

DNA Sequence Similarity Requirements for Interspecific Recombination in *Bacillus*

Jacek Majewski and Frederick M. Cohan

Department of Biology, Wesleyan University, Middletown, Connecticut 06459

Manuscript received February 12, 1999

Accepted for publication August 9, 1999

ABSTRACT

Gene transfer in bacteria is notoriously promiscuous. Genetic material is known to be transferred between groups as distantly related as the Gram positives and Gram negatives. However, the frequency of homologous recombination decreases sharply with the level of relatedness between the donor and recipient. Several studies show that this sexual isolation is an exponential function of DNA sequence divergence between recombining substrates. The two major factors implicated in producing the recombinational barrier are the mismatch repair system and the requirement for a short region of sequence identity to initiate strand exchange. Here we demonstrate that sexual isolation in *Bacillus* transformation results almost exclusively from the need for regions of identity at both the 5' and 3' ends of the donor DNA strand. We show that, by providing the essential identity, we can effectively eliminate sexual isolation between highly divergent sequences. We also present evidence that the potential of a donor sequence to act as a recombinogenic, invasive end is determined by the stability (melting point) of the donor-recipient complex. These results explain the exponential relationship between sexual isolation and sequence divergence observed in bacteria. They also suggest a model for rapid spread of novel adaptations, such as antibiotic resistance genes, among related species.

CURRENT models of gene transfer in bacteria assume that the RecA protein binds to a single-stranded 3' end of donor DNA and then searches the recipient genome for a homologous segment (Rao *et al.* 1995). This search is satisfied by an identical region of ~26 bp, the minimum efficiently processed segment (MEPS) believed to be the unit of homologous pairing (Shen and Huang 1986, 1989; Hsieh *et al.* 1992; Rao *et al.* 1995). After identifying the homologous region, RecA is able to initiate strand invasion, which is followed by branch migration and extension of the donor-recipient heteroduplex molecule. The stringency of the homology search, along with the subsequent action of the mismatch repair system, is believed to cause sexual isolation between divergent bacterial species (Harris-Warwick and Lederberg 1978; Rayssiguier *et al.* 1989; Datta *et al.* 1997; Vulic *et al.* 1997; Majewski and Cohan 1998).

We previously demonstrated that, in *Bacillus* transformation, sexual isolation is an exponential function of DNA sequence divergence between the donor and the recipient (Roberts and Cohan 1993; Zawadzki *et al.* 1995). We subsequently showed that, in our system, mismatch repair is not a significant cause of sexual isolation and proposed a model where sexual isolation between divergent sequences is caused by difficulty in finding regions of identity necessary for strand invasion

(Majewski and Cohan 1998). While the model explained the observed exponential relationship, it was based on molecular mechanisms identified in *Escherichia coli*, which may not necessarily extend to *Bacillus*. Moreover, most of the molecular studies of recombination are performed either *in vitro* or between artificial substrates (such as plasmid and phage). Little is known about actual *in vivo* mechanisms governing the major pathways of genetic exchange in bacteria (transformation, conjugation, and transduction).

In this work, we used an *in vivo* system of natural transformation in *Bacillus subtilis* to study the DNA sequence requirements for successful genetic exchange between divergent *Bacillus* species. We followed the approach of Roberts and Cohan (1993) involving transformation of *B. subtilis* toward resistance to rifampicin (*rif^R*), using mutant *rif^R* alleles of *rpoB* from related *Bacillus* species (Roberts and Cohan 1993; Zawadzki and Cohan 1995; Zawadzki *et al.* 1995; Majewski and Cohan 1998). We used chimeric constructs composed of donor DNA flanked by the recipient's own DNA to show that flanking identity is necessary for successful recombination. We also carried out a detailed analysis of recombinant junctions resulting from transformation with divergent DNA and determined the minimum sequence similarity for junctions of recombination.

MATERIALS AND METHODS

Strains: *B. subtilis* subsp. *subtilis* strain 1A96 was obtained from the *Bacillus* Genetic Stock Center (BGSC; Table 1). The *B. subtilis* mismatch repair deletion mutant, PB1856 (Ginetti

Corresponding author: Jacek Majewski, Laboratory of Statistical Genetics, Box 192, Rockefeller University, 1230 York Ave., New York, NY 10021. E-mail: majewski@complex.rockefeller.edu

TABLE 1
List of strains

Strain	Origin	Genotype
<i>B. subtilis</i> subsp. <i>subtilis</i>		
1A96	BGSC	<i>pheA1 trpC2</i>
PB1856	A. Albertini	<i>pheA1 trpC2 ΔmutSL cat</i>
BG126	J. C. Alonso	<i>addA5(β)</i>
BG125	J. C. Alonso	<i>addB72(β)</i>
RO-NN-1	F. M. Cohan from soil	wt
<i>B. subtilis</i> subsp. <i>spizizeni</i>		
RO-E-2	F. M. Cohan from soil	wt
2A2	BGSC	wt
<i>B. Mojavensis</i>		
RO-C-2	F. M. Cohan from soil	wt
<i>B. atrophaeus</i>		
NRRL NRS-213	ARSCC	wt
<i>B. amyloliquefaciens</i>		
ATCC 23350	ATCC	wt
<i>B. licheniformis</i>		
ATCC 14580	ATCC	wt

wt, wild type.

et al. 1996), was kindly provided by Alessandra Albertini. The isogenic strains 1A96 and PB1856 were used as recipients in transformation experiments. The *B. subtilis addA* and *addB* mutants, BG125 and BG126 (Alonso *et al.* 1992), kindly provided by Juan C. Alonso, were used as recipients in experiments investigating the role of the AddAB enzyme on the length of integrated donor segments. The type strains of *B. amyloliquefaciens* (ATCC 23350) and *B. licheniformis* (ATCC 14580) were obtained from the American Type Culture Collection (ATCC). The type strain of *B. atrophaeus* (NRRL NRS-213) was obtained from the Agricultural Research Service Culture Collection (ARSCC) at the National Center for Agricultural Utilization Research. Naturally occurring strains of *B. subtilis* subsp. *subtilis* (RO-NN-1), *B. subtilis* subsp. *spizizeni* (RO-E-2; Nakamura *et al.* 1999), and *B. Mojavensis* (RO-C-2) were isolated from soil (Cohan *et al.* 1991).

Estimate of sequence divergence at *rpoB*: The reported values of sequence divergence between the recipient strains and the donor strains are based on restriction digest data analysis (Roberts and Cohan 1993). The divergence level between *B. licheniformis* and the recipient was confirmed by sequencing 3320 bp of the *B. licheniformis* gene.

Transformation: The recipient strains were induced to be competent and were transformed toward rifampicin resistance with 3 μg/ml DNA (Cohan *et al.* 1991; Roberts and Cohan 1993). This DNA concentration is above saturating, so as to prevent any effect of errors in concentration measurement on transformation frequencies. For each experimental replicate, sexual isolation (ρ) was calculated as the ratio of the homogamic transformation frequency (*i.e.*, using *rif^R* DNA derived from a mutant of the recipient strain) to the heterogamic transformation frequency (using a divergent donor's *rif^R* DNA). The sexual isolation values were then averaged over all experimental trials. Frequencies of transformation were calculated as the fraction of colony-forming units that were resistant to rifampicin, after accounting for spontaneous mutation toward rifampicin resistance. (The spontaneous rate of resistance was determined for each experimental trial and subtracted from the colony counts obtained after transformation.)

PCR amplification of the *rpoB* gene: The 3361-bp fragment

was amplified using primers extending from 9 to 29 (5'-TCAACTAGTTCAGTATGGACG-3') and 3369 to 3349 (5'-ACCTGGTTCAGGAACATTGTC-3') of the *B. subtilis rpoB* sequence (Boor *et al.* 1995). For the 5' fragment of chimeric constructs the primers were 9–29 (as above) and 1352–1332 (5'-TCAGCTGAGGCGTAATTGTAT-3'; Figure 1b), or 1464–1441 (5'-TGCTGACAGACGACGCTTGTACGT-3'; Figure 1c). For the fragment encompassing the *rif^R* region, the primers were 1322–1352 (5'-ATACAATTACGCCTCAGCAGCTGATCAATAT-3') or 1441–1475 (5'-ACGTACAAGCGTCGTCTGT CAGCAYTWGGNCCNGG-3'), and 2533–2512 (5'-TACCAA CAAGAAGATCTCCGTC-3'). The primers for the 3' end of the gene were 2512–2533 (5'-GACGGAGATCTTCTTGTG GTA-3') and 3369–3349 (as above). All of the above PCRs were carried out using Vent DNA polymerase (New England Biolabs, Beverly, MA), at 52° annealing temperature, under standard conditions, for 25 cycles. The fragments were purified by 0.8% agarose gel electrophoresis and extracted from gels using the Qiaquick (QIAGEN, Chatsworth, CA) gel purification kit. The DNA concentration was determined by comparing to a standardized concentration of λ phage *BsEII* marker.

Overlap PCR and construction of chimeric genes: Chimeric genes were constructed using a method adapted from Upender *et al.* (1995; Ge and Rudolph 1997). The following three overlapping gene fragments were obtained in the primary PCRs (above): 5' end of the recipient gene, *rif^R* region of the donor gene, and the 3' end of the recipient gene. These fragments were mixed in equimolar quantities, and 3 ng of the mixture was used as template in the secondary PCR. This reaction was carried out using *Taq* DNA polymerase (Promega, Madison, WI) for 2 cycles, at 52° and 2.5 min. extension time to allow formation of a chimeric template. The end primers (9–29 and 3369–3349, see above) were then added to the mixture and the reaction was carried out for an additional 25 cycles, with 3.5 min extension time. The chimeric products were gel purified, adjusted to a standard concentration, and used for transformation. The accuracy of the constructs was confirmed by restriction mapping.

Sequencing: We sequenced the *rpoB* gene from the *B. licheniformis* donor strain and obtained 3320 bp of double-stranded

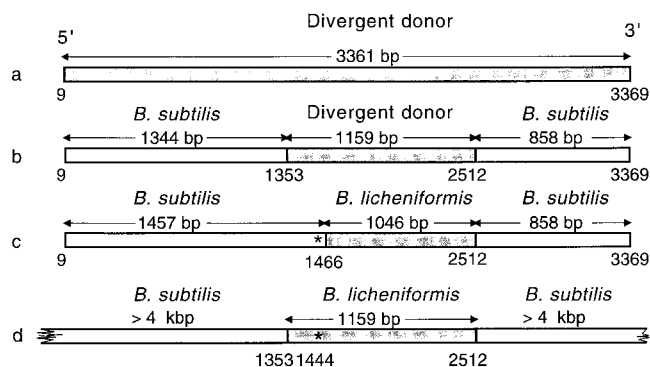


Figure 1.—Substrates used for transformation. Shading indicates the portion of the gene derived from a divergent donor. Simple PCR (a) was used to amplify *rpoB* genes from *B. subtilis* subsp. *subtilis* strain 1A96 (recipient), *B. subtilis* subsp. *subtilis* RO-NN-1 (1% divergent at *rpoB*), *B. subtilis* subsp. *spizizeni* RO-E-2 (3%), *B. mojavensis* RO-C-2 (4.5%), *B. atrophaeus* NRRL NRS-213 (7%), *B. amyloliquefaciens* ATCC 23350 (8%), *B. licheniformis* ATCC 14580 (14.5%). Chimeric genes (b and c) were constructed using a method adapted from Upende *et al.* (1995). The strains used in the chimeric constructs (b) were from *B. mojavensis*, *B. atrophaeus*, and *B. licheniformis*. Only the *B. licheniformis* strain was used for the modified construct (c). Chromosomal DNA (d) was isolated from a 1A96 transformant containing the 1159-bp *rif^R* segment of *B. licheniformis* DNA. The asterisk indicates the position of the *rif^R* mutation, an H → Y substitution at amino-acid position 482 of *B. subtilis* RpoB.

sequence (bases 30–3349 of *rpoB*, GenBank accession no. AF172323). All sequencing reactions were carried out directly from PCR products, at the University of Pennsylvania DNA sequencing facility.

RESULTS

Transformation with chimeric DNA fragments: We used overlap-extension PCR to construct DNA fragments consisting of the *rif^R* region from a divergent donor species, flanked by 5' and 3' ends of the *B. subtilis* recipient (Figure 1). Such chimeric genes provide the recombining *rif^R* fragment with regions of perfect DNA sequence identity necessary for initiation of heteroduplex formation. We transformed two recipients (a wild-type *B. subtilis* and its mismatch-repair-deficient derivative) with chimeric constructs (Figure 1) and purely donor-derived PCR-amplified segments (Figure 1). The results are shown in Tables 2 and 3. Sexual isolation obtained with the constructs was drastically reduced relative to that of simple PCR products (Figure 2). For *B. licheniformis* (14.5% divergence at *rpoB*) as donor, the sexual isolation decreased from $\rho = 471$ (simple PCR fragment) to $\rho = 3.6$ (chimeric construct) for the wild-type recipient and from $\rho = 354$ (simple PCR) to $\rho = 2.4$ (construct) for the mismatch repair mutant.

Retransformation with chimeric genomic DNA: While using the chimeric PCR constructs greatly reduced the recombinational barrier between *B. subtilis* and *B. li-*

cheniformis, sexual isolation was not totally eliminated even in the absence of mismatch repair ($\rho = 2.4$). We believed that the residual sexual isolation resulted from clipping of the construct ends, which occurs before strand exchange (Dubnau 1993; Zawadzki and Cohan 1995; see below); this would remove the flanking identity from the donor DNA molecule. (Note that this exclusion of flanking ends explains why simple PCR products exhibit high levels of sexual isolation, despite being flanked by 20- to 26-bp primers constituting perfect identity to the recipient; the primer regions are always removed prior to integration.) We next isolated chromosomal DNA from a transformant containing the entire 1159 bp of *B. licheniformis* insert (Figure 1d) and used this chimeric DNA to retransform the recipients. The rationale was that chromosomal DNA would be long enough (usually >8 kb; Dubnau 1993) to prevent the removal of flanking identity. Sexual isolation obtained with chromosomal transformant DNA (as compared to recipient's own *rif^R* chromosomal DNA) was $\rho = 2.56$ in the wild-type recipient and $\rho = 1.18$ in the mismatch repair mutant (Table 4). Hence, by providing flanking regions of identity, we can almost entirely remove the recombinational barrier for a divergent DNA segment. By further removing the mismatch repair system, we can eliminate the barrier completely. These results show that sexual isolation in *Bacillus* is predominantly caused by difficulty in strand invasion, with only a slight contribution from mismatch repair.

Analysis of fragments integrated during transformation: To confirm that the residual sexual isolation observed using chimeric PCR constructs results from the construct ends being removed prior to recombination, we analyzed the transformant *rpoB* DNA. We extracted genomic DNA from the transformants obtained with the *B. licheniformis* construct and then PCR amplified the transformant *rif^R* regions (1159 bp). We used restriction mapping to determine the length of foreign donor DNA that actually integrated into the chromosome. We chose *B. licheniformis* for this analysis, since its high degree of sequence divergence from the recipients allowed us to generate detailed restriction maps and to determine the size of the donor insert with high confidence. We analyzed 12 independent transformants of the wild-type recipient (1A96) and 12 of the mismatch repair mutant (PB1856). While there was no difference in the lengths of fragments integrated by the wild-type and mismatch repair deficient recipients, we found that only 12.5% of the transformants contained the entire donor insert, while the majority (82%) integrated only small pieces (<300 bp) directly adjacent to the 5' region of identity. Hence, the 3' region of identity (858 bp) was usually removed prior to recombination and could not facilitate strand invasion.

We next investigated the possibility that, once strand invasion takes place, there might exist a difficulty in extending the heteroduplex past mismatched regions.

TABLE 2

Transformation frequencies and levels of sexual isolation using purely donor-derived PCR-amplified DNA

Donor	Frequency of transformation ($\times 10^{-3}$)		Sexual isolation	
	Wild-type recipient (1A96)	Δ MutSL recipient (PB 1856)	Wild-type recipient (1A96)	Δ MutSL recipient (PB 1856)
<i>B. subtilis</i> subsp. <i>subtilis</i> 1A96	1.575 \pm 0.428 ^a	2.893 \pm 0.401 ^a	1	1
<i>B. subtilis</i> subsp. <i>subtilis</i> RONN1	1.743 \pm 0.617	2.819 \pm 0.491	1.174 \pm 0.197 ^a	1.059 \pm 0.057 ^a
<i>B. subtilis</i> subsp. <i>spizizeni</i> ROE2	0.804 \pm 0.167	1.983 \pm 0.225	1.875 \pm 0.256	1.477 \pm 0.164
<i>B. mojavensis</i> ROC2	0.767 \pm 0.196	1.609 \pm 0.329	2.313 \pm 0.315	2.043 \pm 0.269
<i>B. atrophaeus</i> NRRL NRS-213	0.153 \pm 0.052	0.491 \pm 0.157	14.81 \pm 4.38	5.972 \pm 0.547
<i>B. amyloliquefaciens</i> ATCC 23350	0.025 \pm 0.006	0.173 \pm 0.044	64.89 \pm 4.83	20.30 \pm 3.73
<i>B. licheniformis</i> ATCC 14580	0.0032 \pm 0.0008	0.0069 \pm 0.0017	471.0 \pm 27.9	354.8 \pm 83.7

^a Mean and standard error values for four experimental trials.

Such a mechanism has been proposed as a barrier to transduction between *Salmonella* species (Zahrt and Maloy 1997): while branch migration stalls at mismatched regions, the RecBCD enzyme is believed to digest the free DNA strand, reducing the amount of genetic material available for integration. The *B. subtilis* *addA* and *addB* are homologues of *E. coli* *recB* and *recC*, and the AddAB enzyme is functionally active in *E. coli* strains deficient in RecBCD activity (Kooistra *et al.* 1993). We used the *B. licheniformis* construct to transform *B. subtilis* strains PB124 and PB125, with deletions in the *addA* and *addB* genes, respectively. We analyzed 12 transformants of each mutant recipient and found that the sizes of integrated foreign inserts were not significantly different than those in the wild-type recipient. Thus, AddA and AddB are not responsible for reducing the size of recombining substrates in *Bacillus* transformation.

To test whether some other factor might be responsible for clipping the free end of DNA during branch migration, we created a modified construct, where the 5' region of identity was extended further downstream to include the *rif^R* mutation (Figure 1). We reasoned that if only a small fragment directly adjacent to the identity region can be integrated, then extending the identity toward the 3' end should result in transformants extending further 3'-ward into the *B. licheniformis* insert.

However, on analyzing the transformants obtained with the new constructs, we found that the donor fragments were now smaller (<200 bp) and that most of them (22 out of 24) did not extend past base 1564 of *rpoB*. This was the same cutoff position observed with the previous construct, suggesting that there might be a recombination "hot spot" around position 1564. On examining the donor-recipient sequence alignment in this region, we found two highly conserved sequences (1534–1565 and 1495–1523; Figure 3) that may constitute such a recombination hot spot. These results suggest that regions of identity may be necessary at both ends of the donor fragment, as has been suggested by Harris-Warwick and Lederberg (1978). This requirement, rather than difficulty in heteroduplex extension past mismatched regions, may be responsible for the small size of integrated fragments.

Analysis of recombination junctions and requirements for successful recombination: To determine whether invasive ends need to be present on only the 3' end of the donor strand (as in *E. coli*; Rao *et al.* 1995; Reddy *et al.* 1995) or on both ends, as suggested by the results above, we isolated 43 clones resulting from transforming the wild-type recipient with PCR-amplified segments of purely *B. licheniformis*-derived *rpoB* (Figure 1a). We used restriction mapping to determine the exact size and integration junctions of the 43 clones (Figure

TABLE 3

Transformation frequencies and levels of sexual isolation using chimeric PCR-amplified DNA

Donor (source of chimera's insert)	Frequency of transformation ($\times 10^{-3}$)		Sexual isolation	
	Wild-type recipient (1A96)	Δ MutSL recipient (PB 1856)	Wild-type recipient (1A96)	Δ MutSL recipient (PB 1856)
<i>B. subtilis</i> subsp. <i>subtilis</i> 1A96	1.869 \pm 0.214 ^a	3.663 \pm 0.757 ^a	1	1
<i>B. mojavensis</i> ROC2	1.404 \pm 0.189	3.249 \pm 0.796	1.343 \pm 0.023 ^a	1.210 \pm 0.088 ^a
<i>B. atrophaeus</i> NRRL NRS-213	0.834 \pm 0.069	3.010 \pm 0.723	2.223 \pm 0.076	1.256 \pm 0.054
<i>B. licheniformis</i> ATCC 14580	0.514 \pm 0.031	1.551 \pm 0.353	3.604 \pm 0.204	2.411 \pm 0.067

^a Mean and standard error values for four experimental trials.

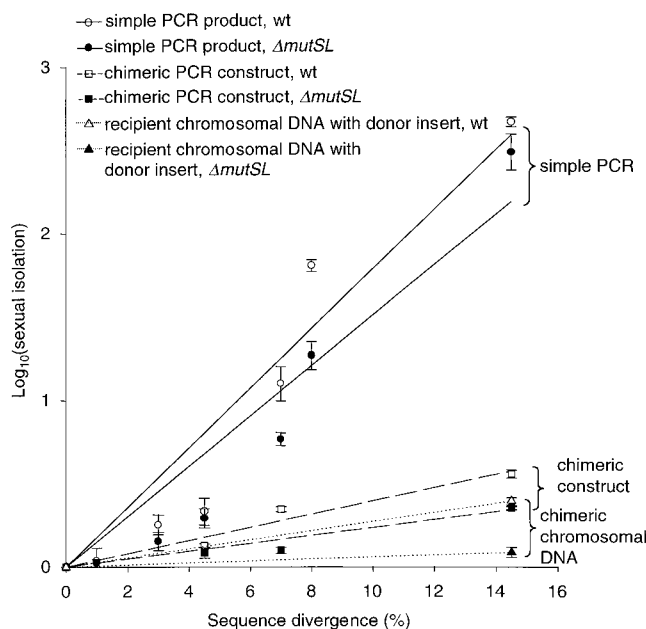


Figure 2.—Sexual isolation using different substrates as donor DNA. The horizontal axis represents the sequence divergence (%) between the recipient and the donor (or donor insert in the case of constructs). Error bars represent standard errors for at least four replicate transformation experiments. Lines represent regression. The mismatch repair genotype of the recipient [wild type (*wt*) or mutant (Δ *mutSL*)] is indicated.

4). We found that 84 out of the 86 junctions analyzed ended in relatively mismatch-free stretches of 20–30 bp containing at most one or two mismatches (Figure 5). These results demonstrate that invasive ends are necessary on both ends of the donor strand. This differs from the findings in *E. coli*, where it is believed that a single-stranded invasive 3' end is sufficient for successful recombination between divergent sequences.

Finally, we determined the sequence similarity requirements for invasive ends. Many of the recombination junctions fell within regions of perfect sequence identity and could be easily explained. However, 44 junctions ended in regions interrupted by one to several mismatches. Moreover, some of the most recombinogenic regions were virtually identical in degree of mismatch to nearby regions producing no junctions (Figure

3). We found that all the mismatched recombinogenic regions were characterized by a relatively high GC content (*i.e.*, GC > 52%, in a gene with 45% GC). Hence, we hypothesized that there might exist a critical stability of the bond (which may be quantified by its melting temperature) between the invasive end and the recipient DNA, necessary to initiate recombination. To determine the critical melting point (T_c), we scanned the donor-recipient sequence alignment, identifying all oligonucleotides between 18 and 27 bp long (which is believed to be the MEPS size in *E. coli*; Hsieh *et al.* 1992) and calculated their melting temperatures. We tested several critical melting temperatures for the degree of association between the theoretical MEPS criterion and the empirically observed transformant junctions. The temperature $T_c = 51.2^\circ$ explained the greatest number of observed junctions, while minimizing the number of unused putative MEPS sequences (Figure 5). This melting temperature criterion, which takes into consideration the number of mismatches, the GC content, and the length of the oligonucleotide, gave a better explanation of the observed data than any criterion involving the length and the number of mismatches alone.

DISCUSSION

Mechanisms and barriers to recombination: The molecular mechanisms of recombination in *Bacillus* have not yet been fully elucidated, yet it is clear that they differ from those observed in *E. coli*. In this article, we show that recombination in *Bacillus* requires short regions of conserved sequences to be present at both ends of the donor-recipient DNA. This is in contrast to the case of *E. coli*, where a conserved sequence at only one end is sufficient (Shen and Huang 1986). We show that the mismatch repair system is relatively ineffective in preventing recombination between divergent sequences and that the main barrier to recombination is caused by the difficulty in locating conserved regions necessary for strand invasion. Thus, highly divergent donor segments (up to 15% in this study) recombine with no resistance, if they are flanked with sequences identical to the recipient. This result implies that recom-

TABLE 4

Transformation frequencies and levels of sexual isolation using genomic transformant DNA

Donor of genomic DNA	Frequency of transformation ($\times 10^{-3}$)		Sexual isolation	
	Wild-type recipient (1A96)	Δ MutSL recipient (PB 1856)	Wild-type recipient (1A96)	Δ MutSL recipient (PB 1856)
1A96 spontaneous <i>rif^R</i> mutant	10.3 \pm 1.8 ^a	10.0 \pm 1.1 ^a	1	1
1A96 containing a 1159-bp <i>B. licheniformis</i> insert	5.6 \pm 0.8	8.4 \pm 1.0	2.56 \pm 0.16 ^a	1.18 \pm 0.06 ^a

^a Mean and standard error values for four experimental trials.

	5' 3'	position bp.	no. of junctions	length/no. of mismatches	% GC	melting temp. °C
1	AACCAATCCGATCGGTTTAAGC AACCAATCCGATCGGTTTAAGC	1252-1275	0	24/1	39	47
2	ACGCCAGCAGCTGATCAATATTCG ACGCCAGCAGCTGATCAATATTCG	1330-1355	8	26/1	52	54
3	ATGGATCAGCGAATCCCGTTGCTGA ATGGATCAGCGAATCCCGTTGCTGA	1411-1436	8	26/1	52	54
4	GCCGGAATGGAATCCGTGACGTTCACTA GCCGGAATGGAATCCGTGACGTTCACTA	1495-1523	10	28/1	56	56
5	GGCCGATGTGTCGGATTGAAACCCCTGAGG GGCCGATGTGTCGGATTGAAACCCCTGAGG	1534-1565	12	32/2	55	54
6	GAAACGCCGATCGCCGCGT GAAACGCCGATCGCCGCGT	1630-1649	2	20/1	63	52
7	AACCTTATAAATACATGCTGCTGAATTTGTCCG AACCTTATAAATACATGCTGCTGAATTTGTCCG	2116-2150	0	35/2	35	51
8	TGCTACAACCGGTCGATCGTAAGGTCGG TGCTACAACCGGTCGATCGTAAGGTCGG	2166-2197	4	32/1	58	59

Figure 3.—Examples of sequences with low numbers of mismatches permitting (segments 2–6 and 8) and not permitting (segments 1 and 7) recombination junctions. Segments as short as 20 bp with one mismatch may be used as invasive ends, provided they have a high GC content (segment 6). In contrast, larger segments with low GC content (segments 1 and 7) produced no transformants.

bination rates at various loci will depend not only on the *average* sequence divergence, but also on the degree of clustering of mismatches. Genes containing highly conserved regions interspersed with divergent stretches will be able to initiate strand invasion and should recombine easily, while genes that are uniformly divergent should recombine less frequently.

The present data allow us to improve our theoretical model explaining the exponential relationship between sequence divergence and sexual isolation (Majewski and Cohan 1998). We previously showed that the probability of finding a conserved sequence of n nucleotides on the 3' invasive end of the donor strand decreases

exponentially with the DNA sequence divergence between the donor and the recipient. In view of the new results, we now require that such conserved sequences be present on both ends of the donor strand. The modified model still leads to an exponential relationship between sexual isolation and sequence divergence. However, we now obtain a different estimate for the length (n) of the MEPS,

$$n = 2\phi_b \times \ln(10),$$

where ϕ_b is the empirically determined sensitivity of sexual isolation to sequence divergence, in the absence of mismatch repair. [This is the same as Equation 5 of Majewski and Cohan (1998), except for the additional factor of 2 to account for the need of identity at both ends]. Using $\phi_b = 17.85$ (Majewski and Cohan 1998), the model predicts that the length of the MEPS should be 21 bp. This prediction is in excellent agreement with the MEPS sequences observed in our analysis of recombination junctions, where we show that regions of identity as short as 20 bp may promote recombination between divergent species.

The requirement for flanking identity can also explain why there is a more significant deviation from the exponential relationship when PCR-amplified DNA fragments are used in transformation, as compared to genomic DNA, which follows the relationship very closely (Roberts and Cohan 1993; Majewski and Cohan 1998). This deviation has been observed by Zawadzki *et al.* (1995) and in this article (Figure 2). We expect that transformation with small fragments, where most of the strand invasion events take place in the immediate vicinity of the mutation conferring antibiotic resistance, will be very sensitive to the position of the mutation; if the mutation occurs within a highly conserved region, transformation rates will be en-

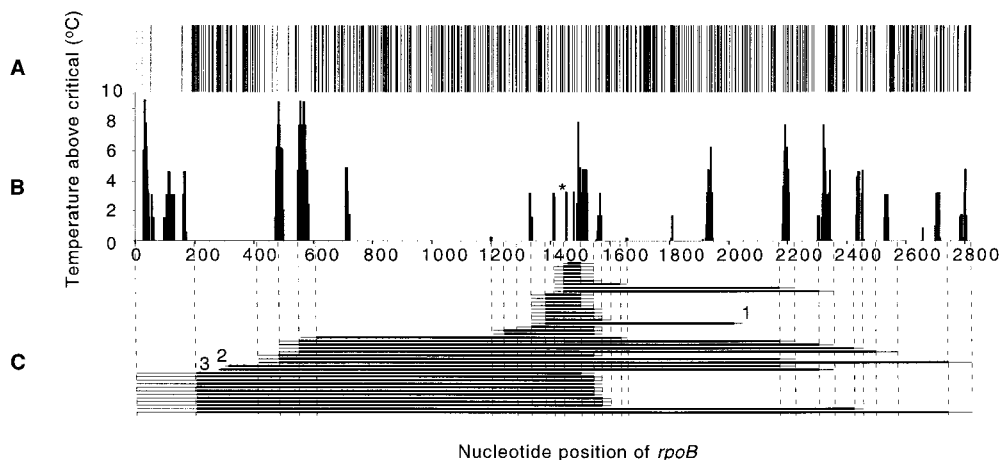


Figure 4.—(A) *rpoB* sequence alignment of *B. subtilis* and *B. licheniformis* for the first 2810 bp (the region where transformation was detected). Dark lines represent mismatches. (B) Extent to which melting temperatures exceed the critical temperature of $T_c = 51.2^\circ$ (see Figure 5 for explanation) for each oligonucleotide (18–27 bp) that may serve as an invasive end. The melting temperature was calculated using the formula $T_m = 81 - 600/\text{length} + 16.6 \log[\text{Na}] + 0.41(\%GC) - 0.8(\% \text{ mismatch})$ and evaluated at 50 mM

Na^+ concentration (Sambrook *et al.* 1989; Malkov *et al.* 1997). (C) Inserts resulting from transformation of wild-type recipient with pure *B. licheniformis* PCR-amplified fragment. The end lines represent confidence intervals for transformant junctions. The asterisk represents the rif^R mutation (nucleotide 1444). All junctions (except those labeled 1, 2, and 3) are associated with high melting point regions. The exceptions, 1 and 2, fall within an AT-rich region containing sequences similar to those implicated in illegitimate recombination in *B. subtilis* (Meima *et al.* 1998). Junction 3 falls in a region with a T_m of 49° (just below critical).

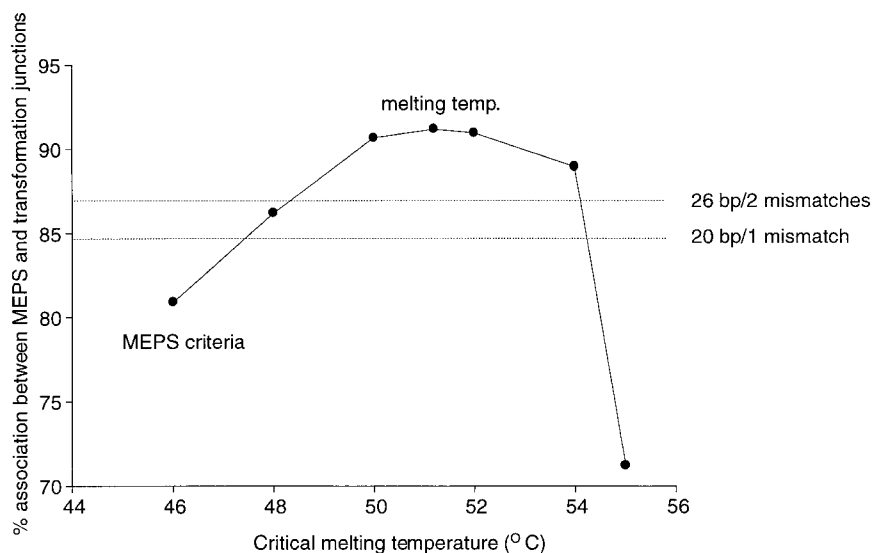


Figure 5.—Estimation of the critical melting temperature of a MEPS. The vertical axis value is calculated as the mean of the following: (1) the percentage of junctions that are associated with a given critical temperature and (2) the percentage of putative MEPS with the given critical melting temperature that are associated with transformant junctions. Note that using the melting point (and thus including GC content) in MEPS determination gives a better explanation of experimental data than using only length and the number of mismatches. No preference was found for GT-rich regions, believed to be favored in recombination in *E. coli* and *Saccharomyces cerevisiae*.

hanced, whereas mutations in less conserved regions will exhibit reduced frequencies of integration. On the other hand, during transformation with genomic DNA, strand invasion takes place within several kilobases of the mutation (Zawadzki and Cohan 1995), and the exact position of the marker is not as relevant. Hence, for transformation with spontaneous mutant markers (as is the case in our experiments), the exponential relationship will be followed more closely by large DNA fragments.

Our results also suggest that the strength of the bonds formed between the endpoints of the donor-recipient complex may determine whether or not recombination will take place. In our system, sequences with a relatively high GC content are more likely to produce recombinants than similar sequences (with respect to length of identity and number of mismatches) with a lower GC content. Although it has been suggested that the bonds formed between the invading donor strand and the recipient DNA duplex are not Watson-Crick in nature (Reddy *et al.* 1995), it is possible that their strength might still be dependent on GC pairing. We did not find any preference for recombination in GT-rich regions, which have been shown to preferentially bind both the *E. coli* RecA and the yeast Rad51 homologue (Tracy *et al.* 1997).

We are currently unable to determine why *two* conserved regions are necessary for successful recombination in *Bacillus*. *In vitro* studies using *E. coli* proteins show that RecA is able to initiate strand invasion at both 3' and 5' free ssDNA ends, but that branch migration can take place only in the 3' → 5' direction. It is possible that in *Bacillus* the 5' bond serves either to stabilize the entire complex during branch migration, or is necessary for termination of the process. It is also possible that the polarity of branch migration is different in *Bacillus* than in *E. coli*, since the yeast RecA homologue, Rad51, promotes branch migration in the 5' → 3' direction.

Interspecies recombination and adaptation: Consider next the possible adaptive value of the barriers to interspecies recombination observed in *Bacillus* and other taxa. We have previously shown that recombination across species poses little fitness cost in bacteria (Cohan 1994). Because recombination in bacteria is extremely rare, the cost of incorporating maladaptive foreign alleles should be very low, even if between-species recombination were to occur at the same rate as recombination within species. Given that recombination occurs within *Bacillus* species, for example, at a rate of 10^{-7} , the incremental fitness increase of further decreasing the between-species recombination rate is likely very small. Nevertheless, the selection disfavoring between-species recombination is not negligible, owing to the large population sizes for bacteria.

On the other hand, it is occasionally beneficial for bacteria to incorporate DNA from other species. The promiscuous nature of bacterial recombination allows a recipient to acquire adaptations from other species. Bacterial genomes are modular enough to accommodate the expression of some foreign alleles, and the size of recombining fragments is frequently small enough so that a generally adaptive allele can pass between species without the cotransfer of maladaptive alleles at other loci (Zawadzki and Cohan 1995). The average benefit of being able to recombine between species at an increased rate is likely to be quite low, however, since only a small fraction of adaptations would be expected to be functional in different species.

In summary, there are low, but nonzero, selection pressures acting to both increase and decrease the rate of between-species recombination. We expect that the optimum level of interspecies recombination will depend on the relative importance of obtaining other species' adaptations *vs.* avoiding uptake of maladaptive foreign alleles.

Can the balance between these selection pressures

explain differences between bacterial species in their resistance to between-species recombination? This study has shown that in *B. subtilis* transformation, resistance to recombination occurs primarily at the strand invasion stage, with little effect of mismatch repair. That recombination appears to be blocked only by the thermodynamic requirements of donor-recipient stability suggests that the recombinational barrier observed may be a minimum that cannot be further reduced. In contrast, conjugation-mediated recombination between *E. coli* and *Salmonella typhimurium* is reduced 100 times more than is the case for transformation between similarly divergent *Bacillus* species (Vulic *et al.* 1997). This recombinational barrier is even more pronounced in transduction between *S. typhimurium* and *S. typhi* (Zahrt and Maloy 1997), where a 1% DNA sequence divergence can lead to an almost complete inhibition of recombination. This is due to a greater inhibitory effect of mismatch repair on recombination between these enteric species.

The difference between *Bacillus* and the enteric bacteria in recombinational barriers may result from a difference in the balance between the positive and negative selection pressures acting on interspecies recombination rates. In *Bacillus* transformation, competence is induced primarily under starvation (Dubnau 1993), a time when the relative importance of taking up new metabolic capabilities might be enhanced. We argue that transformation under unfavorable growth conditions would push the selective balance in favor of increased between-species recombination. Conjugation between *Escherichia* and *Salmonella* is not associated with a particular condition of growth, so the optimal level of between-species recombination may be quite low, requiring the inhibitory effect of mismatch repair.

It will be interesting to investigate whether genetic barriers such as mismatch repair may become more effective in *Bacillus* under nonstressful conditions. Specifically, are the additional barriers actively suppressed during transformation, resulting in the relatively low sexual isolation, or do they simply not exist under any conditions? Experiments such as interspecific transduction between related *Bacillus* species should be useful in resolving the above issue.

Whether genetic barriers in *Bacillus* are facultatively suppressed or constitutively weak, natural transformation presents an efficient system for acquisition of new genes and functions. Our findings have important implications in the spread of new adaptations, such as antibiotic resistance or artificially introduced foreign genes, in the bacterial world. We suggest that once a gene for a novel adaptation transfers into a species from a divergent donor, the gene becomes flanked by recipient DNA, and, in the absence of barriers other than resistance to strand invasion, this allows efficient homologous recombination of the adaptation into related species (see also Miller and Levy 1989). Our results

suggest that even very divergent and heterologous genes may transfer between species at very high rates via homologous recombination.

This research was funded by U.S. Environmental Protection Agency grant R825348-01-0 and research grants from Wesleyan University.

LITERATURE CITED

- Alonso, J. C., G. Luder and T. A. Trautner, 1992 Intramolecular homologous recombination in *Bacillus subtilis*. *Mol. Gen. Genet.* **236**: 60–64.
- 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.
- Cohan, F. M., 1994 Genetic exchange and evolutionary divergence in prokaryotes. *Trends Ecol. Evol.* **9**: 175–180.
- Cohan, F. M., M. S. Roberts and E. C. King, 1991 The potential for genetic exchange by transformation within a natural population of *Bacillus subtilis*. *Evolution* **45**: 1393–1421.
- 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.
- Dubnau, D., 1993 Genetic exchange and homologous recombination, pp. 555–584 in *Bacillus Subtilis and Other Gram-Positive Bacteria: Biochemistry, Physiology, and Molecular Genetics*, edited by A. L. Sonenshein, J. A. Hoch and R. Losick. American Society for Microbiology, Washington, DC.
- Ge, L., and P. Rudolph, 1997 Simultaneous introduction of multiple mutations using overlap extension PCR. *Biotechniques* **22**: 28–30.
- Ginetti, F., M. Perego, A. M. Albertini and A. Galizzi, 1996 *Bacillus subtilis* mutS mutL operon: identification, nucleotide sequence and mutagenesis. *Microbiology* **142**: 2021–2029.
- Harris-Warrick, R. M., and J. Lederberg, 1978 Interspecies transformation in *Bacillus*: mechanism of heterologous intergenote transformation. *J. Bacteriol.* **133**: 1246–1253.
- Hsieh, P., C. S. Camerini-Otero and R. D. Camerini-Otero, 1992 The synapsis event in the homologous pairing of DNAs: RecA recognizes and pairs less than one helical repeat of DNA. *Proc. Natl. Acad. Sci. USA* **89**: 6492–6496.
- Kooistra, J., B. J. Hajjema and G. Venema, 1993 The *Bacillus subtilis* *addAB* genes are fully functional in *Escherichia coli*. *Mol. Microbiol.* **7**: 915–923.
- Majewski, J., and F. M. Cohan, 1998 The effect of mismatch repair and heteroduplex formation on sexual isolation in *Bacillus*. *Genetics* **148**: 13–18.
- Malkov, V. A., L. Sastry and R. D. Camerini-Otero, 1997 RecA protein assisted selection reveals a low fidelity of recognition of homology in a duplex DNA by an oligonucleotide. *J. Mol. Biol.* **271**: 168–177.
- Meima, R., G. J. Haan, G. Venema, S. Bron and S. de Jong, 1998 Sequence specificity of illegitimate plasmid recombination in *Bacillus subtilis*: possible recognition sites for DNA topoisomerase. *Nucleic Acids Res.* **26**: 2366–2373.
- Miller, R. V., and S. B. Levy, 1989 Horizontal gene transfer in relation to environmental release of genetically engineered microorganisms, pp. 405–420 in *Gene Transfer in the Environment*, edited by S. B. Levy and R. V. Miller. McGraw-Hill, New York.
- Nakamura, L. K., M. S. Roberts and F. M. Cohan, 1999 Relationship between the *Bacillus subtilis* clades associated with strains 168 and W23: a proposal for *B. subtilis* subsp. *subtilis* and *B. subtilis* subsp. *spizizeni*. *Int. J. Syst. Bacteriol.* (in press).
- Rao, B. J., S. K. Chiu, L. R. Bazemore, G. Reddy and C. M. Radding, 1995 How specific is the first recognition step of homologous recombination? *Trends Biochem. Sci.* **20**: 109–113.
- 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.
- 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.
- Roberts, M. S., and F. M. Cohan, 1993 The effect of DNA sequence divergence on sexual isolation in *Bacillus*. *Genetics* **134**: 401–408.
- Sambrook, J., E. F. Fritsch and T. Maniatis, 1989 *Molecular Cloning: A Laboratory Manual*, Ed. 2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Shen, P., and H. V. Huang, 1986 Homologous recombination in *Escherichia coli*: dependence on substrate length and homology. *Genetics* **112**: 441–457.
- Shen, P., and H. V. Huang, 1989 Effect of base pair mismatches on recombination via the RecBCD pathway. *Mol. Gen. Genet.* **218**: 358–360.
- Tracy, R. B., J. K. Baumohl and S. C. Kowalczykowski, 1997 The preference for GT-rich DNA by the yeast Rad51 protein defines a set of universal pairing sequences. *Genes Dev.* **24**: 3423–3431.
- Upender, M., L. Raj and M. Weir, 1995 Megaprimer method for in vitro mutagenesis using parallel templates. *Biotechniques* **18**: 29–30, 32.
- 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.
- Zahrt, T. C., and S. Maloy, 1997 Barriers to recombination between closely related bacteria: MutS and RecBCD inhibit recombination between *Salmonella typhimurium* and *Salmonella typhi*. *Proc. Natl. Acad. Sci. USA* **94**: 9786–9791.
- Zawadzki, P., and F. M. Cohan, 1995 The size and continuity of DNA segments integrated in *Bacillus* transformation. *Genetics* **141**: 1231–1243.
- 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.

Communicating editor: W. F. Eanes