genes according to their predicted functions. The selected genes ($n = 193$) have a dysregulation with an adjusted P-value < 0.05 in both the *lomA* mutant versus WT and *lomA* mutant versus complemented strain, and they are not differentially expressed in the complemented strain vs WT (adjusted P-value > 0.05). The number of genes that are differentially expressed are grouped based on COG (Clusters of Orthologous Groups). An odds-ratio (OR) > 1 corresponds to an over-representation in differentially expressed genes and an OR < 1 to an under-representation.

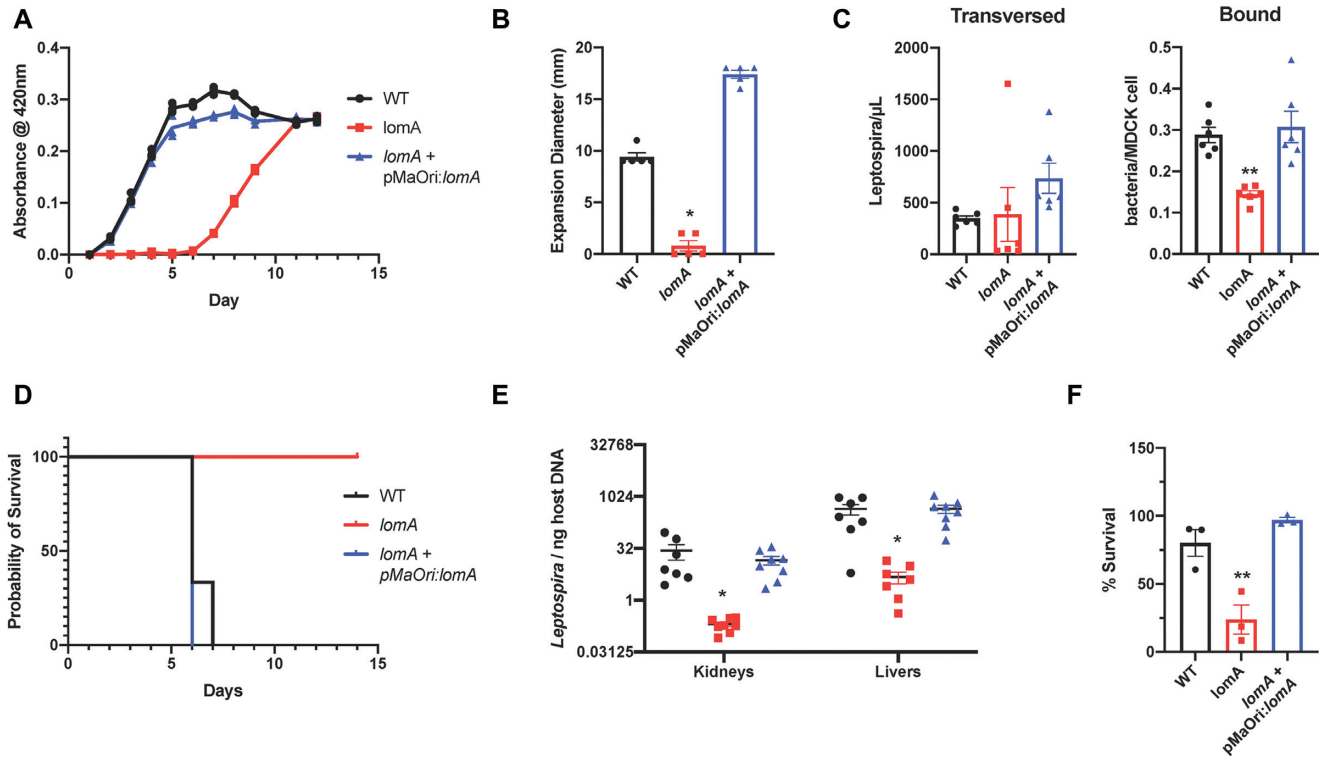

Figure 3. LomA inactivation results in multiple deleterious phenotypes. (**A**) Growth curve in liquid EMJH at 30°C shows a growth defect for the *lomA* mutant in comparison to the wild-type and complemented strains ($n = 3$). (**B**) Colony expansion in semisolid EMJH ($n = 3$). Soft agar plates were inoculated with *L. interrogans* strains and the colony diameter was measured. Inactivation of *lomA* results in decreased motility through the semi-solid medium. (**C**) The quantities of bacteria able to either bind or translocate through polarized MDCK I cells were measured by qPCR. Inactivation of *lomA* prevents wild-type binding to MDCK I cells but the *lomA* mutant does not differ in translocation ($n = 6$ wells). (**D**) Hamsters were challenged with *Leptospira* by intraperitoneal infection with 10^6 leptospires ($n = 3$). Survival curves of infected hamsters show that the *lomA* mutant is attenuated in virulence. (**E**) At D4 post-infection hamster kidneys and livers were harvested for quantification of *Leptospira* tissue burden. The *lomA* mutant was present in lower quantities in both organs ($n = 7$ hamsters for WT and 8 for *lomA* and complemented strains, one *lomA*-infected kidney had a bacterial burden below the detectable limit). (**F**) Cells were incubated with 2.5 µg/ml polymyxin B and cell viability was assessed by their capacity to reduce blue resazurin into pink resorufin data indicated are the results of three independent plates collated ($n = 3$). Inactivation of LomA results in >50% decrease in cell viability when challenged with polymyxin B.

genome-wide predicted frequencies. This analysis revealed an enrichment of dysregulation in genes coding for proteins with roles in signal transduction, cell envelope biogenesis (including LPS), and motility (Figure 2B and Supplementary Figure S4 for the list of genes regulated by a LomA-dependent mechanism) which is consistent with the phenotypes observed in the mutant (Figure 3).

To evaluate whether the methylation of CTAG sites by LomA could play a role in gene regulation, the frequency of CTAG sites and their methylation status was examined in the promoter regions and within coding sequences (Figure 2A). We further compared the frequency of 4mC groups found in each gene and said gene's dysregulation in the *lomA* mutant (Supplementary Figure S6). Despite our efforts, we were not able to find a direct and statistically significant link between CTAG site positions, their methylation status, and gene expression when examining the entire bacterial genome. In addition, upon closer examination of the three genes with the highest change in expression (FC>15; LIMLP_06130, LIMLP_15890 and LIMLP_03790), none had a CTAG site in the 300 bp preceding the coding region (Supplementary Figure S7). Two genes had a CTAG motif in the coding sequence but one gene, LIMLP_03790, had no CTAG motifs in the entire coding sequence (Supplementary Figure S7).

The genome architecture of *lomA*

As the CTAG motifs appear to be unevenly spaced throughout the genome, we considered the possibility that absence of methylation at these sites might lead to re-arrangement of the two chromosomes of *L. interrogans*. To test this, we performed chromosomal conformation capture (3C) (28,49) on exponentially growing cultures of WT, *lomA*, and complemented strains. These contact maps show that, in the absence of LomA, there is an overall loss of mid-range contacts throughout both chromosomes. This disorganization was partially restored in the complemented strain. When we examined genomic areas with exceptional gene dysregulation, such as the LPS locus, we observed no differences in these regions compared to the rest of the chromosome (Supplementary Figure S8). These data suggest that the chromosomal alterations seen in the *lomA* mutant are unlikely to cause the specific alterations of transcription in these cells.

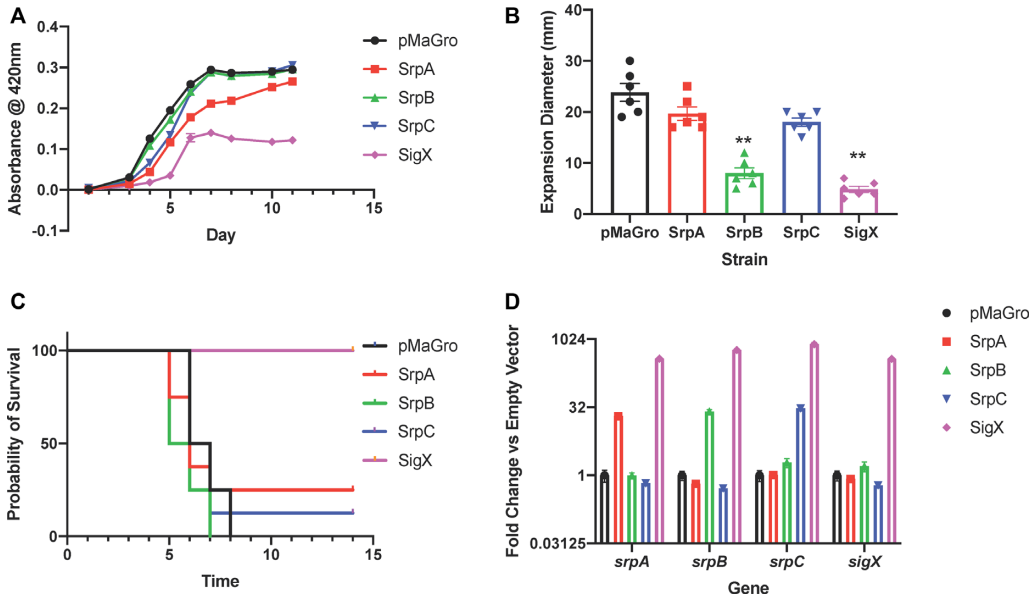

Figure 4. σ^X overproduction is physiologically deleterious. Phenotypes of strains overexpressing *srpA, srpB, srpB* and *sigX* in a wild-type background. An empty vector, pMaGro, was used as a control. (**A**) Growth curve demonstrating similar growth rates for the control, *srpA*, *srpB* and *srpC* overexpressing strains and growth defect for the *sigX* strain. The strains were grown in liquid EMJH at 30°C ($n = 3$) (**B**) Expansion in semisolid EMJH. Soft agar plates were inoculated with *L. interrogans* strains and the colony diameter was measured. The average colony diameter measured for the *sigX and srpB* overexpressing strains was significantly smaller than for the wild-type (empty) and *srpA* and *srpC* overexpressing strains ($n = 6$). (**C**) Survival curve of hamsters infected intraperitoneally with 10^6 leptospires ($n = 8$) showing 100% survival for animals infected with the *sigX* overexpressing strain and survival rates statistically equal to the empty vector control in the other three lipoprotein overexpression mutants. (**D**) RT-qPCR results for the σ^X regulon comparing the *srp* and *sigX* transcript levels in strains overexpressing each gene reveal that σ^X controls the expression of the three *srp* genes as well as its own but the Srp proteins do not affect any of the other genes. Data were obtained from the RNA of three independent cultures from each strain.

An ECF sigma factor up-regulated in a *lomA* mutant controls the most highly dysregulated genes and also affects virulence

As none of the experiments thus far demonstrated a direct link between gene dysregulation and loss of virulence, we decided to examine the potential effects on bacterial physiology of the three genes most highly upregulated in the *lomA* mutant plus one other gene that we selected for its possible importance. Accordingly, we generated a set of strains where these genes were cloned into plasmid under a strong constitutive promoter in WT *L. interrogans* to mimic their overexpression. We first undertook a selection of the optimal promoter by cloning several known promoters from *L. interrogans* in front of the lipoprotein Loa22 tagged with a FLAG epitope into a replicative vector and analyzing protein abundance by immunoblot (Supplementary Figure S9). The selected promoter with the strongest activity was that of *groES*. In total, four genes were selected for overexpression: LIMLP_15890 (fold change: 46), LIMLP_03790 (FC: 43), LIMLP_06130 (FC: 16) and LIMLP_03680 (FC: 5). The most highly expressed genes (LIMLP_06130 and two paralogous genes LIMLP_15890 and LIMLP_03790) are all putative lipoproteins of unknown function. LIMLP_03680 encodes a putative extracytoplasmic function (ECF) sigma factor, which, as expected, is in an operon with its cognate trans-membrane anti-sigma factor (LIMLP_03685). For clarity and ease of reading for the manuscript, we have decided to re-name this gene *sigX*, the anti-sigma factor *rsx,* and the three putative lipoprotein-encoding genes as SigX regulated proteins for reasons that will become clear further on in the text (LIMLP_06130: *srpA*; LIMLP_15890: *srpB*; LIMLP_03790: *srpC*). Notably, these five genes are only present in the pathogenic species of *Leptospira* (19) (Supplementary Table S3).

The four strains—each over-expressing one of the genes—were next subjected to several phenotypic assays. A slight growth defect was measured in the SrpA overexpression mutant, but it was not as dramatic as the *lomA* mutant (Figure 4A). Furthermore, the σ^X overexpression mutant was unable to reach the same density at stationary phase as the strain harboring an empty vector control, which differs from the delayed exit from lag phase seen in the *lomA* mutant. One strain, the SrpB overexpression mutant, also showed a marked decrease in expansion in semisolid EMJH (Figure 4B), however this strain did not possess a difference in growth rate in EMJH medium. In hamsters, only the σ^X-overexpressing strain displayed an attenuation of virulence, no other single construct was able to nullify virulence in the hamster model (Figure 4C). The phenotype in the pMaGro:*sigX* strain could be explained by the exaggerated increase of *sigX* mRNA as observed using RT-qPCR. In the strain possessing this construct, the *sigX* mRNA is present at levels several hundred-fold above WT which is clearly higher than the *lomA* mutant (6-fold change over WT) (Figure 4D). The Srp constructs exhibit significant increases in expression of their cognate genes between 20- and 30-fold over the empty vector, relatively equivalent to the results seen in the *lomA* mutant. Altogether, these results suggest an autoregulatory function for σ^X; a hypothesis supported by the fact that overexpression of the sigma factor also increases the expression of the downstream anti-sigma factor

(Figure 4D) even though this regulator was not included in the overexpression plasmid.

Importantly, in our σ^X-overexpressing strain, all three *srp* genes appear to be controlled by σ^X—hence the selected naming convention—as overexpression of the regulator resulted in massively increased levels of each gene (Figure 4D). This result explains why these genes are highly deregulated in the *lomA* mutant despite the lack of CTAG motifs in the upstream or promoter regions. The same was not true of the other dysregulated genes in the *lomA* mutant, as the overexpression of the ECF sigma factor did not alter the expression of other genes previously identified with LomA-dependent regulation (Supplementary Figure S10). A search for a target promoter sequence in *sigX*, *srpA*, *srpB* and *srpC* enabled the identification of a consensus sequence: TGA/GCAC – N(15/16) – TCGTT in the –10/–35 region (50) of the four genes (Supplementary Figure S11) (51) in agreement with the canonical requirements of ECF-dependent promoters: optimal spacer length of 14–17 bp and a 'CGT' motif in the –10 region (52). *In silico* analyses showed that this consensus sequence is only found at the appropriate position in the promoter region of these four genes, further suggesting an autoregulation of the *sigX* gene as well as a directly controlled expression of genes encoding lipoproteins (*srpA*, *srpB* and *srpC*) by σ^X but not the other LomA-regulated genes (Supplementary Figure S11).

We further tested if the expression of *sigX* is directly regulated by LomA-dependent methylation or indirectly in the *lomA* mutant. For this, we constructed a reporter plasmid in which a CTAG motif found in the promoter region near the putative transcriptional start site and one in the upstream coding region of *sigX* were mutated so the motifs could no longer be methylated (Figure 5A). The mutation in the coding sequence is silent. The resulting allelic constructs containing of the first 150 bp of the *sigX* gene were translationally fused to GFP (23), cloned into the pathogenic *Leptospira* vector pRAT724, and conjugated into wild-type and *lomA L. interrogans*. The resulting fluorescence was measured to compare background expression of the *sigX* gene with or without methyl groups (Figure 5B). Modification of the CTAG in the promoter but not sequence resulted in an statistically significant increase in gene expression in wild type *L. interrogans* (but not in the *lomA* mutant where the levels are similar between all constructs), suggesting that the 4mCTAG modifications at this site are inhibitory, which is in accordance with the transcriptomic data from the *lomA* mutant. Mutation of both CTAG sites did not result in an additive effect, and both sites were identified as methylated under the conditions used for SMRT-seq (Supplementary Figure S7, Supplementary Table S6).

DISCUSSION

The introduction of SMRT sequencing technology has made elucidation of bacterial methylomes easier than ever before; however, the prevalence of 4mC MTases in bacterial genomes is not well understood, and its biological function largely understudied (3,7). Although one other study on *H. pylori* showed a defect in some determinants of pathogenicity *in vitro* in a 4mC methylase mutant, no *in vivo* virulence model has shown a requirement for 4mC methylation in any bacterium using molecular Koch's postulates (53) until now.

The large number of differentially expressed genes in the *lomA* mutant suggests that the phenotype seen in this mutant of *L. interrogans* is likely mediated by multiple genes that are either directly regulated by LomA or indirectly regulated by a transcriptional cascade. Motility in spirochetes is essential for virulence and non-motile mutants in *L. interrogans* no longer cause disease and are quickly cleared from the tissues of infected hosts (44,54,55). While both the *lomA* mutant and σ^X overexpression strain were both deficient in motility in soft agar and avirulent, another bacterial strain overexpressing *srpC* also displayed a statistically significant decrease in motility, but was, nevertheless, able to kill hamster hosts as effectively as the empty vector control. The motility defects observed could simply be an artifact of the slowed growth of the *lomA* and σ^X-overexpressing strains, even though the bacteria were plated during log phase. Furthermore, the fact that *lomA* maintained the ability to translocate across cell monolayers with very high trans-epithelial resistance (25) suggests that the motility differences seen in soft agar are not sufficient to completely attenuate virulence, even though motility genes were preferentially deregulated in the *lomA* mutant.

Although we were initially unable to show a correlation between methylation motif density and gene dysregulation in the *lomA* mutant probably due to confounding factors (such as regulation cascades), *sigX* proved a promising target for further investigation. The three genes in the known regulon of σ^X had no CTAG motifs in their putative promoters (up to 400 bp upstream of start codon), which is less than would be expected by random chance. Conversely, the *sigX* gene contains two CTAG motifs, one in the promoter, and another in the 5′ coding region. Our experiments into the direct regulation of gene transcription showed that 4mC MTases are capable of regulating the transcription of specific genes. To our knowledge, these are the first such experiments verifying this hypothesis. We also postulate that the up-regulation of the *sigX* gene may be further dysregulated by the perturbations of the membrane compartment. These dysbioses could result in the release and activation of σ^X, as is typical of ECF sigma factors (Figure 5C). Our analysis of the functional gene categories most affected transcriptionally supports this hypothesis, as we see that COGs representing proteins involved in signal transduction mechanisms and cell envelope biogenesis are both disproportionately overrepresented. This phenomenon was also observed recently in a 5mC methylase mutant of *Vibrio cholerae* (56).

Another factor potentially influencing the expression of the *sigX* gene could be the CTAG motif density in the surrounding genes. Thus, the binding of the RNA polymerase and/or regulatory proteins may be affected by the methylation of sites in the promoter region as previously (57). However, the chromosomal contact maps generated herein do not support this hypothesis as genomic loci with higher CTAG densities do not appear to preferentially mis-fold, including the *sigX* locus itself. While the 4mCTAG residues may aide in proper folding of the chromosome generally, this phenomenon does not appear to be locus-specific. The DNA misfolding itself is also particularly mild and non-specific when compared to more severe phenotypes such as

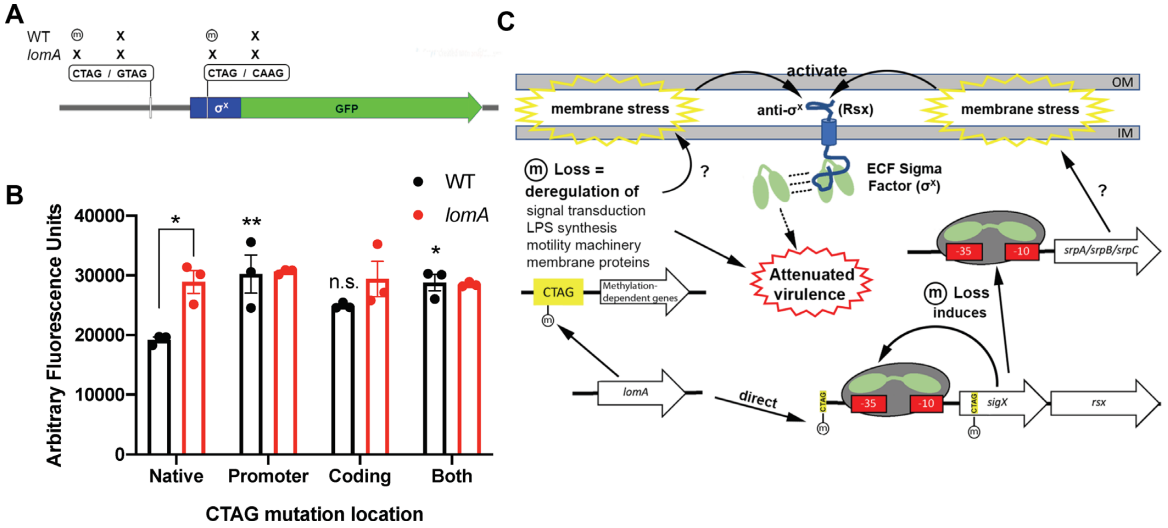

Figure 5. LomA facilitates transcriptional regulation of *sigX* in *L. interrogans*. (**A**) Graphical representation of the *sigX-gfp* translation fusion and site-directed mutants created in this study indicating the placement of CTAG motifs in the promoter and upstream coding sequence of *sigX*. (**B**) Fluorescence of the *sigX-gfp* constructs in the WT and *lomA* background strains. Altering the promoter or both CTAG motifs in the WT strain resulted in increased fluorescence, but the same mutation had no effect in the *lomA* mutant ($n = 3$ independent isolates for each construct). *P*-value is provided versus native WT results. Results are indicative of at least three independent plates with similar results. (**C**) Model detailing the role of LomA and the 4mCTAG modifications it is responsible for in the phenotypes observed in this study. Our fluorescence experiments suggest a direct role in the regulation of the *sigX* gene, which further activates transcription of the srp lipoprotein family. other phenotypes are likely dependent on other mechanisms, such as those listed. Membrane stresses potentially caused by the other phenotypes and the σ^X-controlled *srp* lipoprotein family could further result in the exacerbation of the σ^X response.

that seen in an *E. coli matP* mutant, for example (28). In light of these data, we remain skeptical that rearrangements of DNA structure play a significant role in the transcriptional dysregulation found in the *lomA* mutant.

ECF sigma factors have, likewise, been characterized as essential for bacterial virulence (58). In some pathogens, inactivation of ECF sigma factors is deleterious for bacterial survival, and overexpression of the regulators results in increased bacterial virulence (59–61). In this particular instance, however, it appears that the opposite is true. It should be taken into consideration that *Leptospira* species are atypical pathogens different from many model bacteria, and have a varied lifestyle that might necessitate the use of non-traditional DNA regulation models and may have ECF or ECF-like sigma factors that facilitate survival at different stages of the bacterial life cycle. In fact, strains of *L. interrogans* have been annotated at having as many as 11 separate ECF sigma factors and LIMLP_03680/ σ^X homologues are only found in infectious species and not in saprophytes, suggesting that σ^X is important for the specific life cycle requirements of pathogens (62,63). Although only found in pathogens, the overexpression of the ECF sigma factor in this study was detrimental to bacterial virulence. This may have been due to the extreme nature of the overexpression in this instance and future research with a more focused approach may determine an inverse effect upon more controlled overexpression of the gene. Spirochetes also have a unique infection paradigm and increased virulence may lead to a heightened immune reaction, resulting in the early control of bacterial numbers. The specific stimulus or stimuli activating this particular transcriptional regulator remains unknown, but ECF sigma factors canonically respond to periplasmic stress (64). Accordingly, we found that the *lomA* mutant exhibits a two-fold increased susceptibility to LPS-binding antibiotics, and transcriptomic analysis indicates that membrane proteins and LPS are preferentially affected in the mutant, suggesting a connection between the 4mC-catalyzing MTase and envelope stress in *L. interrogans*.

A study in *L. interrogans* serovar Lai has shown that the transcription of the homologous *sigX/rsx* operon was upregulated 4-fold after a temperature shift from 28°C (the optimal temperature for *in vitro* growth) to 37°C—which represents the mammalian physiological temperature (65). Overall, the σ^X overexpression strain only matched the *lomA* mutant phenotype with regards to virulence attenuation and expansion on semisolid agar, neither of which are highly specific to these mutations. Indeed, even the infection dynamics appear divergent, as the σ^X-overexpressing spirochetes could not be observed by darkfield analysis of tissue slurries at 14 days post-infection as the *lomA* mutant could. This suggests that the root cause(s) of the phenotypes observed in the *lomA* mutant remain(s) elusive, but are likely due to the preferentially impacted COGs (signaling, motility, and envelope biogenesis). This complexity is compounded by the fact that *Leptospira*, and, more generally, all spirochetes, possess many putative genes that code for proteins of unknown function with little to no sequence identity with characterized proteins from model organisms (66).

It is important to note that, while under laboratory conditions, the DNA methylations catalyzed by LomA appear to be largely suppressive, this may not be the case during the infectious cycle of the bacterium. Indeed, genes controlled by CTAG methylation that remain silent during medium-facilitated growth of *Leptospira* may become active during infection or other stressful conditions. For instance, after *in vitro* incubation at 37°C (67), the *srp* genes in the *lomA* mu-

tant became even more dysregulated than they were at 30°C (Supplementary Figure S12). While promoter methylation in eukaryotes is largely thought of as inhibitory, the role of methylated DNA in bacteria appears to be more dynamic. For example, the study on 4mC in *H. pylori* found that the ratio of up to down-regulation of genes in a methylase mutant was nearly 1:1 (15), but this could be due, at least in part, to polar effects from transcription factors, as seen in our study. Further studies may help determine whether *Leptospira lomA* mutants react differently than WT under other growth or survival conditions.

The role of LomA in leptospiral virulence is supported by the fact that every sequenced reference strain from the P clades of *Leptospira* spp. (containing the evolved pathogenic strains) harbors an allele of *lomA*, but the MTase is absent from sequenced members of the S clades (composed of saprophytic strains). In fact, no putative MTase targeting CTAG motifs could be found in the genome of any sequenced saprophyte. This has been supported by the lack of methyl groups on this motif by the few species that have been analyzed by SMRT-seq. In addition, artificial expression of *L. interrogans lomA* in the S species (*L. biflexa*) seems to be toxic. This fact could suggest that the LomA MTase may be a recently evolved essential determinant of bacterial virulence across the genus. Species-specific or clade-specific MTases have been described in other pathogens such as *Neisseria meningitidis* and *Burkholderia pseudomallei* and may have played roles in the emergence of lineage structure (68–70). The origins of the *lomA* gene in *Leptospira* species remain unknown. This orphan MTase could have possibly been horizontally acquired together with its cognate REase that was subsequently lost. The closest well-characterized homologue of LomA is MthZIM, which is part of a plasmid-borne R-M system that likely acts as a defense mechanism against lytic archaeal phages in *M. thermoautotriphicus* (46). However, the transition from R-M system to an orphan MTase is an infrequent event, and recent studies suggests that orphan MTases occur more frequently from transfer via large mobile genetic elements (MGEs) such as plasmids and phages rather than arise *de novo* from decaying R-M systems through the loss of the REase gene (5,71,72). Investigation of the genomic architecture of the *lomA* and *sigX* loci and the surrounding genes has revealed a dichotomy between the species of groups P1 and P2 (Supplementary Figure S13). The fact that *lomA* and *sigX* are found exclusively in species of subclades P1 and P2 could suggest that the two genes were acquired concomitantly at the node of the speciation after the divergence of the P clade. *lomA* and *sigX* are not adjacent which could be due to an ancestral rearrangement event or it may also suggest that the action of LomA on SigX is the result of a recent molecular adaptation. Indeed, the link between LomA and sigma factors may be also emphasised as in all sequenced P2 species, *lomA* is located near another ECF sigma factor which may suggest species-specific associations (Supplementary Figure S13). Still, how and when the gene became incorporated into the leptospiral genome remains unknown. This acquisition, however, may have played an important role in the evolution of *Leptospira* from saprophyte to pathogen.

Overall, the results obtained herein reinforce the need for examining the roles of DNA methylation in non-traditional bacterial models. Leptospirosis, while an important worldwide zoonosis, remains critically understudied (18,73). This may be due, in part, to the fastidious nature of the causative agent; pathogenic *Leptospira* are slow-growing bacteria and genetic tools remain largely inefficient or non-existent in pathogenic species. Nevertheless, the unique nature of the bacterium allows for non-traditional mechanisms of pathogenesis and gene regulation (e.g. the first *in vivo* validation of a 4mC MTase as a virulence factor). As 4mC DNA MTases are unique to prokaryotes, these enzymes could also serve as optimal target for future generations of antibiotics. Future work investigating the methylome and corresponding regulome of this enigmatic bacterial genus should provide further insights into how epigenetic changes may drive adaptation to particular environments.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Drs. Matsunaga and Haake for providing plasmid pRAT724 and Dr. Koszul for feedback on CCMs. We thank Dr. Linda Grillova for aid with bioinformatic analysis and critical reading of the manuscript. We also thank Drs. Simonetta Gribaldo and Nadia Benaroudj for critical reading of the manuscript.

FUNDING

Institut Pasteur [PTR 30-2017 to M.P.]; ANR 18 CE15 0027 LEPTOMOVE to M.P., Pasteur Foundation Fellowship to R.A.G. (April 2017–June 2020), ERC grant agreement 771813 to R.K., Fellowship from the Natural Sciences and Engineering Research Council of Canada to ATV. FJV is a research scholar of the Fonds de Recherche du Québec-Santé. Funding for open access charge: Institut Pasteur.
Conflict of interest statement. None declared.

REFERENCES

1. Bird,A. (2007) Perceptions of epigenetics. *Nature*, **447**, 396–398.
2. 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.
3. Beaulaurier,J., Schadt,E.E. and Fang,G. (2019) Deciphering bacterial epigenomes using modern sequencing technologies. *Nat. Rev. Genet.*, **20**, 157.
4. Vasu,K. and Nagaraja,V. (2013) Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiol. Mol. Biol. Rev. MMBR*, **77**, 53–72.
5. Seshasayee,A.S.N., Singh,P. and Krishna,S. (2012) Context-dependent conservation of DNA methyltransferases in bacteria. *Nucleic Acids Res.*, **40**, 7066–7073.
6. Casadesús,J. and Low,D. (2006) Epigenetic gene regulation in the bacterial world. *Microbiol. Mol. Biol. Rev.*, **70**, 830–856.
7. 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.

8. Krebes,J., Morgan,R.D., Bunk,B., Spröer,C., Luong,K., Parusel,R., Anton,B.P., König,C., Josenhans,C., Overmann,J. *et al.* (2014) The complex methylome of the human gastric pathogen *Helicobacter pylori*. *Nucleic Acids Res.*, **42**, 2415–2432.
9. Murray,I.A, Clark,T.A., Morgan,R.D., Boitano,M., Anton,B.P., Luong,K., Fomenkov,A., Turner,S.W., Korlach,J., Roberts,R.J. *et al.* (2012) The methylomes of six bacteria. *Nucleic Acids Res.*, **40**, 11450–11462.
10. Bendall,M.L, Luong,K., Wetmore,K.M., Blow,M., Korlach,J., Deutschbauer,A. and Malmstrom,R.R. (2013) Exploring the roles of DNA methylation in the metal-reducing bacterium *Shewanella oneidensis* MR-1. *J. Bacteriol.*, **195**, 4966–4974.
11. 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.
12. Seib,K.L, Jen,F.E.-.C., Tan,A., Scott,A.L., Kumar,R., Power,P.M., Chen,L.-.T., Wu,H.-.J., Wang,A.H.-.J., Hill,D.M.C. *et al.* (2015) Specificity of the ModA11, ModA12 and ModD1 epigenetic regulator N6-adenine DNA methyltransferases of *Neisseria meningitidis*. *Nucleic Acids Res.*, **43**, 4150–4162.
13. Estibariz,I., Overmann,A., Ailloud,F., Krebes,J., Josenhans,C. and Suerbaum,S. (2019) The core genome m5C methyltransferase JHP1050 (M.Hpy99III) plays an important role in orchestrating gene expression in *Helicobacter pylori*. *Nucleic Acids Res.*, **47**, 2336–2348.
14. Low,D.A., Weyand,N.J. and Mahan,M.J. (2001) Roles of DNA adenine methylation in regulating bacterial gene expression and virulence. *Infect. Immun.*, **69**, 7197–7204.
15. Kumar,S., Karmakar,B.C., Nagarajan,D., Mukhopadhyay,A.K., Morgan,R.D. and Rao,D.N. (2018) N4-cytosine DNA methylation regulates transcription and pathogenesis in *Helicobacter pylori*. *Nucleic Acids Res.*, **46**, 3429–3445.
16. Oliveira,P.H, Ribis,J.W., Garrett,E.M., Trzilova,D., Kim,A., Sekulovic,O., Mead,E.A., Pak,T., Zhu,S., Deikus,G. *et al.* (2020) Epigenomic characterization of *Clostridioides difficile* finds a conserved DNA methyltransferase that mediates sporulation and pathogenesis. *Nat. Microbiol.*, **5**, 166–180.
17. Costa,F., Hagan,J.E., Calcagno,J., Kane,M., Torgerson,P., Martinez-Silveira,M.S., Stein,C., Abela-Ridder,B. and Ko,A.I. (2015) Global morbidity and mortality of leptospirosis: a systematic review. *PLoS Negl. Trop. Dis.*, **9**, e0003898.
18. Picardeau,M. (2017) Virulence of the zoonotic agent of leptospirosis: still terra incognita? *Nat. Rev. Microbiol.*, **15**, 297–307.
19. Vincent,A.T, Schiettekatte,O., Goarant,C., Neela,V.K., Bernet,E., Thibeaux,R., Ismail,N., Mohd Khalid,M.K.N., Amran,F., Masuzawa,T. *et al.* (2019) Revisiting the taxonomy and evolution of pathogenicity of the genus *Leptospira* through the prism of genomics. *PLoS Negl. Trop. Dis.*, **13**, e0007270.
20. Murray,G.L, Morel,V., Cerqueira,G.M., Croda,J., Srikram,A., Henry,R., Ko,A.I., Dellagostin,O.A., Bulach,D.M., Sermswan,R.W. *et al.* (2009) Genome-wide transposon mutagenesis in pathogenic *Leptospira* species. *Infect. Immun.*, **77**, 810–816.
21. Ellinghausen,H.C. and Mccullough,W.G. (1965) Nutrition of Leptospira pomona and growth of 13 other serotypes: fractionation of oleic albumin complex and a medium of bovine albumin and polysorbate 80. *Am. J. Vet. Res.*, **26**, 45–51.
22. Pappas,C.J., Benaroudj,N. and Picardeau,M. (2015) A replicative plasmid vector allows efficient complementation of pathogenic *Leptospira* strains. *Appl. Environ. Microbiol.*, **81**, 3176–3181.
23. Matsunaga,J. and Haake,D.A. (2018) cis-acting determinant limiting expression of sphingomyelinase gene *sph2* in *Leptospira interrogans*, identified with a gfp reporter plasmid. *Appl. Environ. Microbiol.*, **84**, e02068-18.
24. Picardeau,M. (2008) Conjugative transfer between *Escherichia coli* and *Leptospira spp*. as a New Genetic Tool. *Appl. Environ. Microbiol.*, **74**, 319–322.
25. Dukes,J.D., Whitley,P. and Chalmers,A.D. (2011) The MDCK variety pack: choosing the right strain. *BMC Cell Biol.*, **12**, 43.
26. Stassen,Q.E.M., Riemers,F.M., Reijmerink,H., Leegwater,P.A.J. and Penning,L.C. (2015) Reference genes for reverse transcription quantitative PCR in canine brain tissue. *BMC Res. Notes*, **8**, 761.
27. Eshghi,A., Henderson,J., Trent,M.S. and Picardeau,M. (2015) *Leptospira interrogans lpxD* homologue is required for thermal acclimatization and virulence. *Infect. Immun.*, **83**, 4314–4321.
28. Lioy,V.S, Cournac,A., Marbouty,M., Duigou,S., Mozziconacci,J., Espéli,O., Boccard,F. and Koszul,R. (2018) Multiscale structuring of the *E. coli* chromosome by nucleoid-associated and condensin proteins. *Cell*, **172**, 771–783.
29. Cournac,A., Marie-Nelly,H., Marbouty,M., Koszul,R. and Mozziconacci,J. (2012) Normalization of a chromosomal contact map. *BMC Genomics*, **13**, 436.
30. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
31. Wick,R.R., Judd,L.M., Gorrie,C.L. and Holt,K.E. (2017) Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Comput. Biol.*, **13**, e1005595.
32. Rice,P., Longden,I. and Bleasby,A. (2000) EMBOSS: the european molecular biology open software suite. *Trends Genet. TIG*, **16**, 276–277.
33. Roberts,R.J., Vincze,T., Posfai,J. and Macelis,D. (2007) REBASE—enzymes and genes for DNA restriction and modification. *Nucleic Acids Res.*, **35**, D269–D270.
34. Cokelaer,T., Desvillechabrol,D., Legendre,R. and Cardon,M. (2017) 'Sequana': a set of snakemake NGS pipelines. *J. Open Source Softw.*, **16**, 352.
35. Martin,M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, **17**, 10–12.
36. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
37. Liao,Y., Smyth,G.K. and Shi,W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinforma. Oxf. Engl.*, **30**, 923–930.
38. R Development Core Team (2020) R: a language and environment for statistical computing. ISBN 3-900051-07-0, Vienna, Austria.
39. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
40. Benjamini,Y. and Hochberg,Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.*, **57**, 289–300.
41. Satou,K., Shimoji,M., Tamotsu,H., Juan,A., Ashimine,N., Shinzato,M., Toma,C., Nohara,T., Shiroma,A., Nakano,K. *et al.* (2015) Complete genome sequences of low-passage virulent and high-passage avirulent variants of pathogenic *Leptospira interrogans* serovar Manilae strain UP-MMC-NIID, originally isolated from a patient with severe leptospirosis, determined using PacBio single-molecule real-time technology. *Genome Announc.*, **3**, e00882-15.
42. Eshghi,A., Lourdault,K., Murray,G.L., Bartpho,T., Sermswan,R.W., Picardeau,M., Adler,B., Snarr,B., Zuerner,R.L. and Cameron,C.E. (2012) *Leptospira interrogans* catalase is required for resistance to H_2O_2 and for virulence. *Infect. Immun.*, **80**, 3892–3899.
43. Gomes-Solecki,M., Santecchia,I. and Werts,C. (2017) Animal models of leptospirosis: of mice and hamsters. *Front. Immunol.*, **8**, 58.
44. Lambert,A., Picardeau,M., Haake,D.A., Sermswan,R.W., Srikram,A., Adler,B. and Murray,G.A. (2012) FlaA proteins in *Leptospira interrogans* are essential for motility and virulence but are not required for formation of the flagellum sheath. *Infect. Immun.*, **80**, 2019–2025.
45. Eshghi,A., Becam,J., Lambert,A., Sismeiro,O., Dillies,M.A., Jagla,B., Wunder,E.A. Jr, Ko,A.I., Coppee,J.Y., Goarant,C. *et al.* (2014) A putative regulatory genetic locus modulates virulence in the pathogen *Leptospira interrogans*. *Infect. Immun.*, **82**, 2542–2552.
46. Nölling,J. and de Vos,W.M. (1992) Identification of the CTAG-recognizing restriction-modification systems MthZI and MthFI from *Methanobacterium thermoformicicum* and characterization of the plasmid-encoded *mthZIM* gene. *Nucleic Acids Res.*, **20**, 5047–5052.
47. Wasserfallen,A., Nölling,J., Pfister,P., Reeve,J. and Conway de Macario,E. (2000) Phylogenetic analysis of 18 thermophilic *Methanobacterium* isolates supports the proposals to create a new genus, *Methanothermobacter* gen. nov., and to reclassify several isolates in three species, *Methanothermobacter thermautotrophicus* comb. nov., *Methanothermobacter wolfeii* comb. nov., and *Methanothermobacter marburgensis* sp. nov. *Int. J. Syst. Evol. Microbiol.*, **50**, 43–53.

48. Ouellette,M., Gogarten,J.P., Lajoie,J., Makkay,A.M. and Papke,R.T. (2018) Characterizing the DNA methyltransferases of *Haloferax volcanii* via bioinformatics, gene deletion, and SMRT sequencing. *Genes*, **9**, 129.
49. Marbouty,M., Le Gall,A., Cattoni,D.I., Cournac,A., Koh,A., Fiche,J.B., Mozziconacci,J., Murray,H., Koszul,R. and Nollmann,M. (2015) Condensin- and replication-mediated bacterial chromosome folding and origin condensation revealed by Hi-C and super-resolution imaging. *Mol. Cell*, **59**, 588–602.
50. Zhukova,A., Fernandes,L.G., Hugon,P., Pappas,C.J., Sismeiro,O., Coppée,J.Y., Becavin,C., Malabat,C., Eshghi,A., Zhang,J.J. *et al.* (2017) Genome-wide transcriptional start site mapping and sRNA identification in the pathogen *Leptospira interrogans*. *Front. Cell. Infect. Microbiol.*, **7**, 10.
51. Crooks,G.E., Hon,G., Chandonia,J.-M. and Brenner,S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.*, **14**, 1188–1190.
52. Helmann,J.D. (2002) The extracytoplasmic function (ECF) sigma factors. *Adv. Microb. Physiol.*, **46**, 47–110.
53. Falkow,S. (1988) Molecular Koch's postulates applied to microbial pathogenicity. *Rev. Infect. Dis.*, **10 Suppl 2**, S274–S276.
54. Fontana,C., Lambert,A., Benaroudj,N., Gasparini,D., Gorgette,O., Cachet,N., Bomchil,N. and Picardeau,M. (2016) Analysis of a spontaneous non-motile and avirulent mutant shows that FliM is required for full endoflagella assembly in *Leptospira interrogans*. *PLoS One*, **11**, e0152916.
55. Wunder,E.A, Figueira,C.P., Benaroudj,N., Hu,B., Tong,B.A., Trajtenberg,F., Liu,J., Reis,M.G., Charon,N.W., Buschiazzo,A. *et al.* (2016) A novel flagellar sheath protein, FcpA, determines filament coiling, translational motility and virulence for the *Leptospira* spirochete. *Mol. Microbiol.*, **101**, 457–470.
56. Chao,M.C, Zhu,S., Kimura,S., Davis,B.M., Schadt,E.E., Fang,G. and Waldor,M.K. (2015) A cytosine methytransferase modulates the cell envelope stress response in the cholera pathogen. *PLos Genet.*, **11**, e1005666.
57. Hofmann,N., Wurm,R. and Wagner,R. (2011) The *E. coli* anti-sigma factor Rsd: studies on the specificity and regulation of its expression. *PLoS One*, **6**, e19235.
58. Kazmierczak,M.J., Wiedmann,M. and Boor,K.J. (2005) Alternative sigma factors and their roles in bacterial virulence. *Microbiol. Mol. Biol. Rev.*, **69**, 527–543.
59. Llamas,M.A, van der Sar,A., Chu,B.C., Sparrius,M., Vogel,H.J. and Bitter,W. (2009) A novel extracytoplasmic function (ECF) sigma factor regulates virulence in *Pseudomonas aeruginosa*. *PLoS Pathog.*, **5**, e1000572.
60. Rattanama,P., Thompson,J.R., Kongkerd,N., Srinitiwarawong,K., Vuddhakul,V. and Mekalanos,J.J. (2012) Sigma E regulators control hemolytic activity and virulence in a shrimp pathogenic *Vibrio harveyi*. *PLoS One*, **7**, e32523.
61. Bashyam,M.D. and Hasnain,S.E. (2004) The extracytoplasmic function sigma factors: role in bacterial pathogenesis. *Infect. Genet. Evol.*, **4**, 301–308.
62. Fouts,D.E, Matthias,M.A., Adhikarla,H., Adler,B., Amorim-Santos,L., Berg,D.E., Bulach,D., Buschiazzo,A., Chang,Y.F., Galloway,R.L. *et al.* (2016) What makes a bacterial species pathogenic?: comparative genomic analysis of the genus *Leptospira*. *PLoS Negl. Trop. Dis*, **10**, e0004403.
63. Nascimento,A.L.T.O, Ko,A.I., Martins,E.A., Monteiro-Vitorello,C.B., Ho,P.L., Haake,D.A., Verjovski-Almeida,S., Hartskeerl,R.A., Marques,M.V., Oliveira,M.C. *et al.* (2004) Comparative genomics of two *Leptospira interrogans* serovars reveals novel insights into physiology and pathogenesis. *J. Bacteriol.*, **186**, 2164–2172.
64. Raivio,T.L. and Silhavy,T.J. (2001) Periplasmic stress and ECF sigma factors. *Annu. Rev. Microbiol.*, **55**, 591–624.
65. Qin,J.-H, Sheng,Y.Y., Zhang,Z.M., Shi,Y.Z., He,P., Hu,B.Y., Yang,Y., Liu,S.G., Zhao,G.P. and Guo,X.K. (2006) Genome-wide transcriptional analysis of temperature shift in *L. interrogans* serovar lai strain 56601. *BMC Microbiol.*, **6**, 51.
66. Lobb,B., Tremblay,B.J.-M., Moreno-Hagelsieb,G. and Doxey,A.C. (2020) An assessment of genome annotation coverage across the bacterial tree of life. *Microb. Genomics*, **6**, e000341.
67. Fraser,T. and Brown,P.D. (2017) Temperature and oxidative stress as triggers for virulence gene expression in pathogenic *Leptospira* spp. *Front. Microbiol*, **8**, 783.
68. Oliveira,P.H. and Fang,G. (2020) Conserved DNA methyltransferases: a window into fundamental mechanisms of epigenetic regulation in bacteria. *Trends Microbiol.*, doi:10.1016/j.tim.2020.04.007.
69. Budroni,S., Siena,E., Dunning Hotopp,J.C., Seib,K.L., Serruto,D., Nofroni,C., Comanducci,M., Riley,D.R., Daugherty,S.C., Angiuoli,S.V. *et al.* (2011) *Neisseria meningitidis* is structured in clades associated with restriction modification systems that modulate homologous recombination. *Proc. Natl. Acad. Sci.*, **108**, 4494–4499.
70. Nandi,T., Holden,M.T., Didelot,X., Mehershahi,K., Boddey,J.A., Beacham,I., Peak,I., Harting,J., Baybayan,P., Guo,Y. *et al.* (2015) *Burkholderia pseudomallei* sequencing identifies genomic clades with distinct recombination, accessory, and epigenetic profiles. *Genome Res.*, **25**, 129–141.
71. Oliveira,P.H., Touchon,M. and Rocha,E.P.C. (2014) The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Res.*, **42**, 10618–10631.
72. Fullmer,M.S., Ouellette,M., Louyakis,A.S., Papke,R.T. and Gogarten,J.P. (2019) The patchy distribution of restriction–modification system genes and the conservation of orphan methyltransferases in *Halobacteria*. *Genes*, **10**, 233.
73. Goarant,C., Picardeau,M., Morand,S. and McIntyre,K.M. (2019) Leptospirosis under the bibliometrics radar: evidence for a vicious circle of neglect. *J. Glob. Health*, **9**, 010302.