

Barriers to Genetic Exchange between Bacterial Species: *Streptococcus pneumoniae* Transformation

JACEK MAJEWSKI,^{1*} PIOTR ZAWADZKI,^{2†} PAUL PICKERILL,^{3‡} FREDERICK M. COHAN,¹
AND CHRISTOPHER G. DOWSON³

Department of Biology, Wesleyan University, Middletown, Connecticut 06459,¹ and Department of Biological Sciences,
University of Sussex, Brighton BN1 9QG,² and Department of Biological Sciences, University of Warwick,
Coventry CV4 7AL,³ United Kingdom

Received 9 August 1999/Accepted 24 November 1999

Interspecies genetic exchange is an important evolutionary mechanism in bacteria. It allows rapid acquisition of novel functions by transmission of adaptive genes between related species. However, the frequency of homologous recombination between bacterial species decreases sharply with the extent of DNA sequence divergence between the donor and the recipient. In *Bacillus* and *Escherichia*, this sexual isolation has been shown to be an exponential function of sequence divergence. Here we demonstrate that sexual isolation in transformation between *Streptococcus pneumoniae* recipient strains and donor DNA from related strains and species follows the described exponential relationship. We show that the Hex mismatch repair system poses a significant barrier to recombination over the entire range of sequence divergence (0.6 to 27%) investigated. Although mismatch repair becomes partially saturated, it is responsible for 34% of the observed sexual isolation. This is greater than the role of mismatch repair in *Bacillus* but less than that in *Escherichia*. The remaining non-Hex-mediated barrier to recombination can be provided by a variety of mechanisms. We discuss the possible additional mechanisms of sexual isolation, in view of earlier findings from *Bacillus*, *Escherichia*, and *Streptococcus*.

Bacteria from all major taxa are able to exchange genes across species by homologous recombination (26). While the various bacteria take up donor DNA by a diversity of mechanisms, all studied systems of homologous recombination share at least one homologous feature: recombination depends ultimately on the activity of the RecA protein and its homologues (26). Similarly, there is one system that hinders recombination across species in both the *Proteobacteria* and the gram-positive bacteria (3). This is the mismatch repair system encoded by *mutS*, *mutL*, and their homologues. Because some of the molecular basis for interspecies recombination is shared across disparate taxa, we might expect that recombination between species is constrained in similar ways throughout the bacterial world.

In addition, previous studies have shown that the frequency of homologous recombination decreases with the sequence divergence between donor and recipient, in a manner that is similar across a wide range of organisms. In *Bacillus* transformation as well as in *Escherichia* conjugation, the frequency of recombination decreases exponentially with the degree of DNA sequence divergence between donor and recipient (23, 28, 31). A similar exponential relationship has been observed for the frequency of intrachromosomal crossovers in *Saccharomyces cerevisiae* (5). Nevertheless, the major mechanisms producing recombinational barriers have been shown to differ in each of the above cases. In *Escherichia coli*, the predominant

barrier to recombination is presented by the methylation-directed mismatch repair system (21). In *Bacillus*, mismatch repair is only marginally effective in preventing recombination between divergent sequences; the most significant barrier is that donor strand invasion and initiation of strand exchange require near identity between donor and recipient at both ends of the donor segment (15, 16). In yeast, the effects of mismatch repair and sequence divergence are both significant recombinational barriers (5).

Whereas an exponential relationship between interspecies recombination and sequence divergence appears to be universal, it is striking how much more mismatch repair contributes to recombination barriers in *Escherichia* than in *Bacillus*. It is not clear whether the role of mismatch repair in *Escherichia* or *Bacillus* is typical and under what circumstances mismatch repair is most likely to evolve to be a significant barrier to interspecies recombination. More comparative work is needed to understand the evolution of mismatch repair and its role as an interspecies recombinational barrier. In this paper, we approach this problem by investigating another system of microbial recombination: natural transformation in *Streptococcus pneumoniae*. We test whether the exponential relationship between sequence divergence and sexual isolation, observed in *Bacillus* and *Escherichia*, also holds for transformation in *Streptococcus*. We also investigate the mechanisms of sexual isolation between *Streptococcus* species.

S. pneumoniae is a gram-positive bacterium, naturally competent for transformation. The *Streptococcus* HexAB mismatch repair system is homologous to MutSL (3) and has been shown to correct single mismatches arising during recombination between otherwise identical substrates. The HexA protein recognizes mismatches in heteroduplex DNA, and, for transformation by mismatched DNA, it is able to remove the entire donor strand and thus prevent recombination (3). However, while the role of the Hex system in detecting single mismatches is well

* Corresponding author. Mailing address: Laboratory of Statistical Genetics, Box 192, Rockefeller University, 1230 York Ave., New York, NY 10021. Phone: (212) 327-7922. Fax: (212) 327-7996. E-mail: majewski@complex.rockefeller.edu.

† Present address: Procter & Gamble, 65824 Schwalbach/Ts, Germany.

‡ Present address: Zeneca Agrochemicals, Jeallots Hill Research Station, Bracknell RG42 6ET, United Kingdom.

TABLE 1. List of strains

Species	Strain	Source	Relevant genotype
<i>S. pneumoniae</i>	Pn16	C. G. Dowson	wt ^a
<i>S. pneumoniae</i>	R6	C. G. Dowson	wt
<i>S. pneumoniae</i>	R6Δ3	C. G. Dowson	Δ <i>hexA</i>
<i>S. pneumoniae</i>	96129	C. G. Dowson	wt
<i>S. mitis</i>	NCTC 10712	NCTC	Type strain
<i>S. oralis</i>	NCTC 11427	NCTC	Type strain
<i>S. sanguis</i>	NCTC 7863	NCTC	Type strain
<i>S. parasanguis</i>	ATCC 15912	ATCC	Type strain
<i>S. crista</i>	NCTC 12479	NCTC	Type strain
<i>S. anginosus</i>	ATCC 9895	ATCC	Type strain
<i>S. constellatus</i>	NCTC 11063	NCTC	Type strain
<i>S. intermedius</i>	ATCC 27335	ATCC	Type strain
<i>S. adjacens</i>	ATCC 49175	ATCC	Type strain

^a wt, wild type.

documented, its effectiveness in preventing recombination between divergent species is not yet fully known. There is evidence that the system is easily saturated in the presence of multiply mismatched DNA and may not be able to provide an efficient recombinational barrier (13).

In this study, we investigated the extent to which DNA sequence divergence reduces the rates of recombination between *Streptococcus* species, as well as the role that the HexAB mismatch repair system plays in producing sexual isolation (i.e., resistance to recombination across species). We transformed three recipient strains, two wild-type strains and one mismatch repair-deficient derivative, with chromosomal DNA isolated from related *Streptococcus* strains marked with rifampin resistance. Consequently, we determined the relationship between transformation frequencies and DNA sequence divergence, in the presence and absence of mismatch repair.

MATERIALS AND METHODS

Strains. Strains used in this study are listed in Table 1. We used *S. pneumoniae* strains Pn16 (highly competent wild-type isolate) and R6 (unencapsulated laboratory strain) as recipients to determine the relationship between sequence divergence and sexual isolation in wild-type strains. We used strain R6Δ3, a mismatch repair mutant of strain R6 with a deletion in the *hexA* gene, as a recipient in investigating the role of mismatch repair in sexual isolation. The remaining strains, used as donors, were obtained from the National Collection of Type Cultures (NCTC; Colindale, United Kingdom) and the American Type Culture Collection (ATCC; Manassas, Va.).

Isolation of rifampin-resistant mutants. Strains used as DNA donors were streaked onto plates from glycerol stocks and incubated overnight. A single colony was picked and spread onto a 2- to 3-cm square on a fresh plate. Following a further 24 h of growth, the square was harvested onto a swab and spread over a fresh plate containing 10 μg of rifampin/ml. Resistant colonies appeared within 2 to 3 days.

Isolation of chromosomal DNA. Bacteria were harvested from eight 90-mm-diameter brain heart infusion (BHI)-blood (5%)-agar plates using a plastic loop and resuspended in 1 ml of 50 mM Tris–10 mM EDTA, pH 8.0. After addition of 100 μl of 10-mg/ml lysozyme and 5 μl of RNase (Boehringer), the cells were incubated at 37°C for 30 to 60 min, followed by addition of 150 μl of 100-mg/ml proteinase K and a further 30 min of incubation. Cells were lysed by addition of 100 μl of 20% (wt/vol) sarcosyl and incubated at 37°C until lysis was complete, indicated by the clearing of the cell suspension, subject to a maximum time of 2 h. The resulting lysate was subjected to two extractions with phenol-chloroform and a further extraction with chloroform. The DNA was then precipitated by the addition of 1/10 volume of 3 M sodium acetate, pH 3.5, and 1 volume of isopropanol. The DNA was collected by centrifugation in a microcentrifuge at 13,000 × g for 15 min, and the resulting DNA pellet was washed with 70% ethanol, air dried briefly, and resuspended in 200 μl of deionized water. The quality and quantity of the DNA were determined by spectrophotometry and confirmed by visualization on an agarose gel. All samples had ratio of optical density at 260 nm to that at 280 nm of between 1.8 and 2.0.

Preparation of competent cells and transformation. A single colony was inoculated into 3 to 4 ml of C medium (11a) and grown at 37°C until the solution became slightly cloudy (approximately 3 h). A 100-μl aliquot was then used to

inoculate 7 ml of prewarmed C medium, and the mixture was incubated for 1 3/4 h. At 10-min intervals, 425 μl was removed and added to 75 μl of sterile glycerol; the solution was mixed, and a 20-μl sample was removed to a fresh tube. Both samples were frozen on dry ice. After 2 h, sampling was stopped.

Cells sampled at each time point were tested for competence as follows. Each 20-μl sample was thawed and added to 480 μl of C medium, and the mixture was split into two 250-μl aliquots. One microgram of chromosomal DNA from the rifampin-resistant isogenic strain was added to one sample. Cells were incubated for 2 h at 37°C, and 50 μl was plated onto BHI-agar supplemented with 10 μg of rifampin/ml. Competence was induced over a 40- to 60-min period, shown by an increasing number of transformants during the competent phase. The time of onset of competence depended upon the strain and the optical density of the original inoculum. In subsequent experiments, the competent cells were partially thawed on ice and 80 μl was added to 1,920 μl of C medium. The remaining cells were refrozen on dry ice and stored at –80°C. The diluted competent cells were split into 150-μl aliquots, and 0.6 μg of donor DNA was added (for the negative control, no DNA was added). Following incubation at 37°C for 2 h, cells were serially diluted in BHI broth and plated on selective medium. The sexual isolation between a recipient and a test donor was quantified as the frequency of homogamic transformation (the frequency at which the recipient was transformed by its own Rif^r mutant) divided by the frequency of heterogamic transformation (the frequency at which the recipient was transformed by the test donor).

Estimate of sequence divergence at *rpoB*. We used PCR to amplify two fragments of the *rpoB* genes from all the strains used in the experiments. The first fragment, extending from base 372 to 1160 (corresponding to base numbers of *Streptococcus pyogenes* ATCC 700294 *rpoB* [http://www.genome.ou.edu]), was amplified using degenerate primers 5'-GGG ACG TTC GTN ATH AAY GG-3', for the leading strand, and 5'-CCA AGG TGG TCD ATR TCR TC-3', for the lagging strand. The second fragment, base 1440 to 2380, was amplified using primers 5'-TTG TCA CAR TTY ATG GAY CA-3', for the leading strand, and 5'-TCG CGA GTG ATY TCY TCM GG-3', for the lagging strand. All primers were designed using the *rpoB* alignment of *Bacillus subtilis* (2) and *S. pneumoniae* (C. G. Dowson, unpublished data). PCRs were carried out using *Taq* DNA polymerase (Qiagen) and consisted of 95°C denaturation for 30 s and 57°C annealing for 30 s, followed by extension at 72°C for 1 min, for 25 cycles. The PCR products were purified using Qiaquick PCR purification columns (Qiagen) and adjusted to standard concentrations. They were sequenced using the external primers (described above) at the DNA Sequencing Facility at the University of Pennsylvania Medical Center, by the dRhodamine dye terminator method with ABI PRISM 373XL and 377 sequencers (Perkin-Elmer, Applied Biosystems Division). The result was a full double-stranded sequence of 510 bp for the first fragment, and 872 bp for the second fragment, for all tested strains. The combined sequence divergence of the two fragments was used as the estimate of DNA sequence divergence in the *rpoB* region. We have earlier found that, in *Bacillus*, the above fragments give a good estimate of interspecies divergence across the entire gene (15). In addition, *rpoB* is located within the highly conserved ribosomal gene cluster. In *Streptococcus*, *rpoB* is directly flanked by *rpoC* and is in close proximity to *rpoA* (the other subunits of RNA polymerase), along with several ribosomal protein genes (data from the ongoing *S. pyogenes* sequencing project is available at http://www.genome.ou.edu). Since all the above genes are highly conserved, the sequence divergence at *rpoB* is representative of divergence within the entire region throughout which the recombinant molecules are expected to extend.

Nucleotide sequence accession numbers. The sequences obtained in this study have been assigned GenBank accession no. AF194507 to AF194528.

RESULTS

Sequence analysis of donor and recipient strains. The sequence divergence values and phylogenetic relationships among the strains are shown in Table 2 and Fig. 1. We used maximum parsimony along with bootstrap analysis based on 1,000 replicates to determine the phylogenetic relationships among all strains, based on the partial *rpoB* sequence data. The phylogeny agrees with that determined earlier using 16S rRNA and *sodA1* data (14, 20), with the exception of the clustering within the *Streptococcus anginosus* group. The analysis of *rpoB* shows *Streptococcus constellatus* to be most closely related to *S. anginosus*, whereas according to earlier data (16S rRNA and *sodA1*) *S. anginosus* forms a clade with *Streptococcus intermedius*. This discrepancy is supported by high bootstrap confidence values and might indicate a recombination event in the vicinity of *rpoB*.

We were not able to detect any nucleotide differences at *rpoB* among the three recipient strains (Pn16, R6, and R6Δ3). Among the rifampin-resistant donors, we noted His526 substitutions, also known to confer resistance in *Bacillus*, *Mycobacterium*, and *Escherichia* (8, 30). We identified Asp526 (*S. pneumoniae* Pn16, *S. intermedius*) and Tyr526 (*Streptococcus oralis*, *S. pneumoniae* strain R6, *S. pneumoniae* strain 96129) substitutions.

Relationship between sexual isolation and DNA sequence divergence. The transformation frequencies for each donor and recipient, along with the corresponding sexual isolation values, are given in Tables 3 and 4. The relationship between sexual isolation and sequence divergence between the donor and the recipient is shown in Fig. 2. We find that the fit of the log-transformed sexual isolation data ($R^2 = 0.94$ for the wild type, and $R^2 = 0.93$ for the Hex⁻ recipient) is significantly better than a corresponding fit of raw, non-log-transformed data ($R^2 = 0.35$ for the wild type; $R^2 = 0.31$ for the Hex⁻ recipient). That is, the relationship between sexual isolation (ρ) and sequence divergence (π) in *Streptococcus* is close to exponential and can be described by the formula

$$\rho = 10^{\pi\phi} \quad (1)$$

Values of the sensitivity parameter, ϕ , represented by the regression slopes in Fig. 1, are 17.82 and 18.39 for the Pn16 and R6 strains, respectively. Analysis of covariance shows that these two values are not significantly different ($F = 0.6$; $P = 0.45$). This suggests that the sensitivity of sexual isolation to sequence divergence is constant across wild-type *S. pneumoniae* strains, despite the fact that strain Pn16 exhibits a baseline (homogamic) transformation rate eightfold higher than that of strain R6 (Table 3).

The effect of mismatch repair on heterospecific transformation. The mismatch repair-deficient mutant exhibits sexual isolation consistently lower than that of its isogenic wild-type strain: ϕ for the wild type (ϕ_{wt}) = 18.39; ϕ for the Hex⁻ mutant (ϕ_{Hex^-}) = 11.70. Analysis of covariance shows these coefficients to be statistically significant ($F = 53.8$; $P < 0.0001$). Averaging over all donor strains, mismatch repair is responsible for 34% of the observed sexual isolation (i.e., $[\phi_{wt} - \phi_{Hex^-}]/\phi_{wt} \cdot 100\%$). For every donor, with the exception of *S. oralis*, sexual isolation values are higher for the wild-type recipient than for the mismatch repair mutant. The influence of mismatch repair on integration frequencies varies across strains, increasing the sexual isolation by up to 36-fold, for *Streptococcus sanguis*. Transformation with *Streptococcus adjacens* (*Abiotrophia adjacens*) DNA, which is 27.4% divergent from that of the recipient at *rpoB*, produced no detectable transformants in the presence of mismatch repair, while the

TABLE 2. Sequence identity matrix for *Streptococcus* strains based on a partial *rpoB* sequence

Species or strain	% DNA sequence identity or divergence ^a compared to:												
	<i>S. pneumoniae</i> strain Pn16	<i>S. pneumoniae</i> strain 91629	<i>S. mitis</i>	<i>S. oralis</i>	<i>S. parasanguis</i>	<i>S. crista</i>	<i>S. sanguis</i>	<i>S. anginosus</i>	<i>S. constellatus</i>	<i>S. intermedius</i>	<i>S. adjacens</i>		
<i>S. pneumoniae</i> strain Pn16	0.7	99.3	94.9	93.3	86.4	86.3	85.0	83.9	83.4	81.8	72.6		
<i>S. pneumoniae</i> strain 91629		5.4	94.6	93.3	86.1	86.1	84.7	83.6	83.0	81.7	72.4		
<i>S. mitis</i>		6.7	7.7	92.3	85.2	86.1	84.7	82.7	82.2	80.8	71.8		
<i>S. oralis</i>		13.9	14.8	14.8	85.2	85.6	85.3	83.3	81.9	81.5	72.4		
<i>S. parasanguis</i>		13.9	13.9	14.4	14.4	85.6	84.4	82.2	83.4	81.9	70.4		
<i>S. crista</i>		15.3	15.4	14.7	15.6	7.8	92.2	83.6	83.5	83.7	71.4		
<i>S. sanguis</i>		16.4	17.3	16.7	17.8	16.4	15.1	84.9	85.5	84.2	72.2		
<i>S. anginosus</i>		16.6	17.8	18.1	16.6	16.5	14.6	7.1	92.9	88.4	70.3		
<i>S. constellatus</i>		18.2	19.2	18.5	18.1	16.3	15.8	11.7	11.4	88.6	71.9		
<i>S. intermedius</i>		27.6	28.2	27.6	29.6	28.6	27.8	29.7	28.1	29.5	70.5		
<i>S. adjacens</i>													

^a Percentages above and to the right of the diagonal formed by the blank entries represent DNA sequence identity; those below and to the left represent sequence divergence.

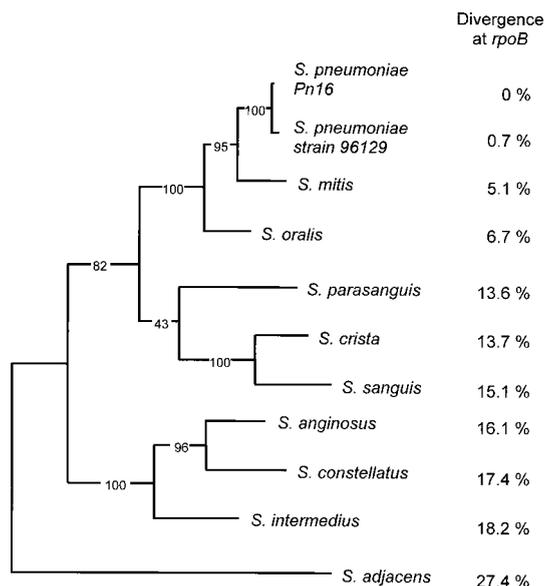


FIG. 1. Phylogeny of *S. pneumoniae* and related species, based on 1,380 bases of the *rpoB* sequence. The tree is the single most parsimonious phylogeny, created using the exhaustive search algorithm of PAUP, version 3.1.1 (D. L. Swofford, PAUP. Phylogenetic analysis using parsimony, version 4, Sinauer Associates, Sunderland, Mass., 1998). Branching confidence values are based on 1,000 bootstrap replicates. DNA sequence divergences between each donor strain and the *S. pneumoniae* recipients are shown on the right. The above phylogeny agrees with results obtained by analysis of 16S rRNA and *sodA* genes, with the notable exception that *S. anginosus* clusters here with *S. constellatus*, whereas earlier results show it to be more closely related to *S. intermedius*. The high bootstrap confidence value of this grouping suggests a possible recombination event at *rpoB*.

mismatch-deficient strain produced transformants at a sexual isolation of 3,800. Thus, mismatch repair plays a role in preventing recombination between divergent *Streptococcus* species, across the entire spectrum of sequence divergence tested (0.7 to 27.4%). However, the effectiveness of the system in detecting multiple mismatches is significantly reduced, compared to its effectiveness against single mismatches (see Discussion). Our results show this reduction to be most pronounced at intermediate sequence divergence levels (*S. oralis*). The lack of a detectable mismatch repair function in preventing integration of *S. oralis* DNA may represent the saturation

effect experienced by the Hex system in the presence of partially divergent DNA, as described by Humbert et al. (13).

DISCUSSION

The relationship between sequence divergence and sexual isolation. Our study shows that *S. pneumoniae* experiences sexual isolation from related strains and species. Sexual isolation in *Streptococcus* increases approximately exponentially with DNA sequence divergence, as has previously been observed within *Bacillus* and *Escherichia*. It is notable that *Streptococcus* and *Bacillus* share not only the exponential form of the relationship but also the same level of sensitivity of sexual isolation to sequence divergence. The sensitivities of sexual isolation to sequence divergence parameters (ϕ) for transformation of wild-type *Streptococcus* and *Bacillus* strains are 18.1 (average for Pn16 and R6 recipients) and 21.37 (15), respectively. These figures are remarkably close. To further illustrate this trend, the sexual isolation (ρ) between *S. pneumoniae* and *S. sanguis* is 1,920 at 15.1% sequence divergence. Within *Bacillus*, the corresponding sexual isolation between *B. subtilis* and *Bacillus licheniformis* is 2,344 at 14.5% divergence (15). It is possible that maintenance of comparable levels of sexual isolation arises from similar evolutionary pressures on recombination frequencies within each genus.

Possible barriers to interspecies recombination. An interspecies recombination event may be broken down into the following steps: (i) donor DNA is taken up by the recipient cell, (ii) donor DNA escapes the recipient cell's restriction system, (iii) a donor-recipient DNA heteroduplex molecule is formed, (iv) the mismatched heteroduplex escapes the action of the mismatch repair system, and (v) the recombinant strand is replicated by the recipient's DNA replication machinery. We present a short discussion of the relevance of the above steps as mechanisms of sexual isolation in *Streptococcus*.

The initial DNA uptake step can be dismissed as a genetic barrier in our system. Although in some bacterial species, such as *Haemophilus influenzae* (26), DNA uptake has been shown to be sequence specific, in the known gram-positive species (*Streptococcus* and *Bacillus*) the rate of uptake is independent of the sequence and is hence not sensitive to the sequence divergence between the donor and the recipient (18, 26). Thus, in *Streptococcus*, DNA uptake is unlikely to be a source of sexual isolation.

The donor DNA must then escape the action of the recipient's restriction endonucleases. The recipient strains used in

TABLE 3. Transformation frequencies, log₁₀ transformed

Source of donor DNA	Transformation frequency ^a for recipient strain:		
	Pn16	R6	R6Δ3
<i>S. pneumoniae</i> Pn16	-2.59 ± 0.32	-3.26 ± 0.24	-2.70 ± 0.34
<i>S. pneumoniae</i> R6	-2.79 ± 0.50	-3.33 ± 0.27	-2.81 ± 0.24
<i>S. pneumoniae</i> 91629	-2.64 ± 0.42	-3.65 ± 0.22	-2.59 ± 0.22
<i>S. mitis</i>	-3.61 ± 0.23	-4.59 ± 0.22	-3.24 ± 0.11
<i>S. oralis</i>	-3.37 ± 0.28	-4.15 ± 0.07	-3.30 ± 0.28
<i>S. parasanguis</i>	-4.73 ± 0.50	-5.97 ± 0.21	-3.92 ± 0.31
<i>S. crista</i>	-4.72 ± 0.62	-5.89 ± 0.14	-4.17 ± 0.50
<i>S. sanguis</i>	-5.63 ± 0.54	-6.80 ± 0.26	-4.71 ± 0.52
<i>S. anginosus</i>	-5.04 ± 0.22	-6.00 ± 0.20	-4.14 ± 0.20
<i>S. constellatus</i>	-6.17 ± 0.29	-6.74 ± 0.35	-4.91 ± 0.34
<i>S. intermedius</i>	-5.49 ± 0.35	-6.26 ± 0.44	-4.50 ± 0.12
<i>S. adjacens</i>	No transformants	No transformants	-5.62 ± 0.12

^a Log₁₀ (frequency of transformation) ± standard error, based on four or more experimental trials.

TABLE 4. Sexual isolation (\log_{10} transformed) between the *S. pneumoniae* recipients and the donor strains

Source of donor DNA	Log_{10} (sexual isolation) \pm SE ^b for recipient strain:		
	Pn16	R6	R6 Δ 3
<i>S. pneumoniae</i> Pn16	0 ^a	-0.08 \pm 0.23	-0.11 \pm 0.16
<i>S. pneumoniae</i> R6	0.21 \pm 0.22	0	0
<i>S. pneumoniae</i> 91629	0.06 \pm 0.10	0.32 \pm 0.11	-0.04 \pm 0.28
<i>S. mitis</i>	1.00 \pm 0.19	0.95 \pm 0.17	0.48 \pm 0.26
<i>S. oralis</i>	1.16 \pm 0.28	0.73 \pm 0.23	0.88 \pm 0.13
<i>S. parasanguis</i>	2.13 \pm 0.23	2.64 \pm 0.44	1.11 \pm 0.23
<i>S. crista</i>	2.07 \pm 0.27	2.56 \pm 0.36	1.80 \pm 0.08
<i>S. sanguis</i>	3.04 \pm 0.42	3.47 \pm 0.24	1.91 \pm 0.28
<i>S. anginosus</i>	2.45 \pm 0.16	2.67 \pm 0.34	1.27 \pm 0.36
<i>S. constellatus</i>	3.58 \pm 0.22	3.41 \pm 0.27	2.10 \pm 0.42
<i>S. intermedius</i>	3.06 \pm 0.33	2.78 \pm 0.29	2.01 \pm 0.40
<i>S. adjacens</i>	No transformants	No transformants	3.39 \pm 0.15

^a The sexual isolation between a recipient and its own Rif^r mutant is defined to be zero.

^b Sexual isolation values and standard errors are based on at least four experimental trials. Note that the values are obtained by dividing the homogamic transformation frequency by the heterogamic frequency for each trial for each donor-recipient pair and then averaged. Since not all donor-recipient comparisons were carried out in the same experimental trial, sexual isolation values will not in general be equal to the results of simply dividing the average transformation frequencies from Table 3.

the experiments described in this paper do not harbor restriction/modification systems. Thus, in this particular case, neutralization of donor DNA by the recipient's restriction enzymes cannot control recombination rates. In general, the effect of restriction systems has been shown to introduce a constant reduction in recombination frequencies, independent of the divergence between the donor and the recipient (31). In *Ba-*

cillus, this effect is small (possibly because the single-stranded donor DNA taken up on transformation is not subject to digestion), amounting to at most a sixfold reduction in recombination frequencies for restricting strains compared to those for nonrestricting strains (31). Similarly, in conjugation between *Salmonella* and *Escherichia* cells, sexual isolation is not significantly influenced by the presence of type II restriction (17).

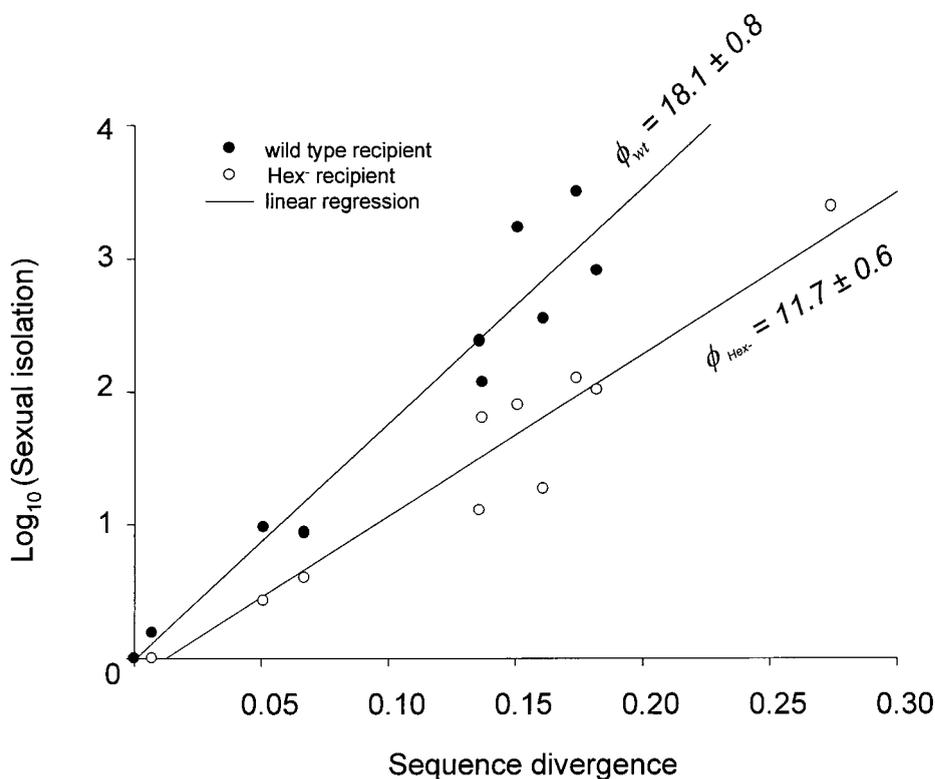


FIG. 2. Log_{10} (sexual isolation) versus sequence divergence between each donor and recipient. Sexual isolation values were calculated by dividing the frequency of homogamic recombination (using the donor's own Rif^r DNA) by the frequency of heterogamic recombination (using a divergent donor's DNA). The sexual isolation shown for the wild-type recipients is the average value for the two wild-type recipient strains, Pn16 and R6. The sexual isolation in the absence of mismatch repair is the result of transformation of the R6 Δ 3 *hexA* mutant. ϕ_{wt} and ϕ_{Hex^-} are the slopes of the regression of \log_{10} (sexual isolation) on sequence divergence for the wild type and the Hex⁻ mutant, respectively.

Therefore, in *Streptococcus*, restriction is not expected to be a source of the exponential relationship between sequence divergence and sexual isolation (18).

The two subsequent steps in the recombination process, namely, the formation of a heteroduplex molecule and the evasion of the recipient's mismatch repair system, present the most likely barrier to interspecies recombination. A wealth of information on the formation of recombinant DNA molecules is available from both in vivo and in vitro studies of *E. coli* and *S. cerevisiae*. In *E. coli* initial processing of the donor recombinant substrate is often necessary to produce free single-stranded ends. This is carried out by the RecBCD enzyme, which unwinds and digests double-stranded DNA until it encounters a characteristic χ sequence. The process results in the creation of free 3' single-stranded DNA ends. All further recombinogenic processes depend on the action of the RecA protein in binding to the single-stranded DNA and initiating invasion into a homologous and highly conserved (24) region of the recipient's double-stranded DNA molecule. (Although 5' donor ends can also be invasive, they cannot be extended and do not initiate recombination [10, 22].) The 3' recombinant joints are then subject to extension and editing (see reference 27 for a recent discussion of the above processes). In *E. coli*, (27) all modes of DNA editing, while involving several different components, appear to be controlled by the MutS protein (homologue of the *Streptococcus* HexA protein). Hence, the HexA mutant used in our experiments is expected to lack all mismatch repair activity.

The importance of RecA and HexAB functions in promoting and preventing recombination between divergent substrates is well documented (18, 26), and we discuss those events in detail, as pertinent to *Streptococcus* transformation, in separate sections below. In fact, we argue that RecA-dependent and HexAB-dependent steps are likely to produce most of the sexual isolation in *Streptococcus*.

On the other hand, since during transformation the donor DNA enters the cell as a single-stranded molecule, no processing by RecBCD is necessary. In fact, no known homologues of the RecD protein exist in gram-positive bacteria, while the function of the AddA and AddB enzymes in *Bacillus* (RecB and RecC homologues) seems to be limited to UV damage repair and might not be required during transformation (31). Thus, it is unlikely that DNA processing prior to recombination is involved in preventing interspecies recombination in *Streptococcus*.

In later stages, postsynaptic events such as branch migration and DNA replication may influence the frequency of successful recombination. Branch migration, which extends the initial heteroduplex molecule, is promoted by the RuvA, RuvB, and RuvC proteins (26). RuvA mutants of *E. coli* exhibit a reduced frequency of conjugal recombination. However, the reduction is relatively small and appears not to be very sensitive to DNA sequence divergence. Matic et al. (17) demonstrated that, in intraspecific crosses, the recombination frequency is reduced by a factor of 4, whereas in interspecific conjugation (between *E. coli* and *Salmonella enterica* serovar Typhimurium) it is reduced by a factor of about 10 (27). This corresponds to a sexual isolation factor due to the RuvA pathway of at most 2.5. In addition, Majewski and Cohan (16) demonstrated that, in *Bacillus*, by providing regions of flanking identity to the recipient, they were able to force a divergent donor insert to recombine at a nearly homogamic frequency. Similarly, the RecG protein, which is involved in the resolution of Holiday junctions, has comparable effects in intraspecific and interspecific crosses in *E. coli* (17). Thus, branch migration is at

most slightly impeded by mismatches and should not contribute considerably to sexual isolation between species.

Lastly, DNA synthesis may be initiated from the early heteroduplex intermediates, preventing dissociation of such potentially unstable joints. Stambuk and Radman (27) showed that mutations in *E. coli priA* and *recF* genes, which are known to be involved in recombination-associated DNA synthesis, have differential effects in interspecific and intraspecific crosses. It is possible that de novo DNA synthesis is necessary to circumvent the potential dissociation of early heteroduplex intermediates. It is possible that such DNA replication events may be impeded by mismatched heteroduplex joints. However, in *E. coli*, disrupting the RecF pathway reduces recombination frequencies by not more than sixfold in interspecies recombination and about twofold in intraspecific events. Hence the sexual isolation produced by mismatches impeding the RecF-mediated processes is at the most a factor of 3. In addition, Zawadzki et al. (31) found no measurable effect of RecF in preventing interspecific transformation in *Bacillus*. Thus, DNA synthesis induced by recombination is likely to be only a minor source of sexual isolation in *Streptococcus*.

Mismatch repair as recombination barrier. Our results show that the Hex mismatch repair system plays a significant role in preventing recombination between divergent DNA sequences. The extent of its significance is intermediate between that observed in *Escherichia*, where mismatch repair is the major recombinational barrier, and *Bacillus*, where it plays only a minor part. It has been suggested (15) that the relative ineffectiveness of the *Bacillus* mismatch repair system may be caused by degradation of mismatch repair proteins, which occurs under starvation conditions (9) necessary for induction of competence. In *Streptococcus*, competence is not induced under starvation, and thus mismatch repair proteins are probably not degraded (13). This may explain the more significant role of mismatch repair in preventing interspecific recombination in *Streptococcus* than in *Bacillus*.

However, the *Streptococcus* mismatch repair is still a rather poor mechanism of sexual isolation. The Hex system has been shown to efficiently correct specific single-base pair mismatches, known as low efficiency (LE) markers, resulting from transformation, with an accuracy of over 90% (4). It is believed that once a mismatch is detected, the entire donor strand, which may be as large as several kilobases, is rejected (3). Let us consider a simple model, assuming that, in a donor strand containing several (m) mismatches, detection of every mismatch is an independent event. Hence the probability of the entire strand escaping detection is equal to the probability of m mismatches being undetected. We further assume that, on average, one in every four mismatches is an LE mismatch (4, 11), escaping detection with a probability of 0.1, while others are high-efficiency (HE) mismatches, always escaping detection. The probability of a foreign strand escaping the mismatch repair system can then be expressed as $(0.1)^{m/4}$ or $(0.1)^{L\pi/4}$, where L is the length of the donor strand and π is the sequence divergence. Taking into account that the length of donor DNA integrated on transformations is often greater than 8 kb (1, 3), we should expect that the reduction in transformation frequency between 15%-divergent DNA strands, caused by mismatch repair, should be of the order of 10^{300} .

In our experiments, the Hex system reduces the recombination frequencies by not more than 36-fold (*S. sanguis*, 15% sequence divergence from recipient). It is obvious that the effectiveness of mismatch repair against multiply mismatched sequences is greatly reduced compared to its effectiveness against single mismatches. Hence, the simple model described above is incorrect, and detection events cannot be considered

independent. We present three explanations for the decrease in efficiency of the Hex system in the presence of multiple mismatches.

First, it is possible, that a saturation of mismatch repair may be reducing the effectiveness of the Hex system. Humbert et al. (13) have previously shown that the concentrations of HexA and HexB proteins become limiting during transformation with divergent DNA. They found that mismatch repair is ineffective in preventing transformation by DNA with intermediate sequence divergence levels (1 to 10%) from *S. pneumoniae* and suggested that saturation of the system may be responsible. Independent recombination events at loci other than the selected locus may reduce the number of mismatch repair proteins available to prevent integration of the selectable marker. This would further suggest that organisms closely related to *S. pneumoniae*, such as *S. oralis*, that are quite likely to possess a wide range of genes with intermediate sequence divergence are potentially important donors in the evolution of the pneumococcal genome (6, 25; S. King, unpublished data). This finding suggests a potential problem for the development of novel pneumococcal targets for vaccination or chemotherapy that may become subject to a strong selective pressure (7).

Second, it has also been shown that the presence of a particular HE marker (namely, a C/C mismatch) in the immediate proximity of an LE marker reduces the probability of detection of the LE marker (11). Thus, some mismatches may have a "masking" effect on adjacent mismatches, possibly by destabilizing the DNA helix conformation. In multiply mismatched strands, extended destabilized regions may increase the overall probability of an entire strand escaping detection.

Finally, multiple mismatches may interfere with the search for nicks in the donor strand. Such nicks allow the Hex system to distinguish between donor and recipient and correct the appropriate strand. The nick-directed search needs to extend from the mismatch all the way to the end of the donor strand. Mismatches present along the way may impede the search. On the other hand, in *Escherichia*, mismatch repair uses methylation patterns to distinguish between the donor and the recipient strands. The search for methylation does not need to extend to the end of the strand and takes, on average, 256 bases. This may explain why, in *E. coli*, mismatch repair is less hampered by multiple mismatches (13).

It has been suggested that the mismatch repair system may be modulated so as to provide a strong barrier to recombination under normal growth conditions and a weak barrier under stressful conditions, when recombination may be particularly favorable (12, 28). Our results show that even under exponential growth, which is necessary for induction of competence in *Streptococcus*, mismatch repair is a poor barrier to interspecific recombination. Moreover, Humbert et al. (13) have shown that levels of mismatch repair proteins are not modulated during competence. In view of the above and the data from transformation in *B. subtilis* (15, 16), we believe that the role of mismatch repair as a barrier to interspecific recombination in gram-positive bacteria is minor and most probably incidental to that involved in the correction of errors introduced during DNA replication.

Recombinant joint formation as a genetic barrier. According to data from recombination in *Escherichia* (27–29) and *Bacillus*, the reluctance to form mismatched DNA heteroduplex molecules is caused primarily by the scarcity of minimum efficiently processed segments (MEPS), short regions of near identity between the donor and the recipient (15, 16). Once such conserved blocks are located, the adjacent sequences seem to undergo recombination despite a high level of divergence between the donor and the recipient. In *Bacillus*, recom-

ination requires mismatch-free segments of about 20 bp or more at both ends of the donor strand (16). Data from *E. coli* indicate that only one highly conserved, invasive end is needed in that system (22, 24). Such sequences are most probably necessary for initiation of RecA-mediated strand invasion, stabilization of the recombinant joint, and possibly the subsequent recombination-associated DNA synthesis.

In the preceding discussion of possible mechanisms of sexual isolation, we suggested that, other than mismatch repair, most intermediate processes are unlikely to have a major sequence divergence-specific effect on sexual isolation. We further propose that the remaining major barrier to recombination between diverged sequences is the availability of MEPS sequences. We can apply the model described by Majewski and Cohan (15) to determine the length of a MEPS in *Streptococcus* from the experimentally determined sensitivity of sexual isolation to sequence divergence, ϕ_{Hex^-} , in the absence of mismatch repair. The model assumes that a mismatch-free region of n nucleotides is necessary for a successful recombination. It can be shown from equation 1 that the length of such a region may be estimated as $n = \phi_{\text{Hex}^-} \cdot \ln(10)$. Thus, in *Streptococcus*, the minimum length of an invasive end would be 27 bp, if one such end was needed, or 14 bp, if two flanking ends were necessary (as for *Bacillus*). More work, for example, analysis of the recombinant junction produced on transformation with divergent DNA, needs to be done to determine which of the above models is correct. We also observe that the resistance to joint formation poses less of a barrier to recombination in *Streptococcus* ($\phi_{\text{Hex}^-} = 11.72$) than in *Bacillus* ($\phi_{\text{MutSL}^-} = 17.95$) (15). This difference is significant by analysis of covariance ($F = 33.4$; $P < 0.0001$).

To conclude, we find that the frequency of transformation in *Streptococcus* is an approximately exponential function of the sequence divergence between the donor and the recipient. As in other recombinational systems, the two major factors producing sexual isolation are the mismatch repair system and the difficulty in heteroduplex formation.

ACKNOWLEDGMENTS

We thank J. P. Claverys for the supply of the *hexA* construct that was used to create the R6 *hexA* mutant strain.

This work was supported by U.S. Environmental Protection Agency grant R82-5348-01-2 and by the Wellcome Trust.

REFERENCES

1. Biswas, G. D., and A. W. Ravin. 1971. Heterospecific transformation of *Pneumococcus* and *Streptococcus*. IV. Variations in hybrid DNA produced by recombination. *Mol. Gen. Genet.* **110**:1–22.
2. Boor, K. J., M. L. Duncan, and C. W. Price. 1995. Genetic and transcriptional organization of the region encoding the beta subunit of *Bacillus subtilis* RNA polymerase. *J. Biol. Chem.* **270**:20329–20336.
3. Claverys, J. P., and S. A. Lacks. 1986. Heteroduplex deoxyribonucleic acid base mismatch repair in bacteria. *Microbiol. Rev.* **50**:133–165.
4. Claverys, J. P., V. Mejean, A. M. Gasc, and A. M. Sicard. 1983. Mismatch repair in *Streptococcus pneumoniae*: relationship between base mismatches and transformation efficiencies. *Proc Natl Acad Sci USA* **80**:5956–5960.
5. Datta, A., M. Hendrix, M. Lipsitch, and S. Jinks-Robertson. 1997. Dual roles for DNA sequence identity and the mismatch repair system in the regulation of mitotic crossing-over in yeast. *Proc. Natl. Acad. Sci. USA* **94**:9757–9762.
6. Dowson, C. G., T. J. Coffey, C. Kell, and R. A. Whitley. 1993. Evolution of penicillin resistance in *Streptococcus pneumoniae*; the role of *Streptococcus mitis* in the formation of a low affinity PBP2B in *S. pneumoniae*. *Mol. Microbiol.* **9**:635–643.
7. Dowson, C. G., V. Barcus, S. King, P. Pickerill, A. Whatmore, and M. Yeo. 1997. Horizontal gene transfer and the evolution of resistance and virulence determinants in *Streptococcus*. *Soc. Appl. Bacteriol. Symp. Ser.* **26**:42–51.
8. Enright, M., P. Zawadzki, P. Pickerill, and C. G. Dowson. 1998. Molecular evolution of rifampicin resistance in *Streptococcus pneumoniae*. *Microb. Drug Resist.* **4**:65–70.
9. Feng, G., H. C. Tsui, and M. E. Winkler. 1996. Depletion of the cellular amounts of the MutS and MutH methyl-directed mismatch repair proteins in

- stationary-phase *Escherichia coli* K-12 cells. *J. Bacteriol.* **178**:2388–2396.
10. **Friedman-Ohana, R., and A. Cohen.** 1998. Heteroduplex joint formation in *Escherichia coli* recombination is initiated by pairing of a 3'-ending strand. *Proc. Natl. Acad. Sci. USA* **95**:6909–6914.
 11. **Gasc, A. M., A. M. Sicard, and J. P. Claverys.** 1989. Repair of single- and multiple-substitution mismatches during recombination in *Streptococcus pneumoniae*. *Genetics* **121**:29–36.
 - 11a. **Gurney, T., Jr., and M. S. Fox.** 1968. Physical and genetic hybrids formed in bacterial transformation. *J. Mol. Biol.* **32**:83–100.
 12. **Harris, R. S., G. Feng, K. J. Ross, R. Sidhu, C. Thulin, S. Longerich, S. K. Szigety, M. E. Winkler, and S. M. Rosenberg.** 1997. Mismatch repair protein MutL becomes limiting during stationary-phase mutation. *Genes Dev.* **11**:2426–2437.
 13. **Humbert, O., M. Prudhomme, R. Hakenbeck, C. G. Dowson, and J. P. Claverys.** 1995. Homeologous recombination and mismatch repair during transformation in *Streptococcus pneumoniae*: saturation of the Hex mismatch repair system. *Proc. Natl. Acad. Sci. USA* **92**:9052–9056.
 14. **Kawamura, Y., X. G. Hou, F. Sultana, H. Miura, and T. Ezaki.** 1995. Determination of 16S rRNA sequences of *Streptococcus mitis* and *Streptococcus gordonii* and phylogenetic relationships among members of the genus *Streptococcus*. *Int. J. Syst. Bacteriol.* **45**:406–408.
 15. **Majewski, J., and F. M. Cohan.** 1998. The effect of mismatch repair and heteroduplex formation on sexual isolation in *Bacillus*. *Genetics* **148**:13–18.
 16. **Majewski, J., and F. M. Cohan.** 1999. DNA sequence similarity requirements for interspecific recombination in *Bacillus*. *Genetics* **153**:1525–1533.
 17. **Matic, I., C. Rayssiguier, and M. Radman.** 1995. Interspecies gene exchange in bacteria: the role of the SOS and mismatch repair systems in evolution of species. *Cell* **80**:507–515.
 18. **Mejean, V., and J. P. Claverys.** 1993. DNA processing during entry in transformation of *Streptococcus pneumoniae*. *J. Biol. Chem.* **268**:5594–5599.
 19. **Mortier-Barriere, I., O. Humbert, B. Martin, M. Prudhomme, and J.-P. Claverys.** 1997. Control of recombination rate during transformation of *Streptococcus pneumoniae*: an overview. *Microb. Drug Resist.* **3**:233–242.
 20. **Poyart, C., G. Quesne, S. Coulon, P. Berche, and P. Trieu-Cuot.** 1998. Identification of streptococci to species level by sequencing the gene encoding the manganese-dependent superoxide dismutase. *J. Clin. Microbiol.* **36**:41–47.
 21. **Rayssiguier, C., D. S. Thaler, and M. Radman.** 1989. The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch-repair mutants. *Nature* **342**:396–401.
 22. **Reddy, G., B. Burnett, and C. M. Radding.** 1995. Uptake and processing of duplex DNA by RecA nucleoprotein filaments: insights provided by a mixed population of dynamic and static intermediates. *Biochemistry* **34**:10194–10204.
 23. **Roberts, M. S., and F. M. Cohan.** 1993. The effect of DNA sequence divergence on sexual isolation in *Bacillus*. *Genetics* **134**:401–408.
 24. **Shen, P., and H. V. Huang.** 1986. Homologous recombination in *Escherichia coli*: dependence on substrate length and homology. *Genetics* **112**:441–457.
 25. **Sibold, C., J. Henrichsen, A. Konig, C. Martin, L. Chalkley, and R. Hakenbeck.** 1994. Mosaic *pbpX* genes of major clones of penicillin-resistant *Streptococcus pneumoniae* have evolved from *pbpX* genes of a penicillin-sensitive *Streptococcus oralis*. *Mol. Microbiol.* **12**:1013–1023.
 26. **Smith, G. R.** 1988. Homologous recombination in prokaryotes. *Microbiol. Rev.* **52**:1–28.
 27. **Stambuk, S., and M. Radman.** 1998. Mechanism and control of interspecies recombination in *Escherichia coli*. I. Mismatch repair, methylation, recombination and replication functions. *Genetics* **150**:533–542.
 28. **Vulic, M., F. Dionisio, F. Taddei, and M. Radman.** 1997. Molecular keys to speciation: DNA polymorphism and the control of genetic exchange in enterobacteria. *Proc. Natl. Acad. Sci. USA* **94**:9763–9767.
 29. **Westmoreland, J., G. Porter, M. Radman, and M. A. Resnick.** 1997. Highly mismatched molecules resembling recombination intermediates efficiently transform mismatch repair proficient *Escherichia coli*. *Genetics* **145**:29–38.
 30. **Williams, D. L., L. Spring, L. Collins, L. P. Miller, L. B. Heifets, P. R. Gangadharam, and T. P. Gillis.** 1998. Contribution of *rpoB* mutations to development of rifamycin cross-resistance in *Mycobacterium tuberculosis*. *Antimicrob. Agents Chemother.* **42**:1853–1857.
 31. **Zawadzki, P., M. S. Roberts, and F. M. Cohan.** 1995. The log-linear relationship between sexual isolation and sequence divergence in *Bacillus* transformation is robust. *Genetics* **140**:917–932.