

Recombination proteins differently control the acquisition of homeologous DNA during *Bacillus subtilis* natural chromosomal transformation

Ester Serrano¹, Cristina Ramos¹, Juan C. Alonso^{1,✉} and Silvia Ayora^{1,✉}

Running title: genetic recombination functions

Abstract

In naturally competent *Bacillus subtilis* cells the acquisition of closely related genes occurs *via* homology-directed chromosomal transformation (CT), and its frequency decreases log-linearly with increased sequence divergence (SD) up to 15%. Beyond this and up to 23% SD the interspecies boundary prevails, the CT frequency marginally decreases, and short (<10-nucleotides) segments are integrated *via* homology-facilitated micro-homologous integration. Both poorly known CT mechanisms are RecA-dependent. Here we identify the recombination proteins required for the acquisition of interspecies DNA. The absence of AddAB, RecF, RecO, RuvAB or RecU, crucial for repair-by-recombination, does not affect CT. However, inactivation of *dprA*, *radA*, *recJ*, *recX* or *recD2* strongly interfered with CT. Interspecies CT was abolished beyond ~8% SD in $\Delta dprA$, ~10% in $\Delta recJ$, $\Delta radA$, $\Delta recX$ and 14% in $\Delta recD2$ cells. We propose that DprA, RecX, RadA/Sms, RecJ and RecD2 help RecA to unconstrain speciation and gene flow. These functions are ultimately responsible for generating genetic diversity and facilitate CT and gene acquisition from bacteria of the same genus.

Keywords: horizontal gene transfer | microbial evolution | kin discrimination | Muller's ratchet

¹Department of Microbial Biotechnology, Centro Nacional de Biotecnología, CNB-CSIC, 3 Darwin Str, 28049 Madrid, Spain.

✉ Juan C. Alonso jcalonso@cnb.csic.es and Silvia Ayora sayora@cnb.csic.es

Introduction

The unidirectional nonsexual incorporation of DNA into the genome of a recipient bacterium (horizontal gene transfer) has been the major evolutionary force that has constantly reshaped genomes and persisted through evolution to maximize adaptation to new ecological conditions and genetic diversity ^{1;Fraser, 2007 #154}. The exchange of chromosomal genes among members of the same or related species occurs *via* RecA-dependent homologous recombination (HR) by Hfr conjugation, viral transduction or natural chromosomal transformation (CT) ²⁻⁴. In Hfr conjugation and transduction the recombination machinery processes double-stranded (ds) DNA, whereas during CT the machinery integrates single-stranded (ss) DNA ^{4,5}. Bacteria have developed mechanisms to fight gene transfer by conjugation and transduction, but cells are less efficient to impose barriers to natural transformation ¹. Restriction-modification systems rapidly fragmentate internalised non-self dsDNA, but ssDNA, internalised by transformation, is refractory to most restriction systems ^{6,7}. Adaptive immune systems, as CRISPR-Cas, are usually absent in naturally competent bacteria ⁸.

Genes devoted to natural competence are encoded in the genome of many bacteria, among them the model organism *Bacillus subtilis*, which occupies a wide range of aquatic and terrestrial niches, and colonizes animal guts ⁹. Upon stress, a small fraction of *B. subtilis* stationary phase cells induce competence. During competence development, ongoing DNA replication is halted, a transcriptional programme is activated, a DNA uptake apparatus is assembled at a cell pole, and lysis of kin is induced, releasing DNA for uptake ^{7,10-12}. This DNA uptake apparatus binds any extracellular dsDNA, degrades one strand, and internalises the complementary strand into the cytosol ¹⁰⁻¹².

Naturally competent cells can integrate fully homologous DNA (homogamic CT), and with lower efficiency homeologous (similar but not identical) DNA (heterogamic CT) ². This interspecies CT can replace the recipient by a homeologous DNA, leading to mosaic structures ¹³⁻¹⁵. The transformation frequency of heterogamic DNA decreases with increased sequence divergence (SD) both in *B. subtilis* and *S. pneumoniae* cells ^{14,15}. In *B. subtilis*, the transformation assays with donor DNAs from different *Bacilli* species revealed a biphasic curve. Interspecies CT decreased log-linear up to 15% SD, and integration in this part of the curve occurred *via* one-step homology-directed RecA-dependent gene replacement ^{16,17}. Beyond 15% SD the CT reached a plateau, and the integration of just few nucleotides, at micro-homologous segments (<10-nucleotides [nt]), was observed at a very low efficiency ($\sim 1 \times 10^{-5}$) ¹⁷, suggesting another recombination mechanism.

Genetic recombination functions

CT also contributes to expand the pan-genome of a species, because it allows the integration of a heterologous DNA sequence, if flanked by two regions identical with recipient, usually >400-nt, but longer homologous regions significantly increase the integration frequency^{7,18}. Here, two homologous recombination (HR) events in the flanks integrate the heterologous DNA via recipient-deletion /donor-insertion with ~10-fold lower efficiency than homologous gene replacement^{7,19}. Another mechanism is observed in replicating competent *Streptococcus pneumoniae* or *Acinetobacter baylyi* cells. Here, homology at only one flank (anchor region) facilitates illegitimate recombination of short (<10-nt) micro-homologous segments with subsequent deletion of the intervening host DNA and integration of long heterologous DNA segments, albeit with very low efficiency when compared to homologous CT^{20,21}. The length of the anchor region affects the efficiency of this homology-facilitated illegitimate recombination (HFIR)¹⁹.

The proteins responsible for the acquisition of natural homeologous DNA remain poorly known. The main player during HR is the RecA recombinase. RecA from a naturally competent bacterium (e.g., *B. subtilis*) has evolved to catalyse strand exchange in either the 5'→3' or 3'→5' direction and to tolerate 1-nt mismatch in an 8-nt region^{16,17}. In naturally competent cells, the essential SsbA and competence-specific SsbB coated the incoming linear ssDNA as soon it leaves the entry channel. *B. subtilis* RecA cannot filament onto SsbA- or SsbB-coated ssDNA^{22,23}. The RecA accessory proteins are divided into four broad classes. First, the mediators that act before RecA-mediated homology search. The DprA mediator facilitates partial disassembly of SsbA and SsbB from the ssDNA, allows RecA binding, and in concert with SsbA assists RecA to catalyse DNA strand exchange²³. In the absence of DprA, RecO assists RecA to filament onto SsbA-coated ssDNA²². Second, the modulators, which act during DNA identity search and strand exchange, and either promote RecA nucleoprotein filament assembly as RecF (in the $\Delta recX$ context) or its disassembly from the ssDNA, as RecX or RecU (in the $\Delta recX$ context)²⁴⁻²⁸. With the help of both, mediators and modulators, a dynamic RecA-ssDNA filament is engaged in sequence identity search. *E. coli* RecA requires ~15-nt of homologous ssDNA to promote strand exchange, defining the *in vitro* minimal efficient processing segment (MEPS)^{29,30}. *In vivo*, *E. coli* RecA significantly recombines two duplex DNAs with 25- to 30-bp MEPS³¹, thus this length was proposed for the *B. subtilis* protein³². *In vivo*, upon finding a MEPS, RecA initiates strand invasion to form a heteroduplex, known as displacement loop (D-loop)³³. Then, branch migration translocases (RecD2, RuvAB, RecG and RadA/Sms) allow D-loop extension, and help to generate a stable heteroduplex^{28,34}. Finally, after DNA strand exchange a structure-specific nuclease must resolve the D-loop. The RecU resolvase is unable to cleave D-

Genetic recombination functions

loops³⁵⁻³⁷, and the enzyme responsible for such activity remains unknown. The contribution of these RecA accessory proteins to homogamic CT has been studied in deep^{7,28,38}, but little is known about their role in the integration of heterogamic DNA. In *B. subtilis*, two exonucleases, the AddAB complex (counterpart of *E. coli* RecBCD) and RecJ are crucial to process double-strand breaks (DSBs) during DNA repair^{33,39}, but their putative role in the degradation of the displaced strand during CT remains to be documented.

DNA transfer occurs in ecologically cohesive communities. In this study we aimed to identify the recombination mechanisms and the proteins that contribute to heterogamic CT. We found that interspecies CT frequency is similar to *rec*⁺ cells in the *addAB*, *recO*, *recF*, *ruvAB* and *recU* context. Our results suggest that for interspecies CT only a subset of recombination proteins is required. RecA, RecX, RecJ, RecD2, DprA and RadA/Sms participate in both, homology-directed HR and homology-facilitated micro-homologous integration mechanisms that are active depending on the degree of SD in *B. subtilis* competent cells.

Results

Experimental design

The rationale to select *B. subtilis* competent cells to analyse how the recombination machinery contributes to heterogamic CT is summarised in Supplementary [Annex 1](#). Competence development is a stochastic process driven by the expression of ComK, and it occurs only in a small fraction (0.1 - 5%) of cells^{11,12}. Interspecies CT, at SD above 8%, is at the limit of detection in wild type cells¹⁴. To overcome such technical difficulty, Rok, which directly represses *comK* expression, was deleted⁴⁰. Inactivation of *rok* increases the subpopulation of cells that develop natural competence⁴¹, and thereby the sensitivity of the assay. The different *rec* mutations were mobilised into the *rok* strain (see [Table 1](#)). All the strains additionally lack Rok, but for simplicity we only state the *rec* gene that is mutated.

The *rpoB* gene encodes for the essential β subunit of the RNA polymerase. The 2997-bp donor DNAs have a mutation at codon 482 that confers rifampicin resistance (Rif^R, *rpoB*482) (Fig. [S1B](#)). The *rpoB*482 DNA from *B. subtilis* 168 with 1 mismatch (*Bsu* 168 *rpoB*482) was used for homogamic CT (*i.e.*, the recipient's own DNA, with just the Rif^R mutation). For heterogamic CT a fixed concentration of purified *rpoB*482 donor DNAs derived from the *B. subtilis* clade (2.47% SD and 8.35% SD), the *B. amyloliquefaciens* clade with 10.12% SD, the *B. licheniformis* clade (14.52% SD and 17% SD), the *B. thuringiensis* clade with 20.83% SD, or a far distant *Bacillus* with 22.74% SD (*B. smithii* DSM4512) (Supplemental [Annex 2](#), Fig. [S1A](#)). A further description of these DNAs is presented in Supplementary material [Annex 2](#). As revealed in Fig. [S1B](#), the mismatches in these natural homeologous DNAs are almost homogeneously distributed^{16,17}, although there are some short regions with higher SD (see below). To avoid fitness costs, the promoter of the *rpoB* gene is provided by the recipient strain, and the sequence identity at the protein level varied from 99% to 89% (Fig. [S1C](#)).

DprA, RecX, RadA/Sms, RecJ and RecD2 contribute to homogamic CT, but not RecF, RecO, RecU, RuvAB and AddAB

Except in $\Delta recD2$ cells, the number of Rif^R spontaneous mutant colonies, which appeared in the absence of *rpoB*482 DNA in the different *rec*⁻ strains was similar to the *wt* control (see Supplementary material [Annex 3](#)). To test how the different *rec*⁻ mutants affect intraspecies CT, competent cells were transformed with *Bsu* 168 *rpoB*482 DNA with direct selection for Rif^R ([Table 1](#), Fig. [1A](#)). The homogamic CT frequency with this donor DNA was similar to the one obtained with *Bsu* 168 *met*⁺ DNA, also with a single donor-recipient mismatch¹⁷. As previously reported, intraspecies CT was blocked in the $\Delta recA$ strain ([Table 1](#), Fig. [1A](#))^{16,42}.

Natural homogamic CT was barely affected in *recF15*, $\Delta recO$, $\Delta recU$, $\Delta ruvAB$, $\Delta addAB$ cells (Table 1, Fig. 1A-B). Their CT frequencies were similar to results previously observed for these recombination mutants in the *rok*⁺ background^{28,43}. However, the frequency of intraspecies CT was significantly reduced in competent $\Delta recJ$ (by ~80-fold), $\Delta recX$ (~500-fold), $\Delta recD2$ (~950-fold), $\Delta radA$ (~1600-fold) and $\Delta dprA$ (~24,000-fold) cells (Table 1, Fig. 1C-D). Except in $\Delta recJ$, intraspecies CT efficiency in these mutants was lower in Δrok than in *rok*⁺^{26,34}. The deleterious effect of additionally mutate *rok* in these backgrounds will be reported elsewhere.

Interspecies CT requires DprA, RecX, RadA/Sms, RecD2 and RecJ

To gain insight into the contribution of these recombination proteins to interspecies CT, we used *rpoB482* DNAs with different degree of SD (Supplementary Annex 2). The frequency of interspecies CT decreased logarithmically with increased SD up to ~15% in competent *recF15*, $\Delta recO$, $\Delta recU$, $\Delta ruvAB$ or $\Delta addAB$ cells (Fig. 1A-B). When SD was further increased, beyond 15% and up to ~23% SD, the interspecies CT frequencies varied <3-fold in *recF15*, $\Delta recO$, $\Delta recU$, *ruvAB* and *addAB* (Fig. 1A-B), as in the *rec*⁺ control (Fig. 1A)¹⁷, suggesting that these functions do not limit genetic recombination in otherwise *rec*⁺ cells.

A different outcome was observed in $\Delta recJ$, $\Delta recX$, $\Delta radA$, $\Delta recD2$ and $\Delta dprA$. The CT frequency was similar to the frequency of spontaneous Rif^R mutations beyond ~8% SD in $\Delta dprA$, ~10% in $\Delta recX$ and $\Delta radA$, and ~15% in $\Delta recJ$ and $\Delta recD2$ competent cells, (Fig. 1C-D). We believe that this strong defect was not due to an impairment in fitness cost. The colony size observed after overnight growth at 37°C under selective pressure was similar in all cases. Furthermore, the sequence analysis of the chimeric *rpoB482* genes revealed a RpoB482 protein only bearing the Rif^R mutation (see below).

The interspecies barrier is altered in $\Delta recJ$, $\Delta recX$, $\Delta recD2$, $\Delta radA$ and $\Delta dprA$

To simplify the analysis of the strains that are significantly impaired in heterogamic CT, we gave a value of 1 to the transformation rate obtained by the given mutant in the intraspecies CT assay, and plotted the data relative to this value. Hence only the impact of SD is evaluated (Fig. 2A-B). Different outcomes were observed. First, in competent $\Delta recX$, $\Delta radA$ and $\Delta recJ$ cells interspecies CT decreased logarithmically with increased SD, but only up to ~10% ($\Delta recJ$, ~14%) SD, to reach a plateau at higher divergence (Fig. 2A). Second, competent *recD2* reached a plateau already at 8% SD (Fig. 2B). Third, upon inactivation of *dprA*, the decline in the rate of

recombination with SD was only observed at ~2% SD, to reach a plateau at higher divergence (Fig. 2B).

RecO, RecF, RecU, RuvAB and AddAB are dispensable in heterogamic transformation

The above results suggest that RecO, RecF, RecU, RuvAB and AddAB play no apparent roles in interspecies CT, but their contribution to the integration length is unknown. To analyse this, we sequenced the *rpoB482* gene from Rif^R clones and calculated the mean integration length.

The maximal integration length that can be detected with the donor DNA of ~2% SD (*Bsu* W23 DNA) is 2628-bp, because the first mismatch is located at position 350 and the last at position 2978. Previous analysis showed that when the *rec*⁺ strain was transformed with this donor DNA, the mean integration length was close to this maximal integration length, around ~2300-bp^{16,17}. Similar results were obtained with the donor DNA of ~2% SD in the $\Delta recO$, $\Delta recU$, $\Delta ruvAB$ and $\Delta addAB$ backgrounds (Fig 3A-B).

Up to 15% divergence range, RecA is believed to integrate the DNA by a homology-directed HR mechanism, initiating recombination in a MEPS^{16,22}. A comparison of the nucleotide sequence of donor with recipient DNA revealed the presence of 22 MEPS at or above 25-nt in donor DNA with ~8% SD, and 21 at ~10% SD. In both donor DNAs there is a long stretch of ~200-nt of sequence identity upstream of the *rpoB482* mutation, and several regions with a MEPS longer than 54-nt downstream of the *rpoB482* mutation. They could define the left and right recombination endpoints, being the region in-between integrated independently of its SD, as it occurs in the insertion of heterologous DNA by two-step HR at the flanks (see Introduction). However, integrated fragments of ~1600-nt (*i.e.*, recombination endpoints at the 157-nt [at position 1-157] and the 81-nt MEPS [at position 1509-1589]) was not observed in *rec*⁺ transformations with ~8% SD donor DNA, and the same was observed with 10% SD (Fig 3A). This result suggests that the two-step deletion/insertion CT may not take place with interspecies DNA, probably because it requires two longer flanking homologous regions (see Introduction), or the sequence in-between plays a relevant role.

The analysis of 10-20 Rif^R clones obtained in the transformation of competent *recF15*, $\Delta recO$, $\Delta recU$, $\Delta ruvAB$ and $\Delta addAB$ cells with donor DNA of ~8% SD revealed that the mean integration length was as in *rec*⁺, 700-900-nt, except in $\Delta ruvAB$, which was ~490-nt (*i.e.*, ~2 times less) (Fig. 3A-B, and Table S1). A sequence analysis of the integrated region revealed eight MEPS (four upstream [50-, 35-, 38- and 36-nt] and four downstream [35-, 81-, 41- and 33-nt]) of the *rpoB482* mutation. One recombination endpoint was usually at one of these MEPS, but

Genetic recombination functions

the other endpoint was usually at other region, with a size below MEPS (Table S1). Several hypotheses can explain these results: i) the MEPS used *in vivo* are shorter, ii) RecA is insensitive to 1-nt mismatch every ~8-nt, but not to higher mismatches. Indeed, if 1-nt mismatch is allowed at the short endpoints, longer MEPS regions can be predicted in all cases, and iii) integration starts at one MEPS and proceeds uni- or bidirectionally until RecA finds a barrier (e.g., 2-nt mismatch every ~8-nt). We noticed that the DNA inserts are usually followed by higher local SD (~20% to ~28%) in the next 25-nt interval, suggesting that recombination ended there because RecA found a heterologous block.

In *recF15*, $\Delta recO$, $\Delta recU$, $\Delta ruvAB$, $\Delta addAB$ and *rec*⁺ cells, when SD is ~10%, the mean integration length was ~300-nt (Fig. 3A-B). In these transformants, the nucleotide sequence changes do not alter the RpoB482 protein sequence (see Fig. S1C). Within a 600-nt interval centred at the *rpoB482* mutation, there are only three sequences at or above MEPS (see Fig. 4A). Analysis of 10 different Rif^R clones in *rec*⁺ showed that in ~70% of the cases, the 3'-endpoint was at the 54-nt MEPS (at position 1536-1590, Fig. 4A). Recombination in this region did not extend beyond, probably because it is followed by a local ~24% SD in the next 25-nt that could act as a strong heterologous barrier. The other endpoints were located in regions below MEPS which are usually followed by barriers with a higher SD (e.g., endpoint at position 1371 is preceded by 25-nt with ~24% SD, at position 1251 by ~24% SD). No patched sequences were observed, discarding that in some transformants multiple recombination events had occurred at different loci. As in *rec*⁺, the recombination endpoints in the *recF15*, $\Delta recO$, $\Delta recU$, $\Delta ruvAB$ and $\Delta addAB$ transformants did not always coincide with the longest MEPS present (Table S1). In these mutants one endpoint was usually at a region where a MEPS was located, whereas the other endpoint was less specific, and often below MEPS. All these results suggest that a MEPS is necessary to initiate strand invasion, but a SD barrier may halt RecA-mediated uni- or bidirectional DNA strand exchange.

Nucleotide sequence analysis of 15-30 Rif^R clones obtained with ~15% SD in $\Delta ruvAB$ and *rec*⁺ cells showed that all were genuine transformants, but these values dropped to ~40-50% in $\Delta recU$, $\Delta recO$, $\Delta addAB$, and *recF15*. The mean integration length was between 400- to 130-nt in *recF15*, $\Delta recO$, $\Delta recU$, $\Delta ruvAB$ and $\Delta addAB$, as well as in the *rec*⁺ control (Table S1 and Fig. 4B). This donor DNA has a MEPS of 104-nt (position 54-157) upstream of the *rpoB482* mutation, and a 38-nt MEPS (position 2295-2332) downstream, but ~2100-nt inserts were not observed, confirming that the region between the MEPS are relevant. Potential MEPS are not present in the integrated region (Fig. 4B). The sequence analysis of *rec*⁺ transformants revealed

a great heterogenicity in the recombination endpoints, as well as in the length of the MEPS. In all cases the MEPS used were short: between 5- to 20-nt without mismatches (Fig. 4B and Table S1). However, if 1-mismatch is allowed a longer MEPS is detected (*e.g.*, the endpoint at position 1329 increases from 5- to 26-nt) (Fig. 4B).

Beyond 15% SD the transformation efficiency and the frequency of spontaneous mutation almost overlap, although there is a ~3-fold difference, and it is higher than in a $\Delta recA$ mutant, suggesting that still this is a RecA-dependent integration event (Fig. 1). Four regions at or above MEPS (74-, 56-, 28- and 26-nt) upstream and one downstream (26-nt) of the *rpoB482* mutation are present in the 17% SD donor, but they were not used to integrate the region in-between.

Nucleotide sequence analyses of Rif^R *rec*⁺ clones obtained with ~17% SD showed that only ~37% were genuine transformants, *i.e.*, they had incorporated two or more nucleotides of donor DNA. Similarly, 20-30% of the Rif^R clones obtained in $\Delta recO$, $\Delta addAB$, $\Delta recU$, $\Delta ruvAB$ and *recF15* were genuine transformants (Fig. 3A-B), and the mean integration length was also 4- to 8-nt, or ~5-fold below MEPS (Table S1). The observed low efficiency of micro-homologous integration cannot be attributed to a defect in the resulting RpoB482 protein, because at ~17% SD, except the amino acid change that confers Rif^R, there is no mutation in the 50 residues intervals up and downstream the Rif^R change (Fig. S1C). The close inspection in donor DNA with 17% SD of the region surrounding the *rpoB482* mutation showed that this mutation is embedded in a region with strong SD with recipient DNA. The 25-nt region upstream of the *rpoB482* mutation has 8 mismatches (*i.e.*, 32% SD) and the one downstream 10 mismatches (40% SD) that could act as heterologous barriers.

At ~21% SD there are two sites at or above MEPS (42-nt [at position 90-131] and 27-nt [at position 2781-2807]), one on each side of the *rpoB482* mutation, but they were not used. After sequencing of transformants obtained in *recF15*, $\Delta recO$, $\Delta recU$, $\Delta ruvAB$, $\Delta addAB$ or *rec*⁺, we found that again just few nucleotides from the donor had been integrated, and that the fraction of genuine transformants was ~20% (Fig 3A-B). Here, the analysis showed that the *rpoB482* mutation is also surrounded by higher SD: 8 mismatches in 25-nt upstream (*i.e.*, ~32% SD) and the one downstream 6 mismatches (*i.e.*, ~24% SD). Finally, at ~23% SD, only one MEPS exists (at position 801-826). The proportion of genuine Rif^R transformants accounts to only ~6% of the sequenced clones in the *rec*⁺ strain. The mean integration length in *rec*⁺ transformants was 4- to 8-nt (Fig. 3A-B). Similar results were obtained in *recF15*, $\Delta recO$, $\Delta recU$, $\Delta ruvAB$ and $\Delta addAB$.

RecD2, RecJ, RecX, RadA/Sms and DprA are crucial for interspecies CT

Genetic recombination functions

Nucleotide sequence analyses of 10-30 Rif^R clones obtained in the $\Delta radA$, $\Delta recJ$, $\Delta recX$, $\Delta recD2$ or $\Delta dprA$ context revealed that when divergence was low (~2%), the mean integration length of the *rpoB482* DNA was similar to the *rec*⁺ control (Fig 3C-D). The proportion of genuine transformants was different in these mutants. At ~8% SD, all sequenced clones obtained in $\Delta radA$ were genuine transformants, but this number was reduced to ~50 % in $\Delta dprA$, $\Delta recJ$ and $\Delta recX$, and even lowered to ~25% in $\Delta recD2$. Representatives of the genuine Rif^R clones are documented in Table S2.

The efficiency of interspecies transformation at ~10% SD was strongly reduced in these mutants, to levels similar to the spontaneous mutation rate (Fig. 1). Only ~25% of the Rif^R clones obtained in $\Delta recX$ were genuine transformants, and these values lowered to ~15% in $\Delta recD2$, and ~5% in $\Delta recJ$ and $\Delta radA$ cells. Furthermore, in the $\Delta dprA$ strain all 25 sequenced Rif^R clones were spontaneous mutants (Table S2). The mapping of recombination endpoints in the $\Delta recJ$, $\Delta recD2$, $\Delta recX$ and $\Delta radA$ transformants showed that, as in the *rec*⁺ transformants, the 3'-endpoint of recombination was located at the 54-nt MEPS, except in some $\Delta recX$ transformants. The 5'-endpoint of recombination was usually located at a larger MEPS when compared with the *rec*⁺ control (Tables S1 and S2).

At ~15% SD, all sequenced Rif^R clones obtained were spontaneous mutants in the $\Delta recX$, $\Delta radA$, $\Delta recJ$ and $\Delta dprA$ strains. In $\Delta recD2$, only one of the sequenced Rif^R clones (1/19) was a genuine transformant with a mean integration length >120-nt (Table S2). At ~17% SD, all sequenced Rif^R clones were spontaneous mutants in the $\Delta recX$, $\Delta radA$, $\Delta recJ$, $\Delta dprA$, and $\Delta recD2$ backgrounds. These results show that these functions are all needed for homology-directed integration when SD is larger than 10%, and also for micro-homologous integration.

Discussion

From the data presented in this work, it can be inferred that recombination proteins differently facilitate adaptation and genetic diversity, and that two different recombination mechanisms occur depending on the SD of the interspecies DNA. Up to 15% SD, homology-directed HR accounts for the integration of divergent sequences longer than 130-nt. Beyond this, integration of just few nucleotides at micro-homologous segments is observed.

Inactivation of the main end-resection complex (*addAB*), a positive mediator (*recO*), a positive (*recF*) or a negative (*recU*) modulator or enzymes essential for Holliday junction processing and cleavage (*ruvAB*, *recU*) barely affects heterogamic CT (Fig. 1A-B), although they are essential for DNA DSB repair ^{39,43,44}. The interpretation of such result may not be so simple, because cells lacking both RecF and AddAB are blocked even in homogamic CT ⁴³, suggesting that they are backup functions during CT (*e.g.*, RecF is essential in the *recX* context) ²⁵. We have observed that a specific subset of HR proteins (RecA, DprA, RecX, RadA/Sms, RecD2 and RecJ) contributes to acquire homeologous DNA. Except DprA which is competence specific and present in all transformable bacteria, and even in non-transformable ones ¹⁰, the recombination proteins identified here also participate, together with another subset of HR proteins, in the accurate repair of lethal DSBs and in the restart of stalled replication forks. In some distantly related competent cells (*e.g.*, Proteobacteria phylum) RecD2 is absent, and RadA/Sms is replaced by ComM ⁴⁵. They differently participate in the recombination mechanisms that may be active during the acquisition of homeologous DNA. In the absence of RecA CT is blocked ¹⁶. Interspecies homology-directed CT was blocked with DNA of the same clade (up to ~8% SD) in $\Delta dprA$, beyond ~10% SD in the $\Delta recX$, $\Delta recJ$ and $\Delta radA$ backgrounds, and beyond ~15% SD in $\Delta recD2$ cells (Fig. 3C-D), showing an essential role for these proteins in the acquisition of interspecies DNA.

We can envision that during homology-directed CT, RecA nucleates in the incoming ssDNA coated by SsbA and SsbB, with the help of the DprA-SsbA two-component mediator. Then, a dynamic RecA nucleoprotein filament, with the contribution of DprA-SsbA and RecX, identifies an identical sequence in the recipient genome. Once a homologous region is found RecA promotes DNA strand invasion to produce a metastable heteroduplex DNA ^{23,27}. RecA at this D-loop interacts with and loads the branch-migration translocase RadA/Sms as documented in Firmicutes ^{34,46}. RadA/Sms, in concert with RecA, facilitates bidirectional branch migration until a region with a SD >20% is found. This is consistent with the *in vitro* observation that homology-directed RecA-mediated strand-exchange halted at DNA patches >16% SD ¹⁶.

Genetic recombination functions

The contribution of RecD2 helicase and RecJ exonuclease to homology-directed recombination during interspecies CT is poorly understood. Recently, it was suggested that RecD2 contributes to branch migrate the heteroduplex DNA in a relaxed molecule, perhaps upon cleavage of the displaced strand²⁸. As RecJ of other naturally competent cells⁴⁷, *B. subtilis* RecJ possesses an extra C-terminal domain, absent in *E. coli* RecJ, critical for protein-protein interaction (e.g., SsbA). Perhaps in concert with the RecD2 helicase, it might degrade the displaced strand and the non-paired tails. Finally, the ends are sealed, leading to the acquisition of homeologous DNA >130-nt.

It was remarkable the heterogeneity of the MEPS at recombination endpoints in homology-directed HR. *In vitro* assays with *E. coli* RecA show that identity search relies on probing tracts of 8-nt homology, based on the transient interactions between the stretched ssDNA within the filament and bases in a locally melted and stretched DNA duplex^{48,49}. It has been shown that RecA evolved to tolerate 1-nt mismatch every ~8-nt region *in vitro*, albeit DNA strand exchange with a short region of 16% SD is delayed^{16,50}. We propose that a sum of delays might compromise DNA strand exchange and determines the length of DNA integrated.

Beyond 15% sequence, integration of <10-nt segments was observed in $\Delta addAB$, $recF$, $\Delta recO$, $\Delta recU$ and $\Delta ruvAB$ (Fig. 3A-B), but it was not detected in $\Delta recA$, $\Delta dprA$, $\Delta recJ$, $\Delta recX$, $\Delta radA$ and $\Delta recD2$ cells (Fig. 3C-D). These results showed that a similar set of recombination functions are also required for this short integration, which occurs up to 23% SD, although with low efficiency. Here we propose a homology-facilitated micro-homologous integration mechanism: As above, a RecA dynamic filament, with the help of DprA-SsbA and RecX, searches for and identifies a MEPS on the *rpoB482* DNA, which is used in this case as an anchor region, to produce a metastable D-loop intermediate, in concert with RadA/Sms. Once the DNA is anchored at MEPS, DprA could mediate the annealing of short stretches of homeologous DNA (3- to 8-nt) around the *rpoB482* mutation²³. Then, the donor ssDNA loop between the anchored region and the micro-homologous paired segment has to be deleted, perhaps by RecJ in concert with RecD2. Finally, the ends of the integrated segment are sealed and rapidly expressed⁴⁵. This mechanism differs from HFIR observed in replicating competent *S. pneumoniae* and *A. baylyi* cells (see Introduction)^{20,21}. In these competent bacteria, inactivation of *recBCD* or *recJ* significantly increased HFIR^{19,21}. In contrast, in non-replicating competent *B. subtilis* cells integration of thousands of nucleotides of the heterologous DNA with the subsequent deletion of the recipient DNA was not observed, and inactivation of *recJ* blocked homology-facilitated micro-homologous integration. Competent *A. baylyi* cells can also integrate short ssDNA (20-nt) in the recipient genome by another mechanism, which occurs with extremely low efficiency.

Genetic recombination functions

This mechanism requires active DNA replication, inactivation of *recJ* and is independent of RecA⁵¹. It is likely that competent *B. subtilis* cells may use homology-facilitated microhomologous integration to restore genes inactivated by mutations and thereby prevent the irreversible deterioration of genomes (Muller's ratchet)⁴. At ~23% SD interspecies CT frequency was similar to spontaneous mutations, suggesting that beyond ~23% SD microhomologous integration might be inefficient, probably because the unique MEPS present is too short to serve as a stable anchor region¹⁹. Similarly, competent *H. pylori* cells cannot be transformed by *Campylobacter jejuni* DNA with ~24% SD⁵².

Can the above observations be extrapolated to interspecies Hfr chromosomal conjugation? The frequency of interspecies Hfr chromosomal conjugation also decreases log-linearly with increased SD up to ~16% SD. Inactivation of *mutSL* alleviates interspecies Hfr conjugation by ~1000-fold, and deletion of *recBCD* or *ruvAB* reduced interspecies Hfr conjugation, but inactivation of *recJ* increases it^{2,3,53,54}, suggesting that interspecies Hfr chromosomal conjugation uses the repair-by-recombination mechanism. In contrast, inactivation of *mutSL* marginally prevents interspecies CT with up to 15% SD in Firmicutes^{15,16,55,56} and inactivation of *recJ*, but not *addAB* or *ruvAB*, inhibits interspecies CT. All these results suggest that bacteria have evolved different genetic recombination mechanisms devoted to interspecies genetic exchange to generate diversity. Bacterial *recA*, *radA* and *recX* genes, which play crucial roles in interspecies CT, perhaps contributed in their transfer from mitochondria or chloroplasts to the nucleus of land plants, green algae and moss⁵⁷, although the evolutionary force and molecular functions that contributed to the transfer of these genes well beyond the species boundaries is poorly understood.

Materials and methods

Bacterial strains and donor DNAs

The parental strain was *B. subtilis* BG1359. The *rec* mutations listed in Table 1 were introduced by SPP1 transduction ⁵⁸.

The *rpoB* gene from different species, which encodes for the β -subunit of RNA polymerase, was used as donor DNA, and the *rpoB*482 mutation, which renders cells Rif^R, was introduced into all the donor DNAs (Annex 2, supplementary material). Plasmid DNA was prepared by Qiagen extraction and extensive dialysis in Tris-EDTA buffer ¹⁷.

Transformation assays

Natural competence was induced as described ⁴². Competent cells were incubated with 0.1 $\mu\text{g}\cdot\text{ml}^{-1}$ of the indicated *rpoB*482 donor DNA (30 min, 37°C), and then plated on Rif (8 $\mu\text{g}\cdot\text{ml}^{-1}$) containing LB-agar plates. A control was performed in which competent cells were treated equally, but with no donor DNA to score the appearance of spontaneous Rif^R mutants. These values (spontaneous Rif^R mutants) were extracted to the number of transformants. Transformation frequency was calculated as the number of Rif^R transformants per colony-forming-unit (CFU).

Mapping of integration endpoints

The integration endpoint is defined by the end of the donor sequence followed by the sequence of the recipient. To map integration endpoints, the *rpoB* gene from the Rif^R transformants was amplified by PCR, and its nucleotide sequence compared with the one of recipient and donor strains. The presence or the absence of the mismatches between the donor and the recipient DNA were used to determine the MEPS. Endpoints are defined as described ⁵⁹, and integration length is calculated as the distance between endpoints ⁵⁹.

Acknowledgements ES and CR thank the Ministerio de Ciencia e Innovación [Agencia Estatal de Investigación] (MCIU[AEI]), BES-2013-063433 to E.S., and BES-2017-080504 to C.R. for the fellowships. This work was partially supported by MCIU/AEI/FEDER, EU PGC2018-097054-B-I00 to J.C.A and S.A. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions S.A. and J.C.A. conceived the project and designed the experiments, E.S., C.R. performed the experiments, E.S., C.R. S.A. and J.C.A. evaluated the results, S.A. and J.C.A. wrote the manuscript.

Compliance with ethical standards

Conflict of interest. The authors declare that they have no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data or in the decision to publish the results.

References

- 1 Gogarten, J. P., Doolittle, W. F. & Lawrence, J. G. Prokaryotic evolution in light of gene transfer. *Mol Biol Evol* **19**, 2226-2238, doi:10.1093/oxfordjournals.molbev.a004046 (2002).
- 2 Fraser, C., Hanage, W. P. & Spratt, B. G. Recombination and the nature of bacterial speciation. *Science* **315**, 476-480, doi:10.1126/science.1127573 (2007).
- 3 Matic, I., Taddei, F. & Radman, M. Genetic barriers among bacteria. *Trends Microbiol* **4**, 69-72, doi:10.1016/0966-842X(96)81514-9 (1996).
- 4 Takeuchi, N., Kaneko, K. & Koonin, E. V. Horizontal gene transfer can rescue prokaryotes from Muller's ratchet: benefit of DNA from dead cells and population subdivision. *G3 (Bethesda)* **4**, 325-339, doi:10.1534/g3.113.009845 (2014).
- 5 Hanage, W. P. Not So Simple After All: Bacteria, Their Population Genetics, and Recombination. *Cold Spring Harb Perspect Biol* **8**, doi:10.1101/cshperspect.a018069 (2016).
- 6 Trautner, T. A. & Spatz, H. C. Transfection in *B. subtilis*. *Curr Top Microbiol Immunol* **62**, 61-88 (1973).
- 7 Kidane, D., Ayora, S., Sweasy, J. B., Graumann, P. L. & Alonso, J. C. The cell pole: the site of cross talk between the DNA uptake and genetic recombination machinery. *Crit Rev Biochem Mol Biol* **47**, 531-555, doi:10.3109/10409238.2012.729562 (2012).
- 8 Oliveira, P. H., Touchon, M. & Rocha, E. P. Regulation of genetic flux between bacteria by restriction-modification systems. *Proc Natl Acad Sci U S A* **113**, 5658-5663, doi:10.1073/pnas.1603257113 (2016).
- 9 Brito, P. H. *et al.* Genetic Competence Drives Genome Diversity in *Bacillus subtilis*. *Genome Biol Evol* **10**, 108-124, doi:10.1093/gbe/evx270 (2018).
- 10 Johnston, C., Martin, B., Fichant, G., Polard, P. & Claverys, J. P. Bacterial transformation: distribution, shared mechanisms and divergent control. *Nat Rev Microbiol* **12**, 181-196, doi:10.1038/nrmicro3199 (2014).
- 11 Dubnau, D. & Blokesch, M. Mechanisms of DNA Uptake by Naturally Competent Bacteria. *Annu Rev Genet* **53**, 217-237, doi:10.1146/annurev-genet-112618-043641 (2019).
- 12 Maier, B. Competence and Transformation in *Bacillus subtilis*. *Curr Issues Mol Biol* **37**, 57-76, doi:10.21775/cimb.037.057 (2020).

Genetic recombination functions

- 1 13 Spratt, B. G., Bowler, L. D., Zhang, Q. Y., Zhou, J. & Smith, J. M. Role of interspecies
2 transfer of chromosomal genes in the evolution of penicillin resistance in pathogenic
3 and commensal *Neisseria* species. *J Mol Evol* **34**, 115-125, doi:10.1007/bf00182388
4 (1992).
- 5 14 Zawadzki, P., Roberts, M. S. & Cohan, F. M. The log-linear relationship between sexual
6 isolation and sequence divergence in *Bacillus* transformation is robust. *Genetics* **140**,
7 917-932 (1995).
- 8 15 Humbert, O., Prudhomme, M., Hakenbeck, R., Dowson, C. G. & Claverys, J. P.
9 Homeologous recombination and mismatch repair during transformation in
10 *Streptococcus pneumoniae*: saturation of the Hex mismatch repair system. *Proc Natl*
11 *Acad Sci U S A* **92**, 9052-9056 (1995).
- 12 16 Carrasco, B., Serrano, E., Martín-González, A., Moreno-Herrero, F. & Alonso, J. C. *Bacillus*
13 *subtilis* MutS modulates RecA-mediated DNA strand exchange between divergent DNA
14 sequences. *Front Microbiol* **10**, 237, doi:doi: 10.3389/fmicb.2019.00237 (2019).
- 15 17 Carrasco, B., Serrano, E., Sanchez, H., Wyman, C. & Alonso, J. C. Chromosomal
16 transformation in *Bacillus subtilis* is a non-polar recombination reaction. *Nucleic Acids*
17 *Res* **44**, 2754-2768, doi:10.1093/nar/gkv1546 (2016).
- 18 18 Simpson, D. J., Dawson, L. F., Fry, J. C., Rogers, H. J. & Day, M. J. Influence of flanking
19 homology and insert size on the transformation frequency of *Acinetobacter baylyi*
20 BD413. *Environ Biosafety Res* **6**, 55-69, doi:10.1051/ebr:2007027 (2007).
- 21 19 Brigulla, M. & Wackernagel, W. Molecular aspects of gene transfer and foreign DNA
22 acquisition in prokaryotes with regard to safety issues. *Appl Microbiol Biotechnol* **86**,
23 1027-1041, doi:10.1007/s00253-010-2489-3 (2010).
- 24 20 Prudhomme, M., Libante, V. & Claverys, J. P. Homologous recombination at the border:
25 insertion-deletions and the trapping of foreign DNA in *Streptococcus pneumoniae*. *Proc*
26 *Natl Acad Sci U S A* **99**, 2100-2105, doi:10.1073/pnas.032262999 (2002).
- 27 21 de Vries, J. & Wackernagel, W. Integration of foreign DNA during natural transformation
28 of *Acinetobacter* sp. by homology-facilitated illegitimate recombination. *Proc Natl Acad*
29 *Sci U S A* **99**, 2094-2099, doi:10.1073/pnas.042263399 (2002).
- 30 22 Carrasco, B., Yadav, T., Serrano, E. & Alonso, J. C. *Bacillus subtilis* RecO and SsbA are
31 crucial for RecA-mediated recombinational DNA repair. *Nucleic Acids Res* **43**, 5984-
32 5997, doi:10.1093/nar/gkv545 (2015).
- 33 23 Yadav, T., Carrasco, B., Serrano, E. & Alonso, J. C. Roles of *Bacillus subtilis* DprA and SsbA
34 in RecA-mediated genetic recombination. *J Biol Chem* **289**, 27640-27652,
35 doi:10.1074/jbc.M114.577924 (2014).
- 36 24 Carrasco, B., Ayora, S., Lurz, R. & Alonso, J. C. *Bacillus subtilis* RecU Holliday-junction
37 resolvase modulates RecA activities. *Nucleic Acids Res* **33**, 3942-3952, doi:
38 10.1093/nar/gki713 (2005).
- 39 25 Cardenas, P. P. *et al.* RecX facilitates homologous recombination by modulating RecA
40 activities. *PLoS Genet* **8**, e1003126, doi:10.1371/journal.pgen.1003126 (2012).
- 41 26 Serrano, E., Carrasco, B., Gilmore, J. L., Takeyasu, K. & Alonso, J. C. RecA Regulation by
42 RecU and DprA During *Bacillus subtilis* Natural Plasmid Transformation. *Front Microbiol*
43 **9**, 1514, doi:10.3389/fmicb.2018.01514 (2018).
- 44 27 Le, S. *et al.* *Bacillus subtilis* RecA with DprA-SsbA antagonizes RecX function during
45 natural transformation. *Nucleic Acids Res* **45**, 8873-8885, doi:10.1093/nar/gkx583
46 (2017).

Genetic recombination functions

- 1 28 Serrano, E., Ramos, C., Ayora, S. & Alonso, J. C. Viral SPP1 DNA is infectious in naturally
2 competent *Bacillus subtilis* cells: inter- and intramolecular recombination pathways.
3 *Environ Microbiol*, doi:10.1111/1462-2920.14908 (2020).
- 4 29 Hsieh, P., Camerini-Otero, C. S. & Camerini-Otero, R. D. The synapsis event in the
5 homologous pairing of DNAs: RecA recognizes and pairs less than one helical repeat of
6 DNA. *Proc Natl Acad Sci U S A* **89**, 6492-6496 (1992).
- 7 30 Yang, D., Boyer, B., Prevost, C., Danilowicz, C. & Prentiss, M. Integrating multi-scale data
8 on homologous recombination into a new recognition mechanism based on simulations
9 of the RecA-ssDNA/dsDNA structure. *Nucleic Acids Res* **43**, 10251-10263,
10 doi:10.1093/nar/gkv883 (2015).
- 11 31 Watt, V. M., Ingles, C. J., Urdea, M. S. & Rutter, W. J. Homology requirements for
12 recombination in *Escherichia coli*. *Proc Natl Acad Sci U S A* **82**, 4768-4772 (1985).
- 13 32 Majewski, J. & Cohan, F. M. DNA sequence similarity requirements for interspecific
14 recombination in *Bacillus*. *Genetics* **153**, 1525-1533 (1999).
- 15 33 Kowalczykowski, S. C. An Overview of the Molecular Mechanisms of Recombinational
16 DNA Repair. *Cold Spring Harb Perspect Biol* **7**, doi:10.1101/cshperspect.a016410 (2015).
- 17 34 Torres, R., Serrano, E. & Alonso, J. C. *Bacillus subtilis* RecA interacts with and loads
18 RadA/Sms to unwind recombination intermediates during natural chromosomal
19 transformation. *Nucleic Acids Res* **47**, 9198-9215, doi:10.1093/nar/gkz647 (2019).
- 20 35 Ayora, S., Carrasco, B., Doncel-Perez, E., Lurz, R. & Alonso, J. C. *Bacillus subtilis* RecU
21 protein cleaves Holliday junctions and anneals single-stranded DNA. *Proc Natl Acad Sci*
22 *U S A* **101**, 452-457, doi:10.1073/pnas.2533829100 (2004).
- 23 36 McGregor, N. *et al.* The structure of *Bacillus subtilis* RecU Holliday junction resolvase
24 and its role in substrate selection and sequence-specific cleavage. *Structure* **13**, 1341-
25 1351, doi:10.1016/j.str.2005.05.011 (2005).
- 26 37 Cañas, C. *et al.* Interaction of branch migration translocases with the Holliday junction-
27 resolving enzyme and their implications in Holliday junction resolution. *J Biol Chem* **289**,
28 17634-17646, doi:10.1074/jbc.M114.552794 (2014).
- 29 38 Kidane, D. *et al.* Evidence for different pathways during horizontal gene transfer in
30 competent *Bacillus subtilis* cells. *PLoS Genet* **5**, e1000630,
31 doi:10.1371/journal.pgen.1000630 (2009).
- 32 39 Ayora, S. *et al.* Double-strand break repair in bacteria: a view from *Bacillus subtilis*. *FEMS*
33 *Microbiol Rev* **35**, 1055-1081, doi:10.1111/j.1574-6976.2011.00272.x (2011).
- 34 40 Hoa, T. T., Tortosa, P., Albano, M. & Dubnau, D. Rok (YkuW) regulates genetic
35 competence in *Bacillus subtilis* by directly repressing comK. *Mol Microbiol* **43**, 15-26
36 (2002).
- 37 41 Maamar, H. & Dubnau, D. Bistability in the *Bacillus subtilis* K-state (competence) system
38 requires a positive feedback loop. *Mol Microbiol* **56**, 615-624, doi:10.1111/j.1365-
39 2958.2005.04592.x (2005).
- 40 42 Alonso, J. C., Tailor, R. H. & Luder, G. Characterization of recombination-deficient
41 mutants of *Bacillus subtilis*. *J Bacteriol* **170**, 3001-3007 (1988).
- 42 43 Alonso, J. C., Stiege, A. C. & Luder, G. Genetic recombination in *Bacillus subtilis* 168:
43 effect of recN, recF, recH and addAB mutations on DNA repair and recombination. *Mol*
44 *Gen Genet* **239**, 129-136 (1993).
- 45 44 Sanchez, H. *et al.* The RuvAB branch migration translocase and RecU Holliday junction
46 resolvase are required for double-stranded DNA break repair in *Bacillus subtilis*.
47 *Genetics* **171**, 873-883, doi:10.1534/genetics.105.045906 (2005).

Genetic recombination functions

- 45 Dalia, A. B. & Dalia, T. N. Spatiotemporal Analysis of DNA Integration during Natural Transformation Reveals a Mode of Nongenetic Inheritance in Bacteria. *Cell* **179**, 1499-1511 e1410, doi:10.1016/j.cell.2019.11.021 (2019).
- 46 Marie, L. *et al.* Bacterial RadA is a DnaB-type helicase interacting with RecA to promote bidirectional D-loop extension. *Nat Commun* **8**, 15638, doi:10.1038/ncomms15638 (2017).
- 47 Cheng, K. *et al.* Structural basis for DNA 5'-end resection by RecJ. *Elife* **5**, e14294, doi:10.7554/eLife.14294 (2016).
- 48 Folta-Stogniew, E., O'Malley, S., Gupta, R., Anderson, K. S. & Radding, C. M. Exchange of DNA base pairs that coincides with recognition of homology promoted by E. coli RecA protein. *Mol Cell* **15**, 965-975, doi:10.1016/j.molcel.2004.08.017 (2004).
- 49 Qi, Z. *et al.* DNA sequence alignment by microhomology sampling during homologous recombination. *Cell* **160**, 856-869, doi:10.1016/j.cell.2015.01.029 (2015).
- 50 Bazemore, L. R., Folta-Stogniew, E., Takahashi, M. & Radding, C. M. RecA tests homology at both pairing and strand exchange. *Proc Natl Acad Sci U S A* **94**, 11863-11868, doi:10.1073/pnas.94.22.11863 (1997).
- 51 Overballe-Petersen, S. *et al.* Bacterial natural transformation by highly fragmented and damaged DNA. *Proc Natl Acad Sci U S A* **110**, 19860-19865, doi:10.1073/pnas.1315278110 (2013).
- 52 Levine, S. M. *et al.* Plastic cells and populations: DNA substrate characteristics in *Helicobacter pylori* transformation define a flexible but conservative system for genomic variation. *FASEB J* **21**, 3458-3467, doi:10.1096/fj.07-8501com (2007).
- 53 Rayssiguier, C., Thaler, D. S. & Radman, M. The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch-repair mutants. *Nature* **342**, 396-401, doi:10.1038/342396a0 (1989).
- 54 Matic, I., Rayssiguier, C. & Radman, M. Interspecies gene exchange in bacteria: the role of SOS and mismatch repair systems in evolution of species. *Cell* **80**, 507-515 (1995).
- 55 Majewski, J. & Cohan, F. M. The effect of mismatch repair and heteroduplex formation on sexual isolation in *Bacillus*. *Genetics* **148**, 13-18 (1998).
- 56 Majewski, J., Zawadzki, P., Pickerill, P., Cohan, F. M. & Dowson, C. G. Barriers to genetic exchange between bacterial species: *Streptococcus pneumoniae* transformation. *J Bacteriol* **182**, 1016-1023 (2000).
- 57 Briebe, L. G. Structure-Function Analysis Reveals the Singularity of Plant Mitochondrial DNA Replication Components: A Mosaic and Redundant System. *Plants (Basel)* **8**, doi:10.3390/plants8120533 (2019).
- 58 Valero-Rello, A., Lopez-Sanz, M., Quevedo-Olmos, A., Sorokin, A. & Ayora, S. Molecular Mechanisms That Contribute to Horizontal Transfer of Plasmids by the Bacteriophage SPP1. *Front Microbiol* **8**, 1816, doi:10.3389/fmicb.2017.01816 (2017).
- 59 Serrano, E. & Carrasco, B. Measurement of the Length of the Integrated Donor DNA during *Bacillus subtilis* Natural Chromosomal Transformation. *Bio-Protocol* **9**, e3338, doi:10.21769/BioProtoc.3338 (2019).
- 60 Torres, R., Serrano, E., Tramm, K. & Alonso, J. C. *Bacillus subtilis* RadA/Sms contributes to chromosomal transformation and DNA repair in concert with RecA and circumvents replicative stress in concert with DisA. *DNA Repair (Amst)* **77**, 45-57, doi:10.1016/j.dnarep.2019.03.002 (2019).

Table 1. Homogamic CT frequency of *rec*-deficient strains

Strain name ^a	Relevant genotype ^a	Homogamic Transformation	source
BG1359	Δrok , rec^+	100 (3.3×10^{-4})	17
BG1641	+ $\Delta recO$	76 ± 25	This work
BG1611	+ $recF15$	78 ± 17	This work
BG1631	+ $\Delta addAB$	49 ± 10	This work
BG1485	+ $\Delta ruvAB$	58 ± 11	This work
BG1653	+ $\Delta recU$	34 ± 12	This work
BG1813	+ $\Delta recJ$	1.2 ± 0.9	This work
BG1397	+ $\Delta recX$	0.2 ± 0.1	This work
BG1549	+ $\Delta recD2$	0.1 ± 0.05	This work
BG1647	+ $\Delta radA$	0.06 ± 0.003	60
BG1811	+ $\Delta dprA$	0.004 ± 0.002	This work
BG1633	+ $\Delta recA$	<0.001	16

^aAll *B. subtilis* *rec*-deficient strains are isogenic with BG1359. The genotype of the BG1359 strain is *trpCE metA5 amyE1 ytsJ1 rsbV37 xre1 xkdA1 att^{SPB} att^{ICEBs1} Δrok*. This strain lacks restriction-modification, CRISPR-Cas systems, different prophages and MGEs that might reduce the transformation rate.

Genetic recombination functions

FIGURES

Figure 1

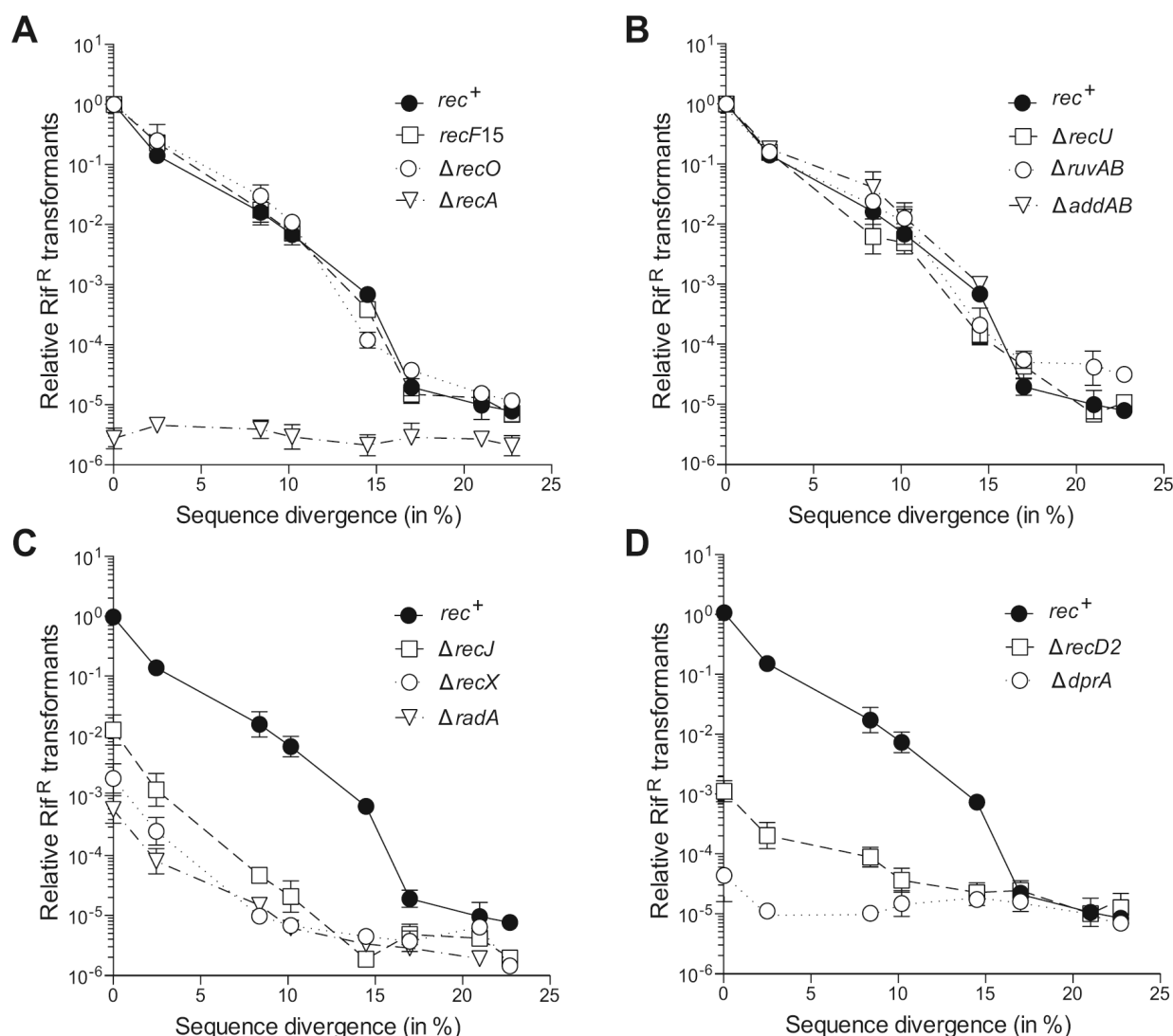


Figure 1. CT frequencies as a function of SD in different *rec* mutants. Donor DNA was a *rpoB482* DNA conferring Rif^R, derived from *B. subtilis* 168 (0.04% SD, homologous DNA), *B. subtilis* W23 (2.47%), *B. atrophaceus* 1942 (8.35%), *B. amyloliquefaciens* DSM7 (10.12%), *B. licheniformis* DSM13 (14.52%), *B. gobiensis* FJAT-4402 (17%), *B. thuringiensis* MC28 (20.83%) and *B. smithii* DSM4216 (22.74% SD). The *rpoB482* DNA (0.1 μ g DNA/ml) from these different *Bacillus* species with the selectable Rif^R mutation was used to transform BG1359 (*rec*⁺, ●) competent cells and its isogenic derivatives. The values are plotted dividing the number of transformants/CFUs obtained in each condition by the number of transformants/CFUs obtained when the *rec*⁺ cells are transformed with *Bsu*168 *rpoB482* DNA. In (A): BG1611 (*recF15*, □), BG1641 (Δ *recO*, ○) and BG1633 (Δ *recA*, ▽); in (B): BG1653 (Δ *recU*, □), BG1485 (Δ *ruvAB*, ○) and BG1631 (Δ *addAB*, ▽); in (C): BG1813 (Δ *recJ*, □), BG1397 (Δ *recX*, ○) and BG1647 (Δ *radA*, ▽); in (D): BG1549 (Δ *recD2*, □) and BG1811 (Δ *dprA*, ○). All data points are mean \pm standard error of the mean (SEM) derived from 3 to 5 independent experiments

Genetic recombination functions

Figure 2

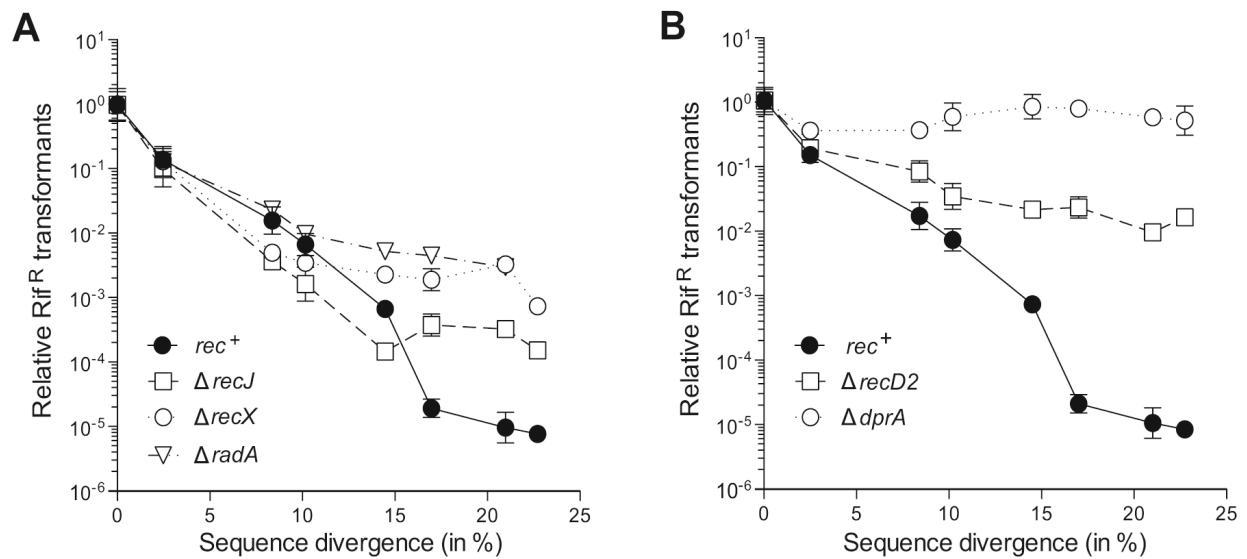


Figure 2. Normalised CT frequencies. The CT frequencies were normalized to give a value of 1 to the transformation frequency of the indicated *rec*-deficient strain when transformed with homologous DNA, and heterologous CTs are plotted relative to these values. In (A): *ΔrecJ*, *ΔrecX* and *ΔradA* (A); in (B): *ΔrecD2* and *ΔdprA* Rif^R cells. For comparison, the values obtained for the *rec*⁺ strain are also plotted.

Genetic recombination functions

Figure 3

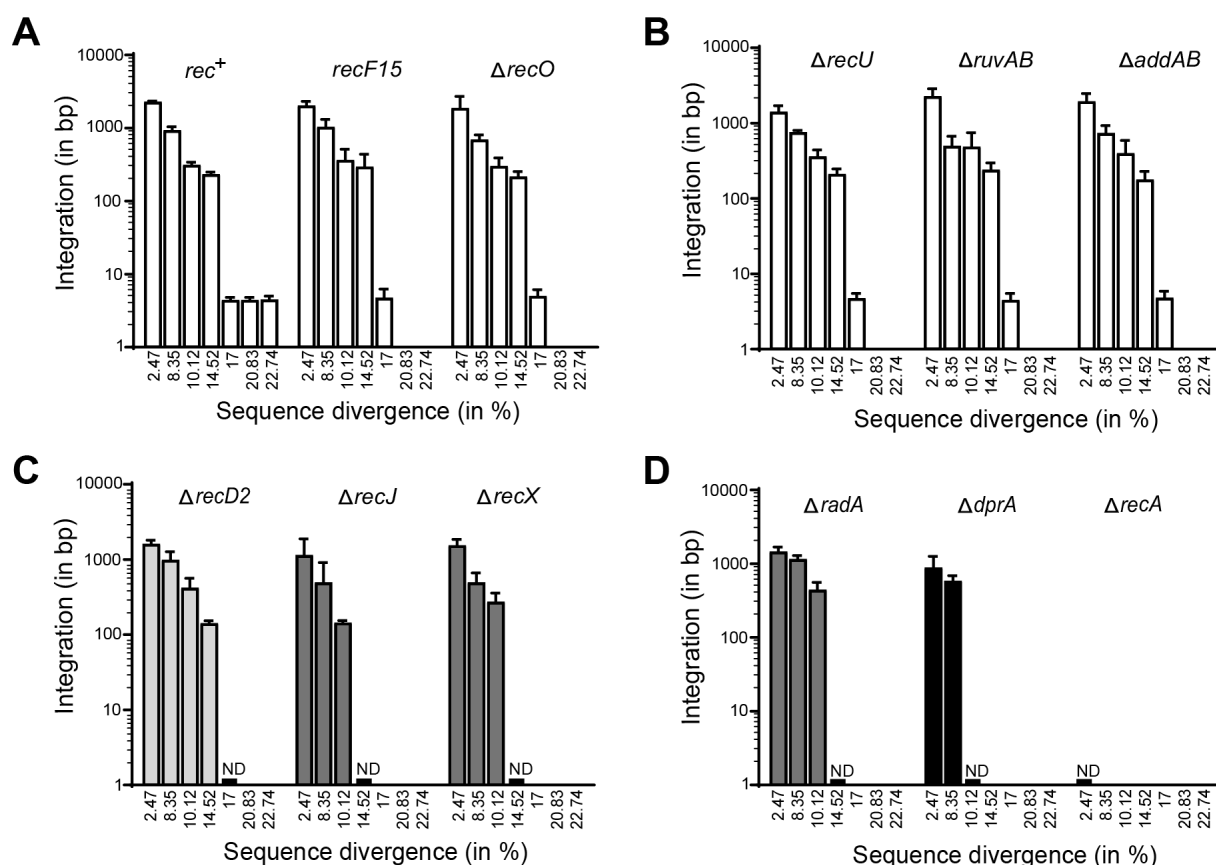


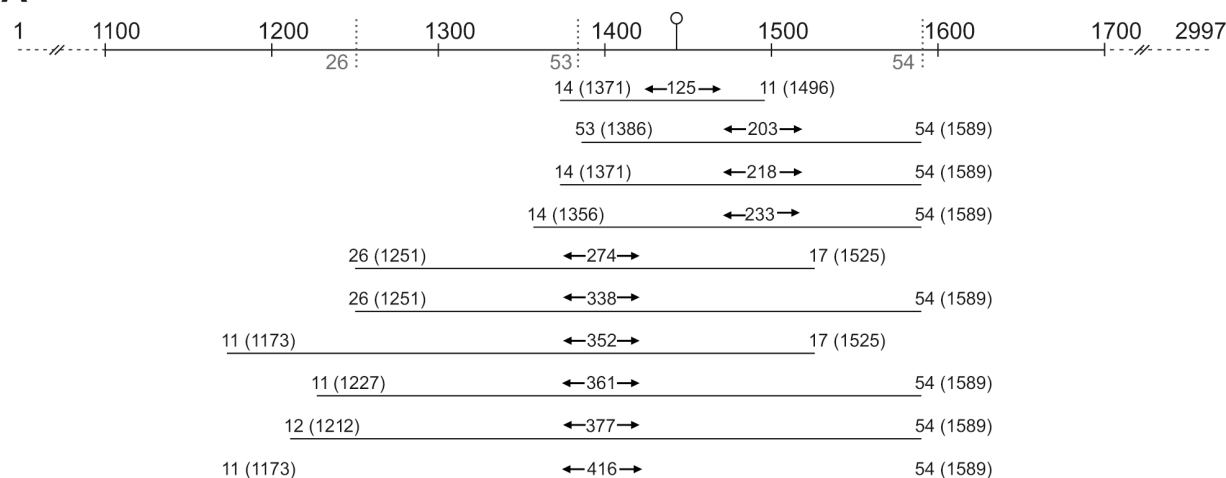
Figure 3. Determination of integration length of donor DNA with increased SD in different *rec* mutants.

The mean length of integration was determined by sequencing the 2997-bp *rpoB*482 DNA from different transformants obtained in *rec⁺*, *recF15* and $\Delta recO$ (A); $\Delta recU$, $\Delta ruvAB$ and $\Delta addAB$ (B); $\Delta recD2$, $\Delta recJ$ and $\Delta recX$ (C) and $\Delta radA$, $\Delta dprA$ and $\Delta recA$ (D). Integration endpoints were defined as the average between the last single nucleotide polymorphism present and the next SNP absent in the sequence of transformants (integration endpoint). The empty bars highlight the values obtained with strains that undergo both homology-directed and homology-facilitated CT, and filled bars denote strains impaired in homology-directed and blocked in homology-facilitated CT. ND, not detected.

Genetic recombination functions

Figure 4

A



B

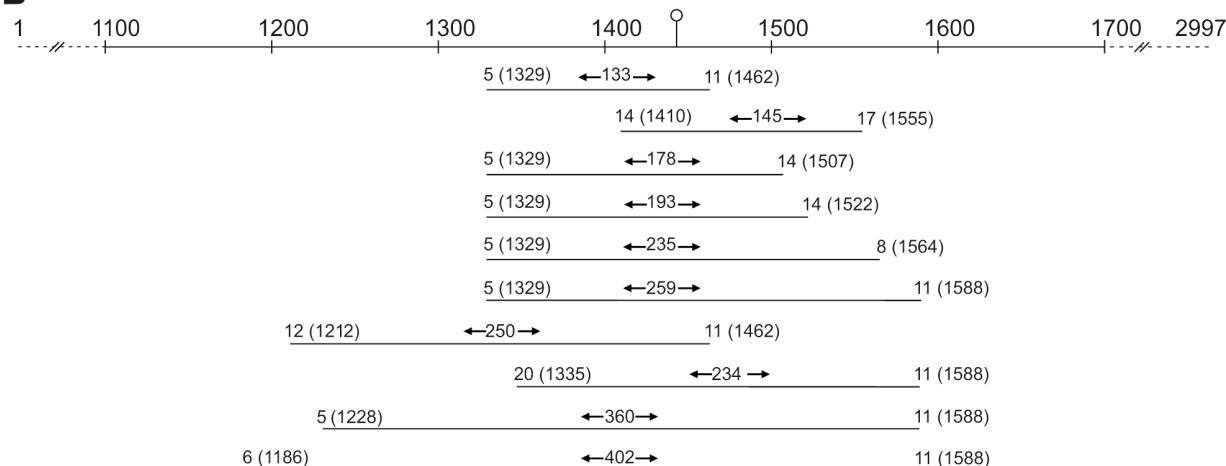


Figure 4. Mapping of recombination endpoints during interspecies CT in *rec*⁺ cells. The length of integrated DNA and recombination endpoints were determined by sequencing ten different Rif^R clones obtained following transformation with *Bam* DSM7 *rpoB*482 donor DNA (10.12% SD, in A) or *Bli* DSM13 *rpoB*482 donor DNA (14.52% SD, in B). In the top of the figure, the features of the donor DNA are indicated: The mutation that confers Rif^R, located at 1443 bp in the *rpoB* genes, is marked by a pin. In *Bam* DSM7 *rpoB*482 DNA (10.12% SD) the regions in the sequence where MEPSs (i.e., fully homologous regions larger than 26-bp) are found are denoted by a vertical dotted line. Their length is indicated also. MEPS regions are not observed in *Bli* DSM13 *rpoB*482 DNA (14.52 % SD). The MEPS and integration lengths of ten different Rif^R clones are shown, in brackets are indicated the location of recombination endpoints.