RESEARCH ARTICLE

Genetic Stabilization of the Drug-Resistant PMEN1 Pneumococcus Lineage by Its Distinctive DpnIII Restriction-Modification System

Rory A. Eutsey,[a] Evan Powell,[a] Janina Dordel,[b] Susannah J. Salter,[b] Tyson A. Clark,[c] Jonas Korlach,[c] Garth D. Ehrlich,[d,e] N. Luisa Hiller[a,f]

Center of Excellence in Biofilm Research, Allegheny Health Network, Pittsburgh, Pennsylvania, USA[a]; Pathogen Genomics, The Wellcome Trust Sanger Institute, Cambridge, United Kingdom[b]; Pacific Biosciences, Menlo Park, California, USA[c]; Department of Microbiology and Immunology, Drexel University College of Medicine, Philadelphia, Pennsylvania, USA[d]; Department of Otolaryngology-Head and Neck Surgery, Drexel University College of Medicine, Philadelphia, Pennsylvania, USA[e]; Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA[f]

ABSTRACT **The human pathogen *Streptococcus pneumoniae* (pneumococcus) exhibits a high degree of genomic diversity and plasticity. Isolates with high genomic similarity are grouped into lineages that undergo homologous recombination at variable rates. PMEN1 is a pandemic, multidrug-resistant lineage. Heterologous gene exchange between PMEN1 and non-PMEN1 isolates is directional, with extensive gene transfer from PMEN1 strains and only modest transfer into PMEN1 strains. Restriction-modification (R-M) systems can restrict horizontal gene transfer, yet most pneumococcal strains code for either the DpnI or DpnII R-M system and neither limits homologous recombination. Our comparative genomic analysis revealed that PMEN1 isolates code for DpnIII, a third R-M system syntenic to the other Dpn systems. Characterization of DpnIII demonstrated that the endonuclease cleaves unmethylated double-stranded DNA at the tetramer sequence 5′ GATC 3′, and the cognate methylase is a C5 cytosine-specific DNA methylase. We show that DpnIII decreases the frequency of recombination under *in vitro* conditions, such that the number of transformants is lower for strains transformed with unmethylated DNA than in those transformed with cognately methylated DNA. Furthermore, we have identified two PMEN1 isolates where the DpnIII endonuclease is disrupted, and phylogenetic work by Croucher and colleagues suggests that these strains have accumulated genomic differences at a higher rate than other PMEN1 strains. We propose that the R-M locus is a major determinant of genetic acquisition; the resident R-M system governs the extent of genome plasticity.**

IMPORTANCE **Pneumococcus is one of the most important community-acquired bacterial pathogens. Pneumococcal strains can develop resistance to antibiotics and to serotype vaccines by acquiring genes from other strains or species. Thus, genomic plasticity is associated with strain adaptability and pneumococcal success. PMEN1 is a widespread and multidrug-resistant highly pathogenic pneumococcal lineage, which has evolved over the past century and displays a relatively stable genome. In this study, we characterize DpnIII, a restriction-modification (R-M) system that limits recombination. DpnIII is encountered in the PMEN1 lineage, where it replaces other R-M systems that do not decrease plasticity. Our hypothesis is that this genomic region, where different pneumococcal lineages code for variable R-M systems, plays a role in the fine-tuning of the extent of genomic plasticity. It is possible that well-adapted lineages such as PMEN1 have a mechanism to increase genomic stability, rather than foster genomic plasticity.**

Received 2 February 2015 **Accepted** 21 May 2015 **Published** 16 June 2015

Citation Eutsey RA, Powell E, Dordel J, Salter SJ, Clark TA, Korlach J, Ehrlich GD, Hiller NL. 2015. Genetic stabilization of the drug-resistant PMEN1 pneumococcus lineage by its distinctive DpnIII restriction-modification system. mBio 6(3):e00173-15. doi:10.1128/mBio.00173-15.

Invited Editor Donald A. Morrison, University of Illinois at Chicago **Editor** Howard A. Shuman, University of Chicago

Address correspondence to N. Luisa Hiller, lhiller@andrew.cmu.edu, or Garth D. Ehrlich, garth.ehrlich@drexelmed.edu.

The Gram-positive bacterium *Streptococcus pneumoniae* (pneumococcus) is an opportunistic pathogen with high rates of asymptomatic colonization. In pediatric day care centers, the rates of colonization have been estimated to be over 60% (1). In some cases, colonization can develop into diseases that range from mucosal and respiratory infections, including otitis media and pneumonia, to severe invasive diseases such as sepsis and meningitis. Global estimates indicate that pneumococcus is responsible for ~850,000 deaths a year of children under the age of 5 years (2). The severity of pneumococcal disease has been exacerbated by the global spread of multidrug-resistant lineages (3). Furthermore, while implementation of the polyvalent pneumococcal vaccines over the past 15 years has led to extensive reductions in deaths in both pediatric and geriatric populations (4–6), the rates of asymptomatic colonization remain almost unchanged and multiple lineages resistant to the vaccine have emerged in the postvaccine era (7, 8).

There is extensive genomic diversity among pneumococcal strains. Approximately half of the supragenome (pangenome) is shared by all strains (core set), and the remainder is unevenly distributed among the various isolates (9, 10). This variability in gene possession leads to different disease phenotypes, as well as variability in drug and vaccine resistance (11). This high level of diversity arises as a direct result of the multiple horizontal gene

transfer mechanisms encoded by pneumococcus; nearly all strains are naturally competent, allowing isolates to take up DNA from the environment and integrate it into their genome by homologous recombination (12–16). In addition, many strains possess conjugative transposons that enable mating to occur (17). Extracellular and intracellular reservoirs of DNA are available to pneumococcus. This bacterium often forms chronic biofilms where cells are embedded in an extracellular polymeric substance rich in DNA (18), and cells can hold internal reservoirs of single-stranded DNA (ssDNA) corresponding to half a genome equivalent (19). This plasticity allows for strain evolution over short time periods and has played a key role in the acquisition of drug resistance, as well as evasion of subtype vaccines (20–24). Population and *in vitro* studies, as well as mathematical models, suggest that different pneumococcal isolates undergo homologous recombination at different rates (25–30).

The PMEN1 (Pneumococcal Molecular Epidemiology Network clone 1) lineage was first identified in Spain in the 1980s as the Spain23F-1 isolate and soon spread worldwide. Now it is one of the most common pneumococcal strains in both carriage and disease (3, 31, 32). Many PMEN1 isolates code for multiple genes associated with drug resistance, and within the lineage, there has been a significant level of serotype switching, which can afford resistance to the pneumococcal vaccine (33–35). Whole-genome comparisons of pneumococcal strains suggest that strains in the PMEN1 lineage are frequent DNA donors (32). Wyres and colleagues demonstrate that two unrelated drug-resistant lineages, PMEN3 and CGSP14, have, respectively, acquired 5.3 and 9.5% of their genomes from the PMEN1 lineage. Furthermore, the PMEN1 alleles for the penicillin binding proteins encoded by *pbp2x*, *pbp1a*, and *pbp2b*, which provide drug resistance, are widely distributed among non-PMEN1 lineages. In contrast, while pneumococcal strains can acquire heterologous sequences via homologous recombination of flanking regions, there is little evidence of such events in PMEN1 isolates. In fact, 95% of the coding sequences from modern PMEN1 isolates are ≥98% similar to a common ancestor isolated in 1967 (23F/4), and many PMEN1 strains show little or no evidence of recombination. It is noteworthy that whole-genome comparisons of 240 PMEN1 strains identified hundreds of recombination events within this lineage but very few of these correspond to heterologous regions; when present, these represent genes under strong selective pressure (33). Thus, together, these studies suggest that the PMEN1 lineage is an extensive gene donor but not a major recipient of novel genes.

The clinical importance of PMEN1 led us to search for genes enriched in this lineage relative to other pneumococcal strains. This comparison revealed the DpnIII restriction-modification (R-M) system, which is present in all PMEN1 strains but rare outside this lineage. R-M systems have been discussed as factors that may regulate bacterial speciation (36). R-M systems consist of a restriction enzyme (RE) and a cognate methyltransferase (MT). The RE recognizes a specific DNA sequence and cleaves at or adjacent to that site, and the MT modifies DNA at the recognition sequence and, in so doing, prevents cleavage by the RE. In *Helicobacter pylori*, an RE (active in ~19% of strains) has been shown to restrict the uptake of unmethylated plasmids and transformation with unmethylated chromosomal DNA. The activity of the cognate MT was highly conserved, suggesting that this function may provide increased fitness under some conditions (37). In *Neisseria*, many R-M systems are specific to phylogenetic clades, consistent with the idea that these systems restrict natural transformation (38). Yet not all R-M systems restrict gene transfer, as evidenced by DpnI and DpnII, the two well-characterized endonucleases in pneumococcus. DpnII codes for two MTs, one of which promotes recombination by methylating foreign ssDNA and protecting the integrated region from postreplicative digestion by DpnII (39). Without the DpnA methylase, the unmethylated imported DNA could undergo recombination, but once the cells replicated, the newly integrated unmethylated double-stranded region would be sensitive to DpnII cleavage. The discovery of DpnIII, an R-M system present in the PMEN1 lineage, combined with the largely unidirectional transfer of heterologous DNA from PMEN1 to non-PMEN1 strains led us to hypothesize that DpnIII decreases the flow of novel gene into this lineage and, in doing so, may be playing a role in stabilizing a well-adapted genome content.

RESULTS

Identification of a lineage-specific R-M system by comparative genomics. To explore the possibility that the directional transfer of PMEN1 genes was genetically determined, we performed a comparative genomic analysis of PMEN1 and non-PMEN1 strains to identify genes unique to the PMEN1 lineage. To this end, we used a set of 59 curated pneumococcal whole-genome sequences (WGS), including four from the PMEN1 lineage (see Table S1 in the supplemental material). The selected non-PMEN1 strains reflect a large variety of multilocus sequence types (MLSTs) and serotypes, as well as strains isolated from different disease states and geographic locations.

These WGS were submitted to RAST for coding sequence (CDS) identification and annotation (40). The 125,612 CDSs were organized into 3,571 clusters of homologous sequences. An uncharacterized R-M system was present in the four PMEN1 isolates but no other genomes. This system is referred to as DpnIII, as it is the third R-M system characterized in *S. pneumoniae* (original name for this species, *Diplococcus pneumoniae*). This R-M system is characterized by two genes, one for an RE, *r.dpnIII* {Spn23FORF18650P [referred to as SPN23F18640 in the original SPN23F annotation (41)]}, and one for a methylase, *m.dpnIII* (M.Spn23FORF18650P) (42). The two genes are on opposite DNA strands and flanked by genes predicted to be involved in xanthine metabolism and galactose/lactose uptake and utilization (Fig. 1). In non-PMEN1 strains, this region codes for either the DpnI R-M system (genes *dpnC* and *dpnD*) or the DpnII R-M system (endonuclease gene *dpnB* and two MT genes [*dpnM* and *dpnA*]) (39, 43–45). Thus, variable pneumococcal R-M systems are syntenic among the pneumococcal genomes.

To determine the distribution of DpnIII within additional pneumococcal strains, we used BLAST to search for *r.dpnIII* and *m.dpnIII* sequences in the set of *S. pneumoniae* sequences in the NCBI nonredundant database. This region was identified in known PMEN1 genomes and eight additional strains with no available additional information. To determine whether the unannotated strains coding for DpnIII are related to PMEN1, we aligned the genomes with the reference PMEN1 strain (SPN23F) and used Mauve to count the single-nucleotide polymorphisms (SNPs) in the aligned genomes. Seven genomes displayed <1,500 SNPs relative to the reference PMEN1 strain, consistent with the known degree of difference among members of this lineage (*Streptococcus pneumoniae* GA44378, NP170, GA41565, GA11663,

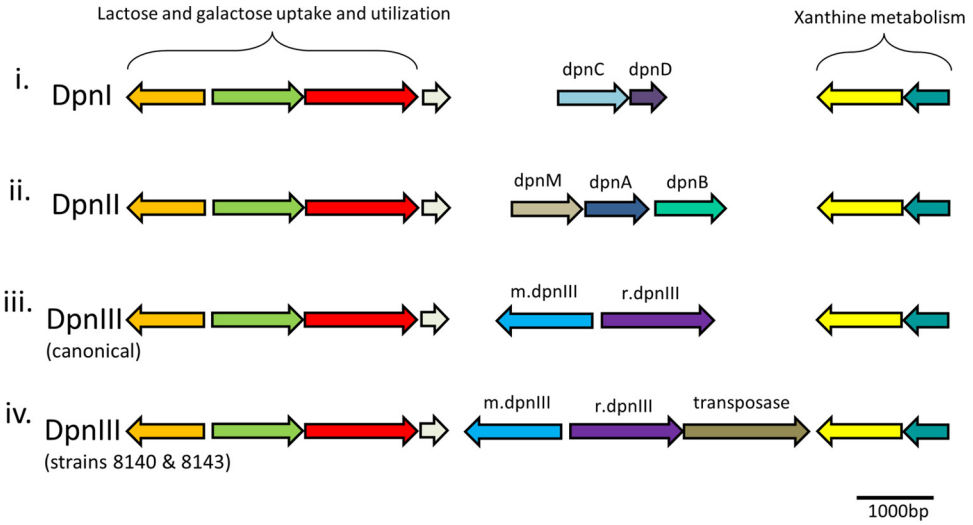

FIG 1 Loci of R-M systems in *S. pneumoniae*. Syntenic pneumococcal R-M systems. (i) DpnI-encoding locus with endonuclease encoded by *dpnC* and protein of unknown function encoded by *dpnD*. (ii) DpnII-encoding locus with methylases encoded by *dpnM* and *dpnA* and endonuclease encoded by *dpnB*. (iii) Canonical DpnIII-encoding locus of PMEN1 strains with methylase encoded by *m.dpnIII* and endonuclease encoded by *r.dpnIII*. (iv) DpnIII-encoding locus in PMEN1 strains 8140 and 8143 where a transposase is inserted at the 3′ end of *r.dpnIII*. All Dpn-encoding loci are flanked by genes predicted to be involved in lactose and galactose uptake and utilization at the 5′ end and by genes predicted to be involved in xanthine metabolism at the 3′ end. This illustration is based on RAST annotation.

GA13430, 2070005, and 357). One strain displayed >25,000 SNPs distributed in many chromosomal regions, suggesting that it represents an instance where this R-M has been acquired by a non-PMEN1 strain (*S. pneumoniae* 2070768, serotype 23F, and MLST 507). Recent work on streptococcal evolution also identified DpnIII and noted its presence in the reference PMEN1 genome, one *Streptococcus pseudopneumoniae* strain, one *Streptococcus mitis* strain, and one *Streptococcus oralis* strain (46, 47). Our phylogenetic analysis suggests that DpnIII is relatively rare across *S. mitis* strains but more common in *S. pseudopneumoniae* (3/3 genomes), where it is adjacent to DpnI-encoding genes that have accumulated stop codons that likely make DpnI nonfunctional. The most closely characterized endonuclease is the plasmid-encoded LlaKR2I RE from *Lactococcus lactis* (76% identity, 90% similarity over the entire protein). R.DpnIII, LlaKR2I, Sau3AI, and the mismatch repair protein MutH have related sequences, and the last three cleave the tetramer GATC.

Next, we sought to determine the distribution of DpnIII within known PMEN1 strains. We used a set of 216 PMEN1 genome sequences published by Croucher and colleagues (33). From the *in silico* search, we identified DpnIII in 176 genomes but not in 40 genomes. The R-M region in these 40 genomes was often in a sequencing gap or a region of low coverage. Thus, these 40 genomes were analyzed by PCR with primers within the flanking regions and the amplimer was sequenced by Sanger sequencing. This work established that the DpnIII R-M region is present in 216/216 PMEN1 genomes (data not shown). For two strains (8140 and 8143), corresponding to Spanish isolates from 2001 (referred to as "clade S" in reference 33), the amplimer was larger than the remaining sequences. Sanger sequencing revealed a 1,709-kb transposon at the 3′ end of the *r.dpnIII* sequence (Fig. 1iv). The same transposon is also present in another chromosomal location in PMEN1 strains (gene SPN23F21520 in model strain SPN23F), consistent with duplication within this genome or transfer from a related strain.

DpnIII cleaves DNA at the sequence 5′ GATC 3′ on double-stranded DNA (dsDNA). The *r.dpnIII* gene displays high sequence similarity to that for the well-characterized Sau3AI endonuclease (34% identity over 86% of its length), suggesting that DpnIII might be a Sau3AI isoschizomer. To test the site specificity of R.DpnIII, we generated a C-terminally histidine-tagged form of the protein expressed from the endogenous promoter (strain SV35-RE6his) and enriched for the recombinant protein with a nickel column. All fractions from the purification process were individually incubated with plasmid pUC19 to identify an elution fraction enriched in endonuclease activity (see Fig. S1 in the supplemental material). The purified fraction was mixed with PCR amplimers of pUC19 or of the spectinomycin resistance gene, and the DNA was submitted to pyrosequencing, as well as visualized on an agarose gel. Sequencing of pUC19 digested with the DpnIII fraction generated 1,352 reads that mapped to regions where a GATC site was present (out of 1,641 reads); of these, 1,026 (76%) ended at 5′ GATC 3′ (see Fig. S2A in the supplemental material). Similarly, sequencing of the spectinomycin gene amplimer digested with DpnIII generated 1,403 reads that mapped to regions where a GATC site was present (out of 1,703 reads); of these, 1,062 (76%) ended at 5′ GATC 3′ (see Fig. S2B). No pattern was observed regarding nucleotides surrounding the various GATC cleavage sites, and no other cleavage site was evident. Consistently, the size of the digestion products of 2,686-bp linearized pUC19 (expected, 9 bands of <100 bp, 3 between 100 and 200 bp, as well as 258, 341, 585, and 839 bp), and the 840-bp spectinomycin resistance gene amplimer (expected, 3 bands of <100, 154, and 602 bp) correspond to the sizes predicted for enzyme cleavage of GATC (Fig. 2A).

M.DpnIII is a C5 cytosine-specific DNA methylase that recognizes the 5′ GATC 3′ tetramer. Prokaryotic methylases that generate a C5-methylcytosine (5mC) are characterized by multiple conserved sequence motifs (48), and these motifs are present in M.DpnIII. To investigate whether M.DpnIII is a 5mC-specific

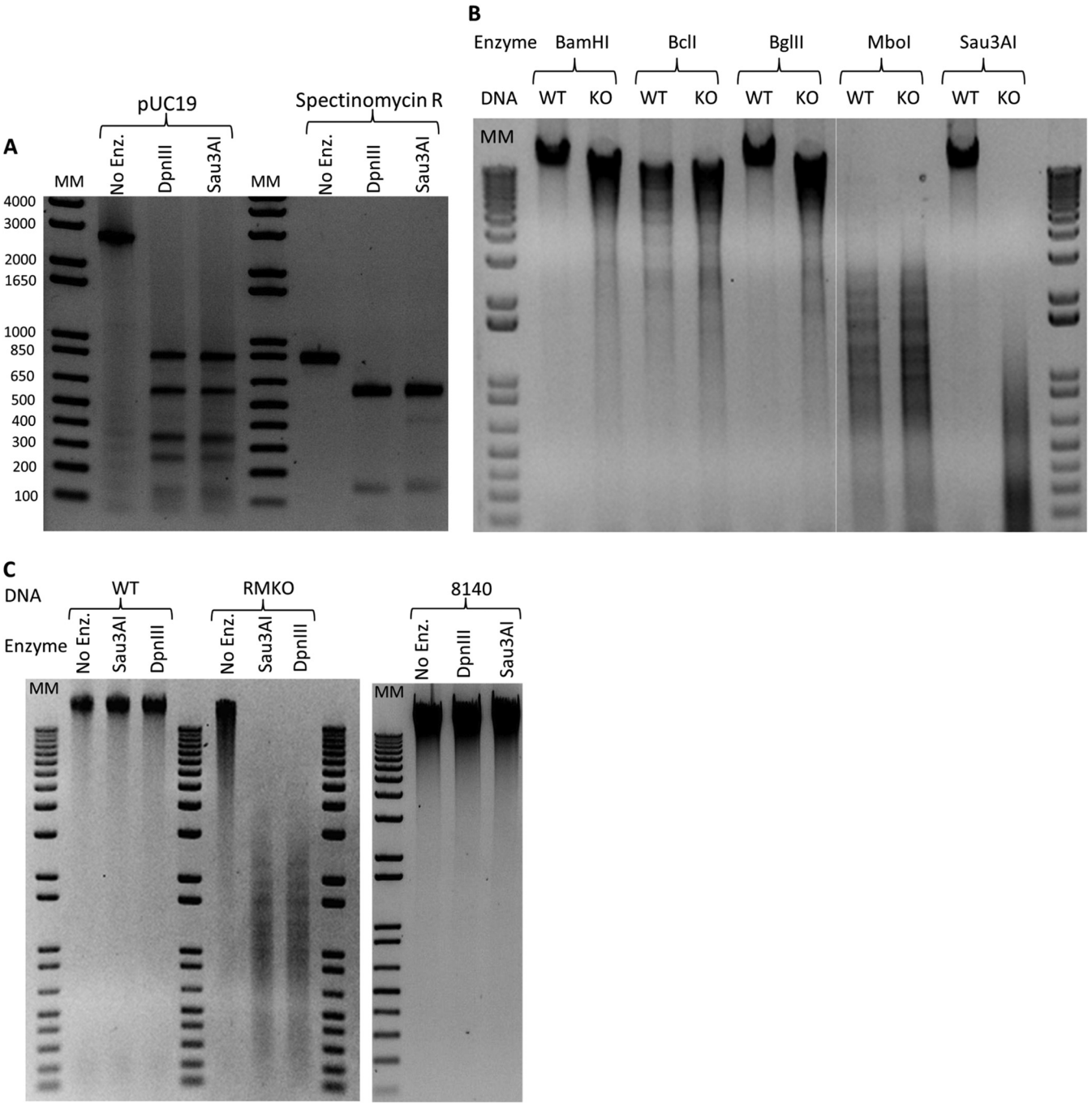

FIG 2 Characterization of (R-M system) DpnIII demonstrating that R.DpnIII cleaves DNA at 5′ GATC 3′ and M.DpnIII methylates DNA at the cytosine. (A) Digestion of pUC19 and spectinomycin R with a histidine-tagged DpnIII-enriched fraction and Sau3AI, showing bands consistent with digestion at GATC. (B) Genomic DNA isolated from the WT and RMKO strains combined with endonucleases that cleave at GATC but are inhibited by methylation at different positions (cleavage by BamHI, BglII, and Sau3AI is inhibited by methylation of the cytosine, and cleavage by BclI and MboI is inhibited by methylation of the adenine). (C) WT and RMKO DNA mixed with Sau3AI and histidine-tagged DpnIII, where only the RMKO is susceptible to digestion. Further, WT DNA of strain 8140 is protected by digestion with Sau3AI and DpnIII. Enz., enzyme; MM, mass markers. The values to the left of panel A are molecular masses in base pairs.

DNA methylase that modifies within the sequence GATC, we made use of a wild-type (WT) PMEN1 strain (SV35-T23), a cognate R-M deletion mutant (SV35-RMKO), and a set of endonucleases with well-characterized activities. The endonucleases were selected such that their cleavage site includes the tetramer 5′ GATC 3′ and they are inhibited by either N^6-methyladenine or 5mC. If M.DpnIII is a 5mC MT, the WT strain should be protected from cleavage by enzymes that recognize GATC and are inhibited by 5mC, but the R-M mutant should be sensitive.

DNA extracted from both the WT and RMKO strains was

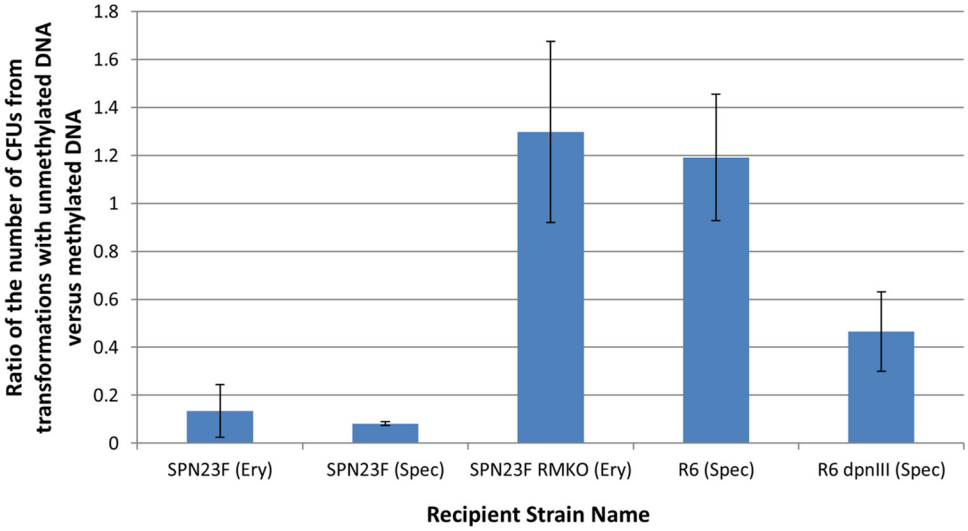

FIG 3 DpnIII protects cells from transformation with unmethylated DNA. WT strain SPN23F was transformed with either unmethylated or methylated PMEN1 DNA. For WT recipients, the number of transformants is lower when the donor DNA is unmethylated than when it is methylated. This was observed for selection with both Ery and spectinomycin. For the RMKO recipient, the number of transformants was independent of the methylation state of the donor. R6 (where DpnIII is absent) and R6-dpnIII was transformed with either unmethylated DNA (from Spain 6B) or methylated DNA (from SPN23F). In R6-dpnIII, but not R6, the number of transformants is lower when the donor DNA is unmethylated than when it is methylated. Transformations with unmethylated and methylated DNA were performed in parallel with the same recipient cells, and the numbers of transformants were compared. Error bars represent six experiments for the SPN23F WT recipients and three experiments for the remaining strains.

tested for sensitivity to the endonucleases selected. A difference in cleavage between the WT and RMKO strains was observed for all of the enzymes that recognize 5′ GATC 3′ and are inhibited by methylation at the cytosine (enzymes and recognition sites: Sau3AI, GATC; BglII, AGATCT; BamHI, GGATCC). In contrast, DNAs from both the WT and RMKO strains were cleaved by enzymes that cleave DNA at 5′ GATC 3′ and are inhibited by methylation of the adenine (enzymes and recognition sites: MboI, GATC; BclI, TGATCA) (Fig. 2B). This differential sensitivity of the WT and RMKO strains to Sau3AI but not MboI was also observed in a different PMEN1 isolate (strain SPN23F; data not shown). Next, we digested the WT and RMKO genomic DNAs with purified DpnIII. To obtain a higher concentration of the purified protein, we expressed histidine-tagged DpnIII from a maltose-inducible promoter on plasmid pLS1-ROM (see Fig. S3 in the supplemental material). When mixed with the RMKO genomic DNA, recombinant DpnIII exhibited the same digestion pattern as Sau3AI. In contrast, when it was mixed with WT genomic DNA, no cleavage was observed (Fig. 2C). These data demonstrate that M.DpnIII is a 5mC MT for the sequence 5′ GA TC 3′ that protects DNA from its cognate endonuclease. To determine whether M.DpnIII is active in the strains where a transposon is inserted downstream of the endonuclease, DNA from strain 8140 was digested with Sau3AI. This DNA is resistant to cleavage, demonstrating that M.DpnIII is also active in this strain (Fig. 2C).

To estimate the percentage of the GATC tetramer that is methylated in the genome during log-phase growth, we used Pacific Biosciences single-molecule real-time (SMRT) sequencing. When used to analyze genome sequence data, the SMRT software can detect differences in the rate of nucleotide incorporation that reflect base modifications such as methylation (49). The signature for 5mC is relatively weak and often hard to detect; thus, to enhance the kinetic signal, we treated the genome with the enzyme Tet1, which converts 5mC to 5-carboxylcytosine (5-caC) (50). In 5-caC modifications, the signature consists of three peaks (positions 6 and 2 before the modified C and the C itself) (see Fig. S4Ai in the supplemental material). The methylomes of both the native and Tet1-treated genomic DNAs were the same, and both display 5mC on the GATC tetramer. The signatures associated with 5mC were not present in the RMKO strain, suggesting that DpnIII is responsible for the 5mC modification. Variations in kinetic signatures that could reflect an additional modification were observed on the GATC tetramer in the WT strain but not in the RMKO strain (see Fig. S4Aii).

Association between R.DpnIII and genomic plasticity. Croucher and colleagues have provided a detailed recombination analysis of PMEN1 strains isolated after 1984 (33). They demonstrated that strain SPN23F (also referred to as ATCC 700669), isolated in the early 1980s during the first documented PMEN1 pandemic, resembles a recent common ancestor (33). The maximum-likelihood phylogeny constructed from SNPs suggests that two isolates (8140 and 8143) have accumulated mutations at a higher rate than other PMEN1 strains. These are the same two strains in which we identified a transposase inserted at the 3′ end of the endonuclease.

We investigated whether the presence of the R-M.DpnIII system could decrease the rate of transformation of an erythromycin (Ery) resistance gene (*ermB*) into the genomes of PMEN1 isolates. To this end, we made use of Ery resistance differences in two naturally occurring PMEN1 isolates that are identical at the DpnIII locus. A WT Ery-sensitive isolate (SPN23F) was mixed with Ery-resistant DNA isolated from either SV35-T23 (present *m.dpnIII* methylated) or SV35-RMKO (absent *m.dpnIII* unmethylated). There were significantly fewer transformant colonies when the donor DNA was unmethylated (Fig. 3). Relative DNA

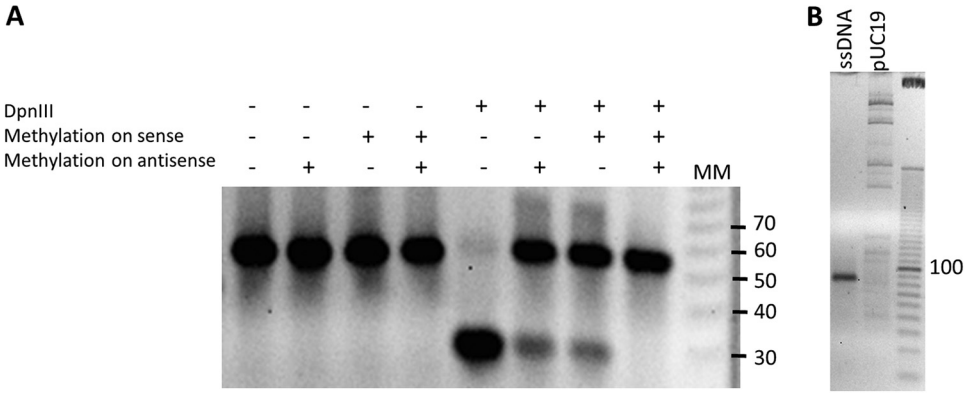

FIG 4 Test of purified DpnIII for cleavage of ssDNA and dsDNA in various methylation states. (A) dsDNA was generated from methylated and/or unmethylated oligonucleotides, DpnIII cleaved unmethylated and, with lower efficiency, hemimethylated DNA but not methylated dsDNA, (B) ssDNA was not cleaved by DpnIII, unlike the pUC19 control. MM, mass markers. The values to the right of panel A are molecular masses in base pairs.

transformation efficiencies were determined by calculating the ratio of the number of colonies from transformations with unmethylated DNA to the number of colonies from transformations with methylated DNA for the same recipient strain (the average number of colonies observed for transformations with methylated Ery-resistant DNA into the WT strain was 86, while the average for unmethylated DNA was 8.6).

To establish whether the difference in transformation efficiency was linked to the presence of DpnIII in the recipient, we transformed the same sets of DNA into a DpnIII deletion mutant of the Ery-sensitive strain (Spn23F-RMKO). In this control, the methylation state of the donor DNA was not significant, as the ratio of the numbers of colonies was essentially 1 (Fig. 3). The ratios are significantly different for the WT and RMKO strains (*t*-test *P* value = 0.002). To ensure that the effect was not specific to the Ery region, we generated a pair of strains with a spectinomycin resistance selection cassette in the SPN23F background. In one strain, the spectinomycin resistance cassette replaced the R-M system (SPN23F-RMKO), such that DNA from this strain is not methylated. In the other strain, the spectinomycin resistance cassette replaced the endonuclease only (SPN23F-REKO), such that DNA from this strain is methylated. DNA from each of these strains was mixed with a WT PMEN1 strain (SPN23F). Once again, the number of colonies was lower when the donor DNA was unmethylated (Fig. 3) (12-fold; the average number of colonies observed for transformations with methylated spectinomycin-resistant DNA into the WT strain was 12, while the average with unmethylated DNA was 1).

Finally, to establish whether the presence of DpnIII is sufficient to account for the difference in transformation, we added the R-M system to the highly transformable pneumococcal R6 laboratory strain, generating R6-dpnIII. Next, we transformed both spectinomycin-sensitive strains R6 and R6-dpnIII with DNA from spectinomycin-resistant strains that was either methylated (SPN23F-spec) or unmethylated (Spain 6B-spec).

The number of colonies of R6-dpnIII was lower when the donor DNA was unmethylated; while this difference was not observed in WT R6 (Fig. 3) (2-fold; the average number of colonies observed for transformations with methylated spectinomycin-resistant DNA into the R6-dpnIII strain was 38,666, while the average for unmethylated DNA was 18,666). Plasmid uptake was a low-frequency event in both the WT and RMKO Spn23F strains, such that the effect of DpnIII could not be assessed (data not shown).

Finally, to determine whether strain 8140, with a transposase inserted at the 3′ end of the endonuclease, was sensitive to the state of methylation of the donor DNA. We performed two transformations for each DNA (methylated SPN23F-REKO and unmethylated SPN23F-RMKO) on the 8140 recipient. The numbers of colonies were similar for both methylated and unmethylated DNAs (average ratio, 1.02). These data fit the prediction that the endonuclease is inactivated by the transposase insertion.

DpnIII cleaves unmethylated and hemimethylated dsDNA. The pneumococcal DNA import machinery degrades one strand of dsDNA during import, such that only ssDNA enters the cytosol. Imported ssDNA can undergo homologous recombination into the chromosome in a RecA-dependent process (51). When the imported material is unmethylated, the newly formed heteroduplex will be hemimethylated and the newly synthesized replication material will be fully unmethylated until it encounters the methylase (52). To determine possible stages at which DpnIII may exert its inhibitory effect on recombination, we tested the specificity of recombinant R.DpnIII to ssDNA and methylated, unmethylated, and hemimethylated dsDNA. dsDNA was generated by mixing complementary 54-bp oligonucleotides that were unmethylated or 5mC methylated on the sense strand or the antisense strand. Purified DpnIII did not cleave ssDNA, efficiently cleaved unmethylated dsDNA, had no effect on fully methylated dsDNA, and partially cleaved hemimethylated dsDNA (Fig. 4). This suggests that DpnIII affects the frequency of recombination primarily by acting after heteroduplex formation, probably postduplication (Fig. 5).

Previous work demonstrated that the *dpnA* methylase in the DpnII R-M systems acts on ssDNA and is upregulated during competence (53). We investigated whether the components of the DpnIII systems are upregulated during competence by quantitative reverse transcription (qRT)-PCR of cells before and after induction of competence. Competence was induced by the addition of competence-inducing peptide CSP2, and gene expression was measured 8 and 13 min postaddition. The genes *comC*, *comX*, *coiA*, and *comF* are known to be upregulated in a temporal fashion during competence and were used as controls. Unlike the expres-

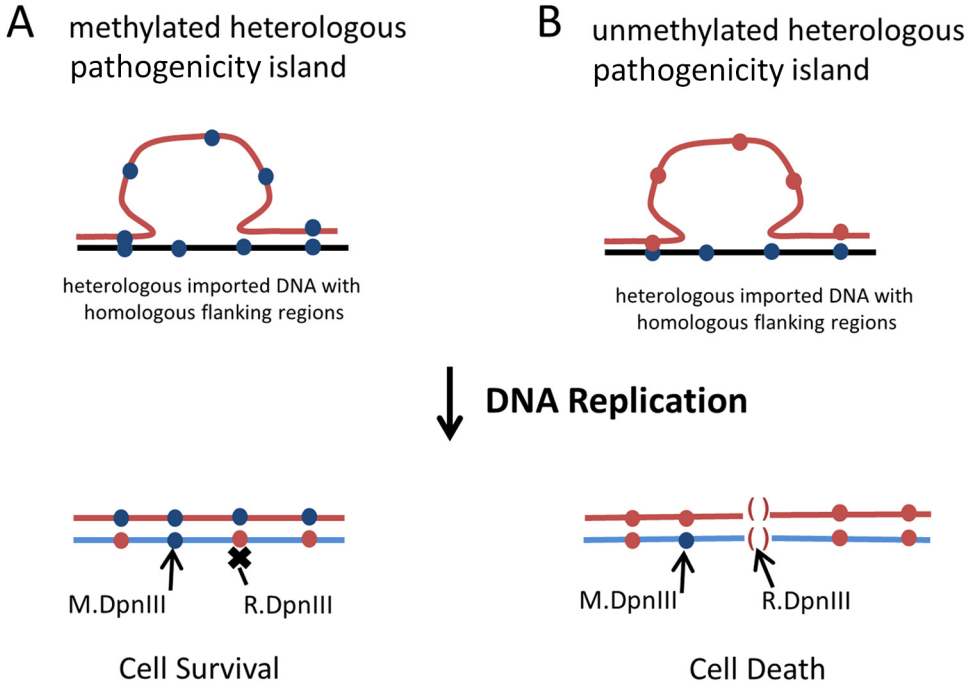

FIG 5 Influence of DpnIII on transformation. The images shown are based on illustrations by C. Johnston and colleagues to allow continuity in the interpretation of the effect of R-M systems on transformation efficiency (39). (A) DpnIII has no effect on the rate of transformation of methylated DNA. Methylated (blue circle) single-stranded transforming DNA (red line) enters the cell and pairs with methylated host DNA (black line with blue circles). Neosynthesized DNA (light blue line) that is still unmethylated (red circles) is generated during replication, producing hemimethylated dsDNA that is not highly sensitive to R.DpnIII (black cross), allowing DNA to replicate and transformants to survive. (B) DpnIII decreases rates of transformation of unmethylated DNA. Unmethylated (red circle) single-stranded transforming DNA (red line) enters the cell and pairs with methylated host DNA (black line with blue circles). Neosynthesized DNA (light blue line) is generated during replication, producing unmethylated dsDNA that is sensitive to R.DpnIII (black arrow pointing to red bracket) such that dsDNA is cleaved (red brackets) and transformants do not survive. Blue circles, DNA methylated at GATC with 5mC; red circles, unmethylated DNA; red line, transforming ssDNA; black line, host chromosome; light blue line, complementary neosynthesized DNA; red brackets, cleavage of dsDNA by R.DpnIII.

sion of the controls, that of the *r.dpnIII* and *m.dpnIII* genes was not upregulated, consistent with a role for these genes after DNA uptake (see Fig. S5 in the supplemental material).

DISCUSSION

Following induction of competence in the pneumococcus, foreign dsDNA enters the cell via a type IV pilus and is processed by the EndA endonuclease into ssDNA (45, 54). Once in the cell, the ssDNA is protected by SsbB (ssDNA binding protein) until RecA-mediated homologous recombination occurs (19). If the imported DNA is unmethylated, after strand displacement, it will form hemimethylated regions in areas of homology with the host DNA and will remain as unmethylated ssDNA in the nonhomologous regions (39). As illustrated in Fig. 5, replication is semiconservative, such that the strand where the imported DNA has been incorporated will serve as a template in a new cell. In this cell and before methylation by host enzymes, the region of the foreign DNA will first form unmethylated dsDNA that is highly susceptible to cleavage by an endonuclease such as DpnIII. In this manner, DpnIII limits the incorporation of unmethylated sequences into the genome. There is evidence of DpnIII-mediated effects in the PMEN1 population, given that the two PMEN1 strains where we observed a transposase inserted at the C-terminal tail of the DpnIII endonuclease are the same ones where Croucher and colleagues have observed increased accumulation of genomic changes (33).

Our *in vitro* experiments in PMEN1 show that the rate of transformation of unmethylated DNA is decreased ~90% in the presence of DpnIII. When DpnIII was expressed in the standard lab strain R6, the difference in transformation dropped to only 50%. Compared to clinical strains, the lab strain R6 has a much higher rate of transformability (often 100 to 1,000 times higher). The reason(s) for this disparity could reflect differences in multiple steps from DNA uptake to growth and survival of strains coding for recombinant DNA. The difference in the extent of restriction likely reflects a combination of the many factors that differ in the recombination steps between these two strains. Alternatively, there may be additional components in PMEN1 that are required for the 90% drop in transformability.

In contrast to DpnIII, the other well-characterized pneumococcal endonucleases do not limit recombination (39). For DpnI, this is because it is an unusual system in that it does not encode an MT and targets only methylated dsDNA. Given that homologous recombination is initiated with ssDNA, a fully methylated sequence sensitive to DpnI will not form on the chromosome and DpnI has no effect on homologous recombination (39). For DpnII, this is because this system codes for the *dpnA* methylase that targets ssDNA. DpnA methylates imported ssDNA and, in doing so, avoids the formation and cleavage of unmethylated dsDNA after replication (39). These enzymes provide protection against methylated and unmethylated phage invasions, respectively (52).

The majority of the pneumococcal genome sequencing projects have not identified plasmids, a fact that is consistent with plasmids being relatively rare in this species. The DpnI/II systems have been shown to limit phage attack but only mildly restrict plasmid entry (55). In contrast, SsbB has been implicated in the limitation of plasmid uptake (15). In our experiments, plasmid uptake was rare in both the WT and R-M deletion mutant strains and more efficient in another non-PMEN1 strain, Sp23BS72. It is likely that pneumococci display multiple mechanisms of plasmid control and that these are in play in PMEN1 strains.

The presence of different R-M systems and/or mobile elements at the same chromosomal location is consistent with a history of gene exchange and deletion. The DpnI, DpnII, and DpnIII systems are syntenic in pneumococci. Furthermore, these R-M systems are present at the same locus in related streptococcal species such as *S. pseudopneumoniae* and *S. mitis* (47). In multiple pneumococcal strains, including 8140 and 8143 described in this work, the R-M locus is disrupted by mobile elements (47). Mobile elements may entirely or partially replace genes in this locus. There may be evolutionary advantages to maintaining a functional methylase while inactivating the endonuclease. In the case of PMEN1 strains, inactivation of *r.dpnIII* would permit gene import and still allow genes from this strain to be taken up without restriction by PMEN1 strains with an active system. Inactivation of the methylase alone would lead to cell death. In addition to a role in plasticity, DpnII and DpnIII could also have a role in the regulation of pneumococcal gene expression. Recombination between the S subunit of the pneumococcal type I SpnD39III R-M system leads to variations in target specificities influencing both gene expression and virulence (56).

The DpnI, DpnII, and DpnIII endonucleases all target the tetramer GATC, yet they vary in specificity. DpnI targets regions methylated on the adenine, DpnII targets regions unmethylated on the adenine, and DpnIII targets regions unmethylated on the cytosine. In some *S. mitis* strains, yet another endonuclease (distinct from DpnI, DpnII, and DpnIII) is present at this locus with sequence similarity to the *Methanocaldococcus jannaschii* MjaIII enzyme that also cleaves at GATC. It seems likely that these streptococcal R-M systems affect the frequency of this tetramer in the pneumococcal genomes.

Pneumococci have evolved multiple mechanisms to increase genomic plasticity; thus, it seems counterintuitive that one of the most widespread lineages would have acquired a mechanism to limit transformation. The PMEN2 lineage also displays decreased plasticity because of inactivation of the competence system. However, unlike PMEN1, this lineage is no longer prevalent (26). This demonstrates a precedent for mutational events modulating plasticity but not necessarily an evolutionary advantage. Other lineages, such as highly stable serotype 3 clonal complex 180, may also have additional mechanisms decreasing genomic plasticity.

The adaptive forces and associated costs and benefits of genomic plasticity are complex, involving intra- and interspecies competition and cooperation, as well as interactions with the host. It is possible that there exist circumstances where there may be advantages to decreased import of novel sequences, while the competence-induced transcriptional changes remain functional as a stress response. PMEN1 is a well-adapted genome, with drug resistance, in some cases vaccine resistance, and very high rates of carriage in the human population. Our findings indicate that the PMEN1 R-M system contributes to its distinctive genome stability relative to strains outside this lineage.

MATERIALS AND METHODS

R-M nomenclature. The newly characterized type II R-M system was named DpnIII. This follows the standard R-M system nomenclature described by Roberts and colleagues (57). There are two previously characterized *S. pneumoniae* REs, DpnI and DpnII. Dpn refers to the pre-1974 name *Diplococcus pneumoniae*.

Bacterial strains. Two WT PMEN1 strains were used for experimental work, Ery-sensitive SPN23F (FM211187) and Ery-resistant SV35-T23 (ADNN); both of these strains have been previously sequenced (15, 41). SV35-T23 is resistant to Ery because of the insertion of a mobile element containing *ermB* (58). Strain SPN23F (also referred to as ATCC 700669) was obtained from Timothy Mitchell. It was isolated from the nasopharynx of a patient in 1984 in Spain and is a representative of the original penicillin-resistant clone linked to the Spanish PMEN1 epidemic of the 1980s. Strain SV35-T23 was recovered from the nasopharynx of a patient attending the AIDS clinic of St. Vincent's Medical Center in Richmond, NY, in 1996 (59). Both isolates code for identical DpnIII regions. In addition, we also used laboratory strain R6 (GenBank accession no. AE007317) and clinical isolate Spain6B (GenBank accession no. AUYK00000000).

Construction of deletion mutants. Deletion mutant strains were created by site-directed homologous recombination to replace a desired region with an antibiotic resistance cassette, as described previously (58, 60). The R-M system was deleted from strains SV35-T23 and SPN23F, creating strains SV35-RMKO and SPN23F-RMKO, respectively. To engineer these mutants, PCR was used to amplify approximately 2,000 bp upstream and downstream of the R-M system, creating the flanking regions. These flanking regions and a spectinomycin resistance cassette (amplified from pR412) were digested with the respective REs and ligated together, and the ligated product was amplified by PCR. This PCR product was transformed into *S. pneumoniae* strain spn23f or SV35-T23, and the clones were selected on Columbia agar plates containing spectinomycin (100 μg/ml) and confirmed by PCR. The same approach was used to create an SPN23F RE deletion mutant (SPN23F-REKO), but in this case, the flanking regions were selected to delete the endonuclease but not affect the methylase. All of the constructs and primers used are listed in Tables S2 and S3 in the supplemental material, respectively.

Construction of R6 containing *dpnIII*. The *dpnIII* R-M system was incorporated into the genome of strain R6 within the *bgaA* gene (61). This was accomplished by PCR amplifying the regions upstream and downstream of *bgaA* of strain SV35-T23, a kanamycin resistance cassette from the Janus cassette (62), and the *dpnIII* R-M system from strain SV35-T23. These primers were designed with NEBuilder software from New England Biolabs to work with their NEBuilder HiFi DNA assembly cloning mixture, which can assemble the four DNA pieces into a single piece *in vitro* by virtue of flanking homologous regions. This assembled DNA was transformed into R6 with CSP1 (sequence, EMRLSKFFRDFILQRKK; from GenScript) at 0.125 μg/ml, and transformants were selected with kanamycin at 100 μg/ml. Clones were confirmed by PCR.

Construction of strains encoding a His-tagged endonuclease. The strains used in this study are listed in Table S2 in the supplemental material. Strain SV35-RE6H codes for a tagged endonuclease within the chromosome at the endogenous locus. This strain was made by PCR amplifying the RE-encoding gene and adding nucleotides that code for six consecutive histidine residues before the stop codon to the reverse primer (see Table S3 in the supplemental material). A spectinomycin resistance cassette was incorporated after the RE-encoding gene for selection. Transformation and selection of the PCR fragment were done as previously described (58).

Strain SPN23FpRE6His codes for a tagged endonuclease that is expressed from an inducible promoter on plasmid pLS1-ROM (donated by Gloria de Solar [63]). The plasmid was amplified into a linear fragment by

PCR. The RE was amplified by PCR from strain SPN23F such that six histidines were incorporated into the reverse primer. The PCR products were digested with their respective REs and ligated together. This ligation was then transformed into SPN23F-REKO with CSP2 (sequence, EMRISRIILDFLFLRKK; purchased from GenScript) at 0.125 μg/ml and selected on Columbia agar plates containing Ery (1 μg/ml).

Transformation experiments. Ery-sensitive strains SPN23F and SPN23F-RMKO were transformed with DNA isolated from Ery-resistant strain SV35-T23 or SV35-RMKO. Five micrograms of DNA was used in each transformation with CSP2 at 0.125 μg/ml, and transformants were incubated for 4 h before being plated on Columbia agar plus Ery (1 μg/ml). For each experiment, transformations were performed in triplicate with plating in duplicate ($n = 6$ evaluations). In a second set of experiments, spectinomycin-sensitive strain SPN23F was transformed with DNA from spectinomycin-resistant strain SPN23F-REKO or SPN23F-RMKO as described above, but selection was done on Columbia agar plates with spectinomycin (100 μg/ml).

Spectinomycin-sensitive strains R6 and R6-dpnIII were transformed with DNA isolated from spectinomycin-resistant strains SPN23F-spec and Spain6B-spec with CSP1 at 0.125 μg/ml. Two micrograms of DNA was used in each transformation, and the transformants were incubated for 2 h before being plated on Columbia agar plus spectinomycin (100 μg/ml). For each experiment, transformations were performed in triplicate with plating in duplicate.

Protein expression and purification. Protein was purified from SPN23F-pRE6His as follows. A 20-ml culture of SPN23F-pRE6His was grown in AGCH medium (64) supplemented with 0.3% sucrose and Ery at 1 μg/ml to an optical density at 600 nm (OD_{600}) of 0.25. To induce the promoter, 5 ml of this culture was added to 475 ml of AGCH medium supplemented with 0.3% maltose and Ery at 1 μg/ml and grown to an OD_{600} of 0.25 (65). Cells were concentrated in native binding buffer (Invitrogen), lysed with lysozyme at 1 mg/ml for 30 min, and sonicated on ice for 10 s, six times (with 10-s intervals between sonications). The lysate was bound to and eluted from an Invitrogen ProBond nickel column in accordance with the manufacturer's instructions. Protein fractions were stored in 50% glycerol at −20°C, and protein content was visualized by SDS-PAGE and colloidal Coomassie staining.

Methylation determination by restriction digestion. The DNA methylation status of SV35-T23 and SV35-RMKO was determined by digesting 800 ng of genomic DNA with BamHI, BclI, BglII, MboI, or Sau3AI from New England Biolabs for 2 h according to the manufacturer's instructions, and digested DNAs were visualized on a 1% agarose gel.

RE activity. To establish the digestion site, purified protein from SV35-RE6His was mixed with 500 ng of pUC19 PCR product (2,686 bp) or 500 ng of spectinomycin PCR products amplified from pR412. Digestion was performed in NEB buffer 1 for 4 h at 37°C and visualized by gel electrophoresis or subjected to genome sequencing (described below). The ability of the endonuclease to digest dsDNA, ssDNA, and hemimethylated dsDNA was investigated by using oligonucleotides. For single-strand cutting, 100-bp oligonucleotides were combined with 5 μl of a purified protein fraction (total concentration, 0.06 mg/ml; purified from SPN23F-pRE6His lysate) in NEB buffer 1 for 4 h at 37°C. For double-strand cutting, 54-bp oligonucleotides were annealed together. Hemimethylated DNA was generated by combining complementary 54-bp oligonucleotides that either have or do not have 5-methylcytosine at a central GATC site. Methylated dsDNA was generated by combining methylated forms of the oligonucleotides, and unmethylated dsDNA was generated by combining the unmethylated oligonucleotides. Oligonucleotides were annealed together in a thermocycler by heating to 90°C and cooling to 25°C at 0.1°C/s. DNA was digested with 2 μl of protein (at 60 μg/ml) for 2 h under the conditions described above. All digests were visualized by gel electrophoresis.

Quantitation of RNA levels pre- and postinduction of competence. Cultures of SV35-T23 were grown in Columbia broth to an OD_{600} of 0.05. Cultures were induced with CSP2 at a concentration of 0.125 μg/ml. Samples (5 ml) were removed and placed into RNAprotect at three time points, pre-CSP, 8 min post-CSP, and 13 min post-CSP. RNA was extracted from samples with Qiagen RNeasy, and cDNA was synthesized with the Roche Transcriptor first-strand cDNA synthesis kit. Quantitative PCR was done with the Roche SYBR green I kit on the Roche LightCycler 480 platform. The data were analyzed with linregPCR, which uses arbitrary fluorescence units to represent the amount of RNA in each sample (65, 66). Data were normalized to the expression level of the 16S rRNA gene and represented as the average expression of test genes in replicates, where the error bars indicate the standard deviation of the expression values within replicates ($n = 3$).

DNA sequencing and analysis of RE digests. The digested spectinomycin resistance cassette and linearized pUC19 amplicon were sequenced by preparing 454 barcoded rapid libraries and following the FLX titanium workflow. The processed reads were aligned with the reference, curated with Sequencher (67), and organized into one image with PowerPoint. The percentage of reads cut at GATC was calculated by comparing the reads to their respective references with BLASTN (68)

Pacific Biosciences SMRT sequencing. Ten micrograms of genomic DNA was extracted from strains SPN23F and SPN23F-RMKO. Briefly, the DNA was cut into 1.5- and 6.0-kb fragments, end repaired, purified with AMPure PB beads, and ligated to SMRTbell hairpin adapters. DNA fragments without adapters were hydrolyzed with two exonucleases. The SMRTbell libraries were further purified with two consecutive AMPure bead purifications and quantified with a NanoDrop spectrophotometer and an Agilent 2100 Bioanalyzer with the Agilent DNA 7500 kit (product no. 5067-1506). Next, primer and polymerase binding steps were performed in accordance with the manufacturer's protocol. Finally, polymerase-bound libraries were loaded onto a PacBio RS and sequenced with four SMRT Cells.

For the analysis with 5mC converted to 5-caC, fragmented DNA was treated with the 5mC Tet1 Oxidation kit (WiseGene) before generation of the SMRTbell library.

Gene annotation and clustering. Genomes were submitted to RAST for CDS prediction and annotation (40). The CDSs were organized into gene clusters as previously described (69). Briefly, similar genes were identified by tfasty36 (FASTA v.3.6 package) for six-frame translation homology searches of all predicted proteins against all possible translations (70). The output was parsed such that genes with at least 70% identity over 70% of their length were grouped into gene clusters

***In silico* search of PMEN1 strains.** Genomes were assembled with an in-house assembly pipeline and Spades, respectively. The in-house pipeline consists of four main steps, (i) initial assembling with Velvet v1.0.12 (71) and VelvetOptimiser with subsequent removal of contigs of <300 bp, (ii) running 16 iterations to scaffold contigs with SSPACE v2.0 (72), (iii) filling of gaps within the scaffolds with GapFiller v1.11 (73), and (iv) mapping of reads against scaffolds with smalt v0.7.5 (http://www.sanger.ac.uk/resources/software/smalt/). As the sequencing coverage was generally low and nonuniform in the DpnIII region, Spades v2.4.0 (74) was used as a second assembly approach by using the "careful" and "rectangles" options. The presence of DpnIII in the assembled PMEN1 genomes was determined by utilizing an *in silico* PCR approach with forward and reverse primers (RESeq_F and RESeq_R) to detect the whole region, as well as screen for *r.dpnIII* and *m.dpnIII*. Manual inspection was carried out by aligning and ordering scaffolds against Spn23F with MUMmer v3.23 (75) and Abacas v1.3.2 (http://abacas.sourceforge.net/) and visualization in ACT v11.0.0 (76).

PCR and Sanger sequencing. PCR of 58 PMEN1 strains (see Table S4 in the supplemental material) was done with forward and reverse primers (RESeq_F and RESeq_R) located 100 to 200 bp outside the restriction endonuclease, producing a 1,750-bp amplicon on the reference genome. PCR was performed with NEB Q5 High-Fidelity polymerase. Products were visualized on a 1% Tris-acetate-EDTA agarose gel and purified with the Affymetrix ExoSAP-IT enzyme mixture for Sanger sequencing by Genewiz on the ABI 3730xl sequencer. To ensure complete coverage of the

amplimer, sequencing was performed with multiple primers (RESeq_F, RESeq_R, RESeq_WalkR, 8140_Walk2F, and 8140_Walk3R) (see Table S3).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.00173-15/-/DCSupplemental.

Figure S1, TIF file, 0.6 MB.
Figure S2, TIF file, 0.7 MB.
Figure S3, TIF file, 0.9 MB.
Figure S4, TIF file, 0.8 MB.
Figure S5, TIF file, 0.2 MB.
Table S1, PDF file, 0.1 MB.
Table S2, PDF file, 0.03 MB.
Table S3, PDF file, 0.04 MB.
Table S4, PDF file, 0.04 MB.

ACKNOWLEDGMENTS

We are obliged to Julian Parkhill and Stephen Bentley at the Sanger Institute for their support with the analysis of PMEN1 genomes and sharing of PMEN1 strains. We are also very grateful for the PMEN1 strains originally acquired from Anne von Gottberg, Mark van der Linden, Lotte Lambertsen, Sylvain Brisse, Lesley McGee, So Hyun Kim, and Jae Hoon Song; Alexander Tomasz and Herminia deLencastre for the SV35-T23 strain; as well as Timothy Mitchell for the SPN23F reference strain. We thank Nicholas Croucher, Dannie Durand, Aaron Mitchell, Adam Retchless, Anagha Kadam, and Xinyu Miao for their support and constructive observations throughout this project and manuscript preparation. We thank Josh Earl and Azad Ahmed for their help with genome sequencing and interpretation of genomic sequences, as well as Gloria de Solar for inducible plasmid pLS1-ROM and Don Morrison for plasmid pR412 and CSP2.

This work was supported by NIH grants R00-DC-011322 to N.L.H. and R01 AI080935 to G.D.E. and Wellcome Trust grant 098051 to J.D., as well as support from the Allegheny Health Network. Tyson A. Clark and Jonas Korlach are full-time employees at Pacific Biosciences, a company commercializing single-molecule, real-time sequencing technologies. The rest of us do not have any conflicts of interest to report.

REFERENCES

1. **Nunes S, Sá-Leão R, Carriço J, Alves CR, Mato R, Avô AB, Saldanha J, Almeida JS, Sanches IS, de Lencastre H.** 2005. Trends in drug resistance, serotypes, and molecular types of *Streptococcus pneumoniae* colonizing preschool-age children attending day care centers in Lisbon, Portugal: a summary of 4 years of annual surveillance. J Clin Microbiol **43:**1285–1293. http://dx.doi.org/10.1128/JCM.43.3.1285-1293.2005.
2. **O'Brien KL, Wolfson LJ, Watt JP, Henkle E, Deloria-Knoll M, McCall N, Lee E, Mulholland K, Levine OS, Cherian T.** 2009. Burden of disease caused by Streptococcus pneumoniae in children younger than 5 years: global estimates. Lancet **374:**893–902. http://dx.doi.org/10.1016/S0140-6736(09)61204-6.
3. **De Lencastre H, Tomasz A.** 2002. From ecological reservoir to disease: the nasopharynx, day-care centres and drug-resistant clones of Streptococcus pneumoniae. J Antimicrob Chemother **50**(Suppl **S2**):75–81. http://dx.doi.org/10.1093/jac/dkf511.
4. **Darkes MJ, Plosker GL.** 2002. Pneumococcal conjugate vaccine (Prevnar; PNCRM7): a review of its use in the prevention of Streptococcus pneumoniae infection. Paediatr Drugs **4:**609–630. http://dx.doi.org/10.2165/00128072-200204090-00005.
5. **Fitzwater SP, Chandran A, Santosham M, Johnson HL.** 2012. The worldwide impact of the seven-valent pneumococcal conjugate vaccine. Pediatr Infect Dis J **31:**501–508. http://dx.doi.org/10.1097/INF.0b013e31824de9f6.
6. **Harboe ZB, Dalby T, Weinberger DM, Benfield T, Mølbak K, Slotved HC, Suppli CH, Konradsen HB, Valentiner-Branth P.** 2014. Impact of 13-valent pneumococcal conjugate vaccination in invasive pneumococcal disease incidence and mortality. Clin Infect Dis **59:**1066–1073. http://dx.doi.org/10.1093/cid/ciu524.
7. **Frazão N, Brito-Avô A, Simas C, Saldanha J, Mato R, Nunes S, Sousa NG, Carriço JA, Almeida JS, Santos-Sanches I, de Lencastre H.** 2005. Effect of the seven-valent conjugate pneumococcal vaccine on carriage and drug resistance of Streptococcus pneumoniae in healthy children attending day-care centers in Lisbon. Pediatr Infect Dis J **24:**243–252. http://dx.doi.org/10.1097/01.inf.0000154326.77617.3e.
8. **Brueggemann AB, Pai R, Crook DW, Beall B.** 2007. Vaccine escape recombinants emerge after pneumococcal vaccination in the United States. PLoS Pathog **3:**e168. http://dx.doi.org/10.1371/journal.ppat.0030168.
9. **Donati C, Hiller NL, Tettelin H, Muzzi A, Croucher NJ, Angiuoli SV, Oggioni M, Dunning Hotopp JC, Hu FZ, Riley DR, Covacci A, Mitchell TJ, Bentley SD, Kilian M, Ehrlich GD, Rappuoli R, Moxon ER, Masignani V.** 2010. Structure and dynamics of the pan-genome of Streptococcus pneumoniae and closely related species. Genome Biol **11:**R107. http://dx.doi.org/10.1186/gb-2010-11-10-r107.
10. **Hiller NL, Janto B, Hogg JS, Boissy R, Yu S, Powell E, Keefe R, Ehrlich NE, Shen K, Hayes J, Barbadora K, Klimke W, Dernovoy D, Tatusova T, Parkhill J, Bentley SD, Post JC, Ehrlich GD, Hu FZ.** 2007. Comparative genomic analyses of seventeen *Streptococcus pneumoniae* strains: insights into the pneumococcal supragenome. J Bacteriol **189:**8186–8195. http://dx.doi.org/10.1128/JB.00690-07.
11. **Forbes ML, Horsey E, Hiller NL, Buchinsky FJ, Hayes JD, Compliment JM, Hillman T, Ezzo S, Shen K, Keefe R, Barbadora K, Post JC, Hu FZ, Ehrlich GD.** 2008. Strain-specific virulence phenotypes of Streptococcus pneumoniae assessed using the Chinchilla laniger model of otitis media. PLoS One **3:**e1969. http://dx.doi.org/10.1371/journal.pone.0001969.
12. **Croucher NJ, Finkelstein JA, Pelton SI, Mitchell PK, Lee GM, Parkhill J, Bentley SD, Hanage WP, Lipsitch M.** 2013. Population genomics of post-vaccine changes in pneumococcal epidemiology. Nat Genet **45:**656–663. http://dx.doi.org/10.1038/ng.2625.
13. **Feil EJ, Smith JM, Enright MC, Spratt BG.** 2000. Estimating recombinational parameters in Streptococcus pneumoniae from multilocus sequence typing data. Genetics **154:**1439–1450. .
14. **Griffith F.** 1928. The significance of pneumococcal types. J Hyg (Lond) **27:**113–159.
15. **Hiller NL, Ahmed A, Powell E, Martin DP, Eutsey R, Earl J, Janto B, Boissy RJ, Hogg J, Barbadora K, Sampath R, Lonergan S, Post JC, Hu FZ, Ehrlich GD.** 2010. Generation of genic diversity among Streptococcus pneumoniae strains via horizontal gene transfer during a chronic polyclonal pediatric infection. PLoS Pathog **6:**e1001108. http://dx.doi.org/10.1371/journal.ppat.1001108.
16. **Martin B, Granadel C, Campo N, Hénard V, Prudhomme M, Claverys J-P.** 2010. Expression and maintenance of ComD-ComE, the two-component signal-transduction system that controls competence of Streptococcus pneumoniae. Mol Microbiol **75:**1513–1528. http://dx.doi.org/10.1111/j.1365-2958.2010.07071.x.
17. **Courvalin P, Carlier C.** 1986. Transposable multiple antibiotic resistance in *Streptococcus pneumoniae*. Mol Gen Genet **205:**291–297.
18. **Hall-Stoodley L, Nistico L, Sambanthamoorthy K, Dice B, Nguyen D, Mershon WJ, Johnson C, Hu FZ, Stoodley P, Ehrlich GD, Post JC.** 2008. Characterization of biofilm matrix, degradation by DNase treatment and evidence of capsule downregulation in Streptococcus pneumoniae clinical isolates. BMC Microbiol **8:**173. http://dx.doi.org/10.1186/1471-2180-8-173.
19. **Attaiech L, Olivier A, Mortier-Barrière I, Soulet A-L, Granadel C, Martin B, Polard P, Claverys J-P.** 2011. Role of the single-stranded DNA-binding protein SsbB in pneumococcal transformation: maintenance of a reservoir for genetic plasticity. PLoS Genet **7:**e1002156. http://dx.doi.org/10.1371/journal.pgen.1002156.
20. **Coffey TJ, Dowson CG, Daniels M, Zhou J, Martin C, Spratt BG, Musser JM.** 1991. Horizontal transfer of multiple penicillin-binding protein genes, and capsular biosynthetic genes, in natural populations of Streptococcus pneumoniae. Mol Microbiol **5:**2255–2260. http://dx.doi.org/10.1111/j.1365-2958.1991.tb02155.x.
21. **Dowson CG, Hutchison A, Brannigan JA, George RC, Hansman D, Liñares J, Tomasz A, Smith JM, Spratt BG.** 1989. Horizontal transfer of penicillin-binding protein genes in penicillin-resistant clinical isolates of Streptococcus pneumoniae. Proc Natl Acad Sci U S A **86:**8842–8846. http://dx.doi.org/10.1073/pnas.86.22.8842.
22. **Engelmoer DJ, Rozen DE.** 2011. Competence increases survival during stress in Streptococcus pneumoniae. Evol Int J Org Evol **65:**3475–3485. http://dx.doi.org/10.1111/j.1558-5646.2011.01402.x.
23. **Prudhomme M, Attaiech L, Sanchez G, Martin B, Claverys J-P.** 2006. Antibiotic stress induces genetic transformability in the human pathogen

Streptococcus pneumoniae. Science **313:**89–92. http://dx.doi.org/10.1126/science.1127912.

24. **Wyres KL, Lambertsen LM, Croucher NJ, McGee L, von Gottberg A, Liñares J, Jacobs MR, Kristinsson KG, Beall BW, Klugman KP, Parkhill J, Hakenbeck R, Bentley SD, Brueggemann AB.** 2013. Pneumococcal capsular switching: a historical perspective. J Infect Dis **207:**439–449. http://dx.doi.org/10.1093/infdis/jis703.
25. **Croucher NJ, Harris SR, Barquist L, Parkhill J, Bentley SD.** 2012. A high-resolution view of genome-wide pneumococcal transformation. PLoS Pathog **8:**e1002745. http://dx.doi.org/10.1371/journal.ppat.1002745.
26. **Croucher NJ, Hanage WP, Harris SR, McGee L, van der Linden M, de Lencastre H, Sá-Leão R, Song J-H, Ko KS, Beall B, Klugman KP, Parkhill J, Tomasz A, Kristinsson KG, Bentley SD.** 2014. Variable recombination dynamics during the emergence, transmission and "disarming" of a multidrug-resistant pneumococcal clone. BMC Biol **12:**49. http://dx.doi.org/10.1186/1741-7007-12-49.
27. **Evans BA, Rozen DE.** 2013. Significant variation in transformation frequency in Streptococcus pneumoniae. ISME J **7:**791–799. http://dx.doi.org/10.1038/ismej.2012.170.
28. **Hanage WP, Fraser C, Tang J, Connor TR, Corander J.** 2009. Hyperrecombination, diversity, and antibiotic resistance in pneumococcus. Science **324:**1454–1457. http://dx.doi.org/10.1126/science.1171908.
29. **Hsieh Y-C, Wang J-T, Lee W-S, Hsueh P-R, Shao P-L, Chang L-Y, Lu C-Y, Lee C-Y, Huang F-Y, Huang L-M.** 2006. Serotype competence and penicillin resistance in Streptococcus pneumoniae. Emerg Infect Dis **12:**1709–1714. http://dx.doi.org/10.3201/eid1211.060414.
30. **Mostowy R, Croucher NJ, Hanage WP, Harris SR, Bentley S, Fraser C.** 2014. Heterogeneity in the frequency and characteristics of homologous recombination in pneumococcal evolution. PLoS Genet **10:**e1004300. http://dx.doi.org/10.1371/journal.pgen.1004300.
31. **McGee L, McDougal L, Zhou J, Spratt BG, Tenover FC, George R, Hakenbeck R, Hryniewicz W, Lefévre JC, Tomasz A, Klugman KP.** 2001. Nomenclature of major antimicrobial-resistant clones of *Streptococcus pneumoniae* defined by the pneumococcal molecular epidemiology network. J Clin Microbiol **39:**2565–2571. http://dx.doi.org/10.1128/JCM.39.7.2565-2571.2001.
32. **Wyres KL, Lambertsen LM, Croucher NJ, McGee L, von Gottberg A, Liñares J, Jacobs MR, Kristinsson KG, Beall BW, Klugman KP, Parkhill J, Hakenbeck R, Bentley SD, Brueggemann AB.** 2012. The multidrug-resistant PMEN1 pneumococcus is a paradigm for genetic success. Genome Biol **13:**R103. http://dx.doi.org/10.1186/gb-2012-13-11-r103.
33. **Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, McGee L, von Gottberg A, Song JH, Ko KS, Pichon B, Baker S, Parry CM, Lambertsen LM, Shahinas D, Pillai DR, Mitchell TJ, Dougan G, Tomasz A, Klugman KP, Parkhill J, Hanage WP, Bentley SD.** 2011. Rapid pneumococcal evolution in response to clinical interventions. Science **331:**430–434. http://dx.doi.org/10.1126/science.1198545.
34. **Pletz MW, McGee L, Jorgensen J, Beall B, Facklam RR, Whitney CG, Klugman KP.** 2004. Levofloxacin-resistant invasive *Streptococcus pneumoniae* in the United States: evidence for clonal spread and the impact of conjugate pneumococcal vaccine. Antimicrob Agents Chemother **48:**3491–3497. http://dx.doi.org/10.1128/AAC.48.9.3491-3497.2004.
35. **Reinert RR, Ringelstein A, van der Linden M, Cil MY, Al-Lahham A, Schmitz F-J.** 2005. Molecular epidemiology of macrolide-resistant *Streptococcus pneumoniae* isolates in Europe. J Clin Microbiol **43:**1294–1300. http://dx.doi.org/10.1128/JCM.43.3.1294-1300.2005.
36. **Jeltsch A.** 2003. Maintenance of species identity and controlling speciation of bacteria: a new function for restriction/modification systems? Gene **317:**13–16. http://dx.doi.org/10.1016/S0378-1119(03)00652-8.
37. **Humbert O, Salama NR.** 2008. The Helicobacter pylori HpyAXII restriction-modification system limits exogenous DNA uptake by targeting GTAC sites but shows asymmetric conservation of the DNA methyltransferase and restriction endonuclease components. Nucleic Acids Res **36:**6893–6906. http://dx.doi.org/10.1093/nar/gkn718.
38. **Budroni S, Siena E, Dunning Hotopp JC, 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. http://dx.doi.org/10.1073/pnas.1019751108.
39. **Johnston C, Polard P, Claverys J-P.** 2013. The DpnI/DpnII pneumococcal system, defense against foreign attack without compromising genetic exchange. Mob Genet Elements **3:**e25582. http://dx.doi.org/10.4161/mge.25582.
40. **Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R.** 2014. The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). Nucleic Acids Res **42:**D206–D214. http://dx.doi.org/10.1093/nar/gkt1226.
41. **Croucher NJ, Walker D, Romero P, Lennard N, Paterson GK, Bason NC, Mitchell AM, Quail MA, Andrew PW, Parkhill J, Bentley SD, Mitchell TJ.** 2009. Role of conjugative elements in the evolution of the multidrug-resistant pandemic clone streptococcus pneumoniaeSpain23F ST81. J Bacteriol **191:**1480–1489. http://dx.doi.org/10.1128/JB.01343-08.
42. **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. http://dx.doi.org/10.1093/nar/gku1046.
43. **Lacks SA, Mannarelli BM, Springhorn SS, Greenberg B.** 1986. Genetic basis of the complementary DpnI and DpnII restriction systems of S. pneumoniae: an intercellular cassette mechanism. Cell **46:**993–1000. http://dx.doi.org/10.1016/0092-8674(86)90698-7.
44. **de la Campa AG, Springhorn SS, Kale P, Lacks SA.** 1988. Proteins encoded by the DpnI restriction gene cassette. Hyperproduction and characterization of the DpnI endonuclease. J Biol Chem **263:**14696–14702.
45. **Lacks S, Greenberg B, Neuberger M.** 1975. Identification of a deoxyribonuclease implicated in genetic transformation of *Diplococcus pneumoniae*. J Bacteriol **123:**222–232.
46. **Croucher NJ, Coupland PG, Stevenson AE, Callendrello A, Bentley SD, Hanage WP.** 2014. Diversification of bacterial genome content through distinct mechanisms over different timescales. Nat Commun **5:**5471. http://dx.doi.org/10.1038/ncomms6471.
47. **Kilian M, Riley DR, Jensen A, Brüggemann H, Tettelin H.** 2014. Parallel evolution of *Streptococcus pneumoniae* and *Streptococcus mitis* to pathogenic and mutualistic lifestyles. mBio **5:**e01490–14. http://dx.doi.org/10.1128/mBio.01490-14.
48. **Pósfai J, Bhagwat AS, Pósfai G, Roberts RJ.** 1989. Predictive motifs derived from cytosine methyltransferases. Nucleic Acids Res **17:**2421–2435. http://dx.doi.org/10.1093/nar/17.7.2421.
49. **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. http://dx.doi.org/10.1093/nar/gks891.
50. **Clark TA, Lu X, Luong K, Dai Q, Boitano M, Turner SW, He C, Korlach J.** 2013. Enhanced 5-methylcytosine detection in single-molecule, real-time sequencing via Tet1 oxidation. BMC Biol **11:**4. http://dx.doi.org/10.1186/1741-7007-11-4.
51. **Mortier-Barrière I, de Saizieu A, Claverys JP, Martin B.** 1998. Competence-specific induction of recA is required for full recombination proficiency during transformation in Streptococcus pneumoniae. Mol Microbiol **27:**159–170. http://dx.doi.org/10.1046/j.1365-2958.1998.00668.x.
52. **Johnston C, Martin B, Fichant G, Polard P, Claverys J-P.** 2014. Bacterial transformation: distribution, shared mechanisms and divergent control. Nat Rev Microbiol **12:**181–196. http://dx.doi.org/10.1038/nrmicro3199.
53. **Lacks SA, Ayalew S, de la Campa AG, Greenberg B.** 2000. Regulation of competence for genetic transformation in Streptococcus pneumoniae: expression of dpnA, a late competence gene encoding a DNA methyltransferase of the DpnII restriction system. Mol Microbiol **35:**1089–1098. http://dx.doi.org/10.1046/j.1365-2958.2000.01777.x.
54. **Laurenceau R, Péhau-Arnaudet G, Baconnais S, Gault J, Malosse C, Dujeancourt A, Campo N, Chamot-Rooke J, Le Cam E, Claverys J-P, Fronzes R.** 2013. A type IV pilus mediates DNA binding during natural transformation in Streptococcus pneumoniae. PLoS Pathog **9:**e1003473. http://dx.doi.org/10.1371/journal.ppat.1003473.
55. **Lacks SA, Springhorn SS.** 1984. Transfer of recombinant plasmids containing the gene for DpnII DNA methylase into strains of *Streptococcus pneumoniae* that produce DpnI or DpnII restriction endonucleases. J Bacteriol **158:**905–909.
56. **Manso AS, Chai MH, Atack JM, Furi L, De Ste Croix M, Haigh R, Trappetti C, Ogunniyi AD, Shewell LK, Boitano M, Clark TA, Korlach J, Blades M, Mirkes E, Gorban AN, Paton JC, Jennings MP, Oggioni MR.** 2014. A random six-phase switch regulates pneumococcal virulence via global epigenetic changes. Nat Commun **5:**5055. http://dx.doi.org/10.1038/ncomms6055.
57. **Roberts RJ, Belfort M, Bestor T, Bhagwat AS, Bickle TA, Bitinaite J,**

Blumenthal RM, Degtyarev SKh, 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, Krüger 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 P-C, Stein DC, Stoddard BL, Szybalski W, Trautner TA, Etten JLV, Vitor JMB, Wilson GG, Xu S. 2003. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. Nucleic Acids Res **31:**1805–1812. http://dx.doi.org/10.1093/nar/gkg274.

58. **Hiller NL, Eutsey RA, Powell E, Earl JP, Janto B, Martin DP, Dawid S, Ahmed A, Longwell MJ, Dahlgren ME, Ezzo S, Tettelin H, Daugherty SC, Mitchell TJ, Hillman TA, Buchinsky FJ, Tomasz A, de Lencastre H, Sá-Leão R, Post JC, Hu FZ, Ehrlich GD.** 2011. Differences in genotype and virulence among four multidrug-resistant Streptococcus pneumoniae isolates belonging to the PMEN1 clone. PLOS One **6:**e28850. http://dx.doi.org/10.1371/journal.pone.0028850.
59. **Nesin M, Ramirez M, Tomasz A.** 1998. Capsular transformation of a multidrug-resistant *Streptococcus pneumoniae in vivo*. J Infect Dis **177:**707–713. http://dx.doi.org/10.1086/514242.
60. **Shankarappa B, Balachandran R, Gupta P, Ehrlich GD.** 1992. Introduction of multiple restriction enzyme sites by in vitro mutagenesis using the polymerase chain reaction. PCR Methods Appl **1:**277–278. http://dx.doi.org/10.1101/gr.1.4.277.
61. **Zähner D, Hakenbeck R.** 2000. The *Streptococcus pneumoniae* beta-galactosidase is a surface protein. J Bacteriol **182:**5919–5921. http://dx.doi.org/10.1128/JB.182.20.5919-5921.2000.
62. **Sung CK, Li H, Claverys JP, Morrison DA.** 2001. An *rpsL* cassette, Janus, for gene replacement through negative selection in *Streptococcus pneumoniae*. Appl Environ Microbiol **67:**5190–5196. http://dx.doi.org/10.1128/AEM.67.11.5190-5196.2001.
63. **Ruiz-Masó JA, López-Aguilar C, Nieto C, Sanz M, Burón P, Espinosa M, del Solar G.** 2012. Construction of a plasmid vector based on the pMV158 replicon for cloning and inducible gene expression in Streptococcus pneumoniae. Plasmid **67:**53–59. http://dx.doi.org/10.1016/j.plasmid.2011.09.001.
64. **Lacks S.** 1966. Integration efficiency and genetic recombination in pneumococcal transformation. Genetics **53:**207–235.
65. **Ruijter JM, Lorenz P, Tuomi JM, Hecker M, van den Hoff MJ.** 2014. Fluorescent-increase kinetics of different fluorescent reporters used for qPCR depend on monitoring chemistry, targeted sequence, type of DNA input and PCR efficiency. Mikrochim Acta **181:**1689–1696. http://dx.doi.org/10.1007/s00604-013-1155-8.
66. **Ramakers C, Ruijter JM, Deprez RH, Moorman AF.** 2003. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neurosci Lett **339:**62–66. http://dx.doi.org/10.1016/S0304-3940(02)01423-4.
67. **Gene Codes Corporation.** 2015. Sequencher version 5.3 sequence analysis software. Gene Codes Corporation, Ann Arbor, MI.
68. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J Mol Biol **215:**403–410. http://dx.doi.org/10.1016/S0022-2836(05)80360-2.
69. **Hogg JS, Hu FZ, Janto B, Boissy R, Hayes J, Keefe R, Post JC, Ehrlich GD.** 2007. Characterization and modeling of the Haemophilus influenzae core and supragenomes based on the complete genomic sequences of Rd and 12 clinical nontypeable strains. Genome Biol **8:**R103. http://dx.doi.org/10.1186/gb-2007-8-6-r103.
70. **Pearson WR, Lipman DJ.** 1988. Improved tools for biological sequence comparison. Proc Natl Acad Sci U S A **85:**2444–2448. http://dx.doi.org/10.1073/pnas.85.8.2444.
71. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res **18:**821–829. http://dx.doi.org/10.1101/gr.074492.107.
72. **Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W.** 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics **27:**578–579. http://dx.doi.org/10.1093/bioinformatics/btq683.
73. **Boetzer M, Pirovano W.** 2012. Toward almost closed genomes with GapFiller. Genome Biol **13:**R56. http://dx.doi.org/10.1186/gb-2012-13-6-r56.
74. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA.** 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol **19:**455–477. http://dx.doi.org/10.1089/cmb.2012.0021.
75. **Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL.** 2004. Versatile and open software for comparing large genomes. Genome Biol **5:**R12. http://dx.doi.org/10.1186/gb-2004-5-2-r12.
76. **Carver TJ, Rutherford KM, Berriman M, Rajandream M-A, Barrell BG, Parkhill J.** 2005. ACT: the Artemis comparison tool. Bioinformatics **21:**3422–3423. http://dx.doi.org/10.1093/bioinformatics/bti553.