

# The *agr* Locus Regulates Virulence and Colonization Genes in *Clostridium difficile* 027

Melissa J. Martin,<sup>a</sup> Simon Clare,<sup>b</sup> David Goulding,<sup>b</sup> Alexandra Faulds-Pain,<sup>a</sup> Lars Barquist,<sup>c,d</sup> Hilary P. Browne,<sup>e</sup> Laura Pettit,<sup>b</sup> Gordon Dougan,<sup>b</sup> Trevor D. Lawley,<sup>e</sup> Brendan W. Wren<sup>a</sup>

Department of Pathogen Molecular Biology, London School of Hygiene and Tropical Medicine, University of London, London, United Kingdom<sup>a</sup>; Microbial Pathogenesis Laboratory, Wellcome Trust Sanger Institute, Hinxton, United Kingdom<sup>b</sup>; Rfam Group, Wellcome Trust Sanger Institute, Hinxton, United Kingdom<sup>c</sup>; EMBL-European Bioinformatics Institute, Hinxton, United Kingdom<sup>d</sup>; Bacterial Pathogenesis Laboratory, Wellcome Trust Sanger Institute, Hinxton, United Kingdom<sup>e</sup>

**The transcriptional regulator AgrA, a member of the LytTR family of proteins, plays a key role in controlling gene expression in some Gram-positive pathogens, including *Staphylococcus aureus* and *Enterococcus faecalis*. AgrA is encoded by the *agrACDB* global regulatory locus, and orthologues are found within the genome of most *Clostridium difficile* isolates, including the epidemic lineage 027/BI/NAP1. Comparative RNA sequencing of the wild type and otherwise isogenic *agrA* null mutant derivatives of *C. difficile* R20291 revealed a network of approximately 75 differentially regulated transcripts at late exponential growth phase, including many genes associated with flagellar assembly and function, such as the major structural subunit, FliC. Other differentially regulated genes include several involved in bis-(3'-5')-cyclic dimeric GMP (c-di-GMP) synthesis and toxin A expression. *C. difficile* 027 R20291 *agrA* mutant derivatives were poorly flagellated and exhibited reduced levels of colonization and relapses in the murine infection model. Thus, the *agr* locus likely plays a contributory role in the fitness and virulence potential of *C. difficile* strains in the 027/BI/NAP1 lineage.**

*Clostridium difficile* is a Gram-positive, anaerobic bacterium that is arguably the most frequent cause of antibiotic-associated colitis and health care-acquired diarrhea worldwide. Disease-causing isolates can produce two exotoxins, TcdA and TcdB, encoded by the 19-kb pathogenicity locus (PaLoc), which interact with the intestinal epithelium and potentially precipitate an acute inflammatory response and even cell death (1–3). Colonization of the intestine by toxigenic *C. difficile* can be asymptomatic, but following antibiotic treatment, a variety of symptoms, including diarrhea or life-threatening pseudomembranous colitis, can ensue (4). Relapsing disease occurs in up to 20% of patients following termination of treatment with certain first-line antimicrobials, such as vancomycin or metronidazole (5). The global increase in incidence and severity of *C. difficile* infection over the last decade is linked to the emergence of certain lineages, including the epidemic 027/BI/NAP1 variants (6–11). Transcontinental dissemination of the 027 variant occurred through at least two distinct fluoroquinolone-resistant lineages, with clinical outcomes resulting in longer colonization duration, increased toxin production, and increased relapse and mortality rates (6, 12, 13). This increased disease severity was typified by multiple outbreaks of *C. difficile* disease between 2003 and 2006, affecting over 300 patients at the Stoke Mandeville hospital, United Kingdom, from which the representative 027 strain, R20291, was isolated (14, 15).

Gene regulatory networks can enable pathogenic bacteria to rapidly adapt to their environment and modulate the expression of virulence-associated factors. So-called two-component systems (TCSs) can play a key role in linking environmental and internal sensing to the control of gene expression. Interestingly, genes with similarity to TCSs, transcriptional regulators, and signaling proteins comprise approximately 10% of *C. difficile* genomes, yet their contribution to the regulatory mechanisms and virulence within *C. difficile* are poorly understood (14, 16). The RolA/B TCS of *C. difficile* has been shown to negatively regulate the *luxS* gene and, consequently, the synthesis of the putative quorum-sensing

signaling molecule, autoinducer 2 (AI-2) (16–18). Other classes of regulators reported in *C. difficile* include the transcriptional regulators CcpA (19), CodY (20), and SigH (21), which influence the expression of the exotoxins, TcdA and TcdB, and Spo0A, which is a key regulator of sporulation and is important for persistence and transmission within the host (22–24). Furthermore, the *C. difficile* flagellar regulon modulates toxin production *in vitro* (25) in addition to having a contributory role in adhesion and colonization *in vivo* (26).

Analysis of the genome sequence of the *C. difficile* 027 isolate R20291 identified a locus with similarity to the *agr* operon, which is a conserved determinant in many Gram-positive bacteria (14). In *Staphylococcus aureus*, the global regulation of virulence genes is coordinated by the *agr* quorum-sensing locus, *agrACDB* (27–29). The *agrD* and *agrB* genes encode the precursor to the small secreted cyclic autoinducing peptide (AIP) and a transmembrane protein involved in processing and exporting of AIP, respectively. Extracellular accumulation of AIP activates a typical bacterial TCS by binding to AgrC, a sensor kinase, subsequently resulting in the phosphorylation of the AgrA response regulator. Phosphorylated AgrA binds to DNA via its C-terminal LytTR domain and can activate the transcription of both the RNAII (*agrBDCA*) transcript, creating a positive feedback loop, and the divergent RNAPIII transcript encoding a regulatory RNA effector molecule (30, 31).

Received 24 April 2013 Accepted 31 May 2013

Published ahead of print 14 June 2013

Address correspondence to Brendan W. Wren, [Brendan.wren@lshtm.ac.uk](mailto:Brendan.wren@lshtm.ac.uk).

Supplemental material for this article may be found at <http://dx.doi.org/10.1128/JB.00473-13>.

Copyright © 2013 Martin et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](http://creativecommons.org/licenses/by/3.0/).

doi:10.1128/JB.00473-13

TABLE 1 Strains and plasmids used in this study

Strain or plasmid	Characteristic(s)	Source
<i>C. difficile</i> strains		
R20291	Epidemic PCR ribotype 027, Stoke Mandeville hospital outbreak, United Kingdom, 2005	14
R20291 <i>agrA</i>	R20291 <i>agrA76a::CT</i>	This study
R20291 <i>agrA</i> complement	<i>C. difficile</i> R20291 <i>agrA</i> complemented with pMTL-84151- <i>agrA</i>	This study
<i>E. coli</i> CA434	Conjugation donor for <i>E. coli</i> HB101 [ $F^-$ <i>mcrB mrr hsdS20</i> ( $r_B^- m_B^-$ ) <i>recA13 leuB6 ara-14 proA2 lacY1 galK2 xyl-5 mtl-1 rpsL20</i> (Sm <sup>r</sup> ) <i>glnV44</i> $\lambda^-$ ] containing plasmid R702	41
Plasmids		
pMTL007-CE2	Clostron plasmid (ColE1, pCD6, <i>catP</i> )	37
pMTL007-CE2- <i>agrA76a</i>	pMTL007-CE2 derivative retargeted to <i>agrA</i>	This study
pMTL-84151	<i>E. coli-C. difficile</i> shuttle plasmid (pCD6, <i>catP</i> , ColE1 + <i>tra</i> )	42
pMTL-84151- <i>agrA</i>	pMTL-84151 containing 0.768-kb <i>agrA</i> coding sequence and 0.372-kb putative upstream promoter region	This study

The *C. difficile agr* locus carries the requisite genes for a functional *agr* operon, *agrACDB*. Interestingly, this complete locus is absent from the first reported *C. difficile* 630 genome, ribotype 012, and it was originally termed “*agr2*”. All analyzed *C. difficile* genomes contain the so-called *agr1* locus, encoding a partial *agr*-like locus of *agrDB* (14).

Originally thought to be specific to 027, a comparative genomic hybridization study identified the complete *agr* locus, *agrACDB*, in other clinical isolates, suggesting that this locus is prevalent within the species (32). Here, we undertake further analysis of *C. difficile* R20291 to determine the network of genes under the regulatory control of the *agr* locus. We identify a number of characteristics, including flagellar biosynthesis, TcdA production, and bis-(3'-5')-cyclic dimeric GMP (c-di-GMP) signaling proteins that are influenced by the *agr* locus and show that it has a contributory role for colonization in the *C. difficile* murine model of infection.

## MATERIALS AND METHODS

**Bacterial growth conditions and strains.** All strains and plasmids used in this study are summarized in Table 1. *C. difficile* strains were routinely cultured at 37°C under anaerobic conditions (Mini-Mac 250; Don Whitley Scientific) using brain heart infusion (BHI; Oxoid) medium or Brazier's CCEY agar supplemented with 4% egg yolk (Bioconnections). Where appropriate, *C. difficile* agar was supplemented with D-cycloserine (250  $\mu\text{g} \cdot \text{ml}^{-1}$ ) and cefoxitin (8  $\mu\text{g} \cdot \text{ml}^{-1}$ ), 15  $\mu\text{g}/\text{ml}$  thiamphenicol (Sigma), or 20  $\mu\text{g}/\text{ml}$  lincomycin (Sigma). *Escherichia coli* strains were cultured aerobically at 37°C using Luria-Bertani (LB) media (Sigma). Where appropriate, media were supplemented with 12.5  $\mu\text{g}/\text{ml}$  chloramphenicol (Alfa Aesar). Spore enumeration was performed by inoculating cultures 1:1 with 100% (vol/vol) ethanol for 1 h at room temperature to kill the vegetative cells. Total CFU were enumerated by serial dilution in phosphate-buffered saline (PBS) and plating on BHI plates supplemented with 0.5% taurocholate (Sigma) to stimulate germination (33).

**Library preparation and genomic DNA sequencing.** Genomic *C. difficile* DNA was isolated as previously described (34). Paired-end multiplexed libraries were created as previously described (35), and the Illumina HiSeq 2000 platform was used for whole-genome sequencing of *C. difficile* R20291 and R20291 *agrA76a::CT*. Sequencing reads were mapped to the R20291 genome (14) using the Burrows-Wheeler Aligner (BWA) (36).

**Mutagenesis and genetic complementation studies.** The Clostron system was used to insertionally inactivate the R20291 *agrA* gene as previously described (37, 38). Briefly, using the Perutka algorithm (39), plasmid pMTL007C-E2 was retargeted to the antisense strand of *agrA* between positions 76 and 77, and the resultant plasmid, pMTL007C-E2-*agrA76a*, was synthesized by DNA 2.0 (Menlo Park). This plasmid was

conjugated into wild-type R20291 as previously described (40) using the electrocompetent *E. coli* CA434 donor strain (41). Transconjugants were selected on agar plates in the presence of thiamphenicol to select for the plasmid and cycloserine/cefoxitin to select against *E. coli*. Clostron integrants then were isolated by their resistance to lincomycin, and loss of the plasmid was confirmed by sensitivity to thiamphenicol. Insertions in *agrA* were confirmed by PCR screening using the primers pairs *agrA76a*-Fw/*agrA76a*-Rv, *agrA76a*-R/EBS, and RAM-Fw/RAM-Rv and sequenced using the Illumina HiSeq platform. For complementation studies, the *agrA* coding sequence and putative promoter region, 372 bp upstream, was PCR amplified from wild-type R20291 template using oligonucleotides *agrAcNdeI*\_Fw and *agrAcHindIII*\_Rv and inserted into the modular vector pMTL84151 by restriction and ligation cloning into the NdeI/HindIII sites (42), creating the complementation vector pMTL-84151-*agrA*. The plasmid was confirmed by DNA sequencing, and pMTL-84151-*agrA* was conjugated into *C. difficile* R20291 *agrA76a::CT* as described previously (40).

**RNA preparation and cDNA synthesis.** Bacterial cells were harvested in RNAprotect bacteria (Qiagen), and total RNA was extracted using the FastRNA Pro blue kit (MP Biomedical) according to the manufacturer's protocol. RNA was eluted in nuclease-free water and quantified using a NanoDrop ND-1000 and 2100 Bioanalyzer (Agilent Technologies). Genomic DNA was removed using one treatment of Turbo DNase (Applied Biosystems), and PCR analyses using primer pairs that amplified housekeeping genes *dxr*, *sigA1*, and *sigB* were performed to confirm DNA depletion. Equal amounts of total DNA-free RNA (5  $\mu\text{g}$ ) were reverse transcribed using random hexamer primers (Invitrogen) and SuperScript III reverse transcriptase (Invitrogen). cDNAs were synthesized in the presence of actinomycin D to prevent spurious second-strand cDNA synthesis, which inhibits DNA-dependent DNA synthesis (43).

**cDNA sequencing and expression profiling.** cDNA sequencing was performed using an Illumina HiSeq platform from 150- to 250-bp multiplexed cDNA libraries. Seventy-five cycles of paired-end sequencing from approximately 169 million cDNAs yielded 12,786 Mb of sequencing data and 15 to 45 million reads per library (see Table S1 in the supplemental material). cDNA sequence reads were aligned to the *C. difficile* R20291 genome reference (14) using BWA with a quality parameter ( $-q$ ) of 15, resulting in 76 to 82% of total reads aligned per library. Reads were mapped to annotated coding sequences (CDSs) and intergenic regions to account for possible unannotated noncoding RNAs. Raw read counts were calculated per nucleotide for each gene and intergenic region from three biological replicates of wild-type R20291 and the R20291 *agrA76a::CT* mutant. Differential expression analyses were performed using R version 2.15.0 and the DESeq statistical analysis package (44). *P* values were corrected for multiple testing using the Benjamini-Hochberg method, and a *q* value threshold of 0.1 was used to define differentially regulated genes with an expected false discovery rate of <10% (see Fig. S2 in the supplemental material).

**Quantitative reverse transcription-PCR (qRT-PCR).** Relative expression levels of target transcripts were determined using Power SYBR green PCR master mix (Invitrogen) by following the manufacturer's protocol. Specific primer pairs for *tcdA*, *fliC*, CDR20291\_1514 (KEGG accession number) and the *rpoA* internal control were designed using Primer3 software (<http://frodo.wi.mit.edu>). RNA from three biological replicates, independent of the RNA-seq samples, was prepared from R20291 and R20291 *agrA76a::CT* isolated from late exponential growth phase. Comparative threshold cycle ( $C_T$ ) analysis was performed, and the mean expression from three biological replicates and three replicates of each for target transcripts was calculated (45). Mean  $C_T$  values were normalized to the internal control housekeeping gene, *rpoA*. Relative mRNA expression was represented by fold change (see Fig. 3).

**Oligonucleotides.** The complete list of oligonucleotides used in this study is provided in Table S2 in the supplemental material.

**Bioinformatics.** Multiple sequence alignments were created using ClustalW2 (46, 47).

**TEM.** Negative staining and transmission electron microscopy (TEM) were performed to visualize *C. difficile* flagella. R20291, R20291 *agrA76a::CT*, and *agrA* complement strains were grown under the same conditions as samples prepared for RNA processing and TcdA quantification. Cultured colonies were mixed with distilled water to create a slightly turbid suspension and applied to Formvar/carbon-coated EM grids. An equal volume of 3% ammonium molybdate with 1% trehalose was added to negative stain. Images were taken on an FEI Spirit Biotwin 120-kV TEM with a Tietz F415 charge-coupled-device (CCD) camera.

**TcdA quantification.** *C. difficile* cultures were grown in BHI broth with shaking to late exponential phase, and culture supernatants were removed. TcdA quantification was performed by sandwich enzyme-linked immunosorbent assay (ELISA) as previously described (23). CFU were determined to ensure equal numbers of vegetative cells in all samples and replicates.

**CI experiments.** The *C. difficile* murine model of infection was used to perform competitive index (CI) experiments as previously described (23). Wild-type C57BL/6 mice ( $n = 5$ ) were infected with  $5 \times 10^6$  spores via gavage in 0.2 ml PBS. Equal amounts of spores from the parental R20291 and isogenic R20291 *agrA76a::CT* mutant derivative were used. Fecal samples were collected and enumerated by plating on *C. difficile* CCYE agar, with and without lincomycin, and incubated for 48 h. Agar supplemented with lincomycin selected for the knockout containing the *ermB* cassette. The CI number was determined using the following ratio:  $(R20291 \text{ } agrA76a::CT/R20291 \text{ wild-type})_{\text{output}} / (R20291 \text{ } agrA76a::CT/R20291 \text{ wild-type})_{\text{input}}$ . Statistical testing was performed using the Mann-Whitney test applied to  $\log_{10}$  values of the CI ratios. All animal infections were performed in accordance with the United Kingdom Home Office Animals (Scientific Procedures) Act of 1986.

**RNA-seq data accession number.** RNA-seq data generated in this work are available online at ArrayExpress under accession number E-ERAD-97.

## RESULTS

**The distribution of the *C. difficile* *agr* locus.** It has previously been shown that the *agr* locus is absent from some *C. difficile* isolates (14, 32). To better understand the prevalence of *agr* within the *C. difficile* species, we carried out comparative genomic analysis to determine its distribution within sequenced and fully annotated *C. difficile* isolates (Table 2) (14, 48). This revealed that the *agr* locus is not limited to the ribotype 027 strains but is present in multiple disease-causing *C. difficile* lineages, including PCR ribotypes 001 and 017.

Analysis of the genome of *C. difficile* R20291 revealed that the *agr* locus included *agrA*, *agrC*, *agrD*, and *agrB*, respectively (14). Interestingly, in the *S. aureus* *agr* operon these genes are in the reverse order (*agrBDCA*) (Fig. 1a). The *C. difficile* R20291 *agrA*-

TABLE 2 Distribution of *agr* locus in fully annotated *C. difficile* genomes

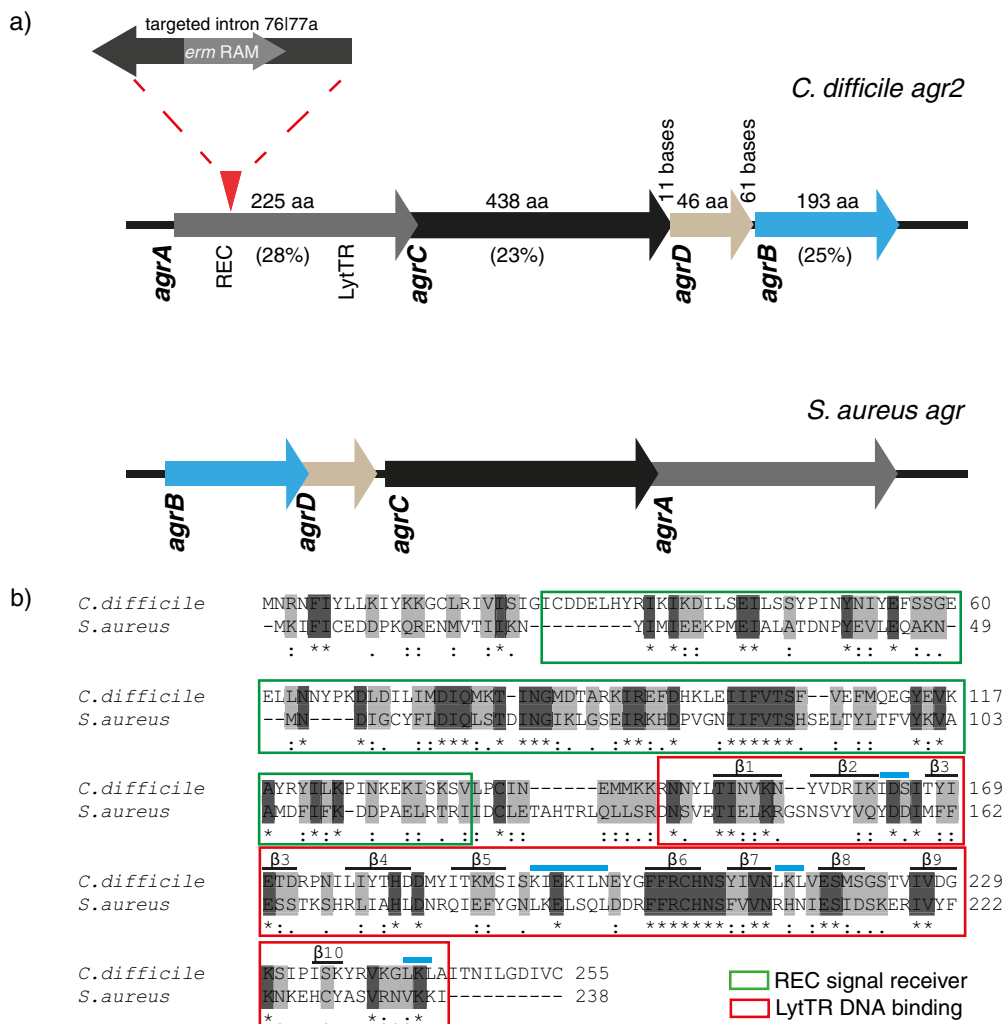
Strain and presence of <i>agr</i> locus	Ribotype	Source	Year and source of origin	Reference	Accession no.
<i>agr</i> locus positive					
BI-9	001	Human	2001, United States	48	FN668944
M68	017	Human	2006, Ireland	48	FN668375
CF5	017	Human	1995, Belgium	48	FN665652
BI-1	027	Human	1988, United States	48	NC_017179
2007855	027	Bovine	2007, United States	48	FN665654
R20291	027	Human	2006, United Kingdom	14	NC_013316
CD196	027	Human	1985, France	14	NC_013315
<i>agr</i> locus negative					
630	012	Human	1982, Switzerland	16	AM180355
M120	078	Human	2007, United Kingdom	48	FN665653

encoded protein shares 28% amino acid identity with the equivalent *S. aureus* AgrA and contains both a predicted N-terminal REC signal receiver domain and C-terminal LytTR-DNA binding domain (Fig. 1b). The *C. difficile* R20291 *agrC*-encoded protein and *agrB*-encoded proteins share 23 and 25% amino acid identity, respectively, with the equivalent *S. aureus* orthologues. The putative 46-amino-acid *C. difficile* R20291 AgrD polypeptide shares no significant similarity with *S. aureus*. This is consistent with the highly variable nature of this peptide observed among *S. aureus* isolates (49). A RNAPIII divergent transcript has not been identified in *C. difficile*.

To determine if the *C. difficile* R20291 *agr* region is transcribed when grown under standard laboratory conditions, RT-PCR analysis of R20291 was performed using primer pairs specific to the *C. difficile* R20291 *agrACDB* coding sequences. This confirmed that the full locus is expressed at exponential (optical density at 600 nm  $[OD_{600}] = 0.3$ ) and late exponential ( $OD_{600} = 0.7$ ) phase in BHI media.

**Insertional inactivation of *C. difficile* R20291 *agrA*.** To study the role of the *agr* locus in *C. difficile* R20291, an isogenic mutant of *agrA* was constructed using the Clostron system (37). The genotype of the *C. difficile* R20291 *agrA76a::CT* mutant derivative was confirmed by PCR analysis exploiting specific primers (data not shown). Illumina sequencing of whole-genome DNA purified from the *C. difficile* R20291 *agrA76a::CT* mutant derivative revealed that no secondary mutations were acquired, and the genetic background was otherwise identical to the parental R20291 aside from the anticipated single intron insertion. Characterization of the growth kinetics of *C. difficile* R20291 *agrA76a::CT* and R20291 revealed no obvious differences under the conditions tested (see Fig. S1 in the supplemental material).

**RNA-seq analysis of the regulon in the *C. difficile* R20291 *agrA76a::CT* mutant.** As the *S. aureus* *agr* locus plays a key role in controlling the coordinated expression of virulence genes, we used RNA-seq analysis to begin to define the regulon under the control of the *C. difficile* R20291 *agr* locus. To this end, total RNA was



**FIG 1** Genetic organization of the *C. difficile* 027 *agr* locus. (a) The complete R20291 *agr* locus (upper panel) compared to the *agr* locus of *S. aureus*. Arrows indicate genes and direction of transcription. The length of each protein is indicated above (aa, amino acids), and the percent identity to the respective *S. aureus* orthologue is indicated below. The location of the group II intron target site, according to the ClosTron delivery system, used to construct the R20291 *agrA76a::CT* mutant strain is shown. (b) ClustalW2 multiple sequence alignment of AgrA from *C. difficile* and *S. aureus*. Symbols: \*, identical amino acids; :, strongly similar; ., weakly similar. AgrA shares 28% amino acid identities, as determined by BLAST search. The two domains of AgrA are outlined: 5' REC signal receiver domain, green box; 3' LytTR-DNA binding domain, red box. The structure of the *S. aureus* AgrA LytTR-DNA binding domain (residues 137 to 238) is indicated (72). The secondary structure is indicated: helices, blue bars; strands, black bars.

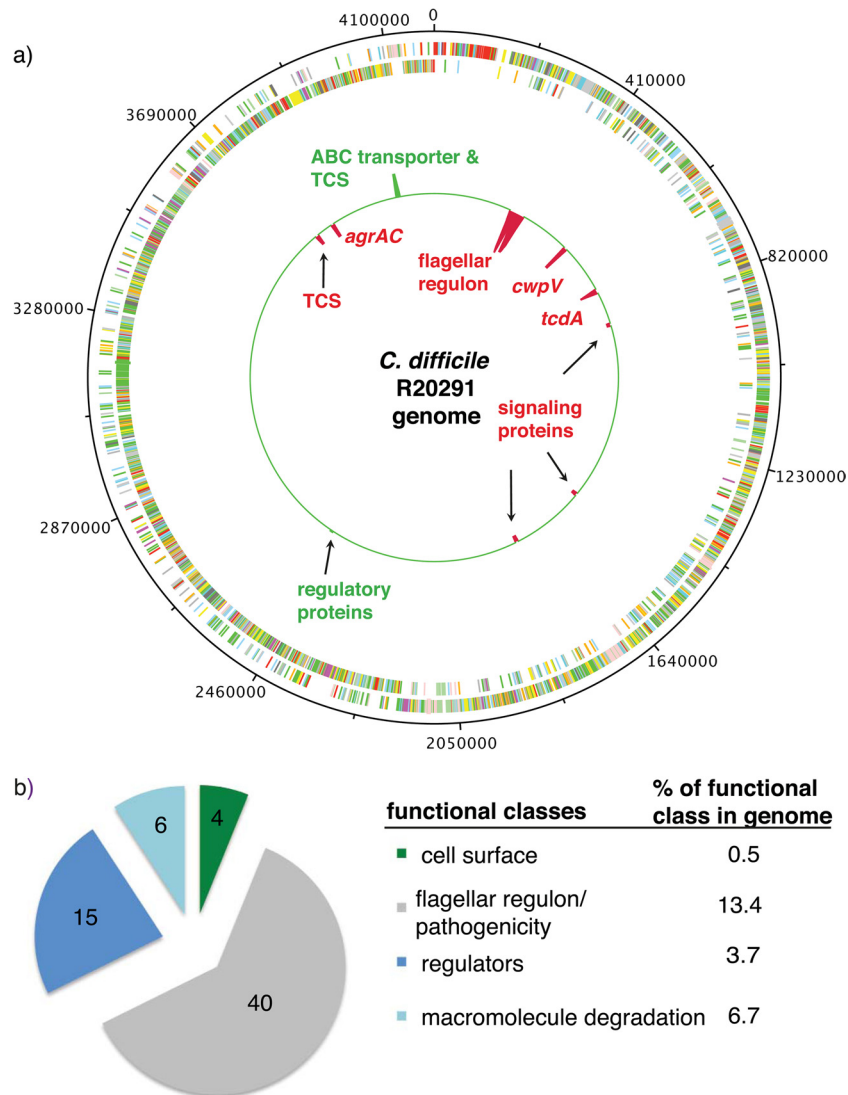
isolated from *C. difficile* R20291 and R20291 *agrA76a::CT* grown to late exponential phase. The log<sub>2</sub> fold change of transcript abundance between these RNA populations was obtained ( $P \leq 0.1$ ) (see Fig. S2 in the supplemental material), and transcripts were visualized by mapping the coverage per base pair to the reference R20291 genome (Fig. 2a). In total, 75 transcripts were found to be significantly differentially expressed in the R20291 *agrA76a::CT* mutant. The differentially regulated transcripts were categorized according to their functional class, and the enrichment of each represented class was determined. The functional classes of flagellar regulon and pathogenicity, regulators, and macromolecule degradation were enriched for differentially expressed genes in R20291 *agrA76a::CT* compared to R20291 ( $P \leq 0.05$  by hypergeometric test) (Fig. 2b). Validation of RNA sequencing results was performed by qRT-PCR. In agreement with the R20291 *agrA76a::CT* transcriptome, *tcdA*, *fliC*, and CDR20291\_1514 tran-

scripts all were underrepresented (-1.39-, -5.64-, and -2.93-fold change, respectively) compared to wild-type R20291 (Fig. 3).

The transcript of *agrC*, encoding the cognate sensor kinase directly downstream of *agrA*, was underexpressed in the R20291 *agrA76a::CT* transcriptome ( $P = 2.17 \times 10^{-18}$ ). There may be a regulatory role of AgrA in *agrC* expression causing this decrease. Alternatively, it is possible that the ClosTron insertion in the *agrA* gene affects the transcription of the subsequent gene in the locus, in this case *agrC*. Polar effects resulting from insertional inactivation using the ClosTron system have been described previously in independent mutants located in the *C. difficile* flagellar region (*fliF*, *fliG*, and *flhB-flhR*), which resulted in 8- to 20-fold down-regulation of the downstream transcript, *fliA* (25, 50).

Of the 75 differentially expressed transcripts, 64 were determined to be underrepresented in the R20291 *agrA76a::CT* transcriptome (see Table S3 in the supplemental material). Most strik-





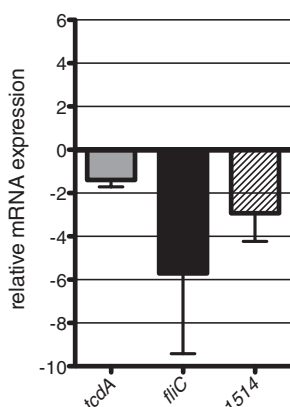
**FIG 2** Chromosomal representation of R20291 *agrA76a::CT* differentially expressed genes. (a) The  $\log_2$  fold change of differentially expressed transcripts mapped per base pair to *C. difficile* R20291 whole-genome sequence in DNA plotter (73). The outermost circles represent the CDSs of the forward strand (first circle) and the reverse strand (second circle) colored by functional classes assigned. The inner circle represents differentially expressed transcripts ( $P \leq 0.1$ ) (red peaks, underexpressed; green peaks, overexpressed). (b) Categorization of differentially expressed transcripts according to functional classes. The percentages of each functional class represented relative to the whole genome are listed. Three functional classes were enriched relative to background distributions: flagellar regulon and pathogenicity ( $P = 6.38 \times 10^{-22}$ ), regulators ( $P = 0.002$ ), and macromolecule degradation ( $P = 0.001$ ) ( $P \leq 0.05$  by hypergeometric test).

ingly, the transcripts of the majority of genes involved in flagellum formation were underexpressed. Of the 54 genes believed to be involved in flagellum formation, 50 were affected, including all of the regulatory and structural genes, with the exception of *fliN*, encoding the hypothetical C-ring protein. The flagellin, the major structural protein of flagella, encoded by *fliC*, was underexpressed 6.5-fold ( $P = 2.05 \times 10^{-20}$ ). The anti-sigma factor encoded by *flgM* and alternative sigma factor encoded by *fliA* were underrepresented 6.2- and 4.2-fold, respectively, and may be direct targets of the transcriptional regulator AgrA. Three of the six flagellin modification genes found in the 027 strains were also found to be underexpressed, including two putative glycosyl transferases, CDR20291\_0242 and CDR20291\_0243 (3.7- and 2.7-fold change, respectively).

Expression of mRNA encoding TcdA was underexpressed 2.4-

fold ( $P = 3.0 \times 10^{-4}$ ) in R20291 *agrA76a::CT* compared to that in R20291 (see Table S3 in the supplemental material). The reason for this altered expression cannot be determined from these data, but it is interesting that the *C. difficile* flagellar regulon has been reported recently to modulate toxin A production (25). TcdB was not as highly expressed as TcdA in the R20291 wild-type and R20291 *agrA76a::CT* mutant transcriptomes and was not differentially regulated in the R20291 *agrA76a::CT* transcriptome.

Ten genes annotated as regulatory proteins were differentially expressed in R20291 *agrA76a::CT* compared to wild-type R20291. These included three genes linked to c-di-GMP signaling, CDR20291\_0685, CDR20291\_1268, and CDR20291\_1514, which were underrepresented in R20291 *agrA76a::CT* and encode proteins that contain GGDEF or EAL domains (see Table S3 in the supplemental material). Furthermore, three transcripts encoding



**FIG 3** Validation of RNA-seq by quantitative real-time PCR analysis. Relative mRNA levels of transcripts corresponding to *tcdA*, *fliC*, and CDR20291\_1514. Data are from three independent experiments performed in triplicate, and error bars indicate standard deviations. Fold change of R20291 *agrA76a::CT* mutant mRNA levels are indicated relative to wild-type R20291 levels.

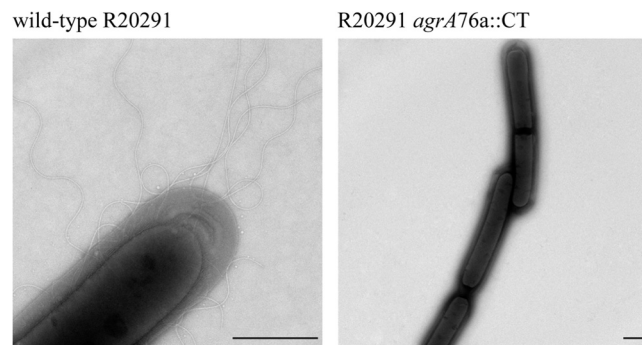
putative TCSs, CDR20291\_3126 to CDR20291\_3128, were also underexpressed in the R20291 *agrA76a::CT* transcriptome (see Table S3 in the supplemental material).

*C. difficile* is predicted to contain 234 putative small noncoding RNAs (51). In this study, 10 intergenic regions were differentially regulated in R20291 *agrA76a::CT* (see Tables S3 and S4 in the supplemental material). Each intergenic sequence was searched against a nonredundant nucleotide collection using NCBI BLAST, but no evidence of sequence conservation outside the *Clostridium* genus was found. Additionally, each sequence was searched against the Rfam database (52), but no Rfam matches were identified for 9/10 differentially regulated intergenic regions. However, the 5' untranslated intergenic region upstream of the flagellar operon where the *C. difficile* c-di-GMP riboswitch, Cd1, is located was underexpressed 5.3-fold ( $P = 1.22 \times 10^{-13}$ ). Due to its location, Cd1 is suggested to regulate flagellar biosynthesis genes and motility by responding to cyclic di-GMP concentrations (53).

**agr locus positively regulates *C. difficile* flagellar biosynthesis and TcdA in vitro.** Based on the RNA-seq transcriptome data, we hypothesized that the *C. difficile* R20291 *agrA76a::CT* mutant would be unable to form flagellar filaments. Consequently, cultures of R20291 and R20291 *agrA76a::CT* were prepared as described for the RNA-sequencing analysis and examined using electron microscopy for the formation of flagella. This analysis revealed that while peritrichous flagellar filaments were abundant on the cell surface of R20291, equivalent structures were absent from similarly treated R20291 *agrA76a::CT* cultures (Fig. 4).

A sandwich ELISA was used to determine the levels of TcdA expressed in R20291 or R20291 *agrA76a::CT* culture supernatants grown under conditions equivalent to those used for RNA preparation (Fig. 5). *C. difficile* R20291 cultures harbored detectable TcdA at 7.42 ng/ml, which is comparable to R20291 *agrA76a::CT* at 3.32 ng/ml when cultured under equivalent conditions.

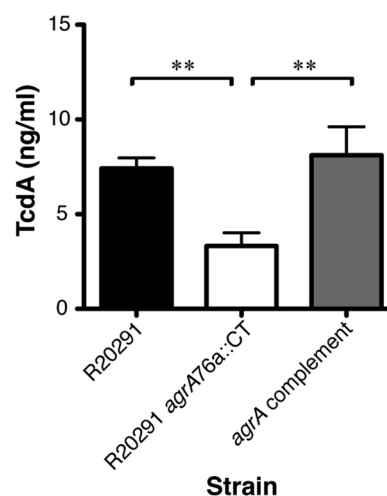
The *agrA* complementation vector (pMTL-84151-*agrA*) was introduced into the R20291 *agrA76a::CT* mutant to confirm that the observed phenotypes were the result of the ClosTron insertion in the *agrA* gene (see Materials and Methods). Growth kinetics for wild-type R20291, R20291 *agrA76a::CT*, and *agrA* complemented



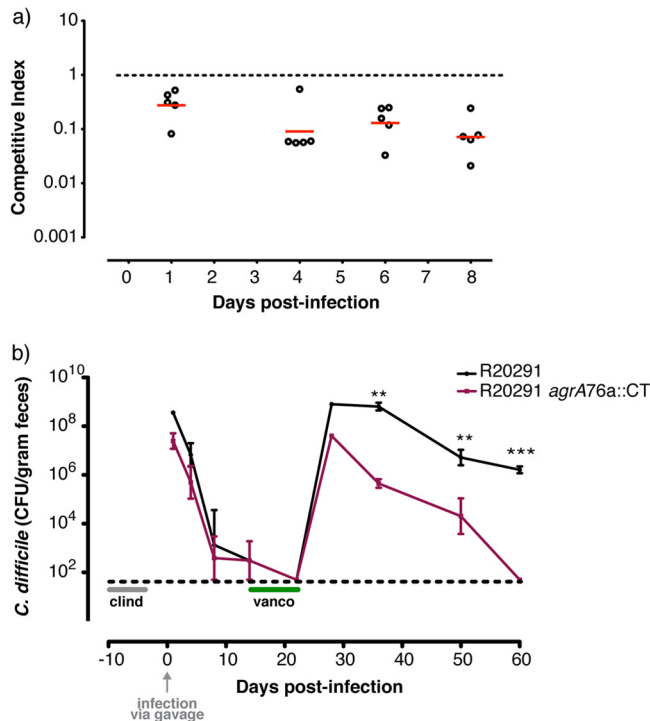
**FIG 4** Visualization of *C. difficile* flagellar filaments. Electron microscopy reveals the absence of flagella in R20291 *agrA76a::CT*. Cultures were grown to late exponential phase in BHI broth. Scale bar, 1  $\mu$ m.

strains revealed similar growth dynamics between all three strains (see Fig. S1 in the supplemental material). The *agrA* complementation vector did not restore the flagellar filaments to the cell surface in the R20291 *agrA76a::CT* mutant strain. It is possible that the complementation vector did not express wild-type levels of *agrA*; therefore, it failed to complement the flagella. However, the *agrA* complemented strain successfully restored TcdA production to 8.11 ng/ml, comparable to wild-type R20291 levels ( $P = 0.008$ ) (Fig. 5).

***C. difficile* R20291 *agrA76a::CT* exhibits an early colonization defect in mice.** *C. difficile* R20291 can establish a chronic intestinal infection in the murine model, with relapse infection consistently occurring after cessation of vancomycin treatment (54–56). To assess any role of the *agr* locus in the murine model of infection, direct fitness comparisons using competitive index (CI) experiments were performed. Here, healthy C57BL/6 mice ( $n = 5$  per group) were orally infected after clindamycin treatment with equal numbers of viable R20291 and R20291 *agrA76a::CT*



**FIG 5** *agr* locus positively regulates TcdA levels. Relative amounts of TcdA produced by *C. difficile* R20291, R20291 *agrA76a::CT*, and *agrA* complement grown to late exponential phase in BHI broth are shown. *C. difficile* R20291 produced TcdA at 7.42 ng/ml, the *agrA* strain produced it at 3.32 ng/ml, and the *agrA* complemented strain produced it at 8.11 ng/ml. Data represent four independent experiments performed in triplicate. Error bars represent standard errors of the means. Asterisks indicate statistically significant differences (\*\*,  $P < 0.01$  by Student's *t* test).



**FIG 6** *C. difficile* R20291 *agr* locus mediates efficient colonization and relapsing infection in the murine model of infection. Mice were treated with clindamycin (clind; represented as a gray line) for 7 days prior to infection via gavage with equal amounts of *C. difficile* R20291 and R20291 *agrA76a::CT*. Fecal shedding of R20291 and the R20291 *agrA76a::CT* mutant was monitored. Data are representative of two independent experiments ( $n = 5$  per group). (a) Competitive index (CI) course of R20291 and R20291 *agrA76a::CT* mutant monitored over 8 days. Open circles indicate CI values from individual mice, and the red horizontal bars represent the geometric means. R20291 *agrA76a::CT* mutant is attenuated at 1, 4, 6, and 8 days postinfection as determined by Mann-Whitney test. (b) To test whether the mutation in the R20291 *agrA* locus influences relapsing infection, mice received a 7-day course of vancomycin (vanco; represented as a green line) following the initial infection, during which fecal shedding of *C. difficile* decreased to below the detection limit (represented as a dashed horizontal line). Fecal shedding of *C. difficile* relapsed in both groups; however, R20291 *agrA76a::CT* levels were consistently reduced. Statistically significant differences are indicated by asterisks (\*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ ) and were determined by two-way analysis of variance.

spores ( $5 \times 10^6$  CFU), and *C. difficile* CFU in fecal sheddings were monitored for 8 days after the challenge. *C. difficile* R20291 *agrA76a::CT* was shed at significantly lower levels than R20291 at 1, 4, 6, and 8 days postchallenge (Fig. 6a). To determine if the *agr* locus influences relapse, the mice from similar CI challenges were treated with a clinically relevant dose of oral vancomycin for 7 days (Fig. 6b), and *C. difficile* fecal shedding was again monitored. Fecal shedding of *C. difficile* returned in both groups within 3 days of vancomycin cessation. However, while the R20291 fecal shedding levels returned to pre-vancomycin-treatment levels, those of R20291 *agrA76a::CT* were consistently reduced. Statistical analysis confirmed that R20291 *agrA76a::CT* is relatively attenuated in terms of relapse infection.

## DISCUSSION

Previous whole-genome sequencing analysis of the epidemic 027, R20291, identified the *C. difficile* *agr* locus (*agrACDB*), which showed similarity to the *agr* genes of *S. aureus*. Comparative

genomic analysis with disease-causing ribotypes 017 and 001 confirmed that this locus is not specific to ribotype 027, as previously thought (48). The *agr* genes of *C. difficile* are in the reverse order from the *agr* genes of *S. aureus*, while the *agr*-like genes of *Enterococcus faecalis* (*fsrABDC*) are also different (57). The genetic organization of *agr* genes varies between different species (57, 58). This work aimed to identify the role of this locus in the virulence of *C. difficile* R20291; therefore, we generated an insertion mutant in the transcriptional regulator, AgrA.

In this study, we report the first RNA sequencing analysis of a *C. difficile* strain to investigate the regulatory network controlled by the *agr* locus. Similar to other studies reporting transcriptional profiling of relevant *agr* loci, we investigated the regulon in the 027 *agrA* mutant when grown to late exponential phase (59). Consequently, we identified 75 genes that were differentially regulated in the R20291 *agrA76a::CT* mutant transcriptome. The genes positively and negatively regulated by the *agr* locus included flagellar biosynthesis genes, *tcdA*, c-di-GMP regulatory protein genes, and uncharacterized two-component regulatory systems.

The majority of the flagellar biosynthesis genes were underexpressed in the *C. difficile* R20291 *agrA76a::CT* transcriptome. This could explain the inability of the R20291 *agrA76a::CT* mutant to produce cell surface-anchored filaments *in vitro* as observed by TEM analysis. Experimental investigations have shown that insertional inactivation of the major flagellin subunit, FliC, results in the inability to produce flagellar filaments, resulting in a nonmotile phenotype and 10-fold less adherence to the murine cecum layer (60) (26). Interestingly, recent studies have shown that the *C. difficile* flagellar regulon has a role in the modulation of toxin A production (25). We have shown that the *tcdA* transcript is underexpressed in the R20291 *agrA76a::CT* mutant, and reduced levels of TcdA were produced *in vitro* compared to those of the parental wild-type strain, R20291. The reduced TcdA phenotype may relate to the differential expression of the flagellar regulon.

The small-molecule bacterial secondary messenger bis-(3'-5')-cyclic dimeric GMP (c-di-GMP) is an important signaling molecule in bacteria, mediating the transition between sessile and motility lifestyles (61). *C. difficile* R20291 is predicted to encode up to 31 proteins involved in c-di-GMP turnover (62). C-di-GMP is synthesized by diguanylate cyclase enzymes that contain GGDEF domains, while phosphodiesterases that contain EAL or HD-GYP domains stimulate degradation of the messenger (63). The three underexpressed c-di-GMP signaling proteins (CDR20291\_0685, CDR20291\_1268, and CDR20291\_1514) in the R20291 *agrA76a::CT* mutant transcriptome are solely phosphodiesterase enzymes due to a predicted catalytically inactive GGDEF domain. The protein orthologues have been experimentally characterized in *C. difficile* strain 630 (CD0757, CD1421, and CD1616, respectively) (62). Purified CD0757 was confirmed to have enzymatic activity when overexpressed in *Vibrio cholerae*, resulting in enhanced motility on soft agar (62). Conversely, mutating the glutamic acid residue of the EVLxR motif, important for enzymatic activity of the phosphodiesterase, abolished this enhanced motility phenotype (62). The *agr*-regulated c-di-GMP EAL-containing proteins, encoded by CD1421 and CD1616 in strain 630, also exhibited enhanced motility phenotype when overexpressed in *V. cholerae*. In agreement, Purcell and colleagues demonstrated that artificial elevation of c-di-GMP levels negatively regulated motility and flagellar biosynthesis genes (*flgB*, *fliA*, and *flgM*) in *C. difficile* 630 (64). Furthermore, our data reveal the



differential expression of novel *C. difficile* transcriptional regulators and two-component systems, such as the sensor histidine kinase and cognate response regulator encoded by CDR20291\_3424 and CDR20291\_3425, respectively. Transcriptional regulators are commonly associated with the signaling network of c-di-GMP; in *Salmonella enterica* serovar Typhimurium, the response regulator CsgD activates the expression of diguanylate cyclase-containing protein, AdrA, in turn triggering cellulose biosynthesis (65, 66). It is possible that the levels of c-di-GMP in the R20291 *agrA76a::CT* mutant are increased due to the underexpression of the EAL-containing regulatory proteins, which stimulate degradation of the small cyclic molecule. The elevated c-di-GMP levels may be negatively affecting the flagellar biosynthesis genes, similar to the findings from Purcell et al. (64).

In addition to *S. aureus*, orthologous *agr* systems have been shown to be relevant for virulence in pathogenic firmicute species (57, 67). The inactivation of the *agrA* gene of *L. monocytogenes* attenuates virulence of the bacterium in the murine model, causing a 50% lethal dose (LD<sub>50</sub>) 10-fold higher than that of the wild-type strain (67). Furthermore, isogenic mutant strains of the *agr*-like locus in *E. faecalis* were attenuated for virulence in the rabbit endophthalmitis model (68, 69), in the nematode *Caenorhabditis elegans* model (70), and in the mouse peritonitis model (57). Here, we have demonstrated that the 027 *agr* locus contributes to colonization and relapsing infection in the *C. difficile* murine infection model. The mechanism by which the R20291 *agrA76a::CT* mutant has reduced colonization and reduced relapse infection is difficult to speculate, as the mutation affects 75 genes *in vitro*. The inability of the R20291 *agrA76a::CT* mutant to form flagellar filaments may contribute to the attenuated colonization and relapse infection observed; similarly, TcdA may be necessary for efficient colonization of *C. difficile* R20291 in the murine model. A direct comparison in the murine infection model between defined *agrA*, *tcdA*, and aflagellate mutants may help to define the relative contribution of these determinants in colonization.

Whole-genome sequencing analysis of the R20291 *agrA76a::CT* mutant confirmed that the observed phenotypic differences were due solely to the insertional inactivation of the *agrA* gene and not acquired secondary mutations. The variation in the effectiveness of complementation may result from relatively low expression of AgrA from the complementing plasmid, allowing AgrA to bind only to the highest affinity sites. Furthermore, polar effects resulting from the AgrA mutation in downstream coding sequences may have inhibited effective complementation studies. The insertional deletion of *agrA* resulted in the underexpression of the downstream coding region of *agrC*, suggesting that *agrAC* form a single transcriptional unit. Similarly, the *C. acetobutylicum* *agr* cluster is also predicted to comprise two transcriptional units, *agrBD* and *agrCA* (71). However, the complete *C. difficile* *agr* locus is significantly underrepresented in the R20291 *agrA76a::CT* mutant at exponential phase in BHI broth (data not shown), suggesting that *agrACDB* is a single transcript, similar to *S. aureus* *agr* RNAlI.

In conclusion, we demonstrate that the *agr* locus modulates known *C. difficile* virulence factors *in vitro* and has a contributory role in colonization and relapse of epidemic *C. difficile* 027 *in vivo*. We propose that this is due to the transcriptional regulatory control of the *agr* locus and have demonstrated its effect on the expression of multiple determinants by RNA sequencing.

## ACKNOWLEDGMENTS

This project was funded by the Wellcome Trust (grant numbers 086418/Z, 098051, and 086418) and a Medical Research Council New Investigator Research Grant (T.D.L.; grant 93614).

We are grateful to John Heap and Nigel Minton for plasmids pMTL007C-E2 and pMTL-84151.

## REFERENCES

1. Lyras D, O'Connor JR, Howarth PM, Sambol SP, Carter GP, Phu-moonna T, Poon R, Adams V, Vedantam G, Johnson S, Gerding DN, Rood JI. 2009. Toxin B is essential for virulence of *Clostridium difficile*. *Nature* 458:1176–1179.
2. Kuehne SA, Cartman ST, Heap JT, Kelly ML, Cockayne A, Minton NP. 2010. The role of toxin A and toxin B in *Clostridium difficile* infection. *Nature* 467:711–713.
3. Shen A. 2012. *Clostridium difficile* toxins: mediators of inflammation. *J. Innate Immun.* 4:149–158.
4. Johnson S, Gerding DN. 1998. *Clostridium difficile*-associated diarrhea. *Clin. Infect. Dis.* 26:1027–1034.
5. Kelly CP, LaMont JT. 2008. *Clostridium difficile*—more difficult than ever. *N. Engl. J. Med.* 359:1932–1940.
6. McDonald LC, Killgore GE, Thompson A, Owens RC, Jr, Kazakova SV, Sambol SP, Johnson S, Gerding DN. 2005. An epidemic, toxin gene-variant strain of *Clostridium difficile*. *N. Engl. J. Med.* 353:2433–2441.
7. Hubert B, Loo VG, Bourgault AM, Poirier L, Dascal A, Fortin E, Dionne M, Lorange M. 2007. A portrait of the geographic dissemination of the *Clostridium difficile* North American pulsed-field type 1 strain and the epidemiology of *C. difficile*-associated disease in Quebec. *Clin. Infect. Dis.* 44:238–244.
8. Loo VG, Poirier L, Miller MA, Oughton M, Libman MD, Michaud S, Bourgault AM, Nguyen T, Frenette C, Kelly M, Vibien A, Brassard P, Fenn S, Dewar K, Hudson TJ, Horn R, Rene P, Monczak Y, Dascal A. 2005. A predominantly clonal multi-institutional outbreak of *Clostridium difficile*-associated diarrhea with high morbidity and mortality. *N. Engl. J. Med.* 353:2442–2449.
9. Redelings MD, Sorvillo F, Mascola L. 2007. Increase in *Clostridium difficile*-related mortality rates, United States, 1999–2004. *Emerg. Infect. Dis.* 13:1417–1419.
10. Goorhuis A, Van der Kooi T, Vaessen N, Dekker FW, Van den Berg R, Harmanus C, van den Hof S, Notermans DW, Kuijper EJ. 2007. Spread and epidemiology of *Clostridium difficile* polymerase chain reaction ribotype 027/toxinotype III in The Netherlands. *Clin. Infect. Dis.* 45:695–703.
11. Kuijper EJ, van Dissel JT, Wilcox MH. 2007. *Clostridium difficile*: changing epidemiology and new treatment options. *Curr. Opin. Infect. Dis.* 20:376–383.
12. Warny M, Pepin J, Fang A, Killgore G, Thompson A, Brazier J, Frost E, McDonald LC. 2005. Toxin production by an emerging strain of *Clostridium difficile* associated with outbreaks of severe disease in North America and Europe. *Lancet* 366:1079–1084.
13. He M, Miyajima F, Roberts P, Ellison L, Pickard DJ, Martin MJ, Connor TR, Harris SR, Fairley D, Bamford KB, D'Arc S, Brazier J, Brown D, Coia JE, Douce G, Gerding D, Kim HJ, Koh TH, Kato H, Senoh M, Louie T, Michell S, Butt E, Peacock SJ, Brown NM, Riley T, Songer G, Wilcox M, Pirmohamed M, Kuijper E, Hawkey P, Wren BW, Dougan G, Parkhill J, Lawley TD. 2012. Emergence and global spread of epidemic healthcare-associated *Clostridium difficile*. *Nat. Genet.* 45:109–113.
14. Stabler RA, He M, Dawson L, Martin M, Valiente E, Corton C, Lawley TD, Sebahia M, Quail MA, Rose G, Gerding DN, Gibert M, Popoff MR, Parkhill J, Dougan G, Wren BW. 2009. Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provides insight into the evolution of a hypervirulent bacterium. *Genome Biol.* 10:R102. doi:10.1186/gb-2011-12-10-r102.
15. Healthcare Commission. 2007. Investigation into outbreaks of *Clostridium difficile* at Maidstone and Tunbridge Wells NHS Trust. Commission for Healthcare Audit and Inspection, London, United Kingdom.
16. Sebahia M, Wren BW, Mullany P, Fairweather NE, Minton N, Stabler R, Thomson NR, Roberts AP, Cerdeno-Tarraga AM, Wang H, Holden MT, Wright A, Churcher C, Quail MA, Baker S, Bason N, Brooks K, Chillingworth T, Cronin A, Davis P, Dowd L, Fraser A, Feltwell T, Hance Z, Holroyd S, Jagels K, Moule S, Mungall K, Price C, Rabbitt



- owitsch E, Sharp S, Simmonds M, Stevens K, Unwin L, Whithead S, Dupuy B, Dougan G, Barrell B, Parkhill J. 2006. The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nat. Genet.* 38:779–786.
17. Dapa T, Leuzzi R, Ng YK, Baban ST, Adamo R, Kuehne SA, Scarselli M, Minton NP, Serruto D, Unnikrishnan M. 2013. Multiple factors modulate biofilm formation by the anaerobic pathogen *Clostridium difficile*. *J. Bacteriol.* 195:545–555.
  18. Lee AS, Song KP. 2005. LuxS/autoinducer-2 quorum sensing molecule regulates transcriptional virulence gene expression in *Clostridium difficile*. *Biochem. Biophys. Res. Commun.* 335:659–666.
  19. Antunes A, Martin-Verstraete I, Dupuy B. 2011. CcpA-mediated repression of *Clostridium difficile* toxin gene expression. *Mol. Microbiol.* 79:882–899.
  20. Dineen SS, Villapakkam AC, Nordman JT, Sonenshein AL. 2007. Repression of *Clostridium difficile* toxin gene expression by CodY. *Mol. Microbiol.* 66:206–219.
  21. Saujet L, Monot M, Dupuy B, Soutourina O, Martin-Verstraete I. 2011. The key sigma factor of transition phase, SigH, controls sporulation, metabolism, and virulence factor expression in *Clostridium difficile*. *J. Bacteriol.* 193:3186–3196.
  22. Rosenbusch KE, Bakker D, Kuijper EJ, Smits WK. 2012. *C. difficile* 630Deltaerm Spo0A regulates sporulation, but does not contribute to toxin production, by direct high-affinity binding to target DNA. *PLoS One* 7:e48608. doi:10.1371/journal.pone.0048608.
  23. Deakin LJ, Clare S, Fagan RP, Dawson LF, Pickard DJ, West MR, Wren BW, Fairweather NF, Dougan G, Lawley TD. 2012. The *Clostridium difficile* *spo0A* gene is a persistence and transmission factor. *Infect. Immun.* 80:2704–2711.
  24. Underwood S, Guan S, Vijayasubhash V, Baines SD, Graham L, Lewis RJ, Wilcox MH, Stephenson K. 2009. Characterization of the sporulation initiation pathway of *Clostridium difficile* and its role in toxin production. *J. Bacteriol.* 191:7296–7305.
  25. Aubry A, Hussack G, Chen W, KuoLee R, Twine SM, Fulton KM, Foote S, Carrillo CD, Tanha J, Logan SM. 2012. Modulation of toxin production by the flagellar regulon in *Clostridium difficile*. *Infect. Immun.* 80:3521–3532.
  26. Tasteyre A, Barc MC, Collignon A, Boureau H, Karjalainen T. 2001. Role of FlhC and FlhD flagellar proteins of *Clostridium difficile* in adherence and gut colonization. *Infect. Immun.* 69:7937–7940.
  27. Recsei P, Kreiswirth B, O'Reilly M, Schlievert P, Gruss A, Novick RP. 1986. Regulation of exoprotein gene expression in *Staphylococcus aureus* by agar. *Mol. Gen. Genet.* 202:58–61.
  28. Novick RP, Ross HF, Projan SJ, Kornblum J, Kreiswirth B, Moghazeh S. 1993. Synthesis of staphylococcal virulence factors is controlled by a regulatory RNA molecule. *EMBO J.* 12:3967–3975.
  29. Tegmark K, Morfeldt E, Arvidson S. 1998. Regulation of *agr*-dependent virulence genes in *Staphylococcus aureus* by RNAlII from coagulase-negative staphylococci. *J. Bacteriol.* 180:3181–3186.
  30. Koenig RL, Ray JL, Maleki SJ, Smeltzer MS, Hurlburt BK. 2004. *Staphylococcus aureus* AgrA binding to the RNAlII-*agr* regulatory region. *J. Bacteriol.* 186:7549–7555.
  31. Leonard PG, Bezar IF, Sidote DJ, Stock AM. 2012. Identification of a hydrophobic cleft in the LytTR domain of AgrA as a locus for small molecule interactions that inhibit DNA binding. *Biochemistry* 51:10035–10043.
  32. Marsden GL, Davis IJ, Wright VJ, Sebahia M, Kuijper EJ, Minton NP. 2010. Array comparative hybridisation reveals a high degree of similarity between UK and European clinical isolates of hypervirulent *Clostridium difficile*. *BMC Genomics* 11:389. doi:10.1186/1471-2164-11-389.
  33. Wilson KH. 1983. Efficiency of various bile salt preparations for stimulation of *Clostridium difficile* spore germination. *J. Clin. Microbiol.* 18:1017–1019.
  34. Wren BW, Tabaqchali S. 1987. Restriction endonuclease DNA analysis of *Clostridium difficile*. *J. Clin. Microbiol.* 25:2402–2404.
  35. Harris SR, Feil EJ, Holden MT, Quail MA, Nickerson EK, Chantratita N, Gardete S, Tavares A, Day N, Lindsay JA, Edgeworth JD, de Lencastre H, Parkhill J, Peacock SJ, Bentley SD. 2010. Evolution of MRSA during hospital transmission and intercontinental spread. *Science* 327:469–474.
  36. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.
  37. Heap JT, Kuehne SA, Ehsaan M, Cartman ST, Cooksley CM, Scott JC, Minton NP. 2010. The Clostron: mutagenesis in *Clostridium* refined and streamlined. *J. Microbiol. Methods* 80:49–55.
  38. Heap JT, Pennington OJ, Cartman ST, Carter GP, Minton NP. 2007. The Clostron: a universal gene knock-out system for the genus *Clostridium*. *J. Microbiol. Methods* 70:452–464.
  39. Perutka J, Wang W, Goerlitz D, Lambowitz AM. 2004. Use of computer-designed group II introns to disrupt *Escherichia coli* DExH/D-box protein and DNA helicase genes. *J. Mol. Biol.* 336:421–439.
  40. Purdy D, O'Keefe TA, Elmore M, Herbert M, McLeod A, Bokori-Brown M, Ostrowski A, Minton NP. 2002. Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier. *Mol. Microbiol.* 46:439–452.
  41. Williams DR, Young DI, Young M. 1990. Conjugative plasmid transfer from *Escherichia coli* to *Clostridium acetobutylicum*. *J. Gen. Microbiol.* 136:819–826.
  42. Heap JT, Pennington OJ, Cartman ST, Minton NP. 2009. A modular system for *Clostridium* shuttle plasmids. *J. Microbiol. Methods* 78:79–85.
  43. Perochi F, Xu Z, Clauder-Munster S, Steinmetz LM. 2007. Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D. *Nucleic Acids Res.* 35:e128.
  44. Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol.* 11:R106. doi:10.1186/gb-2010-11-10-r106.
  45. Schmittgen TD, Livak KJ. 2008. Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.* 3:1101–1108.
  46. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948.
  47. Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R. 2010. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.* 38:W695–W699.
  48. He M, Sebahia M, Lawley TD, Stabler RA, Dawson LF, Martin MJ, Holt KE, Seth-Smith HM, Quail MA, Rance R, Brooks K, Churcher C, Harris D, Bentley SD, Burrows C, Clark L, Corton C, Murray V, Rose G, Thurston S, van Tonder A, Walker D, Wren BW, Dougan G, Parkhill J. 2010. Evolutionary dynamics of *Clostridium difficile* over short and long time scales. *Proc. Natl. Acad. Sci. U. S. A.* 107:7527–7532.
  49. Ji G, Beavis R, Novick RP. 1997. Bacterial interference caused by auto-inducing peptide variants. *Science* 276:2027–2030.
  50. Cartman ST, Heap JT, Kuehne SA, Cockayne A, Minton NP. 2010. The emergence of “hypervirulence” in *Clostridium difficile*. *Int. J. Med. Microbiol.* 300:387–395.
  51. Chen Y, Indurthi DC, Jones SW, Papoutsakis ET. 2011. Small RNAs in the genus *Clostridium*. *mBio* 2:e00340-00310. doi:10.1128/mBio.00340-10.
  52. Burge SW, Daub J, Eberhardt R, Tate J, Barquist L, Nawrocki EP, Eddy SR, Gardner PP, Bateman A. 2013. Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res.* 41:D226–D232.
  53. Sudarsan N, Lee ER, Weinberg Z, Moy RH, Kim JN, Link KH, Breaker RR. 2008. Riboswitches in eubacteria sense the second messenger cyclic di-GMP. *Science* 321:411–413.
  54. Lawley TD, Clare S, Walker AW, Goulding D, Stabler RA, Croucher N, Mastroeni P, Scott P, Raisen C, Mottram L, Fairweather NF, Wren BW, Parkhill J, Dougan G. 2009. Antibiotic treatment of *Clostridium difficile* carrier mice triggers a supershedder state, spore-mediated transmission, and severe disease in immunocompromised hosts. *Infect. Immun.* 77:3661–3669.
  55. Chen X, Katchar K, Goldsmith JD, Nanthakumar N, Cheknis A, Gerdling DN, Kelly CP. 2008. A mouse model of *Clostridium difficile*-associated disease. *Gastroenterology* 135:1984–1992.
  56. Lawley TD, Clare S, Walker AW, Stares MD, Connor TR, Raisen C, Goulding D, Rad R, Schreiber F, Brandt C, Deakin LJ, Pickard DJ, Duncan SH, Flint HJ, Clark TG, Parkhill J, Dougan G. 2012. Targeted restoration of the intestinal microbiota with a simple, defined bacteriotherapy resolves relapsing *Clostridium difficile* disease in mice. *PLoS Pathog.* 8:e1002995. doi:10.1371/journal.ppat.1002995.
  57. Qin X, Singh KV, Weinstock GM, Murray BE. 2000. Effects of *Enterococcus faecalis* *fsr* genes on production of gelatinase and a serine protease and virulence. *Infect. Immun.* 68:2579–2586.
  58. Wuster A, Babu MM. 2008. Conservation and evolutionary dynamics of the *agr* cell-to-cell communication system across firmicutes. *J. Bacteriol.* 190:743–746.
  59. Bourgoigne A, Hilsenbeck SG, Dunny GM, Murray BE. 2006. Compar-

- ison of OG1RF and an isogenic *fsrB* deletion mutant by transcriptional analysis: the Fsr system of *Enterococcus faecalis* is more than the activator of gelatinase and serine protease. *J. Bacteriol.* **188**:2875–2884.
60. Dingle TC, Mulvey GL, Armstrong GD. 2011. Mutagenic analysis of the *Clostridium difficile* flagellar proteins, FliC and FliD, and their contribution to virulence in hamsters. *Infect. Immun.* **79**:4061–4067.
  61. Simm R, Morr M, Kader A, Nimtz M, Romling U. 2004. GGDEF and EAL domains inversely regulate cyclic di-GMP levels and transition from sessility to motility. *Mol. Microbiol.* **53**:1123–1134.
  62. Bordeleau E, Fortier LC, Malouin F, Burrus V. 2011. c-di-GMP turnover in *Clostridium difficile* is controlled by a plethora of diguanylate cyclases and phosphodiesterases. *PLoS Genet.* **7**:e1002039. doi:10.1371/journal.pgen.1002039.
  63. Srivastava D, Waters CM. 2012. A tangled web: regulatory connections between quorum sensing and cyclic Di-GMP. *J. Bacteriol.* **194**:4485–4493.
  64. Purcell EB, McKee RW, McBride SM, Waters CM, Tamayo R. 2012. Cyclic diguanylate inversely regulates motility and aggregation in *Clostridium difficile*. *J. Bacteriol.* **194**:3307–3316.
  65. Garcia B, Latasa C, Solano C, Garcia-del Portillo F, Gamazo C, Lasa I. 2004. Role of the GGDEF protein family in *Salmonella* cellulose biosynthesis and biofilm formation. *Mol. Microbiol.* **54**:264–277.
  66. Zogaj X, Nimtz M, Rohde M, Bokranz W, Romling U. 2001. The multicellular morphotypes of *Salmonella typhimurium* and *Escherichia coli* produce cellulose as the second component of the extracellular matrix. *Mol. Microbiol.* **39**:1452–1463.
  67. Autret N, Raynaud C, Dubail I, Berche P, Charbit A. 2003. Identification of the *agr* locus of *Listeria monocytogenes*: role in bacterial virulence. *Infect. Immun.* **71**:4463–4471.
  68. Engelbert M, Mylonakis E, Ausubel FM, Calderwood SB, Gilmore MS. 2004. Contribution of gelatinase, serine protease, and *fsr* to the pathogenesis of *Enterococcus faecalis* endophthalmitis. *Infect. Immun.* **72**:3628–3633.
  69. Mylonakis E, Engelbert M, Qin X, Sifri CD, Murray BE, Ausubel FM, Gilmore MS, Calderwood SB. 2002. The *Enterococcus faecalis* *fsrB* gene, a key component of the *fsr* quorum-sensing system, is associated with virulence in the rabbit endophthalmitis model. *Infect. Immun.* **70**:4678–4681.
  70. Sifri CD, Mylonakis E, Singh KV, Qin X, Garsin DA, Murray BE, Ausubel FM, Calderwood SB. 2002. Virulence effect of *Enterococcus faecalis* protease genes and the quorum-sensing locus *fsr* in *Caenorhabditis elegans* and mice. *Infect. Immun.* **70**:5647–5650.
  71. Steiner E, Scott J, Minton NP, Winzer K. 2012. An *agr* quorum sensing system that regulates granule formation and sporulation in *Clostridium acetobutylicum*. *Appl. Environ. Microbiol.* **78**:1113–1122.
  72. Sidote DJ, Barbieri CM, Wu T, Stock AM. 2008. Structure of the *Staphylococcus aureus* AgrA LytTR domain bound to DNA reveals a beta fold with an unusual mode of binding. *Structure* **16**:727–735.
  73. Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. 2009. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* **25**:119–120.