AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology®

EVOLUTIONARY AND GENOMIC MICROBIOLOGY

A Type I Restriction-Modification System Associated with *Enterococcus faecium* Subspecies Separation

Wenwen Huo,[a] Hannah M. Adams,[a] Cristian Trejo,[a] Rohit Badia,[a] Kelli L. Palmer[a]

[a]Department of Biological Sciences, University of Texas at Dallas, Richardson, Texas, USA

ABSTRACT The gastrointestinal colonizer *Enterococcus faecium* is a leading cause of hospital-acquired infections. Multidrug-resistant (MDR) *E. faecium* isolates are particularly concerning for infection treatment. Previous comparative genomic studies revealed that subspecies referred to as clade A and clade B exist within *E. faecium*. MDR *E. faecium* isolates belong to clade A, while clade B consists of drug-susceptible fecal commensal *E. faecium* isolates. Isolates from clade A are further grouped into two subclades, clades A1 and A2. In general, clade A1 isolates are hospital-epidemic isolates, whereas clade A2 isolates are isolates from animals and sporadic human infections. Such phylogenetic separation indicates that reduced gene exchange occurs between the clades. We hypothesize that endogenous barriers to gene exchange exist between *E. faecium* clades. Restriction-modification (R-M) systems are such barriers in other microbes. We utilized a bioinformatics analysis coupled with second-generation and third-generation deep-sequencing platforms to characterize the methylomes of two representative *E. faecium* strains, one from clade A1 and one from clade B. We identified a type I R-M system that is clade A1 specific, is active for DNA methylation, and significantly reduces the transformability of clade A1 *E. faecium*. Based on our results, we conclude that R-M systems act as barriers to horizontal gene exchange in *E. faecium* and propose that R-M systems contribute to *E. faecium* subspecies separation.

IMPORTANCE *Enterococcus faecium* is a leading cause of hospital-acquired infections around the world. Rising antibiotic resistance in certain *E. faecium* lineages leaves fewer treatment options. The overarching aim of this work was to determine whether restriction-modification (R-M) systems contribute to the structure of the *E. faecium* species, wherein hospital-epidemic and non-hospital-epidemic isolates have distinct evolutionary histories and highly resolved clade structures. R-M provides bacteria with a type of innate immunity to horizontal gene transfer (HGT). We identified a type I R-M system that is enriched in the hospital-epidemic clade and determined that it is active for DNA modification activity and significantly impacts HGT. Overall, this work is important because it provides a mechanism for the observed clade structure of *E. faecium* as well as a mechanism for facilitated gene exchange among hospital-epidemic *E. faecium* isolates.

KEYWORDS restriction-modification, *Enterococcus faecium*, genome defense, antibiotic resistance

Enterococcus faecium is a Gram-positive opportunistic pathogen that normally resides in the gastrointestinal tracts of humans and other animals (1, 2). *E. faecium* can cause life-threatening infections such as endocarditis and is among the leading causes of catheter-associated bloodstream and urinary tract infections in clinical settings (3).

Previous comparative genomic studies revealed that subspecies exist within the *E. faecium* species (4–7). Different names have been used by different groups to describe these clades; in this study, we use the clade A/B nomenclature. Generally speaking,

Citation Huo W, Adams HM, Trejo C, Badia R, Palmer KL. 2019. A type I restriction-modification system associated with *Enterococcus faecium* subspecies separation. Appl Environ Microbiol 85:e02174-18. https://doi.org/10.1128/AEM.02174-18.

Editor Christopher A. Elkins, Centers for Disease Control and Prevention

Address correspondence to Kelli L. Palmer, kelli.palmer@utdallas.edu.

Received 6 September 2018
Accepted 26 October 2018
Accepted manuscript posted online 2 November 2018
Published 9 January 2019

multidrug-resistant (MDR) *E. faecium* strains belong to clade A, while clade B consists of drug-susceptible fecal commensal *E. faecium* strains (8). Clade A is further split into two subclades, clades A1 and A2, with hospital-endemic strains generally clustering in clade A1 and sporadic-infection isolates and animal isolates generally clustering in clade A2 (8). Specific phenotypes and genomic features are enriched in clade A1 isolates relative to clade A2 and B isolates (8). Specifically, clade A1 isolates have significantly higher mutation rates and larger overall genome sizes, including a larger core genome, and possess more mobile elements. On the other hand, clade A2 possesses a larger pangenome than clades A1 and B, possibly reflective of the broader host origins of these strains. Given that clade A and clade B strains would be expected to comingle in certain environments (for example, in a hospital and municipal sewage), the phylogenetic separation among the *E. faecium* clades suggests that they are not sharing genetic information freely because of endogenous barriers to genetic exchange.

Horizontal gene transfer (HGT) is the exchange of genetic material between cells rather than the vertical inheritance of genetic material from a parental cell. Bacteria can encode genome defense mechanisms that can act in opposition to HGT. Two examples of these mechanisms are clustered regularly interspaced short palindromic repeat (CRISPR) and associated protein (CRISPR-Cas) systems and restriction-modification (R-M) systems. CRISPR-Cas is a dynamic immune system that utilizes sequence complementarity between self (CRISPR RNAs) and foreign nucleic acid to carry out its restrictive function, whereas R-M discriminates self from foreign DNA by DNA methylation patterns. If the *E. faecium* clades encode different defense mechanisms, they may not exchange genetic information freely, thereby facilitating and maintaining phylogenetic separation. However, little is known about CRISPR-Cas and R-M in *E. faecium*. Genomic analysis suggests that these systems could contribute to the observed clade structure of *E. faecium*. For example, CRISPR-Cas systems have been identified exclusively in clade B *E. faecium* strains and in sporadic clade A-clade B recombinant strains (8). For R-M, a predicted methyl-directed restriction endonuclease (REase) is enriched in clade A2 and B *E. faecium* genomes relative to clade A1 genomes (8).

Here, we focused on R-M systems and their roles in regulating gene exchange in *E. faecium* because little is known about R-M defense in this species. Moreover, there is precedent in the literature for R-M systems contributing to bacterial clade structure, as has been observed in *Burkholderia* (9) and *Neisseria* (10). Our overarching hypothesis is that the *E. faecium* clades encode different R-M systems, thereby inhibiting genetic exchange between them. In general, R-M systems are composed of cognate methyltransferase (MTase) and REase activities and are classified into different types based on the specific number and types of enzymes in the system as well as characteristics such as methylation type and pattern, cofactor requirement, and restriction activity (11). A MTase recognizes specific sequences in the bacterial genome and transfers a methyl group to either an adenine or a cytosine residue, resulting in 6-methyladenine (m6A), 4-methylcytosine (m4C), or 5-methylcytosine (m5C). A REase may recognize the same sequence as a MTase and cleave that region if the sequence is unmethylated (or, in some cases, if it is methylated). With the activities of MTases and REases, bacteria can use R-M to impede the entry of nonself DNA.

In this study, we used single-molecule real-time (SMRT) sequencing and whole-genome bisulfite sequencing to characterize the methylomes of representative *E. faecium* strains from clade A1 (*E. faecium* 1,231,502 or Efm502) and clade B (*E. faecium* 1,141,733 or Efm733). Two unique m6A methylation patterns were identified, one in each strain. These patterns were asymmetric and bipartite, which is characteristic of type I R-M methylation motifs (12). Bioinformatic analyses were performed to identify candidate genes responsible for methylation. A unique type I R-M system is encoded by each strain. We have named these systems Efa502I (for Efm502) and Efa733I (for Efm733). Expression of these candidate systems in *Enterococcus faecalis* heterologous hosts followed by SMRT sequencing confirmed that they are responsible for the methylation patterns observed in Efm502 and Efm733. A functional analysis was performed in order to assess the abilities of these systems to reduce *E. faecium* HGT by

TABLE 1 Distribution of predicted DNA MTases and R-M systems[a]

Strain Name	Clade	Type I R-M System			Type I R-M System			Type I R-M System			Type II MTase	Unspecified MTase
		S subunit	M subunit	R subunit	S subunit	M subunit	R subunit	S subunit	M subunit	R subunit		
1,230,933	A1	EFPG_01270	EFPG_01269	EFPG_01271	EFPG_05435; EFPG_05434	EFPG_05433	EFPG_05436				EFPG_01821	
1,231,410	A1	EFTG_00783	EFTG_00782	EFTG_00784	EFTG_02031	EFTG_02032	EFTG_02030				EFTG_02364; EFTG_02653	EFTG_02333
1,231,502	A1	EFQG_01131	EFQG_01132	EFQG_01130							EFQG_01609, EFQG_02270	
1,231,501	A2							EFRG_00106	EFRG_00107	EFRG_00105		EFRG_02239
1,231,408	Hybrid A1/B	EFUG_01527	EFUG_01526	EFUG_01528								EFUG_01512
1,141,733	B	EFSG_05028	EFSG_05029	EFSG_05027							EFSG_00659	
Com12	B											EFVG_00502
Com15	B	EFWG_02518	EFWG_02519	EFWG_02517								

[a]Loci that are in black are strain specific. Loci that are the same color are conserved in their protein sequences based on a >90% sequence identity threshold. An empty cell indicates that the system was not detected.

transformation. In a comparative analysis among 73 *E. faecium* genomes, we found that Efa502I is significantly enriched among clade A1 isolates, while the type I R-M system of Efm733 appears to be strain specific. Overall, this study is a first step toward understanding the role of R-M in regulating HGT in *E. faecium* and the potential for R-M as one mechanism for the clade structure of *E. faecium*.

RESULTS

Identification of a clade A1-specific putative type I R-M system in *E. faecium*. We previously reported that a type II R-M system significantly reduces HGT via conjugation (13) and transformation (14) in *E. faecalis*. Here, we hypothesize that the *E. faecium* clades encode distinct R-M systems that reduce the exchange of genetic information between them. We utilized an approach that we previously developed for *E. faecalis* R-M analysis (13) to predict potential R-M systems in eight previously sequenced *E. faecium* genomes. The 8 genomes included 3 genomes from clade A1, 3 genomes from clade B, 1 genome from clade A2, and 1 recombinant clade A1-clade B hybrid (5, 8). Because REases are difficult to identify with bioinformatics, and MTase prediction is comparatively straightforward, as was previously reported by New England Biolabs (NEB) (15), we first identified predicted DNA MTases in *E. faecium* genomes and then analyzed surrounding genes for predicted R-M-related activities. The complete list of candidates for the eight strains is shown in Table 1.

Interestingly, we predicted at least one putative type I R-M system for seven of the eight *E. faecium* strains (Table 1). Type I R-M systems are multisubunit complexes comprised of a specificity (S) subunit, a methylation (M) subunit, and a restriction (R) subunit (11, 16–18). The S subunit is responsible for the specific DNA recognition motif and associates with the DNA to bring the M and R subunits into contact. The system has two conformations: M_2S_1, which is capable of methylating DNA based on the recognition sequence, and $R_2M_2S_1$, which is capable of restricting DNA (18, 19). One predicted *E. faecium* type I R-M system is comprised of highly conserved (>90% amino acid sequence identity) M and R subunits in six of eight genomes across both clade A1 and clade B (Table 1; see also Data Set S1 in the supplemental material). The specificity

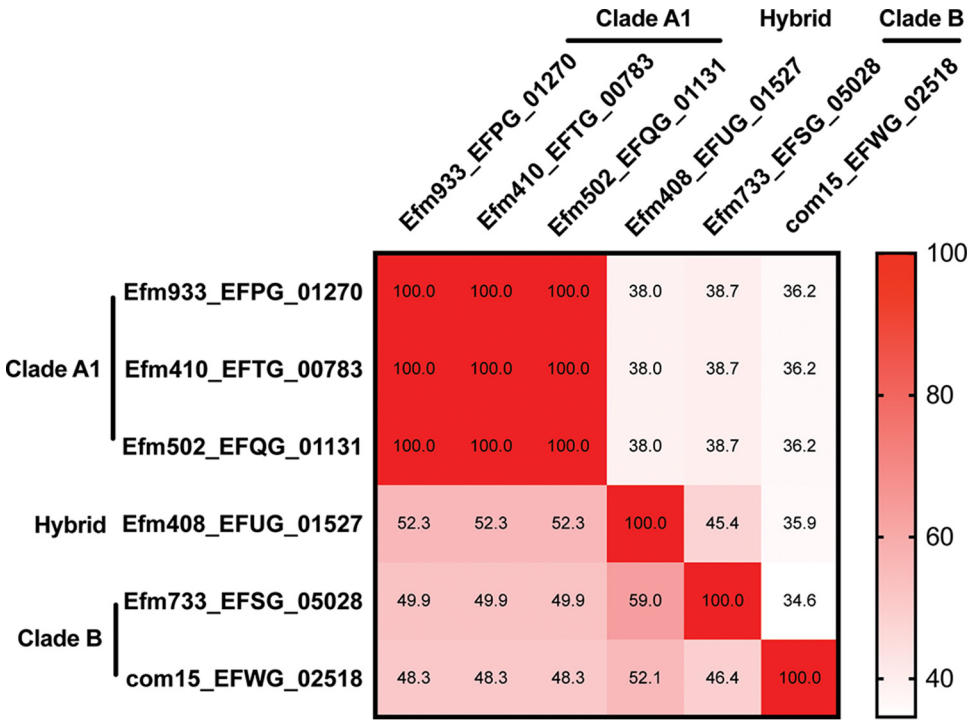

FIG 1 Conservation and variability of S subunits. The protein sequences of predicted S subunits from 6 (out of 8) representative *E. faecium* genomes were pairwise aligned using MacVector. The percent identity of each pair is shown. White to red, low to high percent identities. Each number represents the percent identity of one protein sequence (row name) to another (column name).

subunit from this system, however, is highly conserved in clade A1 genomes but not in clade B (Table 1, Fig. 1, and Fig. S1). S subunits possess two target recognition domains (TRDs) that determine the nucleotide sequence that the subunit binds to (20, 21). The variation in amino acid sequence between the S subunits occurs within these TRDs (Fig. S1), suggesting that these S subunits recognize different DNA sequences. Notably, the S subunits from clade A1 strains are identical to each other, indicating that they utilize the same recognition sequence.

To examine the distribution of this putative clade A1-specific system in a larger collection of *E. faecium* strains, we analyzed 73 *E. faecium* genomes of mostly draft status that were reported previously (8). This list includes 15 clade B isolates, 21 clade A1 isolates, 35 clade A2 isolates, and 2 hybrid isolates (Table S1). We selected Efm502 (clade A1) as our representative clade A1 strain for this analysis and used type I R-M sequences from this genome as references for analysis against the broader collection of *E. faecium* strains. The M and R subunits of the putative clade A1-specific type I system were detected in 51 and 52, respectively, of 73 *E. faecium* genomes, including both clade A and clade B strains (Fig. S2a and b). However, the distribution of the S subunits varied (Fig. S2a and b). The S subunit present in Efm502, EFQG_01131, was significantly enriched within clade A1 isolates (14/21; P value of <0.0001 using a Fisher exact test) (Fig. S2a) and absent from all other clades with the exception of strain EnGen002, which is classified as a clade A1-clade B hybrid strain. Interestingly, the S subunits present in most other *E. faecium* strains are strain specific by the strict thresholds applied here. Given that the Efm502 S subunit is enriched in clade A1, we hypothesize that many clade A1 strains exchange genetic information freely with each other, while exchange with other *E. faecium* strains is restricted.

SMRT sequencing for *E. faecium* methylome analysis. We analyzed the Efm502 and Efm733 genomes by SMRT sequencing. SMRT sequencing measures the kinetics of DNA polymerase as it synthesizes DNA in order to identify bases that have been modified (22–24). It has been extensively utilized for bacterial methylome analysis

TABLE 2 SMRT sequencing results

Strain	Motif[a]	Type	% of motifs detected	No. of motifs detected	No. of motifs in genome	Mean motif coverage (fold coverage)
Efm502	5′-R^{m}AYCNNNNNNTTRG-3′	m6A	98.8	905	916	53.7
	3′-YTRGNNNNNN^{m}AAYC-5′	m6A	97.9	897	916	54.5
Efm733	5′-^{m}AGAWNNNNATTA-3′	m6A	78.5	278	354	22.3
	3′-TCTWNNNNT^{m}AAT-5′	m6A	73.2	259	354	23.6
OG1RF::*efa502I*	5′-R^{m}AYCNNNNNNTTRG-3′	m6A	99.9	795	796	162.8
	3′-YTRGNNNNNN^{m}AAYC-5′	m6A	99.0	788	796	163.2
OG1SSp::*efa733I*	5′-^{m}AGAWNNNNATTA-3′	m6A	99.0	323	326	180.4
	3′-TCTWNNNNT^{m}AAT-5′	m6A	98.5	321	326	180.8

[a]The underlined base indicates a modified base.

(25–33). With SMRT sequencing, 6-methyladenine (m6A) and 4-methylcytosine (m4C) can be easily detected with modest sequence coverage (~25× per strand), while 5-methylcytosine (m5C) detection requires high coverage (~250× per strand) (32, 34). Using SMRT sequencing, we identified two unique m6A methylation motifs in Efm502 and Efm733. Efm502 possessed m6A methylation at the underlined position of the motif 5′-RAYCNNNNNNTTRG-3′ (and its complementary strand 5′-CYAANNNNNNGRTY-3′), and Efm733 possessed m6A methylation at the underlined position of the motif 5′-AGAWNNNNATTA-3′ (and its complementary strand 5′-TAATNNNNWTCT-3′) (Table 2 and Data Set S2). These sequences are asymmetric and bipartite, which is characteristic of type I R-M methylation (12). Due to the coverage of our SMRT sequencing, m5C modification could not be accurately detected. The two unique m6A methylation patterns indicate that DNA from one strain would be recognized as foreign should it cross the strain barrier.

Expression in heterologous hosts links methylation activity to genes in Efm502 and Efm733. According to our predictions (Table 1), there is only one complete type I R-M system encoded by each of the strains Efm502 and Efm733. To determine if these systems are responsible for the methylation patterns identified by SMRT sequencing, we expressed the respective S and M subunits (EFQG_01131-EFQG_01132 [EFQG_01131-01132] for Efm502 and EFSG_05028-05027 for Efm733) in the heterologous host *E. faecalis* OG1RF or its spectinomycin/streptomycin-resistant relative OG1SSp. Previous work in our laboratory had characterized the methylome of OG1RF using SMRT and bisulfite sequencing (14). This allowed us to attribute any new methylation patterns observed during SMRT sequencing to the *E. faecium* genes that were expressed in the OG1RF background. SMRT sequencing of these strains detected the same methylation patterns originally identified in Efm502 and Efm733 (Table 2 and Data Set S2). These data demonstrate that EFQG_01131-01132 is responsible for 5′-RAYCNNNNNNTTRG-3′ methylation in Efm502 and that EFSG_05028-05027 is responsible for 5′-AGAWNNNNATTA-3′ methylation in Efm733. Because we have confirmed the function of these genes, we have named them Efa502I and Efa733I, which is consistent with the R-M system nomenclature convention established by New England Biolabs (12).

Efa502I and Efa733I reduce transformation efficiency in *E. faecium*. To determine whether the type I R-M systems in Efm733 and Efm502 actively defend against exogenous DNA, we constructed null strains (Efm733ΔRM and Efm502ΔRM) (Table 3) and evaluated their transformation efficiencies relative to their wild-type parent strains. Here, we utilized the broad-host-range plasmid pAT28 (Table 3) (35). The pAT28 sequence has motifs recognized by the type I R-M systems in both wild-type Efm733 (1 occurrence) and Efm502 (1 occurrence). The transformation of pAT28 into Efm733 and Efm502 served as a baseline for the experiment. If Efa733I and Efa502I are active, we expect to see higher efficiencies of pAT28 transformation into Efm733ΔRM and Efm502ΔRM, respectively. Indeed, we observed significantly higher efficiencies of transformation into type I R-M system-null strains (P value of <0.05 using one-tailed

TABLE 3 Strains used in this study

Strain or plasmid	Description[a]	Reference or source
Strains		
Enterococcus faecium		
1,231,502	Clade A1 isolate; also referred to as Efm502	39
1,230,933	Clade A1 isolate	39
1,231,410	Clade A1 isolate	39
1,231,408	Hybrid clade A1-clade B isolate	39
1,231,501	Clade A2 isolate	39
1,141,733	Clade B isolate; also referred to as Efm733	39
Com12	Clade B isolate	39
Com15	Clade B isolate	39
Efm733ΔRM	Efm733 with deletion of Efa733I (EFSG_05027-05029)	This study
Efm502ΔRM	Efm502 with deletion of Efa502I (EFQG_01130-01132)	This study
Enterococcus faecalis		
OG1RF	Rifampicin- and fusidic acid-resistant derivative of *E. faecalis* OG1	46
OG1SSp	Streptomycin- and spectinomycin-resistant derivative of *E. faecalis* OG1	46
OG1RF::*efa502I*	OG1RF with Efa502I inserted at the GISE	This study
OG1SSp::*efa733I*	OG1SSp with Efa733I inserted at the GISE	This study
Escherichia coli		
EC1000	Plasmid propagation host	42
DH5α	Plasmid propagation host	
STBL4	Plasmid propagation host	Fisher
STBL4(pGEM-T)	STBL4 with pGEM-T Easy	This study
STBL4(pRB01)	STBL4 with pGEM-T Easy vector containing EFSG_00659	This study
Plasmids		
pGEM-T Easy	Commercial plasmid for gene propagation	Promega
pLT06	Temperature-sensitive plasmid	41
pAT28	Shuttle vector for *E. faecalis*	35
pWH03	Used for gene insertion in the enterococci	14
pRB01	pGEM-T Easy vector with EFSG_00659	This study
pHA102	pWH03 containing NotI-digested fragments of Efa733I M and S loci (EFSG_05028-05029)	This study
pHA103	pWH03 containing NotI-digested fragments of Efa502I M and S loci (EFQG_01131-01132)	This study
pWH16	pLT06 containing fragments upstream and downstream of Efa502I	This study
pWH17	pLT06 containing fragments upstream and downstream of Efa733I	This study

[a]EFSG_05027-05029 refers to EFSG_05027 through EFSG_05029 (etc.).

Student's *t* test) (Fig. 2). These data demonstrate that the type I R-M systems in Efm733 and Efm502 actively function as mechanisms of genome defense.

m5C methylation occurs in Efm733. As described above, our SMRT sequencing had insufficient coverage depth for m5C methylome characterization. Hence, we used REase protection assays with commercially available methylation-sensitive REases to query the presence of m5C methylation in our eight *E. faecium* strains. Table 4 summarizes the recognition sequences and modifications of the enzymes used in this study. Only Efm733 showed evidence of cytosine modification, as it was digested by MspJI (Table 4 and Fig. S3).

To determine the exact cytosine methylation motif present in Efm733, genomic DNA (gDNA) was subjected to whole-genome bisulfite sequencing. Whole-genome-amplified DNA was used as a negative control since whole-genome amplification (WGA) removes all modifications. During bisulfite treatment, cytosine bases are converted to thymine unless they are protected by either m4C or m5C methylation. Additionally, our laboratory previously reported a method of distinguishing between m4C and m5C methylation using thymine conversion ratios of sequencing reads after bisulfite treatment (36). m5C methylation is sufficient to protect the cytosine residue completely from bisulfite conversion so that most sequencing reads at that position contain the original cytosine base. However, m4C methylation provides only partial protection from bisulfite conversion, so a thymine conversion rate of 0.5 at a particular position within the sequencing reads suggests the presence of m4C methylation. Bisulfite conversion and subsequent sequencing revealed that Efm733 possesses an m5C modification at the motif 5′-R^{m}CCGGY-3′ (Fig. 3 and Fig. S4); methylation occurs at the underlined position (position 2), which overlaps the MspJI recognition site

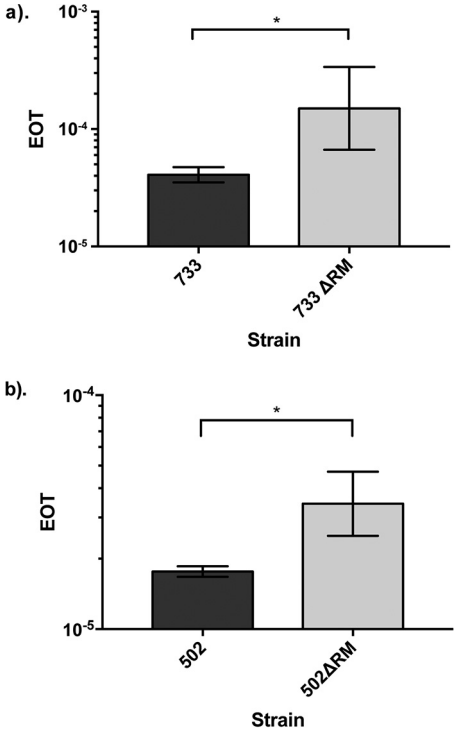

FIG 2 Type I R-M systems reduce transformability of Efm733 and Efm502. Three independent transformation experiments were performed. There is a statistical difference between the efficiencies of transformation of pAT28 into wild-type and R-M-null strains of Efm733 and Efm502. EOT, efficiency of transfer. *, $P < 0.05$.

(5′-CNNR-3′) and hence supports the evidence of methylation obtained from the MspJI digestion assays. To summarize this analysis, of the 780 5′-RCCGGY-3′ motifs in the Efm733 genome, 750 passed the coverage filter applied, and 748 of these were detected as being modified (the number of motifs detected is defined as the number of motifs with a coverage depth of more than 10× and a methylation ratio of 0.35). The average methylation ratio (defined as the mean of the methylation ratios from all motifs in the genome) was 96.5% (standard deviation, 6.2%), and the mean motif coverage was 61.5×. Based on the cytosine conversion ratio of close to 1.0, m5C modification is supported, which is consistent with why it was not detected by our SMRT sequencing.

EFSG_00659 is responsible for m5C methylation in Efm733. Based on the bioinformatics analyses, we hypothesized that EFSG_00659 was responsible for the m5C methylation found in Efm733 (Table 1). EFSG_00659 possessed no homologs in the other seven strains analyzed, making it a good candidate for the unique methylation found in Efm733. We queried the EFSG_00659 protein sequence against the REBASE gold-standard list and identified that it has high sequence similarity to M.AvaIX, M.VchO395I, and M.VchAI (E value of $\leq 3e^{-125}$; recognition sites are 5′-RCCGGY-3′). Interestingly, BLASTP identified no significant hits when EFSG_00659 was queried against the larger collection of 73 *E. faecium* genomes, indicating its unique presence in Efm733.

TABLE 4 Methylation-sensitive REase digestion reaction results[a]

REase	Recognition sequence	Recognition methylation(s)	No. of strains digested
McrBC	5′-$R^mC(N_{40}$-$N_{3000})R^mC$-3′	m5C, m4C	0/8
FspEI	5′-C^mC-3′	m5C	0/8
MspJI	5′-^{m}CNNR-3′	m5C	Efm733

[a]Recognition sequences and methylation patterns were retrieved from NEB.

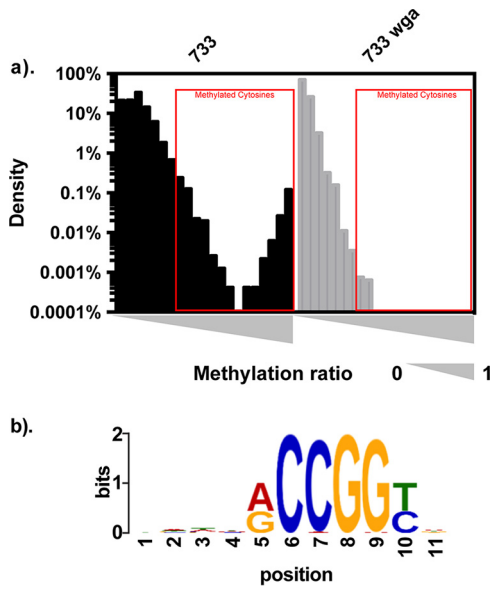

FIG 3 Bisulfite sequencing results for Efm733. Efm733 gDNA and its whole-genome-amplified control DNA were bisulfite converted and deep sequenced. The methylation ratio for each cytosine site was calculated. (a) The methylation ratio was plotted against the density of cytosine sites with that methylation ratio. The presence of cytosine sites with methylation ratios near 100% in Efm733 native gDNA but not the WGA control indicates the presence of m5C methylation. All cytosine sites with a methylation ratio of ≥0.35, as indicated by red boxes, from native gDNA samples were extracted, together with sequences 5 bp upstream and 5 bp downstream. The sequences were subjected to consensus motif analysis using MEME. (b) The consensus motif identified by this analysis is shown, and the center position (position 6) indicates where the modification was detected.

In order to link EFSG_00659 with the m5C methylation identified during bisulfite sequencing, we expressed it in the heterologous host *Escherichia coli* STBL4 and performed a REase protection assay. The REase AgeI recognizes the motif 5′-A̲CCGGT-3′, and its enzymatic activity is blocked if m5C methylation is present at the underlined position. This motif overlaps the m5C methylation motif in Efm733 identified by bisulfite sequencing. If the motif is methylated, DNA will be protected against digestion. We cloned EFSG_00659 into the vector pGEM-T and transformed it into *E. coli* STBL4, generating strain *E. coli* STBL4(pRB01). Genomic DNAs from Efm733, STBL4 (pGEM-T), and STBL4(pRB01) were treated with AgeI according to the manufacturer's instructions. EcoRI was used as a positive control for digestion. Figure 4 shows representative results of the digestions on a 1% agarose gel. As expected, Efm733 was digested by EcoRI and protected against digestion from AgeI. STBL4(pGEM-T) was digested by both EcoRI and AgeI, indicating that the original host and the empty vector

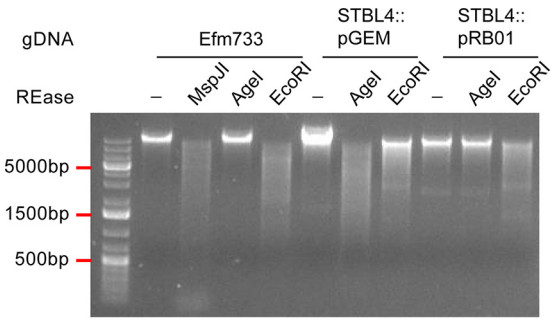

FIG 4 EFSG_00659 confers protection against AgeI digestion. AgeI digestion reaction products were analyzed by agarose gel electrophoresis with ethidium bromide staining. Bacterial gDNA was used as the substrate for REase reactions. Expression of the Efm733 gene EFSG_00659 in *E. coli* STBL4 protects *E. coli* gDNA from AgeI digestion. EcoRI is a positive control for digestion. −, no enzyme added.

pGEM-T did not possess the appropriate m5C methylation. STBL4(pRB01) was protected against digestion by AgeI, demonstrating that EFSG_00659 is responsible for the 5′-R ^{m}CCGGY-3′ methylation found in Efm733.

DISCUSSION

In this study, we used a combination of genomic and genetic approaches to identify a functional type I R-M system that is enriched in clade A1 *E. faecium* and that significantly alters the transformability of a model clade A1 strain. We propose that this R-M system impacts HGT rates among *E. faecium* mixed-clade communities, thereby helping to maintain the observed phylogenetic structure of *E. faecium* and facilitating HGT specifically among clade A1 strains. Mixed communities of *E. faecium* clades are expected to occur in environments where healthy and ill human hosts, human and animal hosts, and/or the feces of any of these hosts comingle (i.e., in sewage). In future studies, we plan to assess the impact of R-M on conjugative plasmid transfer, which is a major mode of HGT in enterococci and was not assessed in the present study.

An interesting observation from our study is the sequence diversity of type I S subunits encoded within *E. faecium* type I R-M systems having nearly identical R and M subunits (see Fig. S2a and b in the supplemental material). After further investigation into those alignments, we found that those S subunits sharing 50% to 70% overall amino acid sequence identities possess sequence diversity within one TRD, where the other TRD and the central conserved domain are conserved. This suggests that these systems share partial recognition sequences. Previous research showed that the diversification of type I R-M recognition sequences is driven by TRD exchanges, permutation of the dimerization domain, and circular permutation of TRDs (37). Our observation suggests that TRD recombination and reorganization events occur for *E. faecium* type I R-M systems outside clade A1. Future studies will use genomics to further explore the relationship between S subunit sequence diversity and its impact on *E. faecium* methylomes and interstrain and interclade HGT.

MATERIALS AND METHODS

Bacterial strains and growth conditions. The strains used in this study are shown in Table 3. All enterococci were grown in brain heart infusion (BHI) broth or agar at 37°C, unless otherwise stated. *Escherichia coli* strains were grown in Luria broth (LB) at 37°C and with shaking at 225 rpm unless otherwise stated. Antibiotic concentrations for enterococcal strains were as follows: 50 μg/ml for rifampin, 25 μg/ml for fusidic acid, 500 μg/ml for spectinomycin, 500 μg/ml for streptomycin, and 15 μg/ml for chloramphenicol. Antibiotic concentrations for *E. coli* strains were as follows: 15 μg/ml for chloramphenicol and 100 μg/ml for ampicillin. All REases were purchased from New England Biolabs (NEB) and used according to the manufacturer's instructions. PCR was performed using *Taq* polymerase (NEB) or Phusion (Fisher). Sanger sequencing to validate all genetic constructs was performed at the Massachusetts General Hospital DNA Core facility (Boston, MA).

Isolation of genomic DNA. Enterococcal strains were cultured overnight in BHI broth prior to genomic DNA (gDNA) extraction. Extraction was performed using a Qiagen blood and tissue DNeasy kit using a previously reported protocol (38). To isolate *E. coli* gDNA, bacteria were grown overnight in LB prior to extraction using either the blood and tissue DNeasy kit (Qiagen) or the UltraClean microbial DNA isolation kit (Qiagen) according to the manufacturer's instructions. Whole-genome amplification (WGA) control DNA was generated by amplification of native gDNA using the REPLI-g kit (Qiagen) according to the manufacturer's instructions.

SMRT sequencing and methylome detection in *E. faecium*. SMRT sequencing of *E. faecium* gDNAs and their WGA controls was performed by the Johns Hopkins Medical Institute Deep Sequencing and Microarray Core. After sequencing, the reads were aligned to existing references (NCBI RefSeq accession numbers NZ_GG688486 to NZ_GG688546 for Efm502 and NZ_GG688461 to NZ_GG688485 for Efm733) and analyzed using the RS modification and motif detection protocol in SMRT portal v1.3.3. WGA controls were used as methylation baselines.

Bioinformatic analysis of R-M systems in eight *E. faecium* genomes. The entire protein complement for eight previously sequenced *E. faecium* isolates (39) was analyzed. To identify potential MTases, the REBASE gold-standard list (15) was used as a reference. This list is comprised of biochemically verified MTases and REases. Each protein sequence from *E. faecium* genomes was analyzed using BLASTP against the REBASE gold-standard list. The protein sequences with significant (E value of $<1e^{-3}$) homology to REBASE gold-standard proteins were further filtered by protein size. If an *E. faecium* query protein length was less than half of its subject's length, the match was removed from the prediction list. Due to the sequence diversity of REases, which complicates their bioinformatic identification (15), guilt by association was used to identify full R-M systems, as we previously described (14). The proteins encoded near candidate DNA MTases were analyzed using BLAST and Pfam for conserved domains consistent with

TABLE 5 Primers used in this study

Primer	Sequence[a]
502SMR_A1F_XbaI	CTAGTCTAGACCCTTGTTCGATATAGACCC
502SMR_A1R_BamHI	CGCGGATCCCCTAATATGAAAGCAATTATCAAC
502SMR_A2F_BamHI	CGCGGATCCGAGATTATCTGCCGCACTAAAT
502SMR_A2R_XmaI	TCCCCCCGGGGTAGAAACATACACAGCTATAC
733SMR_A1F_XbaI	CTAGTCTAGAGGACGTAAACCTTCGACA
733SMR_A1R_BamHI	CGCGGATCCATGTTAGATGATGAGTTGATTCC
733SMR_A2F_BamHI	CGCGGATCCGCTTACTGTGTCATATTCCTC
733SMR_A2R_XmaI	TCCCCCCGGGCTCTACGAACTTGATTGAATCC
733_T1A_SM_F	TACGATGCGGCCGCGGGAAGAACCATTTACACAA
733_T1A_SM_R	ACATTGGCGGCCGCCGAGCATCATTTATTAACATGTTT
502_T1A_SM_F	TTGCATGCGGCCGCCGGATTTGGAGCCCGACT
502_T1A_SM_R	TACGATGCGGCCGCCCCATAGGTAAAGTGCCAC
EFSG_00659_F	CGTTACCTGCAGGCAGCGAAAACTAGTTGGG
EFSG_00659_R	GAATCAGGATCCGCCTCTAAAGAATAAAGATCC

[a]Underlined sequences are restriction enzyme recognition sites.

REase activities and/or sequence identity to confirmed REases. The amino acid sequence of each R-M candidate was then pairwise compared among all the eight strains to identify putative orthologs. If two protein sequences shared an amino acid identity of ≥90% with a query coverage of ≥90%, they were considered to be orthologous.

Expression of R-M systems in *E. faecalis* heterologous hosts. Genes encoding the specificity and methylation subunits of Efa733I (EFSG_05028-EFSG_05027) were PCR amplified in their entirety, including the upstream region to retain the native promoter, using primers 733_T1A_SM_F and 733_T1A_SM_R (see Table 5 for primer sequences). The PCR product was digested with NotI and ligated into NotI-digested pWH03 (14) using T4 DNA ligase (NEB), generating pHA102. pWH03 is a pLT06 derivative for expression of genes from a previously validated neutral genomic insertion site (EF2238-EF2239) for expression (GISE) (14, 40). pHA102 constructs were then introduced into *E. coli* DH5α via heat shock for propagation and sequence confirmation. pHA102 was electroporated into *E. faecalis* OG1SSp using a previously described method (13). An *E. faecalis* OG1SSp derivative with a chromosomal integration of Efa733I, referred to as OG1SSp::*efa733I*, was generated by temperature shifts and *p*-chlorophenylalanine counterselection, as previously described (41).

Genes encoding the specificity and methylation subunits of Efa502I (EFQG_01131-EFQG_01132) were PCR amplified using primers 502_T1A_SM_F and 502_T1A_SM_R. The PCR product was then TA cloned into the pGEM-T Easy vector (Promega) and introduced into DH5α via heat shock to generate pGEM-SMA1. pGEM-SMA1 was then digested with NotI, and the digestion reaction product was used as the insert for ligation into NotI-digested pWH03. The ligation reaction product was then introduced into DH5α, and colonies were screened for chloramphenicol resistance and ampicillin susceptibility to ensure that the pGEM backbone was not ligated into pWH03. Once the construct, referred to as pHA103, was confirmed via Sanger sequencing, it was introduced into electrocompetent *E. faecalis* OG1RF cells. An *E. faecalis* OG1RF derivative with a chromosomal integration of Efa502I, referred to as OG1RF::*efa502I*, was generated by temperature shifts and *p*-chlorophenylalanine counterselection. All plasmids and strains for heterologous expression were validated by PCR and Sanger sequencing.

SMRT sequencing of *E. faecalis* OG1 derivatives expressing Efa502I or Efa733I was performed by the University of Michigan sequencing core facility. Reads were mapped to the *E. faecalis* OG1RF reference sequence (GenBank accession number NC_017316), and the methylation motifs were detected using the RS modification and motif detection protocol in SMRT portal v.2.3.2. *In silico* controls were used as modification baselines.

Generation of *E. faecium* R-M deletion mutants. Regions up- and downstream of Efa502I and Efa733I were PCR amplified using primers listed in Table 5, ligated into pLT06, and transformed into *E. coli* EC1000 (42), generating pWH16 and pWH17 (Table 3). Insert sequences were confirmed using Sanger sequencing. *E. faecium* strains were made electrocompetent using a previously reported protocol (43). Two micrograms of sequence-confirmed plasmids was electroporated into electrocompetent Efm733 and Efm502. The generation of deletion mutants was accomplished using temperature shifts and *p*-chlorophenylalanine counterselection, as previously described (41). The successful deletion mutants were sequence confirmed by PCR and Sanger sequencing.

Transformation efficiency test. Efm733, Efm502, and their respective R-M deletion mutants were made electrocompetent using a modified version of a previously reported protocol (43). Briefly, cultures grown overnight were diluted 10-fold in BHI medium and cultured to an optical density at 600 nm (OD_{600}) of ~0.6. The bacteria were then pelleted and treated with filter-sterilized lysozyme buffer (10 mM Tris-HCl [pH 8.0], 10 mM EDTA [pH 8.0], 50 mM NaCl) supplemented with 83 μl of a 2.5-kU/ml mutanolysin stock for 30 min at 37°C. The cells were then pelleted and washed three times with ice-cold filter-sterilized electroporation buffer (0.5 M sucrose and 10% glycerol). Finally, the cells were pelleted, resuspended in electroporation buffer, and aliquoted for storage at −80°C for future use. One microgram of pAT28 (35) was electroporated into the electrocompetent *E. faecium* cells. The counts of total viable cells and spectinomycin-resistant cells were determined by serial dilution and plating. The transformation efficiency was expressed as a percentage of transformed (spectinomycin-resistant) cells per total viable

cells. Three independent experiments were performed, and statistical significance was assessed using the unpaired one-tailed Student *t* test.

Distribution analysis of putative R-M systems and orphan MTases. The amino acid sequences for select R-M system and orphan MTase candidates were queried against a collection of 73 *E. faecium* isolates previously analyzed by Lebreton et al. (8) using BLASTP. Any proteins that shared >90% query coverage and amino acid identity were considered orthologs. Fisher's exact test was used to determine if an orphan MTase or R-M system was significantly over- or underrepresented in a particular clade.

REase protection assays. To identify m5C methylation, gDNA was treated with the methylation-sensitive REases McrBC, FspEI, and MspJI (NEB). A total of 500 ng gDNA was incubated with each REase at 37°C for 3 h (McrBC) or 6 h (FspEI and MspJI), followed by analysis by electrophoresis on a 1% agarose gel with ethidium bromide.

Bisulfite sequencing. Whole-genome bisulfite sequencing libraries were constructed using the Illumina TruSeq LT PCR Free kit and the Qiagen EpiTect bisulfite kit. Native DNA was isolated as described above. Whole-genome-amplified control DNA was generated by amplification of native gDNA using the Qiagen REPLI-g kit, according to the manufacturer's instructions. For bisulfite sequencing, briefly, 2 μg each of native and WGA control DNA was fragmented using NEB fragmentase. DNA fragments ranging from 200 bp to 700 bp were gel extracted and end repaired. After A tailing of DNA fragments, Illumina TruSeq adapters were ligated. Bisulfite conversion was then performed using the Qiagen EpiTect bisulfite kit, according to the manufacturer's instructions. An 8-cycle PCR enrichment with Illumina primer mix was performed, followed by size selection and gel purification. The libraries were sequenced using Illumina MiSeq with 2- by 75-bp paired-end chemistry.

Whole-genome bisulfite sequencing analysis. The sequencing reads were analyzed using Bismark (44), with additional quality control and filtering as described previously (14). Briefly, the Illumina reads were mapped to the *in silico* bisulfite-converted references (44). We then quantified the conversion rate of each mapped read by calculating the percentage of converted C bases (which will result in T) to the total number of C bases in the reference within the mapped region. The mapped reads with a ≤80% conversion rate were filtered out from the analysis (14). Next, the coverage depth and methylation ratio were calculated for each C site. The methylation ratio was calculated by dividing the total number of C bases by the coverage depth at each C site. A fully methylated C residue thus protected from bisulfite conversion will have a methylation ratio near 1. An unmethylated C residue will have a methylation ratio near zero. To identify consensus methylation motifs, C sites with a methylation ratio of ≥0.35 and a coverage depth of ≥10×, along with the sequences 5 bp upstream and 5 bp downstream, were extracted. The extracted sequences were subjected to a MEME motif search (45).

Confirmation of m5C MTase activity. Primers EFSG_00659_F and EFSG_00659_R (Table 5) were used to amplify the entire Efm733 EFSG_00659 coding region and its upstream predicted promoter. The PCR product was then cloned into the pGEM-T Easy vector (Promega) according to the manufacturer's instructions and transformed into *E. coli* STBL4 (Fisher) to generate pRB01. REase digestion assays with methylation-sensitive enzymes were performed on purified *E. coli* and *E. faecium* gDNAs as described above.

Accession number(s). DNA sequence data generated in this study have been deposited in the Sequence Read Archive under accession numbers PRJNA397049 (for SMRT sequencing data) and PRJNA488088 (for Illumina bisulfite sequencing data).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/AEM.02174-18.

SUPPLEMENTAL FILE 1, PDF file, 0.8 MB.

SUPPLEMENTAL FILE 2, XLSX file, 0.01 MB.

SUPPLEMENTAL FILE 3, XLSX file, 0.01 MB.

ACKNOWLEDGMENTS

This work was supported by Public Health Service grants K22AI099088 and R01AI116610 to K.L.P.

REFERENCES

1. Noble CJ. 1978. Carriage of group D streptococci in the human bowel. J Clin Pathol 31:1182–1186. https://doi.org/10.1136/jcp.31.12.1182.
2. Chenoweth C, Schaberg D. 1990. The epidemiology of enterococci. Eur J Clin Microbiol Infect Dis 9:80–89. https://doi.org/10.1007/BF01963631.
3. Agudelo Higuita NI, Huycke MM. 2014. Enterococcal disease, epidemiology, and implications for treatment. *In* Gilmore MS, Clewell DB, Ike Y, Shankar N (ed), Enterococci: from commensals to leading causes of drug resistant infection. Massachusetts Eye and Ear Infirmary, Boston, MA. https://www.ncbi.nlm.nih.gov/books/NBK190429/.
4. Louis E, Galloway-Peña J, Roh JH, Latorre M, Qin X, Murray BE. 2012. Genomic and SNP analyses demonstrate a distant separation of the hospital and community-associated clades of *Enterococcus faecium*. PLoS One 7:e30187. https://doi.org/10.1371/journal.pone.0030187.
5. Palmer KL, Godfrey P, Griggs A, Kos VN, Zucker J, Desjardins C, Cerqueira G, Gevers D, Walker S, Wortman J, Feldgarden M, Haas B, Birren B, Gilmore MS. 2012. Comparative genomics of enterococci: variation in *Enterococcus faecalis*, clade structure in *E. faecium*, and defining characteristics of *E. gallinarum* and *E. casseliflavus*. mBio 3:e00318-11. https://doi.org/10.1128/mBio.00318-11.
6. Willems RJL, Top J, van Schaik W, Leavis H, Bonten M, Siren J, Hanage WP, Corander J. 2012. Restricted gene flow among hospital subpopulations of *Enterococcus faecium*. mBio 3:e00151-12. https://doi.org/10.1128/mBio.00151-12.
7. van Schaik W, Top J, Riley DR, Boekhorst J, Vrijenhoek JE, Schapendonk CM, Hendrickx AP, Nijman IJ, Bonten MJ, Tettelin H, Willems RJ. 2010. Pyrosequencing-based comparative genome analysis of the nosocomial

pathogen *Enterococcus faecium* and identification of a large transferable pathogenicity island. BMC Genomics 11:239. https://doi.org/10.1186/1471-2164-11-239.
8. Lebreton F, van Schaik W, McGuire AM, Godfrey P, Griggs A, Mazumdar V, Corander J, Cheng L, Saif S, Young S, Zeng Q, Wortman J, Birren B, Willems RJ, Earl AM, Gilmore MS. 2013. Emergence of epidemic multidrug-resistant Enterococcus faecium from animal and commensal strains. mBio 4:e00534-13. https://doi.org/10.1128/mBio.00534-13.
9. Nandi T, Holden MTG, Didelot X, Mehershahi K, Boddey JA, Beacham I, Peak I, Harting J, Baybayan P, Guo Y, Wang S, How LC, Sim B, Essex-Lopresti A, Sarkar-Tyson M, Nelson M, Smither S, Ong C, Aw LT, Hoon CH, Michell S, Studholme DJ, Titball R, Chen SL, Parkhill J, Tan P. 2015. *Burkholderia pseudomallei* sequencing identifies genomic clades with distinct recombination, accessory, and epigenetic profiles. Genome Res 25:129–141. https://doi.org/10.1101/gr.177543.114.
10. Budroni S, Siena E, Hotopp JCD, Seib KL, Serruto D, Nofroni C, Comanducci M, Riley DR, Daugherty SC, Angiuoli SV, Covacci A, Pizza M, Rappuoli R, Moxon ER, Tettelin H, Medini D. 2011. *Neisseria meningitidis* is structured in clades associated with restriction modification systems that modulate homologous recombination. Proc Natl Acad Sci U S A 108:4494–4499. https://doi.org/10.1073/pnas.1019751108.
11. Tock MR, Dryden DT. 2005. The biology of restriction and anti-restriction. Curr Opin Microbiol 8:466–472. https://doi.org/10.1016/j.mib.2005.06.003.
12. Roberts RJ, Belfort M, Bestor T, Bhagwat AS, Bickle TA, Bitinaite J, Blumenthal RM, Degtyarev S, Dryden DT, Dybvig K, Firman K, Gromova ES, Gumport RI, Halford SE, Hattman S, Heitman J, Hornby DP, Janulaitis A, Jeltsch A, Josephsen J, Kiss A, Klaenhammer TR, Kobayashi I, Kong H, Kruger DH, Lacks S, Marinus MG, Miyahara M, Morgan RD, Murray NE, Nagaraja V, Piekarowicz A, Pingoud A, Raleigh E, Rao DN, Reich N, Repin VE, Selker EU, Shaw PC, Stein DC, Stoddard BL, Szybalski W, Trautner TA, Van Etten JL, Vitor JM, Wilson GG, Xu SY. 2003. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. Nucleic Acids Res 31:1805–1812. https://doi.org/10.1093/nar/gkg274.
13. Price VJ, Huo W, Sharifi A, Palmer KL. 2016. CRISPR-Cas and restriction-modification act additively against conjugative antibiotic resistance plasmid transfer in *Enterococcus faecalis*. mSphere 1:e00064-16. https://doi.org/10.1128/mSphere.00064-16.
14. Huo W, Adams HM, Zhang MQ, Palmer KL. 2015. Genome modification in *Enterococcus faecalis* OG1RF assessed by bisulfite sequencing and single-molecule real-time sequencing. J Bacteriol 197:1939–1951. https://doi.org/10.1128/JB.00130-15.
15. Roberts RJ, Vincze T, Posfai J, Macelis D. 2015. REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res 43:D298–D299. https://doi.org/10.1093/nar/gku1046.
16. Luria SE, Human ML. 1952. A nonhereditary, host-induced variation of bacterial viruses. J Bacteriol 64:557–569.
17. Bertani G, Weigle JJ. 1953. Host controlled variation in bacterial viruses. J Bacteriol 65:113–121.
18. Murray NE. 2000. Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). Microbiol Mol Biol Rev 64:412–434. https://doi.org/10.1128/MMBR.64.2.412-434.2000.
19. Taylor I, Patel J, Firman K, Kneale G. 1992. Purification and biochemical characterisation of the EcoR124 type I modification methylase. Nucleic Acids Res 20:179–186. https://doi.org/10.1093/nar/20.2.179.
20. Gough JA, Murray NE. 1983. Sequence diversity among related genes for recognition of specific targets in DNA molecules. J Mol Biol 166:1–19. https://doi.org/10.1016/S0022-2836(83)80047-3.
21. Gann AA, Campbell AJ, Collins JF, Coulson AF, Murray NE. 1987. Reassortment of DNA recognition domains and the evolution of new specificities. Mol Microbiol 1:13–22. https://doi.org/10.1111/j.1365-2958.1987.tb00521.x.
22. Clarke J, Wu H-C, Jayasinghe L, Patel A, Reid S, Bayley H. 2009. Continuous base identification for single-molecule nanopore DNA sequencing. Nat Nanotechnol 4:265–270. https://doi.org/10.1038/nnano.2009.12.
23. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, et al. 2009. Real-time DNA sequencing from single polymerase molecules. Science 323:133–138. https://doi.org/10.1126/science.1162986.
24. Korlach J, Bjornson KP, Chaudhuri BP, Cicero RL, Flusberg BA, Gray JJ, Holden D, Saxena R, Wegener J, Turner SW. 2010. Real-time DNA sequencing from single polymerase molecules. Methods Enzymol 472:431–455. https://doi.org/10.1016/S0076-6879(10)72001-2.
25. Clark TA, Murray IA, Morgan RD, Kislyuk AO, Spittle KE, Boitano M, Fomenkov A, Roberts RJ, Korlach J. 2012. Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. Nucleic Acids Res 40:e29. https://doi.org/10.1093/nar/gkr1146.
26. Fang G, Munera D, Friedman DI, Mandlik A, Chao MC, Banerjee O, Feng Z, Losic B, Mahajan MC, Jabado OJ, Deikus G, Clark TA, Luong K, Murray IA, Davis BM, Keren-Paz A, Chess A, Roberts RJ, Korlach J, Turner SW, Kumar V, Waldor MK, Schadt EE. 2012. Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing. Nat Biotechnol 30:1232–1239. https://doi.org/10.1038/nbt.2432.
27. Murray IA, Clark TA, Morgan RD, Boitano M, Anton BP, Luong K, Fomenkov A, Turner SW, Korlach J, Roberts RJ. 2012. The methylomes of six bacteria. Nucleic Acids Res 40:11450–11462. https://doi.org/10.1093/nar/gks891.
28. Bendall ML, Luong K, Wetmore KM, Blow M, Korlach J, Deutschbauer A, Malmstrom RR. 2013. Exploring the roles of DNA methylation in the metal-reducing bacterium *Shewanella oneidensis* MR-1. J Bacteriol 195:4966–4974. https://doi.org/10.1128/JB.00935-13.
29. Davis BM, Chao MC, Waldor MK. 2013. Entering the era of bacterial epigenomics with single molecule real time DNA sequencing. Curr Opin Microbiol 16:192–198. https://doi.org/10.1016/j.mib.2013.01.011.
30. Kozdon JB, Melfi MD, Luong K, Clark TA, Boitano M, Wang S, Zhou B, Gonzalez D, Collier J, Turner SW, Korlach J, Shapiro L, McAdams HH. 2013. Global methylation state at base-pair resolution of the *Caulobacter* genome throughout the cell cycle. Proc Natl Acad Sci U S A 110:E4658–E4667. https://doi.org/10.1073/pnas.1319315110.
31. Richardson PM, Lluch-Senar M, Luong K, Lloréns-Rico V, Delgado J, Fang G, Spittle K, Clark TA, Schadt E, Turner SW, Korlach J, Serrano L. 2013. Comprehensive methylome characterization of *Mycoplasma genitalium* and *Mycoplasma pneumoniae* at single-base resolution. PLoS Genet 9:e1003191. https://doi.org/10.1371/journal.pgen.1003191.
32. Roberts RJ, Carneiro MO, Schatz MC. 2013. The advantages of SMRT sequencing. Genome Biol 14:405. https://doi.org/10.1186/gb-2013-14-6-405.
33. Krebes J, Morgan RD, Bunk B, Spröer C, Luong K, Parusel R, Anton BP, König C, Josenhans C, Overmann J, Roberts RJ, Korlach J, Suerbaum S. 2014. The complex methylome of the human gastric pathogen *Helicobacter pylori*. Nucleic Acids Res 42:2415–2432. https://doi.org/10.1093/nar/gkt1201.
34. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, Korlach J, Turner SW. 2010. Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat Methods 7:461–465. https://doi.org/10.1038/nmeth.1459.
35. Trieu-Cuot P, Carlier C, Poyart-Salmeron C, Courvalin P. 1990. A pair of mobilizable shuttle vectors conferring resistance to spectinomycin for molecular cloning in *Escherichia coli* and in Gram-positive bacteria. Nucleic Acids Res 18:4296. https://doi.org/10.1093/nar/18.14.4296.
36. Huo W. 2017. *Enterococcus faecalis* genome defense systems and their impact on conjugative antibiotic resistance plasmid transfer. PhD thesis. The University of Texas at Dallas, Richardson, TX. http://hdl.handle.net/10735.1/5776.
37. Loenen WAM, Dryden DTF, Raleigh EA, Wilson GG. 2014. Type I restriction enzymes and their relatives. Nucleic Acids Res 42:20–44. https://doi.org/10.1093/nar/gkt847.
38. Adams HM, Li X, Mascio C, Chesnel L, Palmer KL. 2015. Mutations associated with reduced surotomycin susceptibility in *Clostridium difficile* and *Enterococcus* species. Antimicrob Agents Chemother 59:4139–4147. https://doi.org/10.1128/AAC.00526-15.
39. Palmer KL, Carniol K, Manson JM, Heiman D, Shea T, Young S, Zeng Q, Gevers D, Feldgarden M, Birren B, Gilmore MS. 2010. High-quality draft genome sequences of 28 *Enterococcus* sp. isolates. J Bacteriol 192:2469–2470. https://doi.org/10.1128/JB.00153-10.
40. DebRoy S, van der Hoeven R, Singh KV, Gao P, Harvey BR, Murray BE, Garsin DA. 2012. Development of a genomic site for gene integration and expression in *Enterococcus faecalis*. J Microbiol Methods 90:1–8. https://doi.org/10.1016/j.mimet.2012.04.011.
41. Thurlow LR, Thomas VC, Hancock LE. 2009. Capsular polysaccharide

production in *Enterococcus faecalis* and contribution of CpsF to capsule serospecificity. J Bacteriol 191:6203–6210. https://doi.org/10.1128/JB.00592-09.
42. Leenhouts K, Buist G, Bolhuis A, ten Berge A, Kiel J, Mierau I, Dabrowska M, Venema G, Kok J. 1996. A general system for generating unlabelled gene replacements in bacterial chromosomes. Mol Gen Genet 253:217–224. https://doi.org/10.1007/s004380050315.
43. Bhardwaj P, Ziegler E, Palmer KL. 2016. Chlorhexidine induces VanA-type vancomycin resistance genes in enterococci. Antimicrob Agents Chemother 60:2209–2221. https://doi.org/10.1128/AAC.02595-15.
44. Krueger F, Andrews SR. 2011. Bismark: a flexible aligner and methylation caller for bisulfite-Seq applications. Bioinformatics 27:1571–1572. https://doi.org/10.1093/bioinformatics/btr167.
45. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res 37:W202–W208. https://doi.org/10.1093/nar/gkp335.
46. Gold OG, Jordan HV, van Houte J. 1975. The prevalence of enterococci in the human mouth and their pathogenicity in animal models. Arch Oral Biol 20:473–477. https://doi.org/10.1016/0003-9969(75)90236-8.