

# High-Throughput Analysis of Gene Essentiality and Sporulation in *Clostridium difficile*

Marcin Dembek,<sup>a,b</sup> Lars Barquist,<sup>c</sup> Christine J. Boinett,<sup>d</sup> Amy K. Cain,<sup>d</sup> Matthew Mayho,<sup>d</sup> Trevor D. Lawley,<sup>d</sup> Neil F. Fairweather,<sup>a</sup> Robert P. Fagan<sup>e</sup>

MRC Centre for Molecular Bacteriology and Infection, Imperial College London, London, United Kingdom<sup>a</sup>; Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne, United Kingdom<sup>b</sup>; Institute for Molecular Infection Biology, University of Würzburg, Würzburg, Germany<sup>c</sup>; Wellcome Trust Sanger Institute, Hinxton, United Kingdom<sup>d</sup>; Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield, United Kingdom<sup>e</sup>

**ABSTRACT** *Clostridium difficile* is the most common cause of antibiotic-associated intestinal infections and a significant cause of morbidity and mortality. Infection with *C. difficile* requires disruption of the intestinal microbiota, most commonly by antibiotic usage. Therapeutic intervention largely relies on a small number of broad-spectrum antibiotics, which further exacerbate intestinal dysbiosis and leave the patient acutely sensitive to reinfection. Development of novel targeted therapeutic interventions will require a detailed knowledge of essential cellular processes, which represent attractive targets, and species-specific processes, such as bacterial sporulation. Our knowledge of the genetic basis of *C. difficile* infection has been hampered by a lack of genetic tools, although recent developments have made some headway in addressing this limitation. Here we describe the development of a method for rapidly generating large numbers of transposon mutants in clinically important strains of *C. difficile*. We validated our transposon mutagenesis approach in a model strain of *C. difficile* and then generated a comprehensive transposon library in the highly virulent epidemic strain R20291 (027/BI/NAP1) containing more than 70,000 unique mutants. Using transposon-directed insertion site sequencing (TraDIS), we have identified a core set of 404 essential genes, required for growth *in vitro*. We then applied this technique to the process of sporulation, an absolute requirement for *C. difficile* transmission and pathogenesis, identifying 798 genes that are likely to impact spore production. The data generated in this study will form a valuable resource for the community and inform future research on this important human pathogen.

**IMPORTANCE** *Clostridium difficile* is a common cause of potentially fatal intestinal infections in hospital patients, particularly those who have been treated with antibiotics. Our knowledge of this bacterium has been hampered by a lack of tools for dissecting the organism. We have developed a method to study the function of every gene in the bacterium simultaneously. Using this tool, we have identified a set of 404 genes that are required for growth of the bacteria in the laboratory. *C. difficile* also produces a highly resistant spore that can survive in the environment for a long time and is a requirement for transmission of the bacteria between patients. We have applied our genetic tool to identify all of the genes required for production of a spore. All of these genes represent attractive targets for new drugs to treat infection.

Received 26 November 2014 Accepted 13 January 2015 Published 24 February 2015

**Citation** Dembek M, Barquist L, Boinett CJ, Cain AK, Mayho M, Lawley TD, Fairweather NF, Fagan RP. 2015. High-throughput analysis of gene essentiality and sporulation in *Clostridium difficile*. mBio 6(2):02383-14. doi:10.1128/mBio.02383-14.

**Editor** R. John Collier, Harvard Medical School

**Copyright** © 2015 Dembek et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

Address correspondence to Neil F. Fairweather, fairweather@imperial.ac.uk.

*Clostridium difficile* is a Gram-positive, spore-forming anaerobe and the leading cause of nosocomial, antibiotic-associated diarrhea (1). Infection typically occurs among patients whose intestinal microbiota has been disrupted by prolonged treatment with broad-spectrum antibiotics, allowing the pathogen to colonize the compromised gastrointestinal tract (2). *C. difficile* infection (CDI) is most common in the nosocomial setting, but the incidence of community-associated infections is increasing (3). Two large cytotoxins are responsible for the clinical manifestations of CDI, which range from mild, self-limiting diarrhea to severe, often fatal inflammatory complications, such as pseudomembranous colitis or toxic megacolon (4).

The spore is the primary infectious agent, and recent studies have shown that a mutant strain of *C. difficile* unable to produce Spo0A, the master regulator of sporulation, is unable to efficiently

persist in the environment and transmit disease (5). Due to their multilayered structure, spores are extremely robust and resistant to both chemical and physical insults, thus providing the mechanism by which *C. difficile* evades the potentially fatal consequences of exposure to heat, oxygen, alcohol, and certain disinfectants (6). Spores shed in feces are therefore difficult to eradicate and can persist in health care facilities for extended periods of time leading to infection or reinfection of individuals through inadvertent ingestion of contaminated material (7).

Sporulation is an ancient bacterial cell differentiation program that is largely conserved among *Clostridiales* and *Bacillales*, particularly with regard to the key regulatory components Spo0A and the four sporulation-specific sigma factors,  $\sigma^E$ ,  $\sigma^F$ ,  $\sigma^G$ , and  $\sigma^K$  (reviewed in reference 8). However, recent studies have highlighted several notable differences in the sporulation programs of

*Bacillus* and *Clostridium* species. While both Spo0A and  $\sigma^H$  are present and required for sporulation in *C. difficile* (9), phosphorylation of Spo0A in *C. difficile* involves a simple two-component system unlike the complex phosphorelay that modulates Spo0A activity in *Bacillus subtilis* (8, 10). In *C. difficile*, the main periods of activity of the 4 cell type-specific sigma factors are similar but not identical to those in the *B. subtilis* model, with  $\sigma^F$  and  $\sigma^E$  controlling early stages of development and  $\sigma^G$  and  $\sigma^K$  governing late developmental events. However, the temporal segregation between the activities of the early and late-stage sigma factors is less stringent, and the cross talk between the forespore and the mother cell appears to be weaker, as the activity of  $\sigma^E$  is partially independent of  $\sigma^F$ , and activation of  $\sigma^G$  and  $\sigma^K$  does not require  $\sigma^E$  and  $\sigma^G$ , respectively (11–13). The *C. difficile*  $\sigma^E$ ,  $\sigma^F$ ,  $\sigma^G$ , and  $\sigma^K$  regulons have now been identified (12, 13), each containing key representatives of the homologous pathways in *B. subtilis* (14). This core set of 228 sporulation genes corresponds to about half the number of genes under the control of cell type-specific sigma factors in *B. subtilis* (15–17). This may be a reflection of the more ancestral mechanism of sporulation proposed for clostridia (8, 18).

Understanding the genetic and molecular bases of *C. difficile* pathogenicity is a crucial step in the development of effective therapeutics. A number of methods for directed gene inactivation are now available (19–22), allowing for reverse genetic studies, in which the exact role of a gene, hypothesized to be important in a particular phenotype, can be elucidated experimentally. However, reverse genetic studies are limited in that they are always based on pre-made assumptions regarding the function of a particular gene. An alternative approach involves using random mutant libraries aimed at identifying the genetic basis of a particular phenotype without making any assumptions about the genes involved (23). This is typically achieved through transposon-mediated mutagenesis, creating random mutant pools that can then be screened to identify genes involved in a particular phenotype across the whole genome. Random transposon mutagenesis can then be combined with high-throughput sequencing methods to allow simultaneous screening of saturating transposon libraries without the need to isolate individual mutants (reviewed in references 24 and 25). The *mariner* transposable element inserts randomly into TA target sites via a cut-and-paste mechanism (26), making it particularly useful in low-GC organisms such as *C. difficile*. A *mariner*-based transposon delivery system has been developed for use in *C. difficile* (27). However, the relative inefficiency of plasmid delivery into *C. difficile* by conjugation and a lack of effective control of the timing of transposition limited the practical size of the transposon library that could be generated.

Here we describe the construction of the first comprehensive high-density transposon mutant library in *C. difficile* using a novel conditional *mariner* delivery vector. We have applied transposon-directed insertion site sequencing (TraDIS [28]) to identify the core *C. difficile* essential genome required for *in vitro* growth and have also identified a large number of genes required for sporulation.

## RESULTS

**Construction of an efficient transposon mutagenesis system for *C. difficile*.** Early versions of a tetracycline-inducible expression system for *C. difficile* (29) displayed plasmid instability upon addition of anhydrotetracycline, a nonantibiotic analogue of tetra-

cycline. In this study, we have exploited this serendipitous plasmid instability to generate a conditional *C. difficile* plasmid for the delivery of a *mariner* transposon. The tetracycline-inducible system consists of a pair of divergent promoters, each with an overlapping *tet* operator sequence (TetO). One of the promoters,  $P_{tetR}$ , drives expression of *tetR*, encoding a transcriptional regulator that binds to the *tet* operators to repress both promoters. The second promoter,  $P_{tet}$ , is available to drive the expression of any gene of interest. Tetracycline (or anhydrotetracycline) induces a conformational change in TetR that prevents binding to the operator and relieves repression (30). Under normal conditions, TetR is expressed at low levels, maintaining repression of the system, and induction with tetracycline greatly increases transcription from both promoters. In plasmid pRPF177, the  $P_{tetR}$  promoter is oriented toward the *C. difficile* pCD6 origin of replication (see Fig. S1A in the supplemental material [31]), and this plasmid displays induction-dependent instability. Addition of a transcriptional terminator sequence between the *tetR* gene and the CD6 origin of replication (pRPF185 [29]) prevents this tetracycline-dependent instability, demonstrating that transcriptional read-through from  $P_{tetR}$  into the origin of replication causes the instability.

The conditional plasmid pRPF177 was used as the basis for the construction of a *mariner* delivery system. A codon-optimized *Himar1* transposase gene was cloned downstream of the  $P_{tet}$  promoter, and a custom transposon was assembled by adding the *mariner* inverted terminal repeat sequences (32) either side of an *ermB* erythromycin resistance gene. The resulting *mariner* delivery plasmid, pRPF215, retains tetracycline-dependent conditional replication while allowing tightly regulated transposition of the *ermB* transposon (see Fig. S1B in the supplemental material).

To confirm tetracycline-dependent conditionality, a single culture of *C. difficile* 630 $\Delta erm$  carrying pRPF215 was divided in two and then grown in TY broth (48) with and without anhydrotetracycline. After 13 generations, bacteria were spread on BHI agar with or without thiamphenicol to determine the proportion of cells retaining the plasmid. In the culture without anhydrotetracycline, more than 40% of bacteria were thiamphenicol resistant after 13 generations, whereas in the anhydrotetracycline-induced culture, no thiamphenicol-resistant bacteria were detectable (limit of detection, 10 CFU/ml; data not shown).

In order to determine the frequency of transposition, *C. difficile* strain 630 $\Delta erm$  carrying pRPF215 was grown to mid-logarithmic phase in TY broth, and dilutions were spread on BHI agar plates containing either thiamphenicol to select for the plasmid or erythromycin and anhydrotetracycline to select for transposon mutants. Erythromycin-resistant mutants arose with a frequency of  $1.18 \times 10^{-4}$ . It was also observed that induction of transposition resulted in the production of a wide range of colony morphologies, suggesting successful random transfer of the transposon (see Fig. S1C and S1D in the supplemental material). The frequency of transposition was higher in strain R20291 ( $1.5 \times 10^{-3}$ ) than in strain 630 $\Delta erm$ . To determine the frequency of spontaneous erythromycin resistance or mutation of the conditional plasmid replicon, an analogous experiment was carried out with *C. difficile* strain 630 $\Delta erm$  carrying plasmid pRPF222, encoding a nonfunctional *Himar1* transposase. Erythromycin-resistant colonies arose with a frequency of  $5.9 \times 10^{-8}$ ; therefore, the frequency of spontaneous *Erm*<sup>r</sup> mutants in a transposon library would be expected to be of the order of  $4 \times 10^{-4}$ .

### Generation of a large *C. difficile* transposon mutant library.

In order to test the suitability of pRPF215 to generate large transposon libraries in *C. difficile*, a proof-of-principle library was constructed in model strain 630 $\Delta$ *erm*. Individual mutants were generated on solid media, and approximately 85,000 erythromycin-resistant colonies were pooled. Following genomic DNA (gDNA) isolation from the pooled library, transposon insertion sites were identified using TraDIS following the method developed by Langridge et al. (28). A total of 44,102 unique insertion sites were identified, an average of one insertion every 97 bp. Of the 3,897 annotated coding DNA sequences (CDSs), 736 showed no transposon insertions; however, the insertion density attained did not appear sufficient to saturate nonessential genes (Fig. 1A and B). Despite not reaching saturation, this proof-of-principle experiment clearly demonstrated the ease with which large transposon libraries can be created in *C. difficile*. To identify genes involved in sporulation and germination among the 3,161 CDSs containing transposon insertions, the library was sporulated on solid media, and the spores were purified on a HistoDenz gradient (see Fig. S2A in the supplemental material) and germinated in broth containing the bile acid germinant taurocholate. gDNA was extracted from the spore, and germinated libraries and transposon insertion sites were identified. Mutants incapable of sporulation would be absent from the purified spore population. Likewise, mutants that were capable of producing spores but incapable of germination would be absent from the germination library (Fig. S2B).

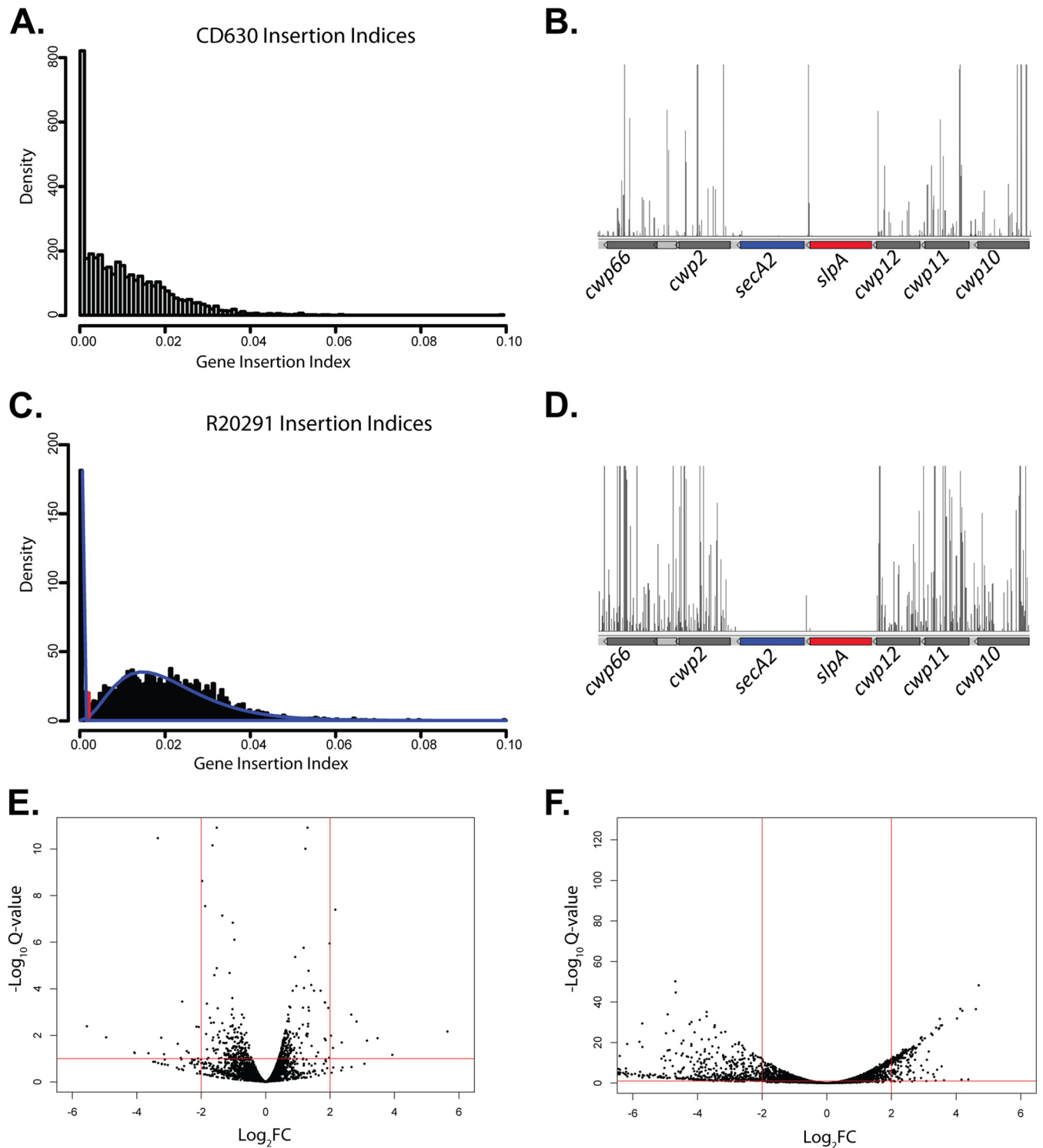
**Analysis of individual mutants.** In order to validate the identification of genes required for sporulation and germination, genes with fewer observed transposon insertions following sporulation or germination were identified in the 630 $\Delta$ *erm* library by manual inspection of insertion plots using Artemis (33). Candidate genes with clear differences in insertion frequency were further analyzed using BLAST. Three candidate genes (CD630\_01250, CD630\_34940, and CD630\_35670) that appeared to play a role in sporulation and one candidate gene (CD630\_01060) that appeared to play a role in germination were chosen for further analysis. Each of these genes was annotated as or displayed homology to genes implicated in sporulation or germination in other species.

CD630\_01060 encodes a putative *N*-acetylmuramoyl-L-alanine amidase that shares 37% amino acid sequence identity with CwlD from *Bacillus subtilis* strain 168. CwlD has been shown to be essential for muramic- $\delta$ -lactam formation during sporulation in *Bacillus* (34). CD630\_01250 encodes a putative endopeptidase that shares 21.4% amino acid sequence identity with *B. subtilis* SpoIIQ, a  $\sigma^F$ -dependent membrane protein that localizes to the forespore membrane and is essential for forespore integrity and late forespore gene expression in *B. subtilis* (35). CD630\_34940 encodes a putative spore coat-associated protein sharing 32.6% amino acid sequence identity with *B. subtilis* YabP, disruption of which causes a late spore morphogenesis defect (36). Finally, CD630\_35670 (*sipL*) encodes a 58.7-kDa protein, identified as a functional homologue of SpoVID, a key morphogenetic protein that directly interacts with SpoIVA and is required for the correct attachment of the spore coat to the forespore (37).

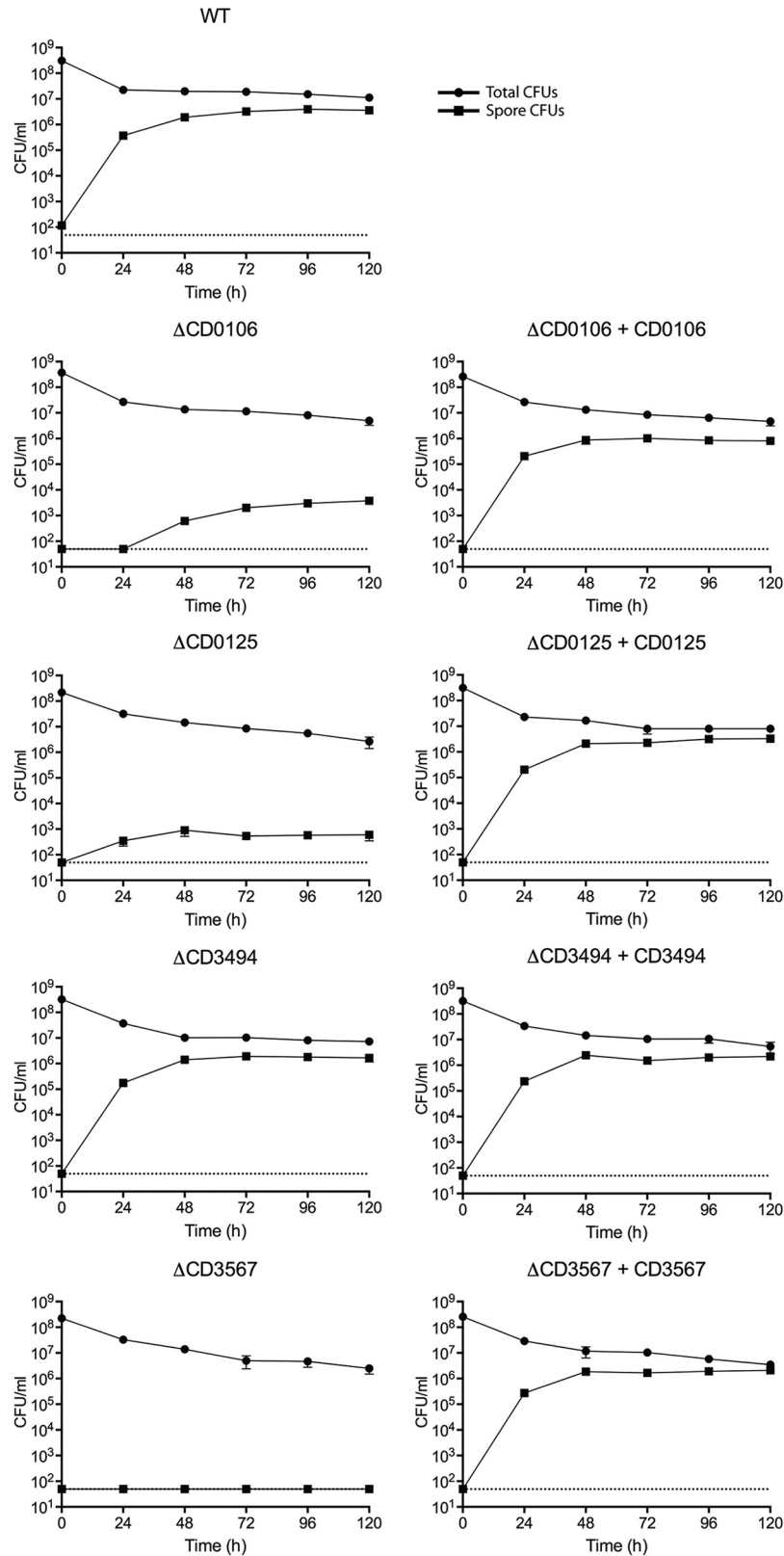
To confirm the role of each gene, four isogenic deletion strains were constructed in strain 630 $\Delta$ *erm*, and each mutant was complemented by restoration of the native gene onto the chromosome by integration at the *pyrE* locus. No significant differences in growth rate were observed for the deletion strains compared to the

parental strain 630 $\Delta$ *erm* (see Fig. S3 in the supplemental material). The ability of each mutant to produce viable spores was then analyzed using a broth-based sporulation assay (Fig. 2), counting the total number of CFU and CFU following heat treatment (viable spores). Surprisingly, no significant decrease in sporulation efficiency was observed for the  $\Delta$ CD630\_34940 mutant despite a lack of transposon insertions in this gene in the 630 $\Delta$ *erm* spore library. All remaining mutants produced significantly fewer viable spores than the wild-type strain. The  $\Delta$ CD630\_01060 and  $\Delta$ CD630\_01250 mutants produced 1,000-fold fewer spores than the wild type, and there was also an apparent delay in the initiation of sporulation in the  $\Delta$ CD630\_01060 mutant, with no spores detected before the 48-h time point. The phenotype of the  $\Delta$ CD630\_35670 mutant was even more severe, with no viable spores detected for the duration of the experiment. In each case, complementation at the *pyrE* locus fully restored sporulation to the wild-type level. In order to differentiate between defects in sporulation and germination, production of spores was followed using phase-contrast and fluorescence microscopy (Fig. 3). As predicted from the broth-based sporulation assay, the  $\Delta$ CD630\_34940 mutant appeared to sporulate as normal. In contrast, the  $\Delta$ CD630\_01250 and  $\Delta$ CD630\_35670 mutants both appeared to initiate spore morphogenesis, producing phase-dark immature spores, but neither mutant was capable of producing normal phase-bright mature spores. This phenotype has been previously described for an insertional disruption mutant of CD630\_35670 (37). Interestingly, the  $\Delta$ CD630\_01060 mutant appeared to sporulate normally despite the dramatic decrease in viable spore numbers observed previously. This suggested that the  $\Delta$ CD630\_01060 mutant was defective in germination rather than sporulation. To examine this in greater detail, spores were purified, and taurocholate-induced germination was monitored by measuring the decrease in optical density resulting from Ca<sup>2+</sup>-dipicolinic acid (DPA) release and using phase-contrast microscopy (Fig. 4). The  $\Delta$ CD630\_01060 mutant showed no response to taurocholate induction, confirming a germination defect. Germination was restored to wild-type levels upon complementation with the CD630\_01060 gene at the *pyrE* locus.

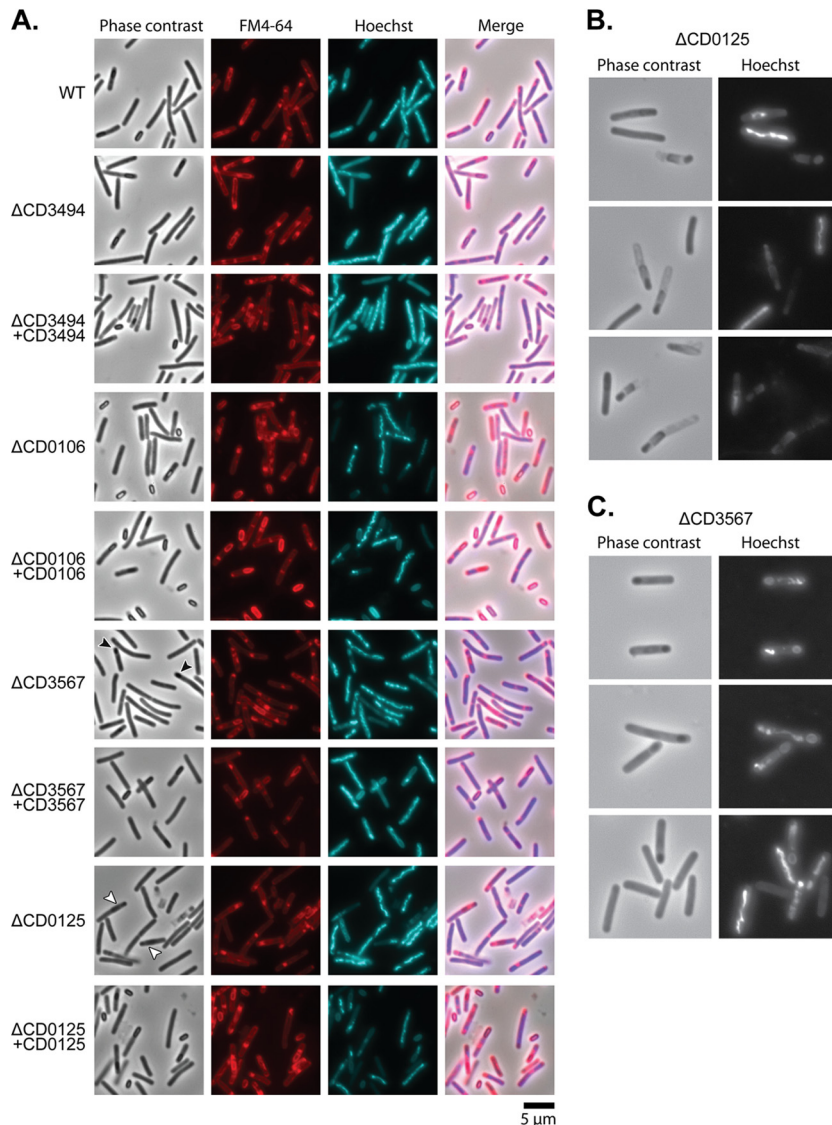
**Identification of essential genes in *C. difficile* R20291.** A higher-density transposon library was required to allow identification of essential genes with greater statistical power. The strain chosen was a representative of the clinically significant 027/NAP1 lineage, R20291, isolated during the first major ribotype 027 outbreak in the United Kingdom (38). In order to increase the insertion density, a total of approximately 750,000 erythromycin-resistant transposon mutants were pooled. TraDIS analysis of the library identified 77,256 unique insertion sites, an average of one insertion every 54 bp. Genes essential for growth *in vitro* under the conditions of this assay would be expected to have no transposon insertions in the library (e.g., Fig. 1D). To identify essential genes across the whole genome, an insertion index was calculated for each gene, normalizing the number of insertions to the length of the gene. A plot of these insertion indices displays a clear bimodal distribution, with the first, narrow peak indicating essential genes and the elongated second peak indicating genes that tolerate transposon insertions (Fig. 1C). Statistical analysis of these data (39) identified 404 essential genes and a further 33 genes that fell into the ambiguous region between essential and nonessential cutoffs (see Table S1 in the supplemental material). KEGG (Kyoto Encyclopedia of Genes and Genomes) analysis identified 25 path-



**FIG 1** Identification of sporulation-specific genes in *C. difficile*. (A and C) Insertion index distribution plots from the *C. difficile* 630 $\Delta$ erm (A) and R20291 (C) transposon libraries. The clear bimodal distribution in the R20291 plot (fitted gamma distributions in blue) allows differentiation of genes that do not tolerate transposon insertions in the sharp left-hand peak and genes containing insertions in the broad right-hand peak. (B and D) S-layer biogenesis locus. Each vertical line indicates the location of a unique transposon insertion, with height indicating the total number of Illumina reads at each insertion point. The *secA2* and *slpA* genes show no transposon insertions, which suggests that they are required for growth *in vitro*. (E and F) Changes in prevalence of the mutants from the initial transposon library after overnight growth (E) or sporulation (F). Red lines show the cutoff criteria of 10% false discovery rate (horizontal) and a log<sub>2</sub> fold change (Log<sub>2</sub>FC) of 2 (vertical).



**FIG 2** Sporulation efficiency of individual isogenic mutants. CD0106, CD0125, CD3494, and CD3567 deletion mutants were compared to strains complemented on the chromosome at the *pyrE* locus and the parental 630 $\Delta$ *erm* strain (wild type [WT]). Stationary-phase cultures in BHIS broth were incubated anaerobically for 5 days, and samples were taken at 24-h intervals for analysis. Total cell numbers were determined by counting the number of CFU on BHIS agar containing the germinant taurocholate. Spore numbers were determined by the same method following incubation at 70°C for 30 min. Experiments were carried out in triplicate on biological duplicate samples. The means  $\pm$  standard deviations (error bars) are shown.

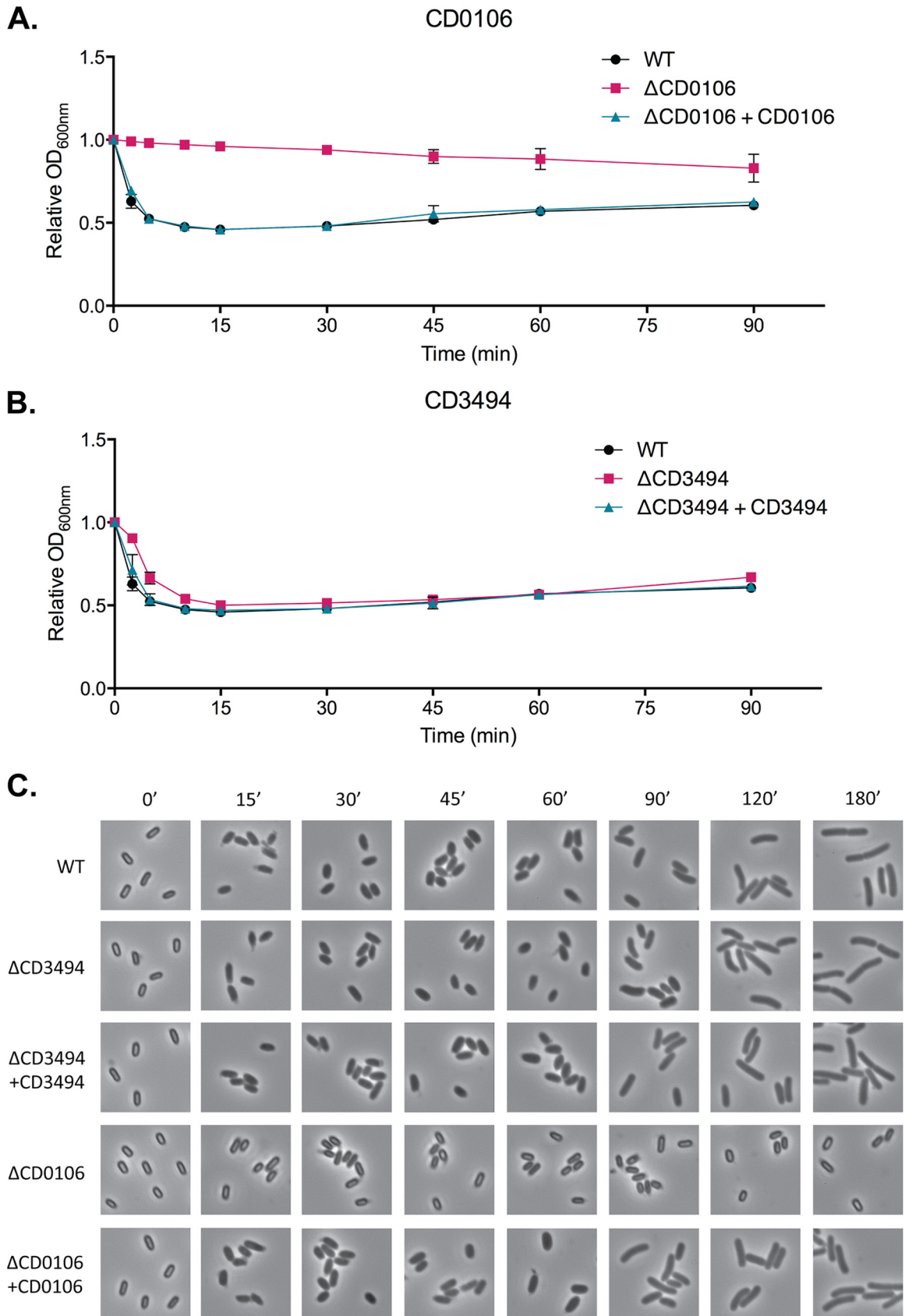


**FIG 3** Microscopic examination of deletion mutants and complemented strains. (A) Cultures were sporulated on SMC agar for 24 h, harvested in PBS, and fluorescently labeled with a membrane stain (FM4-64) and DNA stain (Hoechst 33258). Both the  $\Delta$ CD0106 and  $\Delta$ CD3494 mutants appeared to produce normal phase-bright spores. The  $\Delta$ CD0125 mutant failed to produce visible mature spores, and sporulation appeared to halt at forespore formation (white arrowheads). The  $\Delta$ CD3567 mutant also did not produce mature spores, and only small, phase-dark spores were observed (black arrowheads). Complementation of the  $\Delta$ CD0125 and  $\Delta$ CD3567 mutants restored normal spore production. (B and C) Higher magnification micrographs of  $\Delta$ CD0125 and  $\Delta$ CD3567 mutants showing severe sporulation defects.

ways that were significantly enriched in essential genes (Table 1). As expected, these included genes involved in DNA replication (9 of 13 are essential genes) and mismatch repair (7 of 18 are essential genes), transcription (all 4 RNA polymerase subunits), and translation machinery (46 of 52 ribosomal genes), and many core metabolic pathways, including, for example, the genes required for tRNA biosynthesis (23 of 27 are essential genes), fatty acid biosynthesis (all 11 are essential genes), and peptidoglycan biosynthesis (12 of 19 are essential genes). The conserved core components of the Sec secretion machinery were also identified as essential, including the genes encoding both the housekeeping translocation ATPase SecA1 and the accessory ATPase SecA2. We have previously reported that *secA2* is essential (29) and identified the S-layer precursor as the major substrate of the *C. difficile* accessory Sec system. Interestingly, *slpA*, encoding the S-layer precursor, was

also found to be essential in this screen, suggesting that the essentiality of the accessory Sec system may be due to the essentiality of its major substrate. No insertions were seen in a region of the genome encompassing R20291\_2657 to R20291\_2671, located downstream of the locus encoding *slpA* and other cell wall proteins. This cluster encodes many putative glycosyltransferases and suggests that *C. difficile* elaborates a surface polysaccharide that is required for cell growth. Intriguingly, 53 of the identified essential genes encode proteins with no obvious function based on homology, including 25 hypothetical proteins, 17 conserved hypothetical proteins, 8 hypothetical lipoproteins or membrane proteins, and 3 uncharacterized proteins.

**Identification of genes required for sporulation.** In order to identify genes required for sporulation from among the nonessential genes, the pooled R20291 library was grown overnight, sporu-



**FIG 4** Germination efficiency of CD0106 and CD3494 isogenic deletion mutants. Spores of the parental strain 630Δerm (WT), ΔCD0106, ΔCD3494, and complemented strains were produced on solid media and purified. (A and B) Purified spores were resuspended in BHIS supplemented with the germinant  
(Continued)

TABLE 1 Pathway analysis of essential genes

KEGG pathway	No. of essential genes	Total no. of genes	Adjusted <i>P</i> value
Ribosomal	46	52	$9.26 \times 10^{-37}$
Aminoacyl-tRNA biosynthesis	23	27	$3.00 \times 10^{-17}$
Metabolic pathways	104	489	$2.53 \times 10^{-11}$
Fatty acid metabolism	12	16	$6.41 \times 10^{-8}$
Peptidoglycan biosynthesis	12	19	$1.08 \times 10^{-6}$
DNA replication	9	13	$1.44 \times 10^{-5}$
Purine metabolism	18	52	$5.85 \times 10^{-5}$
Terpenoid backbone biosynthesis	9	14	$3.20 \times 10^{-4}$
RNA polymerase	4	4	$1.26 \times 10^{-3}$
Protein export	7	13	$1.8 \times 10^{-3}$
Lysine biosynthesis	9	21	$1.41 \times 10^{-3}$
Pyrimidine metabolism	14	47	$2.22 \times 10^{-3}$
Homologous recombination	8	18	$2.22 \times 10^{-3}$
Secondary metabolite biosynthesis	37	200	$5.11 \times 10^{-3}$
Oxidative phosphorylation	8	21	$6.33 \times 10^{-3}$
Mismatch repair	7	18	$1.07 \times 10^{-2}$
Folate biosynthesis	5	11	$2.08 \times 10^{-2}$
Carbon metabolism	18	86	$2.25 \times 10^{-2}$
D-Glutamine and D-glutamate metabolism	3	4	$2.25 \times 10^{-2}$
Glycerophospholipid metabolism	6	16	$2.27 \times 10^{-2}$
Microbial metabolism in diverse environments	25	137	$2.80 \times 10^{-2}$
Biotin metabolism	4	8	$2.80 \times 10^{-2}$
Propanoate metabolism	6	17	$2.81 \times 10^{-2}$
Pantothenate and coenzyme A biosynthesis	5	13	$3.51 \times 10^{-2}$
Nicotinate and nicotinamide metabolism	4	9	$4.09 \times 10^{-2}$

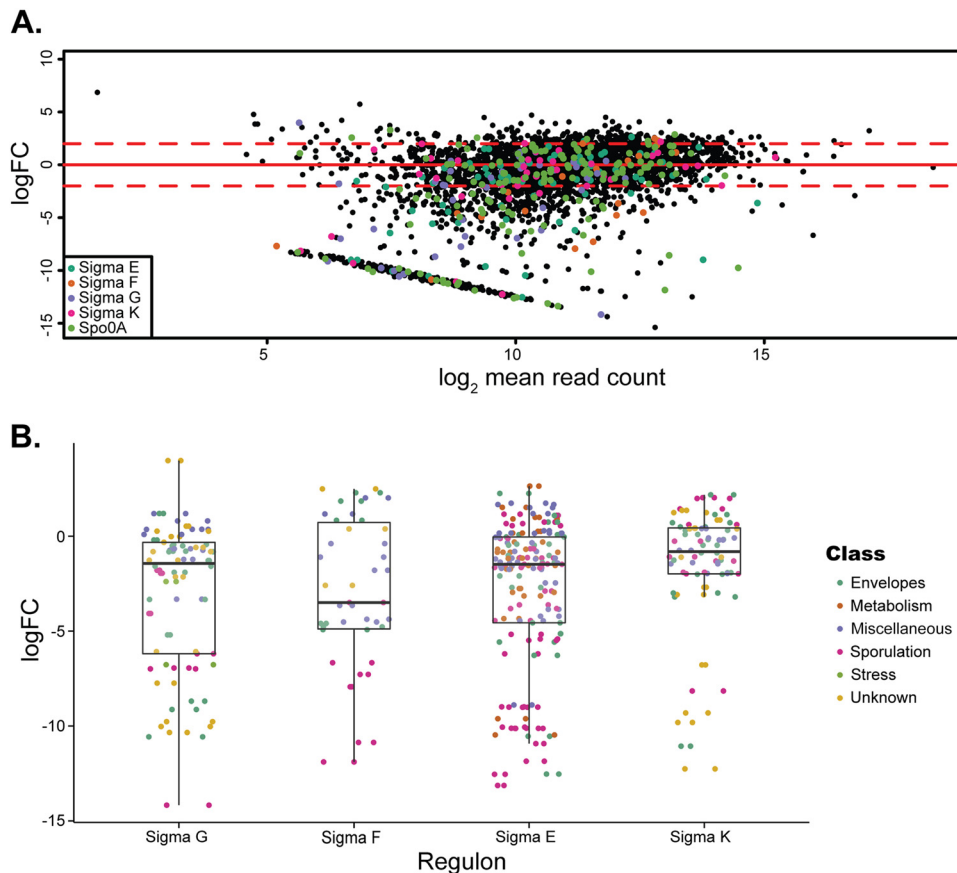
lated on solid media, spores were purified on a HistoDenz gradient, and gDNA was extracted in duplicate from the overnight culture and pure spores. Transposon insertion sites were then identified using TraDIS and analyzed in comparison to the initial library. During production of spores from the initial transposon library, mutants could be lost through either selective or stochastic processes. To limit false discoveries, a stringent cutoff was applied to the analysis; only mutants with at least 20 reads present in both replicates were included. This cutoff limited the number of genes that could be assayed in each sample; of the 3,222 genes with transposon insertions in the initial library, 3,219 in the overnight culture samples and 3,168 in the spore samples satisfied these criteria and were assayable. Following overnight growth, only 392 genes showed statistically significant changes in the prevalence of the mutant at a false discovery rate (FDR) of 10%; however, in general, the magnitudes of these changes were not large, only 36 genes had a statistically significant absolute  $\log_2$  fold change ( $\log_2$ FC) in the prevalence of mutants greater than 2 (Fig. 1E). In contrast, the prevalence of mutants in the purified spore samples indicated massive changes in the mutant population structure during spore formation. Fully 2,388 genes showed statistically significant changes in the prevalence of mutants over the experiment, 798 with a  $\log_2$  fold change lower than or equal to  $-2$ , which suggests that they strongly affect sporulation (Fig. 1F; see Table S2 in the supplemental material).

#### Figure Legend Continued

taurocholate (0.5%), and initiation of germination was monitored by measuring the drop in  $OD_{600}$  due to  $Ca^{2+}$ -DPA release from the core. Data are expressed as the ratio of  $OD_{600}$  at each time point to  $t_0$ . The  $\Delta$ CD0106 mutant did not appear to germinate during the course of the experiment, and complementation restored germination to wild-type levels. The  $\Delta$ CD3494 mutant appeared to germinate with a small delay compared to the parental strain but otherwise appeared to germinate normally. (C) Germinating suspensions were sampled, fixed in formaldehyde, and visualized by phase-contrast microscopy. In agreement with panel A, the  $\Delta$ CD0106 mutant did not germinate in response to taurocholate, whereas the  $\Delta$ CD3494 mutant appeared to germinate normally. Time (in minutes) is shown above the photos.

Among the 798 genes required for sporulation, we identified *spo0A*, encoding the master regulator of sporulation, and *sigH*, encoding  $\sigma^H$ , the key sigma factor of transition phase, both of which have been previously shown to be required for sporulation in *C. difficile* (9). Genes directly controlled by Spo0A and  $\sigma^H$  and thus involved in initiation of sporulation were also identified in this screen, including both the forespore and mother cell-specific early stage RNA polymerase sigma factors ( $\sigma^F$  and  $\sigma^E$ , respectively) together with proteins controlling their activity (SpoIIAA, SpoIIAB, SpoIIE, and SpoIIGA). Many genes in the  $\sigma^F$  and  $\sigma^E$  regulons also appear to be required for sporulation, including genes involved in stage II (*spoIIQ*, *spoIID*, *spoIIP*, and *spoIIR*), stage III (*spoIIIAA*, *spoIIIAB*, *spoIIIAD*, *spoIIIAE*, *spoIIIAF*, *spoIIIAH*, *spoIIIAI*, and *spoIIID*), and stage IV (*spoIVA*, *spoIVB'*, and *sipL*) of sporulation. Among these genes, *spoIIR* has been previously characterized (12), and *spoIID*, *spoIIP*, *spoIIQ*, and genes belonging to the *spoIIIA* locus are of particular note, as they encode homologues of *B. subtilis* proteins involved in forespore engulfment. SpoIIID is a crucial component of the mother cell regulatory network, activating *sigK* expression and playing several auxiliary roles during spore morphogenesis (12, 40), while *spoIVA* and *sipL* encode key spore morphogenetic proteins involved in targeting coat proteins to the surface of the forespore (37). Even though little is known about the molecular mechanisms controlling the later steps in the signaling cascade regulating sporulation





**FIG 5** Comparison of genes required for sporulation with published regulons for Spo0A,  $\sigma^E$ ,  $\sigma^F$ ,  $\sigma^G$ , and  $\sigma^K$ . The 798 genes identified in our screen were cross-referenced with the previously identified regulons of Spo0A (41) and  $\sigma^E$ ,  $\sigma^F$ ,  $\sigma^G$ , and  $\sigma^K$  (12). (A) Each gene is shown as a single point and colored according to the regulon; those genes not identified in the 5 published regulons are shown as black points. Dashed red lines indicate a  $\log_2FC$  of 2 and  $-2$ . (B)  $\log_2FC$  in the spore transposon library of every gene previously identified as belonging to the  $\sigma^E$ ,  $\sigma^F$ ,  $\sigma^G$ , and  $\sigma^K$  regulons colored by functional class, as defined in reference 12. The box-and-whisker plots show the interquartile range (IQR), with the median marked (black line). The whiskers indicate the highest/lowest point within  $1.5 \times$  IQR of the upper/lower quartile.

in *C. difficile*, a number of genes thought to be involved in these processes were identified in our screen. These genes included *sigG*, encoding the late-stage forespore-specific sigma factor  $\sigma^G$ , and genes belonging to the  $\sigma^G$  regulon, *spoIVB*, *spoVAC*, *spoVAD*, and *spoVAE*. Other notable genes involved in sporulation were identified in this group, including genes encoding two small acid-soluble proteins, SspA and SspB, although the former did not meet our minimal fold change criterion ( $\log_2FC = 1.95$ ;  $Q = 0.03$ ).

Overall, Gene Set Enrichment Analysis (GSEA) of sporulation-related regulons determined in previous studies (12, 13, 41) revealed an overrepresentation of genes required for sporulation within the  $\sigma^E$ ,  $\sigma^F$ , and  $\sigma^G$  regulons (Fig. 5A and B). However, as a whole, only the  $\sigma^F$  regulon showed a median  $\log_2FC$  value significantly below zero (Fig. 5B). Most genes belonging to the  $\sigma^K$  regulon, including those encoding spore coat proteins (CotA, CotCB, CotD, CotE, and CotF), exosporial glycoproteins (BclA1, BclA2, and BclA3), and the spore cortex lytic enzyme SleC, required for germination (42), showed little change in the number of insertions and clustered around a  $\log_2FC$  of 0. Despite this and the fact that a *sigK* mutant has been reported to produce low titers of heat-resistant spores devoid of the coat but with an apparently normal cortex layer (11, 40), the *sigK* gene itself was found to be

required for sporulation. This could be due to the relatively low levels of sporulation of the mutant (4-log-unit drop in spore titers compared to that of the wild type) or loss of the aberrant spores during purification on density gradients. Interestingly, of the genes from the SigK regulon that were identified as having a strong effect on sporulation, the majority have no known function.

Other functional groups of genes that affect sporulation included genes encoding ABC transporters (57 genes), phosphotransferase system (37 genes), two-component systems (21 genes), and a large number of conserved hypothetical proteins (70 genes) and putative membrane proteins (54 genes). Importantly, the R20291 homologues of the two sporulation genes confirmed in the 630 $\Delta erm$  mutant (CDR20291\_0124 and CDR20291\_3404) were also found to be required for normal sporulation (see Table S2 in the supplemental material).

## DISCUSSION

*C. difficile* is the main cause of antibiotic-associated diarrhea in nosocomial settings and is a significant cause of morbidity and mortality in Europe and the United States (1). Despite its importance as a human pathogen, our understanding of *C. difficile* lifestyle and virulence has lagged behind that of many other pathogenic bacteria due to a lack of genetic tools for the dissection of the

organism. Recent advances in clostridial genetics (19, 20, 43) have led to key insights into many aspects of *C. difficile* biology but have also demonstrated that there are fundamental differences between *C. difficile* and other members of the *Firmicutes*, particularly in sporulation and germination. Here we have combined our recently developed tightly regulated inducible gene expression system (29) with a novel conditional plasmid origin to produce a *mariner* mutagenesis system suitable for generating saturating transposon libraries in *C. difficile*. We produced a proof-of-principle transposon library in the model strain 630 $\Delta$ *erm* containing more than 44,000 unique mutants, confirming the utility of our mutagenesis system for high-throughput gene function analysis in *C. difficile*. In order to validate the library, we chose three genes predicted to play a role in sporulation and one gene with a role in germination for further study. Three of the isogenic mutants displayed the predicted phenotype; deletion of CD630\_01250 or CD630\_35670 resulted in severe sporulation defects, and a CD630\_01060 mutant sporulated as normal but was incapable of germinating in response to the germinant taurocholate. Surprisingly, a CD630\_34940 deletion mutant sporulated at levels equivalent to that of the parental strain despite clear differences in transposon insertion density in the gene before and after sporulation. CD630\_34940 encodes a homologue of a *B. subtilis* spore coat-associated protein, YabP, and disruption of this gene would be expected to have an effect on coat assembly. CD630\_34940 is the first gene in an apparent 3-gene operon CD630\_34920 to CD630\_34940. It is possible that the polar effects of a transposon insertion may explain the differences observed, although elucidating the exact role of this operon in sporulation will require further study.

Although confirming the validity of our transposon mutagenesis approach, the size and density of the 630 $\Delta$ *erm* library was insufficient to perform a statistical analysis of gene essentiality. However, our method is readily scalable. Rather than attempt to generate a saturating library in the same genetic background, we chose to focus on the most clinically important *C. difficile* lineage currently in circulation, ribotype 027 (BI and NAP1 [38]). We generated a transposon library containing 77,256 unique mutants in strain R20291. TraDIS analysis of this library identified 404 essential genes, 11% of the annotated genes, and a further 33 genes that fell in an ambiguous area between essential and nonessential gene cutoffs. This is of the same order of magnitude as in other organisms; for example, recent studies identified 505 essential genes in *Burkholderia pseudomallei* out of a total of 5,942 genes (8.5% [44]) and 353 essential genes in *Salmonella enterica* serovar Typhimurium out of a total of 4,530 (7.8% [39]). As expected, the list of essential genes in *C. difficile* includes those known to be involved in central cellular processes, such as DNA replication and repair, transcription, translation, protein secretion, cell envelope biogenesis, cell division, and many core metabolic pathways. Although not surprising, identification of these genes serves to validate our transposon screen. Of greater interest will be the 53 essential genes that have no known or predicted function, as these potentially represent attractive targets for future species-specific therapeutics. Of the 351 essential genes with functions predicted from homology, most have not been experimentally confirmed in *C. difficile*, and for many, the cellular processes they contribute to are completely unknown. This is the first time a high-throughput forward genetic screen has been carried out in *C. difficile*, and these data should form an invaluable resource for the community.

Sporulation is an ancient and highly complex cell differentiation process that has been studied in great detail in *B. subtilis* (45). As an obligate anaerobe, the process of sporulation is absolutely critical for the transmission of *C. difficile* from one host to the next (5). Despite overlaps in the regulatory cascades leading to spore formation, there are significant differences in *C. difficile*, including the mechanism of cascade activation via Spo0A (10) and the order and timing of sigma factor activation (11, 13). The spore structural proteins are also poorly conserved between species (46). We have identified 798 genes, mutation of which was strongly associated with a sporulation defect, either by direct disruption of the sporulation process or indirectly due to a more general fitness defect that results in disrupted or delayed sporulation. Only a small number of these genes have been studied in *C. difficile*, but several have well-characterized homologues in other spore-forming species, including *B. subtilis*. The subset of genes with known or predicted roles in sporulation includes the master regulator *spo0A* gene and the four sporulation-specific sigma factor genes *sigE*, *sigF*, *sigG*, and *sigK*. Of greater interest are the large number of genes that have no previously characterized role in sporulation. The vast majority of these have not been studied in *C. difficile*; indeed, fully 334 genes have gene ontology terms including the word “putative,” 68 have “conserved hypothetical,” 28 have “hypothetical,” and 19 have “uncharacterized.” Also of interest are the large number of identified genes encoding ABC transporters, the phosphotransferase system, and two-component systems. Considerable analysis will be required to elucidate the roles of these genes in sporulation, but the data presented here should act as a road map for future research into *C. difficile* sporulation.

One of the greatest challenges in the clinical management of *C. difficile* infection (CDI) is our reliance on antibiotic therapy for treating the infection. As CDI is essentially a consequence of dysbiosis, treatment with broad-spectrum antibiotics, such as metronidazole or vancomycin, leaves the patient acutely susceptible to reinfection, and as a result, relapse is common (47). It would be preferable to treat infection with narrow-spectrum, or even species-specific, therapeutics to avoid disturbance of the intestinal microflora. To develop such therapeutics, we need a detailed understanding of the differences between *C. difficile* and other species, particularly in sporulation, and also of the complement of essential genes, as their products are attractive targets for therapeutic intervention.

## MATERIALS AND METHODS

**Strains and growth conditions.** *C. difficile* strains 630 $\Delta$ *erm* (ribotype 012) and R20291 (ribotype 027) were routinely grown on BHI agar or in TY broth without thioglycolate (48), except where otherwise stated. *Escherichia coli* strains NEB5 $\alpha$  (New England Biolabs) and CA434 (HB101 carrying R702) were grown on LB agar or in LB broth (VWR). Cultures were supplemented with thiamphenicol (15  $\mu$ g/ml), erythromycin (5  $\mu$ g/ml), lincomycin (20  $\mu$ g/ml), chloramphenicol (15  $\mu$ g/ml), taurocholate (0.5%), or anhydrotetracycline (100 ng/ml) where appropriate.

**Construction of a *mariner* delivery plasmid for *C. difficile*.** The *HimarI mariner* transposase gene was synthesized and cloned into a *C. difficile*-*E. coli* shuttle vector, yielding pMar1. The gene was subcloned into the *C. difficile*-*E. coli* shuttle plasmid pRPF177 (29) between the Sacl and BamHI sites, placing expression of the transposase under the control of the tetracycline-inducible P<sub>tet</sub> promoter. The resulting plasmid, pRPF213, was further modified by the addition of an *ermB*-based transposon. The *ermB* gene with its native promoter was amplified from plasmid

pMTL82254 (43), a transcriptional terminator (from the *fdx* gene) was added downstream of *ermB*, and *mariner* inverted terminal repeats (ITRs) were added to each end. The resulting transposon was cloned into the BstXI site of pRPF213, yielding the final *mariner* delivery plasmid pRPF215. A transposition negative-control plasmid, pRPF222, was constructed by deleting the nucleotides corresponding to residues 244 to 291 of the *Himar1* transposase. The resulting plasmid retains the functional *ermB* marker, but the transposon should be incapable of transferring to the chromosome.

**Transposon mutagenesis of *C. difficile*.** Cultures of *C. difficile* carrying pRPF215 were grown overnight in TY broth supplemented with thiamphenicol, subcultured to an optical density at 600 nm ( $OD_{600}$ ) of 0.05, and grown to an  $OD_{600}$  of 0.3. The mid-logarithmic-phase cultures were then spread on BHIS (31) agar plates containing anhydrotetracycline and erythromycin (strain 630 $\Delta$ *erm*) or lincomycin (strain R20291). After 18-h growth, each colony (approximately 85,000 for strain 630 $\Delta$ *erm* and approximately 750,000 for strain R20291) represented a unique transposition event. All colonies were pooled and resuspended in TY broth supplemented with anhydrotetracycline and erythromycin or lincomycin to produce the initial transposon mutant library. Samples were harvested for DNA extraction (see below). The pooled library was then sporulated as described previously (49). Briefly, the pooled library was subcultured to an  $OD_{600}$  of 0.1 in TGY broth (49) supplemented with anhydrotetracycline and erythromycin/lincomycin and grown overnight. Samples were harvested for DNA extraction. The TGY culture was then diluted 1:10 in SMC broth, (49) grown to an  $OD_{600}$  of 0.6, and then spread on SMC agar plates. After a week on SMC agar, spores were recovered in H<sub>2</sub>O and purified on a 20 to 50% HistoDenz (Sigma-Aldrich) gradient (50). Spores from the 630 $\Delta$ *erm* library were germinated in BHIS broth supplemented with erythromycin and taurocholate (50). Samples of the TGY overnight culture, spores, and germinated library were retained for DNA extraction.

**DNA extraction.** For extraction of genomic DNA (gDNA) from vegetative cells, frozen bacterial samples were thawed and resuspended in 2-ml lysis buffer (200 mM NaCl, 50 mM EDTA, 20 mM Tris-HCl [pH 8.0], 2 mg/ml lysozyme) and incubated at 37°C for 1 h. Lysed bacterial samples were then sequentially treated with pronase (0.5 mg/ml; 55°C overnight), *N*-lauroylsarcosine (2%; 1 h at 37°C) and RNase (0.05 mg/ml; 1 h at 37°C). Samples were then extracted twice with phenol-chloroform-isoamyl alcohol (25:24:1) and twice with chloroform-isoamyl alcohol (24:1) using heavy phase lock gel tubes (5 Prime). gDNA was then precipitated with ethanol overnight at -20°C, washed with 70% ethanol, and finally resuspended in 50  $\mu$ l of nuclease-free water. For extraction of gDNA from *C. difficile* spores, samples were resuspended in lysis buffer as described above and then broken over glass beads using a FastPrep-24 instrument (MP Biomedicals) at 6.5 m/s for three 45-s cycles, with cooling on ice between cycles. DNA was then purified as described above.

**Illumina sequencing and analysis.** gDNA was extracted, fragmented, and sequenced on a HiSeq2500 or HiSeq2000 Illumina platform, generating approximately 2 million 50-bp single-end reads per sample. The resulting FASTQ files were filtered for 10 bases matching the 3' end of the transposon. These transposon tags were stripped from the resulting reads, and the stripped reads were then mapped to their corresponding reference sequences using SMALT-0.7.2. The precise insertion site of the transposon was determined using the Bio::Tradis tools (<https://github.com/sanger-pathogens/Bio-Tradis>). Insertion sites and read counts were tabulated per gene, and further analysis was conducted using R.

Gene essentiality was assayed as described previously (39). Briefly, an insertion index (number of insertion sites/gene length) was calculated for each gene. The observed distribution of insertion indices was bimodal, and gamma distributions were fitted to the putative essential (mode at 0) and nonessential peaks of the empirical distribution. Log<sub>2</sub> likelihood ratios (LLR) were calculated between the fitted distributions, and a gene was classified as essential if it had an LLR of less than -2, leading to an essentiality cutoff at an insertion index of 0.0015. Similarly, a gene was classified as nonessential if it had an LLR of >2, giving an insertion index cutoff of

0.0021. Insertion indices falling between these two values were classified as “ambiguous.” Enrichment of essential genes in KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways (51) was determined using a hypergeometric test.

To identify genes important for overnight growth and sporulation, edgeR (52) was used to identify significant differences in read counts over genes before and after selection. The TMM (trimmed mean of M values) normalization was applied, and tagwise dispersion was estimated. Only genes exhibiting greater than 20 reads in both replicates of at least one of the conditions being assayed were tested for differences in the prevalence of mutants. *P* values were corrected for multiple testing by the Benjamini-Hochberg method, and genes with a corrected *P* value (*Q* value) of <0.1 (a hypothetical 10% false discovery rate [FDR]) and an absolute log fold change (logFC) of >2 were considered significant. To determine enrichment of sporulation-related regulons, we used transcriptomic data from previous publications (Spo0A [41] and  $\sigma^E$ ,  $\sigma^F$ ,  $\sigma^G$ , and  $\sigma^K$  [12]) as gene sets for Gene Set Enrichment Analysis (GSEA) analysis (53). GSEA version 2.1.0 was run in ranked-list mode with enrichment statistic “weighted” on the per-gene logFCs calculated by edgeR.

**Construction of individual deletion mutants and complementation.** In-frame deletions of individual genes were generated in *C. difficile* strain 630 $\Delta$ *erm* $\Delta$ *pyrE* as described previously (20). Briefly, homologous recombination plasmids were generated by adding appropriate 750-bp upstream and downstream homology arms to pMTL-YN3 using Gibson assembly. The resulting plasmids were conjugated into strain 630 $\Delta$ *erm* $\Delta$ *pyrE*, and deletion mutants were isolated as described previously (20). Putative mutants were confirmed by PCR. The *pyrE* gene was subsequently restored using plasmid pMTL-YN1 as described previously (20). Each mutant was complemented on the chromosome by cloning a fragment containing the wild-type gene with its native promoter (identified using BPROM [SoftBerry]) into pMTL-YN1C for integration into the *pyrE* locus as described previously (20).

**Sporulation and germination efficiency.** Overnight cultures of *C. difficile* in BHIS broth were diluted to an  $OD_{600}$  of 0.01 in fresh broth, grown to an  $OD_{600}$  of approximately 0.6, and diluted again to an  $OD_{600}$  of 0.0001 before growth overnight to stationary phase. This allowed synchronization of growth and minimized the carryover of spores from initial cultures. The relative proportions of vegetative cells and spores were then monitored at 24-h intervals for 5 days. At each time point, the total number of bacteria was determined by counting the number of CFU on BHIS agar supplemented with 0.1% taurocholate. To determine the number of spores, samples were incubated at 70°C for 30 min, prior to counting CFU.

To determine germination efficiency, purified spores were resuspended to an  $OD_{600}$  of 1.0 in 10 ml BHIS supplemented with 0.5% taurocholate. Spore germination was determined by monitoring the changes in  $OD_{600}$  due to Ca<sup>2+</sup>-dipicolinic acid (DPA) release and outgrowth. Data are expressed as the ratio of  $OD_{600}$  at each time point to time zero ( $t_0$ ). Germination was also monitored by phase-contrast microscopy.

**Microscopy.** One milliliter of culture was stained with FM4-64 membrane dye at a concentration of 0.5  $\mu$ g/ml, centrifuged at 5,000  $\times$  g for 10 min at 4°C, and washed with 1 ml of phosphate-buffered saline (PBS). DNA staining was carried out by adding Hoechst 33258 dye to the mounting medium to a final concentration of 20  $\mu$ g/ml. Images were captured using an Eclipse E600 microscope (Nikon) with a Retiga-400R charge-coupled device (CCD) (Q-Imaging).

**Nucleotide sequence accession numbers.** The GenBank accession numbers for the raw sequencing data for the *C. difficile* 630 $\Delta$ *erm* transposon library are ERR245853 to ERR245855 and ERR237766, and those for the R20291 transposon library are ERR377408 to ERR377413 and ERR377416 to ERR377421 (see Table S3 in the supplemental material).

## SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.02383-14/-/DCSupplemental>.

Figure S1, TIF file, 1.8 MB.

Figure S2, TIF file, 1.9 MB.  
 Figure S3, TIF file, 1 MB.  
 Table S1, PDF file, 0.1 MB.  
 Table S2, PDF file, 0.1 MB.  
 Table S3, DOCX file, 0.1 MB.

## ACKNOWLEDGMENTS

Sequencing was supported by the Wellcome Trust (grant 098051). C.J.B. and A.K.C. were supported by the Medical Research Council (grant G1100100). L.B. was supported by a research fellowship from the Alexander von Humboldt Foundation. M.D. was supported by a Wellcome Trust Ph.D. studentship (grant 089875/Z/09/A). R.P.F. was funded by the Medical Research Council (grant G0800170).

## REFERENCES

- Rupnik M, Wilcox MH, Gerding DN. 2009. *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. *Nat Rev Microbiol* 7:526–536. <http://dx.doi.org/10.1038/nrmicro2164>.
- Carroll KC, Bartlett JG. 2011. Biology of *Clostridium difficile*: implications for epidemiology and diagnosis. *Annu Rev Microbiol* 65:501–521. <http://dx.doi.org/10.1146/annurev-micro-090110-102824>.
- Eyre DW, Cule ML, Wilson DJ, Griffiths D, Vaughan A, O'Connor L, Ip CL, Golubchik T, Batty EM, Finney JM, Wyllie DH, Didelot X, Piazza P, Bowden R, Dingle KE, Harding RM, Crook DW, Wilcox MH, Peto TE, Walker AS. 2013. Diverse sources of *C. difficile* infection identified on whole-genome sequencing. *N Engl J Med* 369:1195–1205. <http://dx.doi.org/10.1056/NEJMoa1216064>.
- Dubberke ER, Olsen MA. 2012. Burden of *Clostridium difficile* on the healthcare system. *Clin Infect Dis* 55(Suppl 2):S88–S92. <http://dx.doi.org/10.1093/cid/cis335>.
- 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. <http://dx.doi.org/10.1128/IAI.00147-12>.
- Lawley TD, Clare S, Deakin LJ, Goulding D, Yen JL, Raisen C, Brandt C, Lovell J, Cooke F, Clark TG, Dougan G. 2010. Use of purified *Clostridium difficile* spores to facilitate evaluation of health care disinfection regimens. *Appl Environ Microbiol* 76:6895–6900. <http://dx.doi.org/10.1128/AEM.00718-10>.
- Gerding DN, Muto CA, Owens RC, Jr. 2008. Measures to control and prevent *Clostridium difficile* infection. *Clin Infect Dis* 46(Suppl 1):S43–S49. <http://dx.doi.org/10.1086/521861>.
- Paredes CJ, Alsaker KV, Papoutsakis ET. 2005. A comparative genomic view of clostridial sporulation and physiology. *Nat Rev Microbiol* 3:969–978. <http://dx.doi.org/10.1038/nrmicro1288>.
- 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. <http://dx.doi.org/10.1128/JB.00272-11>.
- 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. <http://dx.doi.org/10.1128/JB.00882-09>.
- Pereira FC, Saujet L, Tomé AR, Serrano M, Monot M, Couture-Tosi E, Martin-Verstraete I, Dupuy B, Henriques AO. 2013. The spore differentiation pathway in the enteric pathogen *Clostridium difficile*. *PLoS Genet* 9:e1003782. <http://dx.doi.org/10.1371/journal.pgen.1003782>.
- Saujet L, Pereira FC, Serrano M, Soutourina O, Monot M, Shelyakin PV, Gelfand MS, Dupuy B, Henriques AO, Martin-Verstraete I. 2013. Genome-wide analysis of cell type-specific gene transcription during spore formation in *Clostridium difficile*. *PLoS Genet* 9:e1003756. <http://dx.doi.org/10.1371/journal.pgen.1003756>.
- Fimlaid KA, Bond JP, Schutz KC, Putnam EE, Leung JM, Lawley TD, Shen A. 2013. Global analysis of the sporulation pathway of *Clostridium difficile*. *PLoS Genet* 9:e1003660. <http://dx.doi.org/10.1371/journal.pgen.1003660>.
- de Hoon MJ, Eichenberger P, Vitkup D. 2010. Hierarchical evolution of the bacterial sporulation network. *Curr Biol* 20:R735–R745. <http://dx.doi.org/10.1016/j.cub.2010.06.031>.
- Steil L, Serrano M, Henriques AO, Völker U. 2005. Genome-wide analysis of temporally regulated and compartment-specific gene expression in sporulating cells of *Bacillus subtilis*. *Microbiology* 151:399–420. <http://dx.doi.org/10.1099/mic.0.27493-0>.
- Eichenberger P, Fujita M, Jensen ST, Conlon EM, Rudner DZ, Wang ST, Ferguson C, Haga K, Sato T, Liu JS, Losick R. 2004. The program of gene transcription for a single differentiating cell type during sporulation in *Bacillus subtilis*. *PLoS Biol* 2:e328. <http://dx.doi.org/10.1371/journal.pbio.0020328>.
- Wang ST, Setlow B, Conlon EM, Lyon JL, Imamura D, Sato T, Setlow P, Losick R, Eichenberger P. 2006. The forespore line of gene expression in *Bacillus subtilis*. *J Mol Biol* 358:16–37. <http://dx.doi.org/10.1016/j.jmb.2006.01.059>.
- Galperin MY, Mekhedov SL, Puigbo P, Smirnov S, Wolf YI, Rigden DJ. 2012. Genomic determinants of sporulation in *Bacilli* and *Clostridia*: towards the minimal set of sporulation-specific genes. *Environ Microbiol* 14:2870–2890. <http://dx.doi.org/10.1111/j.1462-2920.2012.02841.x>.
- 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. <http://dx.doi.org/10.1016/j.jmimet.2009.10.018>.
- Ng YK, Ehsaan M, Philip S, Coltery MM, Janoir C, Collignon A, Cartman ST, Minton NP. 2013. Expanding the repertoire of gene tools for precise manipulation of the *Clostridium difficile* genome: allelic exchange using *pyrE* alleles. *PLoS One* 8:e56051. <http://dx.doi.org/10.1371/journal.pone.0056051>.
- Cartman ST, Kelly ML, Heeg D, Heap JT, Minton NP. 2012. Precise manipulation of the *Clostridium difficile* chromosome reveals a lack of association between the *tcdC* genotype and toxin production. *Appl Environ Microbiol* 78:4683–4690. <http://dx.doi.org/10.1128/AEM.00249-12>.
- Faulds-Pain A, Wren BW. 2013. Improved bacterial mutagenesis by high-frequency allele exchange, demonstrated in *Clostridium difficile* and *Streptococcus suis*. *Appl Environ Microbiol* 79:4768–4771. <http://dx.doi.org/10.1128/AEM.01195-13>.
- Akerley BJ, Rubin EJ, Camilli A, Lampe DJ, Robertson HM, Mekalanos JJ. 1998. Systematic identification of essential genes by in vitro mariner mutagenesis. *Proc Natl Acad Sci U S A* 95:8927–8932. <http://dx.doi.org/10.1073/pnas.95.15.8927>.
- Barquist L, Boinett CJ, Cain AK. 2013. Approaches to querying bacterial genomes with transposon-insertion sequencing. *RNA Biol* 10:1161–1169. <http://dx.doi.org/10.4161/rna.24765>.
- van Opijnen T, Camilli A. 2013. Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nat Rev Microbiol* 11:435–442. <http://dx.doi.org/10.1038/nrmicro3033>.
- Lampe DJ, Grant TE, Robertson HM. 1998. Factors affecting transposition of the *Himar1* mariner transposon *in vitro*. *Genetics* 149:179–187.
- Cartman ST, Minton NP. 2010. A mariner-based transposon system for in vivo random mutagenesis of *Clostridium difficile*. *Appl Environ Microbiol* 76:1103–1109. <http://dx.doi.org/10.1128/AEM.02525-09>.
- Langridge GC, Phan MD, Turner DJ, Perkins TT, Parts L, Haase J, Charles I, Maskell DJ, Peters SE, Dougan G, Wain J, Parkhill J, Turner AK. 2009. Simultaneous assay of every *Salmonella* Typhi gene using one million transposon mutants. *Genome Res* 19:2308–2316. <http://dx.doi.org/10.1101/gr.097097.109>.
- Fagan RP, Fairweather NF. 2011. *Clostridium difficile* has two parallel and essential Sec secretion systems. *J Biol Chem* 286:27483–27493. <http://dx.doi.org/10.1074/jbc.M111.263889>.
- Geissendörfer M, Hillen W. 1990. Regulated expression of heterologous genes in *Bacillus subtilis* using the Tn10 encoded *tet* regulatory elements. *Appl Microbiol Biotechnol* 33:657–663. <http://dx.doi.org/10.1007/BF00604933>.
- 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. <http://dx.doi.org/10.1046/j.1365-2958.2002.03134.x>.
- Lampe DJ, Churchill ME, Robertson HM. 1996. A purified mariner transposase is sufficient to mediate transposition *in vitro*. *EMBO J* 15:5470–5479.
- Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. 2000. Artemis: sequence visualization and annotation. *Bioinformatics* 16:944–945. <http://dx.doi.org/10.1093/bioinformatics/16.10.944>.
- Sekiguchi J, Akeo K, Yamamoto H, Khasanov FK, Alonso JC, Kuroda A. 1995. Nucleotide sequence and regulation of a new putative cell wall

- hydrolase gene, *cwlD*, which affects germination in *Bacillus subtilis*. *J Bacteriol* 177:5582–5589.
35. Doan T, Morlot C, Meisner J, Serrano M, Henriques AO, Moran CP, Jr, Rudner DZ. 2009. Novel secretion apparatus maintains spore integrity and developmental gene expression in *Bacillus subtilis*. *PLoS Genet* 5:e1000566. <http://dx.doi.org/10.1371/journal.pgen.1000566>.
  36. Fawcett P, Eichenberger P, Losick R, Youngman P. 2000. The transcriptional profile of early to middle sporulation in *Bacillus subtilis*. *Proc Natl Acad Sci U S A* 97:8063–8068. <http://dx.doi.org/10.1073/pnas.140209597>.
  37. Putnam EE, Nock AM, Lawley TD, Shen A. 2013. SpoIVA and SipL are *Clostridium difficile* spore morphogenetic proteins. *J Bacteriol* 195:1214–1225. <http://dx.doi.org/10.1128/JB.02181-12>.
  38. 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. 2013. Emergence and global spread of epidemic healthcare-associated *Clostridium difficile*. *Nat Genet* 45:109–113. <http://dx.doi.org/10.1038/ng.2478>.
  39. Barquist L, Langridge GC, Turner DJ, Phan MD, Turner AK, Bateman A, Parkhill J, Wain J, Gardner PP. 2013. A comparison of dense transposon insertion libraries in the *Salmonella* serovars Typhi and Typhimurium. *Nucleic Acids Res* 41:4549–4564. <http://dx.doi.org/10.1093/nar/gkt148>.
  40. Pishdadian K, Fimlaid KA, Shen A. 2015. SpoIID-mediated regulation of sigma function during *Clostridium difficile* sporulation. *Mol Microbiol* 95:189–208. <http://dx.doi.org/10.1111/mmi.12856>.
  41. Pettit LJ, Browne HP, Yu L, Smits WK, Fagan RP, Barquist L, Martin MJ, Goulding D, Duncan SH, Flint HJ, Dougan G, Choudhary JS, Lawley TD. 2014. Functional genomics reveals that *Clostridium difficile* Spo0A coordinates sporulation, virulence and metabolism. *BMC Genomics* 15:160. <http://dx.doi.org/10.1186/1471-2164-15-160>.
  42. Burns DA, Heap JT, Minton NP. 2010. SleC is essential for germination of *Clostridium difficile* spores in nutrient-rich medium supplemented with the bile salt taurocholate. *J Bacteriol* 192:657–664. <http://dx.doi.org/10.1128/JB.01209-09>.
  43. Heap JT, Pennington OJ, Cartman ST, Minton NP. 2009. A modular system for *Clostridium* shuttle plasmids. *J Microbiol Methods* 78:79–85. <http://dx.doi.org/10.1016/j.mimet.2009.05.004>.
  44. Moule MG, Hemsley CM, Seet Q, Guerra-Assunção JA, Lim J, Sarkar-Tyson M, Clark TG, Tan PB, Titball RW, Cucui J, Wren BW. 2014. Genome-wide saturation mutagenesis of *Burkholderia pseudomallei* K96243 predicts essential genes and novel targets for antimicrobial development. *mBio* 5:e00926-13. <http://dx.doi.org/10.1128/mBio.00926-13>.
  45. Higgins D, Dworkin J. 2012. Recent progress in *Bacillus subtilis* sporulation. *FEMS Microbiol Rev* 36:131–148. <http://dx.doi.org/10.1111/j.1574-6976.2011.00310.x>.
  46. Permpoonpattana P, Tolls EH, Nadem R, Tan S, Brisson A, Cutting SM. 2011. Surface layers of *Clostridium difficile* endospores. *J Bacteriol* 193:6461–6470. <http://dx.doi.org/10.1128/JB.05182-11>.
  47. Figueroa I, Johnson S, Sambol SP, Goldstein EJ, Citron DM, Gerding DN. 2012. Relapse versus reinfection: recurrent *Clostridium difficile* infection following treatment with fidaxomicin or vancomycin. *Clin Infect Dis* 55(Suppl 2):S104–S109. <http://dx.doi.org/10.1093/cid/cis357>.
  48. Dupuy B, Sonenshein AL. 1998. Regulated transcription of *Clostridium difficile* toxin genes. *Mol Microbiol* 27:107–120. <http://dx.doi.org/10.1046/j.1365-2958.1998.00663.x>.
  49. Permpoonpattana P, Hong HA, Phetcharaburanin J, Huang JM, Cook J, Fairweather NF, Cutting SM. 2011. Immunization with *Bacillus* spores expressing toxin A peptide repeats protects against infection with *Clostridium difficile* strains producing toxins A and B. *Infect Immun* 79:2295–2302. <http://dx.doi.org/10.1128/IAI.00130-11>.
  50. Dembek M, Stabler RA, Witney AA, Wren BW, Fairweather NF. 2013. Transcriptional analysis of temporal gene expression in germinating *Clostridium difficile* 630 endospores. *PLoS One* 8:e64011. <http://dx.doi.org/10.1371/journal.pone.0064011>.
  51. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2014. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res* 42:D199–D205. <http://dx.doi.org/10.1093/nar/gkt1076>.
  52. Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140. <http://dx.doi.org/10.1093/bioinformatics/btp616>.
  53. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. 2005. Gene Set Enrichment Analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102:15545–15550. <http://dx.doi.org/10.1073/pnas.0506580102>.