bioRxiv preprint doi: https://doi.org/10.1101/2020.05.13.090589; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 **Recombination proteins differently control the acquisition of homeologous DNA during**
2 ***Bacillus subtilis* natural chromosomal transformation**
3
4 Ester Serrano[1], Cristina Ramos[1], Juan C. Alonso[1, ✉] and Silvia Ayora[1, ✉]
5
6 Running title: genetic recombination functions
7
8 **Abstract**
9 In naturally competent *Bacillus subtilis* cells the acquisition of closely related genes occurs *via*
10 homology-directed chromosomal transformation (CT), and its frequency decreases log-linearly
11 with increased sequence divergence (SD) up to 15%. Beyond this and up to 23% SD the
12 interspecies boundary prevails, the CT frequency marginally decreases, and short (<10-
13 nucleotides) segments are integrated *via* homology-facilitated micro-homologous integration.
14 Both poorly known CT mechanisms are RecA-dependent. Here we identify the recombination
15 proteins required for the acquisition of interspecies DNA. The absence of AddAB, RecF, RecO,
16 RuvAB or RecU, crucial for repair-by-recombination, does not affect CT. However, inactivation
17 of *dprA*, *radA*, *recJ*, *recX* or *recD2* strongly interfered with CT. Interspecies CT was abolished
18 beyond ~8% SD in Δ*dprA*, ~10% in Δ*recJ*, Δ*radA*, Δ*recX* and 14% in Δ*recD2* cells. We propose
19 that DprA, RecX, RadA/Sms, RecJ and RecD2 help RecA to unconstrain speciation and gene
20 flow. These functions are ultimately responsible for generating genetic diversity and facilitate
21 CT and gene acquisition from bacteria of the same genus.
22
23 Keywords: horizontal gene transfer | microbial evolution | kin discrimination | Muller's ratchet
24

[1]Department 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

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.13.090589; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 **Introduction**

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

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

23 Naturally competent cells can integrate fully homologous DNA (homogamic CT), and with
24 lower efficiency homeologous (similar but not identical) DNA (heterogamic CT) [2]. This
25 interspecies CT can replace the recipient by a homeologous DNA, leading to mosaic structures
26 [13-15]. The transformation frequency of heterogamic DNA decreases with increased sequence
27 divergence (SD) both in *B. subtilis* and *S. pneumoniae* cells [14,15]. In *B. subtilis*, the transformation
28 assays with donor DNAs from different *Bacilli* species revealed a biphasic curve. Interspecies
29 CT decreased log-linear up to 15% SD, and integration in this part of the curve occurred *via* one-
30 step homology-directed RecA-dependent gene replacement [16,17]. Beyond 15% SD the CT
31 reached a plateau, and the integration of just few nucleotides, at micro-homologous segments
32 (<10-nucleotides [nt]), was observed at a very low efficiency (~1 x 10^{-5}) [17], suggesting another
33 recombination mechanism.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.13.090589; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 CT also contributes to expand the pan-genome of a species, because it allows the integration
2 of a heterologous DNA sequence, if flanked by two regions identical with recipient, usually
3 >400-nt, but longer homologous regions significantly increase the integration frequency [7,18].
4 Here, two homologous recombination (HR) events in the flanks integrate the heterologous DNA
5 *via* recipient-deletion /donor-insertion with ~10-fold lower efficiency than homologous gene
6 replacement [7,19]. Another mechanism is observed in replicating competent *Streptococcus*
7 *pneumoniae* or *Acinetobacter baylyi* cells. Here, homology at only one flank (anchor region)
8 facilitates illegitimate recombination of short (<10-nt) micro-homologous segments with
9 subsequent deletion of the intervening host DNA and integration of long heterologous DNA
10 segments, albeit with very low efficiency when compared to homogamic CT [20,21]. The length of
11 the anchor region affects the efficiency of this homology-facilitated illegitimate recombination
12 (HFIR) [19].

13 The proteins responsible for the acquisition of natural homeologous DNA remain poorly
14 known. The main player during HR is the RecA recombinase. RecA from a naturally competent
15 bacterium (*e.g.*, *B. subtilis*) has evolved to catalyse strand exchange in either the 5′→3′ or 3′→5′
16 direction and to tolerate 1-nt mismatch in an 8-nt region [16,17]. In naturally competent cells, the
17 essential SsbA and competence-specific SsbB coated the incoming linear ssDNA as soon it
18 leaves the entry channel. *B. subtilis* RecA cannot filament onto SsbA- or SsbB-coated ssDNA
19 [22,23]. The RecA accessory proteins are divided into four broad classes. First, the mediators that
20 act before RecA-mediated homology search. The DprA mediator facilitates partial disassembly
21 of SsbA and SsbB from the ssDNA, allows RecA binding, and in concert with SsbA assists RecA
22 to catalyse DNA strand exchange [23]. In the absence of DprA, RecO assists RecA to filament onto
23 SsbA-coated ssDNA [22]. Second, the modulators, which act during DNA identity search and
24 strand exchange, and either promote RecA nucleoprotein filament assembly as RecF (in the
25 Δ*recX* context) or its disassembly from the ssDNA, as RecX or RecU (in the Δ*recX* context) [24-
26 28]. With the help of both, mediators and modulators, a dynamic RecA-ssDNA filament is
27 engaged in sequence identity search. *E. coli* RecA requires ~15-nt of homologous ssDNA to
28 promote strand exchange, defining the *in vitro* minimal efficient processing segment (MEPS)
29 [29,30]. *In vivo, E. coli* RecA significantly recombines two duplex DNAs with 25- to 30-bp MEPS
30 [31], thus this length was proposed for the *B. subtilis* protein [32]. I*n vivo*, upon finding a MEPS,
31 RecA initiates strand invasion to form a heteroduplex, known as displacement loop (D-loop) [33].
32 Then, branch migration translocases (RecD2, RuvAB, RecG and RadA/Sms) allow D-loop
33 extension, and help to generate a stable heteroduplex [28,34]. Finally, after DNA strand exchange a
34 structure-specific nuclease must resolve the D-loop. The RecU resolvase is unable to cleave D-

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.13.090589; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 loops [35-37], and the enzyme responsible for such activity remains unknown. The contribution of
2 these RecA accessory proteins to homogamic CT has been studied in deep [7,28,38], but little is
3 known about their role in the integration of heterogamic DNA. In *B. subtilis*, two exonucleases,
4 the AddAB complex (counterpart of *E. coli* RecBCD) and RecJ are crucial to process double-
5 strand breaks (DSBs) during DNA repair [33,39], but their putative role in the degradation of the
6 displaced strand during CT remains to be documented.

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.13.090589; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 **Results**

2 **Experimental design**

3 The rational to select *B. subtilis* competent cells to analyse how the recombination machinery
4 contributes to heterogamic CT is summarised in Supplementary *Annex 1*. Competence
5 development is a stochastic process driven by the expression of ComK, and it occurs only in a
6 small fraction (0.1 - 5%) of cells [11,12]. Interspecies CT, at SD above 8%, is at the limit of detection
7 in wild type cells [14]. To overcome such technical difficulty, Rok, which directly represses *comK*
8 expression, was deleted [40]. Inactivation of *rok* increases the subpopulation of cells that develop
9 natural competence [41], and thereby the sensitivity of the assay. The different *rec* mutations were
10 mobilised into the *rok* strain (see Table 1). All the strains additionally lack Rok, but for simplicity
11 we only state the *rec* gene that is mutated.

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

25

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

28 Except in Δ*recD2* cells, the number of Rif^R spontaneous mutant colonies, which appeared in the
29 absence of *rpoB*482 DNA in the different rec^- strains was similar to the *wt* control (see
30 Supplementary material *Annex 3*). To test how the different rec^- mutants affect intraspecies CT,
31 competent cells were transformed with *Bsu* 168 *rpoB*482 DNA with direct selection for Rif^R
32 (Table 1, Fig. 1A). The homogamic CT frequency with this donor DNA was similar to the one
33 obtained with *Bsu* 168 met^+ DNA, also with a single donor-recipient mismatch [17]. As previously
34 reported, intraspecies CT was blocked in the Δ*recA* strain (Table 1, Fig. 1A) [16,42].

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.13.090589; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 Natural homogamic CT was barely affected in *recF*15, Δ*recO*, Δ*recU*, Δ*ruvAB*, Δ*addAB*
2 cells (Table 1, Fig. 1A-B). Their CT frequencies were similar to results previously observed for
3 these recombination mutants in the *rok*$^+$ background [28,43]. However, the frequency of intraspecies
4 CT was significantly reduced in competent Δ*recJ* (by ~80-fold), Δ*recX* (~500-fold), Δ*recD*2
5 (~950-fold), Δ*radA* (~1600-fold) and Δ*dprA* (~24,000-fold) cells (Table 1, Fig. 1C-D). Except
6 in Δ*recJ*, intraspecies CT efficiency in these mutants was lower in Δ*rok* than in *rok*$^+$ [26,34]. The
7 deleterious effect of additionally mutate *rok* in these backgrounds will be reported elsewhere.
8
9 **Interspecies CT requires DprA, RecX, RadA/Sms, RecD2 and RecJ**
10 To gain insight into the contribution of these recombination proteins to interspecies CT, we used
11 *rpoB*482 DNAs with different degree of SD (Supplementary *Annex 2*). The frequency of
12 interspecies CT decreased logarithmically with increased SD up to ~15% in competent *recF*15,
13 Δ*recO*, Δ*recU*, Δ*ruvAB* or Δ*addAB* cells (Fig. 1A-B). When SD was further increased, beyond
14 15% and up to ~23% SD, the interspecies CT frequencies varied <3-fold in *recF*15, Δ*recO*,
15 Δ*recU*, *ruvAB* and *addAB* (Fig. 1A-B), as in the *rec*$^+$ control (Fig. 1A) [17], suggesting that these
16 functions do not limit genetic recombination in otherwise *rec*$^+$ cells.
17 A different outcome was observed in Δ*recJ*, Δ*recX*, Δ*radA*, Δ*recD2* and Δ*dprA*. The CT
18 frequency was similar to the frequency of spontaneous Rif^R mutations beyond ~8% SD in Δ*dprA*,
19 ~10% in Δ*recX* and Δ*radA*, and ~15% in Δ*recJ* and Δ*recD2* competent cells, (Fig. 1C-D). We
20 believe that this strong defect was not due to an impairment in fitness cost. The colony size
21 observed after overnight growth at 37ºC under selective pressure was similar in all cases.
22 Furthermore, the sequence analysis of the chimeric *rpoB*482 genes revealed a RpoB482 protein
23 only bearing the Rif^R mutation (see below).
24
25 **The interspecies barrier is altered in Δ*recJ*, Δ*recX*, Δ*recD2*, Δ*radA* and Δ*dprA***
26 To simplify the analysis of the strains that are significantly impaired in heterogamic CT, we gave
27 a value of 1 to the transformation rate obtained by the given mutant in the intraspecies CT assay,
28 and plotted the data relative to this value. Hence only the impact of SD is evaluated (Fig. 2A-B).
29 Different outcomes were observed. First, in competent Δ*recX*, Δ*radA* and Δ*recJ* cells
30 interspecies CT decreased logarithmically with increased SD, but only up to ~10% (Δ*recJ*,
31 ~14%) SD, to reach a plateau at higher divergence (Fig. 2A). Second, competent *recD2* reached
32 a plateau already at 8% SD (Fig. 2B). Third, upon inactivation of *dprA*, the decline in the rate of

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.13.090589; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 recombination with SD was only observed at ~2% SD, to reach a plateau at higher divergence
2 (Fig. 2B).
3
4 **RecO, RecF, RecU, RuvAB and AddAB are dispensable in heterogamic transformation**
5 The above results suggest that RecO, RecF, RecU, RuvAB and AddAB play no apparent roles
6 in interspecies CT, but their contribution to the integration length is unknown. To analyse this,
7 we sequenced the *rpoB*482 gene from Rif^R clones and calculated the mean integration length.
8 The maximal integration length that can be detected with the donor DNA of ~2% SD
9 (*Bsu* W23 DNA) is 2628-bp, because the first mismatch is located at position 350 and the last at
10 position 2978. Previous analysis showed that when the *rec*$^+$ strain was transformed with this
11 donor DNA, the mean integration length was close to this maximal integration length, around
12 ~2300-bp [16,17]. Similar results were obtained with the donor DNA of ~2% SD in the Δ*recO*,
13 Δ*recU*, Δ*ruvAB* and Δ*addAB* backgrounds (Fig 3A-B).
14 Up to 15% divergence range, RecA is believed to integrate the DNA by a homology-
15 directed HR mechanism, initiating recombination in a MEPS [16,22]. A comparison of the
16 nucleotide sequence of donor with recipient DNA revealed the presence of 22 MEPS at or above
17 25-nt in donor DNA with ~8% SD, and 21 at ~10% SD. In both donor DNAs there is a long
18 stretch of ~200-nt of sequence identity upstream of the *rpoB*482 mutation, and several regions
19 with a MEPS longer than 54-nt downstream of the *rpoB*482 mutation. They could define the left
20 and right recombination endpoints, being the region in-between integrated independently of its
21 SD, as it occurs in the insertion of heterologous DNA by two-step HR at the flanks (see
22 Introduction). However, integrated fragments of ~1600-nt (*i.e*, recombination endpoints at the
23 157-nt [at position 1-157] and the 81-nt MEPS [at position 1509-1589]) was not observed in *rec*$^+$
24 transformations with ~8% SD donor DNA, and the same was observed with 10% SD (Fig 3A).
25 This result suggests that the two-step deletion/insertion CT may not take place with interspecies
26 DNA, probably because it requires two longer flanking homologous regions (see Introduction),
27 or the sequence in-between plays a relevant role.
28 The analysis of 10-20 Rif^R clones obtained in the transformation of competent *recF*15,
29 Δ*recO*, Δ*recU*, Δ*ruvAB* and Δ*addAB* cells with donor DNA of ~8% SD revealed that the mean
30 integration length was as in *rec*$^+$, 700-900-nt, except in Δ*ruvAB*, which was ~490-nt (*i.e*, ~2 times
31 less) (Fig. 3A-B, and Table S1). A sequence analysis of the integrated region revealed eight
32 MEPS (four upstream [50-, 35-, 38- and 36-nt] and four downstream [35-, 81-, 41- and 33-nt])
33 of the *rpoB*482 mutation. One recombination endpoint was usually at one of these MEPS, but

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.13.090589; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

9 In *recF*15, Δ*recO*, Δ*recU*, Δ*ruvAB*, Δ*addAB* and *rec*$^+$ cells, when SD is ~10%, the mean
10 integration length was ~300-nt (Fig. 3A-B). In these transformants, the nucleotide sequence
11 changes do not alter the RpoB482 protein sequence (see Fig. S1C). Within a 600-nt interval
12 centred at the *rpoB*482 mutation, there are only three sequences at or above MEPS (see Fig. 4A).
13 Analysis of 10 different Rif^R clones in *rec*$^+$ showed that in ~70% of the cases, the 3'-endpoint
14 was at the 54-nt MEPS (at position 1536-1590, Fig. 4A). Recombination in this region did not
15 extend beyond, probably because it is followed by a local ~24% SD in the next 25-nt that could
16 act as a strong heterologous barrier. The other endpoints were located in regions below MEPS
17 which are usually followed by barriers with a higher SD (*e.g*., endpoint at position 1371 is
18 preceded by 25-nt with ~24% SD, at position 1251 by ~24% SD). No patched sequences were
19 observed, discarding that in some transformants multiple recombination events had occurred at
20 different loci. As in *rec*$^+$, the recombination endpoints in the *recF*15, Δ*recO*, Δ*recU*, Δ*ruvAB*
21 and Δ*addAB* transformants did not always coincide with the longest MEPS present (Table S1).
22 In these mutants one endpoint was usually at a region where a MEPS was located, whereas the
23 other endpoint was less specific, and often below MEPS. All these results suggest that a MEPS
24 is necessary to initiate strand invasion, but a SD barrier may halt RecA-mediated uni- or
25 bidirectional DNA strand exchange.

26 Nucleotide sequence analysis of 15-30 Rif^R clones obtained with ~15% SD in Δ*ruvAB* and
27 *rec*$^+$ cells showed that all were genuine transformants, but these values dropped to ~40-50% in
28 Δ*recU*, Δ*recO*, Δ*addAB*, and *recF*15. The mean integration length was between 400- to 130-nt
29 in *recF*15, Δ*recO*, Δ*recU*, Δ*ruvAB* and Δ*addAB,* as well as in the *rec*$^+$ control (Table S1 and Fig.
30 4B). This donor DNA has a MEPS of 104-nt (position 54-157) upstream of the *rpoB*482
31 mutation, and a 38-nt MEPS (position 2295-2332) downstream, but ~2100-nt inserts were not
32 observed, confirming that the region between the MEPS are relevant. Potential MEPS are not
33 present in the integrated region (Fig. 4B). The sequence analysis of *rec*$^+$ transformants revealed

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.13.090589; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

5 Beyond 15% SD the transformation efficiency and the frequency of spontaneous mutation
6 almost overlap, although there is a ~3-fold difference, and it is higher than in a Δ*recA* mutant,
7 suggesting that still this is a RecA-dependent integration event (Fig. 1). Four regions at or above
8 MEPS (74-, 56-, 28- and 26-nt) upstream and one downstream (26-nt) of the *rpoB*482 mutation
9 are present in the 17% SD donor, but they were not used to integrate the region in-between.

10 Nucleotide sequence analyses of Rif^R *rec*$^+$ clones obtained with ~17% SD showed that only
11 ~37% were genuine transformants, *i.e*, they had incorporated two or more nucleotides of donor
12 DNA. Similarly, 20-30% of the Rif^R clones obtained in Δ*recO*, Δ*addAB*, Δ*recU*, Δ*ruvAB* and
13 *recF*15 were genuine transformants (Fig. 3A-B), and the mean integration length was also 4- to
14 8-nt, or ~5-fold below MEPS (Table S1). The observed low efficiency of micro-homologous
15 integration cannot be attributed to a defect in the resulting RpoB482 protein, because at ~17%
16 SD, except the amino acid change that confers Rif^R, there is no mutation in the 50 residues
17 intervals up and downstream the Rif^R change (Fig. S1C). The close inspection in donor DNA
18 with 17% SD of the region surrounding the *rpoB*482 mutation showed that this mutation is
19 embedded in a region with strong SD with recipient DNA. The 25-nt region upstream of the
20 *rpoB*482 mutation has 8 mismatches (*i.e*, 32% SD) and the one downstream 10 mismatches (40%
21 SD) that could act as heterologous barriers.

22 At ~21% SD there are two sites at or above MEPS (42-nt [at position 90-131] and 27-nt [at
23 position 2781-2807]), one on each side of the *rpoB*482 mutation, but they were not used. After
24 sequencing of transformants obtained in *recF*15, Δ*recO*, Δ*recU*, Δ*ruvAB*, Δ*addAB* or *rec*$^+$, we
25 found that again just few nucleotides from the donor had been integrated, and that the fraction
26 of genuine transformants was ~20% (Fig 3A-B). Here, the analysis showed that the *rpoB*482
27 mutation is also surrounded by higher SD: 8 mismatches in 25-nt upstream (*i.e*, ~32% SD) and
28 the one downstream 6 mismatches (*i.e*, ~24% SD). Finally, at ~23% SD, only one MEPS exists
29 (at position 801-826). The proportion of genuine Rif^R transformants accounts to only ~6% of the
30 sequenced clones in the *rec*$^+$ strain. The mean integration length in *rec*$^+$ transformants was 4- to
31 8-nt (Fig. 3A-B). Similar results were obtained in *recF*15, Δ*recO*, Δ*recU*, Δ*ruvAB* and Δ*addAB*.
32

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.13.090589; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 Nucleotide sequence analyses of 10-30 Rif^R clones obtained in the Δ*radA*, Δ*recJ*, Δ*recX*, Δ*recD*2
2 or Δ*dprA* context revealed that when divergence was low (~2%), the mean integration length of
3 the *rpoB*482 DNA was similar to the *rec*$^+$ control (Fig 3C-D). The proportion of genuine
4 transformants was different in these mutants. At ~8% SD, all sequenced clones obtained in
5 Δ*radA* were genuine transformants, but this number was reduced to ~50 % in Δ*dprA*, Δ*recJ* and
6 Δ*recX*, and even lowered to ~25% in Δ*recD2*. Representatives of the genuine Rif^R clones are
7 documented in Table S2.

8 The efficiency of interspecies transformation at ~10% SD was strongly reduced in these
9 mutants, to levels similar to the spontaneous mutation rate (Fig. 1). Only ~25% of the Rif^R clones
10 obtained in Δ*recX* were genuine transformants, and these values lowered to ~15% in Δ*recD2,*
11 and ~5% in Δ*recJ* and Δ*radA* cells. Furthermore, in the Δ*dprA* strain all 25 sequenced Rif^R clones
12 were spontaneous mutants (Table S2). The mapping of recombination endpoints in the Δ*recJ*,
13 Δ*recD*2, Δ*recX* and Δ*radA* transformants showed that, as in the *rec*$^+$ transformants, the 3'-
14 endpoint of recombination was located at the 54-nt MEPS, except in some Δ*recX* transformants.
15 The 5'-endpoint of recombination was usually located at a larger MEPS when compared with
16 the *rec*$^+$ control (Tables S1 and S2).

17 At ~15% SD, all sequenced Rif^R clones obtained were spontaneous mutants in the Δ*recX*,
18 Δ*radA*, Δ*recJ* and Δ*dprA* strains. In Δ*recD*2, only one of the sequenced Rif^R clones (1/19) was a
19 genuine transformant with a mean integration length >120-nt (Table S2). At ~17% SD, all
20 sequenced Rif^R clones were spontaneous mutants in the Δ*recX*, Δ*radA*, Δ*recJ*, Δ*dprA,* and
21 Δ*recD*2 backgrounds. These results show that these functions are all needed for homology-
22 directed integration when SD is larger than 10%, and also for micro-homologous integration.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.13.090589; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 **Discussion**
2 From the data presented in this work, it can be inferred that recombination proteins differently
3 facilitate adaptation and genetic diversity, and that two different recombination mechanisms
4 occur depending on the SD of the interspecies DNA. Up to 15% SD, homology-directed HR
5 accounts for the integration of divergent sequences longer than 130-nt. Beyond this, integration
6 of just few nucleotides at micro-homologous segments is observed.
7 Inactivation of the main end-resection complex (*addAB*), a positive mediator (*recO*), a
8 positive (*recF*) or a negative (*recU*) modulator or enzymes essential for Holliday junction
9 processing and cleavage (*ruvAB*, *recU*) barely affects heterogamic CT (Fig. 1A-B), although
10 they are essential for DNA DSB repair [39,43,44]. The interpretation of such result may not be so
11 simple, because cells lacking both RecF and AddAB are blocked even in homogamic CT [43],
12 suggesting that they are backup functions during CT (*e.g.*, RecF is essential in the *recX* context)
13 [25]. We have observed that a specific subset of HR proteins (RecA, DprA, RecX, RadA/Sms,
14 RecD2 and RecJ) contributes to acquire homeologous DNA. Except DprA which is competence
15 specific and present in all transformable bacteria, and even in non-transformable ones [10], the
16 recombination proteins identified here also participate, together with another subset of HR
17 proteins, in the accurate repair of lethal DSBs and in the restart of stalled replication forks. In
18 some distantly related competent cells (*e.g.*, Proteobacteria phylum) RecD2 is absent, and
19 RadA/Sms is replaced by ComM [45]. They differently participate in the recombination
20 mechanisms that may be active during the acquisition of homeologous DNA. In the absence of
21 RecA CT is blocked [16]. Interspecies homology-directed CT was blocked with DNA of the same
22 clade (up to ~8% SD) in Δ*dprA*, beyond ~10% SD in the Δ*recX*, Δ*recJ* and Δ*radA* backgrounds,
23 and beyond ~15% SD in Δ*recD2* cells (Fig. 3C-D), showing an essential role for these proteins
24 in the acquisition of interspecies DNA.
25 We can envision that during homology-directed CT, RecA nucleates in the incoming ssDNA
26 coated by SsbA and SsbB, with the help of the DprA-SsbA two-component mediator. Then, a
27 dynamic RecA nucleoprotein filament, with the contribution of DprA-SsbA and RecX, identifies
28 an identical sequence in the recipient genome. Once a homologous region is found RecA
29 promotes DNA strand invasion to produce a metastable heteroduplex DNA [23,27]. RecA at this D-
30 loop interacts with and loads the branch-migration translocase RadA/Sms as documented in
31 Firmicutes [34,46]. RadA/Sms, in concert with RecA, facilitates bidirectional branch migration until
32 a region with a SD >20% is found. This is consistent with the *in vitro* observation that homology-
33 directed RecA-mediated strand-exchange halted at DNA patches >16% SD [16].

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.13.090589; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 The contribution of RecD2 helicase and RecJ exonuclease to homology-directed
2 recombination during interspecies CT is poorly understood. Recently, it was suggested that
3 RecD2 contributes to branch migrate the heteroduplex DNA in a relaxed molecule, perhaps upon
4 cleavage of the displaced strand [28]. As RecJ of other naturally competent cells [47], *B. subtilis* RecJ
5 possesses an extra C-terminal domain, absent in *E. coli* RecJ, critical for protein-protein
6 interaction (*e.g.*, SsbA). Perhaps in concert with the RecD2 helicase, it might degrade the
7 displaced strand and the non-paired tails. Finally, the ends are sealed, leading to the acquisition
8 of homeologous DNA >130-nt.

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

16 Beyond 15% sequence, integration of <10-nt segments was observed in Δ*addAB*, *recF*,
17 Δ*recO*, Δ*recU* and Δ*ruvAB* (Fig. 3A-B), but it was not detected in Δ*recA*, Δ*dprA*, Δ*recJ*, Δ*recX*,
18 Δ*radA* and Δ*recD2* cells (Fig. 3C-D). These results showed that a similar set of recombination
19 functions are also required for this short integration, which occurs up to 23% SD, although with
20 low efficiency. Here we propose a homology-facilitated micro-homologous integration
21 mechanism: As above, a RecA dynamic filament, with the help of DprA-SsbA and RecX,
22 searches for and identifies a MEPS on the *rpoB*482 DNA, which is used in this case as an anchor
23 region, to produce a metastable D-loop intermediate, in concert with RadA/Sms. Once the DNA
24 is anchored at MEPS, DprA could mediate the annealing of short stretches of homeologous DNA
25 (3- to 8-nt) around the *rpoB*482 mutation [23]. Then, the donor ssDNA loop between the anchored
26 region and the micro-homologous paired segment has to deleted, perhaps by RecJ in concert
27 with RecD2. Finally, the ends of the integrated segment are sealed and rapidly expressed [45]. This
28 mechanism differs from HFIR observed in replicating competent *S. pneumoniae* and *A. baylyi*
29 cells (see Introduction) [20,21]. In these competent bacteria, inactivation of *recBCD* or *recJ*
30 significantly increased HFIR [19,21]. In contrast, in non-replicating competent *B. subtlis* cells
31 integration of thousands of nucleotides of the heterogamic DNA with the subsequent deletion of
32 the recipient DNA was not observed, and inactivation of *recJ* blocked homology-facilitated
33 micro-homologous integration. Competent *A. baylyi* cells can also integrate short ssDNA (20-
34 nt) in the recipient genome by another mechanism, which occurs with extremely low efficiency.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.13.090589; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 This mechanism requires active DNA replication, inactivation of *recJ* and is independent of
2 RecA [51]. It is likely that competent *B. subtilis* cells may use homology-facilitated micro-
3 homologous integration to restore genes inactivated by mutations and thereby prevent the
4 irreversible deterioration of genomes (Muller's ratchet) [4]. At ~23% SD interspecies CT
5 frequency was similar to spontaneous mutations, suggesting that beyond ~23% SD micro-
6 homologous integration might be inefficient, probably because the unique MEPS present is too
7 short to serve as a stable anchor region [19]. Similarly, competent *H. pylori* cells cannot be
8 transformed by *Campylobacter jejuni* DNA with ~24% SD [52].
9 Can the above observations be extrapolated to interspecies Hfr chromosomal conjugation?
10 The frequency of interspecies Hfr chromosomal conjugation also decreases log-linearly with
11 increased SD up to ~16% SD. Inactivation of *mutSL* alleviates interspecies Hfr conjugation by
12 ~1000-fold, and deletion of *recBCD* or *ruvAB* reduced interspecies Hfr conjugation, but
13 inactivation of *recJ* increases it [2,3,53,54], suggesting that interspecies Hfr chromosomal
14 conjugation uses the repair-by-recombination mechanism. In contrast, inactivation of *mutSL*
15 marginally prevents interspecies CT with up to 15% SD in Firmicutes [15,16,55,56] and inactivation
16 of *recJ*, but not *addAB* or *ruvAB,* inhibits interspecies CT. All these results suggest that bacteria
17 have evolved different genetic recombination mechanisms devoted to interspecies genetic
18 exchange to generate diversity. Bacterial *recA*, *radA* and *recX* genes, which play crucial roles in
19 interspecies CT, perhaps contributed in their transfer from mitochondria or chloroplasts to the
20 nucleus of land plants, green algae and moss [57], although the evolutionary force and molecular
21 functions that contributed to the transfer of these genes well beyond the species boundaries is
22 poorly understood.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.13.090589; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 $\mu g \cdot ml^{-1}$ of the indicated *rpoB*482 donor DNA (30 min, 37ºC), and then plated on Rif (8 $\mu g \cdot ml^{-1}$) 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.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.13.090589; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.13.090589; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.13.090589; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.13.090589; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 45 Dalia, A. B. & Dalia, T. N. Spatiotemporal Analysis of DNA Integration during Natural
2 Transformation Reveals a Mode of Nongenetic Inheritance in Bacteria. *Cell* **179**, 1499-
3 1511 e1410, doi:10.1016/j.cell.2019.11.021 (2019).
4 46 Marie, L. *et al.* Bacterial RadA is a DnaB-type helicase interacting with RecA to promote
5 bidirectional D-loop extension. *Nat Commun* **8**, 15638, doi:10.1038/ncomms15638
6 (2017).
7 47 Cheng, K. *et al.* Structural basis for DNA 5 -end resection by RecJ. *Elife* **5**, e14294,
8 doi:10.7554/eLife.14294 (2016).
9 48 Folta-Stogniew, E., O'Malley, S., Gupta, R., Anderson, K. S. & Radding, C. M. Exchange of
10 DNA base pairs that coincides with recognition of homology promoted by E. coli RecA
11 protein. *Mol Cell* **15**, 965-975, doi:10.1016/j.molcel.2004.08.017 (2004).
12 49 Qi, Z. *et al.* DNA sequence alignment by microhomology sampling during homologous
13 recombination. *Cell* **160**, 856-869, doi:10.1016/j.cell.2015.01.029 (2015).
14 50 Bazemore, L. R., Folta-Stogniew, E., Takahashi, M. & Radding, C. M. RecA tests homology
15 at both pairing and strand exchange. *Proc Natl Acad Sci U S A* **94**, 11863-11868,
16 doi:10.1073/pnas.94.22.11863 (1997).
17 51 Overballe-Petersen, S. *et al.* Bacterial natural transformation by highly fragmented and
18 damaged DNA. *Proc Natl Acad Sci U S A* **110**, 19860-19865,
19 doi:10.1073/pnas.1315278110 (2013).
20 52 Levine, S. M. *et al.* Plastic cells and populations: DNA substrate characteristics in
21 Helicobacter pylori transformation define a flexible but conservative system for
22 genomic variation. *FASEB J* **21**, 3458-3467, doi:10.1096/fj.07-8501com (2007).
23 53 Rayssiguier, C., Thaler, D. S. & Radman, M. The barrier to recombination between
24 Escherichia coli and Salmonella typhimurium is disrupted in mismatch-repair mutants.
25 *Nature* **342**, 396-401, doi:10.1038/342396a0 (1989).
26 54 Matic, I., Rayssiguier, C. & Radman, M. Interspecies gene exchange in bacteria: the role
27 of SOS and mismatch repair systems in evolution of species. *Cell* **80**, 507-515 (1995).
28 55 Majewski, J. & Cohan, F. M. The effect of mismatch repair and heteroduplex formation
29 on sexual isolation in Bacillus. *Genetics* **148**, 13-18 (1998).
30 56 Majewski, J., Zawadzki, P., Pickerill, P., Cohan, F. M. & Dowson, C. G. Barriers to genetic
31 exchange between bacterial species: Streptococcus pneumoniae transformation. *J*
32 *Bacteriol* **182**, 1016-1023 (2000).
33 57 Brieba, L. G. Structure-Function Analysis Reveals the Singularity of Plant Mitochondrial
34 DNA Replication Components: A Mosaic and Redundant System. *Plants (Basel)* **8**,
35 doi:10.3390/plants8120533 (2019).
36 58 Valero-Rello, A., Lopez-Sanz, M., Quevedo-Olmos, A., Sorokin, A. & Ayora, S. Molecular
37 Mechanisms That Contribute to Horizontal Transfer of Plasmids by the Bacteriophage
38 SPP1. *Front Microbiol* **8**, 1816, doi:10.3389/fmicb.2017.01816 (2017).
39 59 Serrano, E. & Carrasco, B. Measurement of the Length of the Integrated Donor DNA
40 during Bacillus subtilis Natural Chromosomal Transformation. *Bio-Protocol* **9**, e3338,
41 doi:10.21769/BioProtoc.3338 (2019).
42 60 Torres, R., Serrano, E., Tramm, K. & Alonso, J. C. Bacillus subtilis RadA/Sms contributes
43 to chromosomal transformation and DNA repair in concert with RecA and circumvents
44 replicative stress in concert with DisA. *DNA Repair (Amst)* **77**, 45-57,
45 doi:10.1016/j.dnarep.2019.03.002 (2019).

46

47

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.13.090589; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 **Table 1.** Homogamic CT frequency of *rec*-deficient strains
2

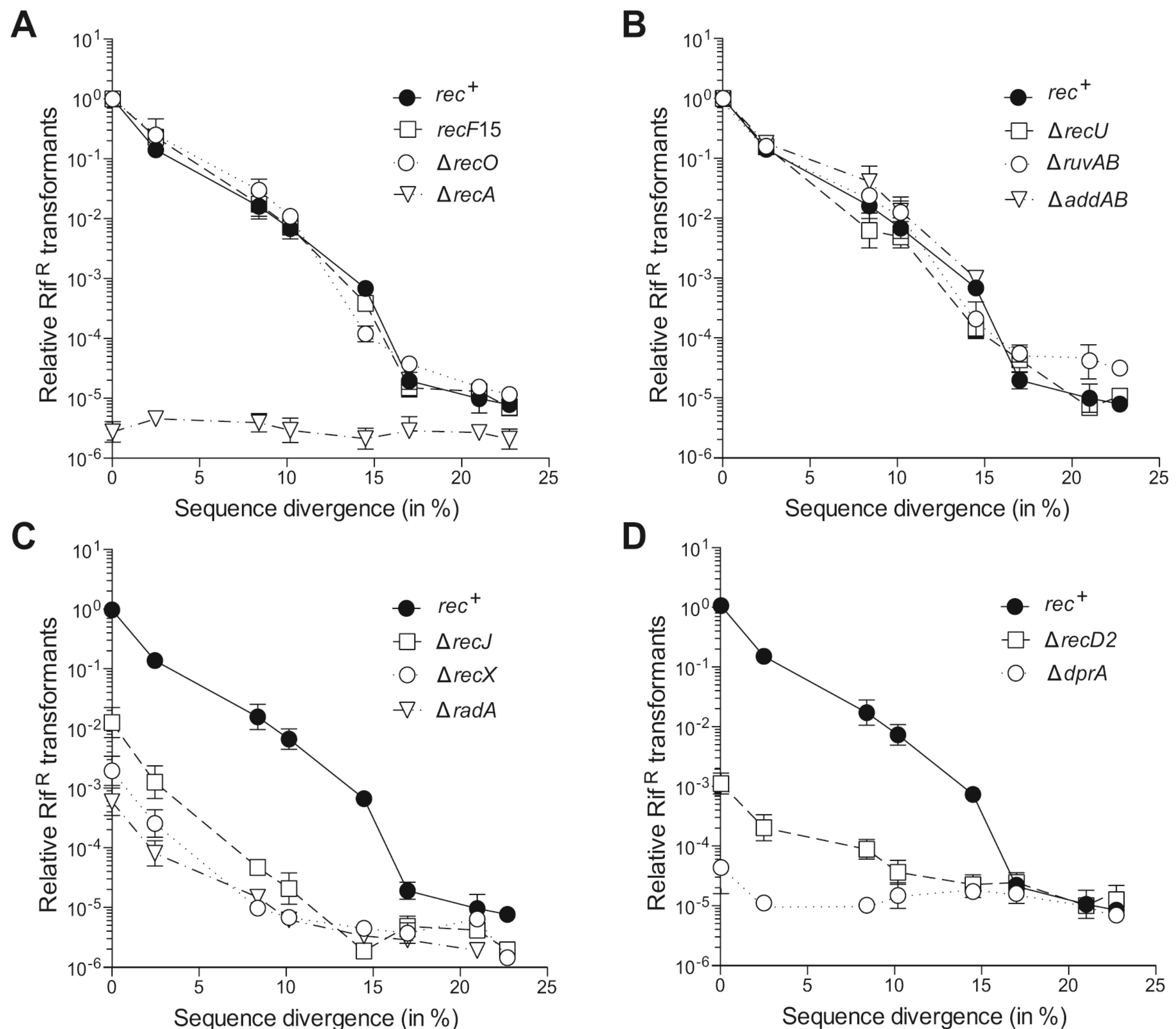
Strain name[a]	Relevant genotype[a]	Homogamic Transformation	source
BG1359	Δ*rok*, *rec*$^{+}$	100 (3.3 x 10^{-4})	17
BG1641	+ Δ*recO*	76 ± 25	This work
BG1611	+ *recF*15	78 ± 17	This work
BG1631	+ Δ*addAB*	49 ± 10	This work
BG1485	+ Δ*ruvAB*	58 ± 11	This work
BG1653	+ Δ*recU*	34 ± 12	This work
BG1813	+ Δ*recJ*	1.2 ± 0.9	This work
BG1397	+ Δ*recX*	0.2 ± 0.1	This work
BG1549	+ Δ*recD2*	0.1 ± 0.05	This work
BG1647	+ Δ*radA*	0.06 ± 0.003	60
BG1811	+ Δ*dprA*	0.004 ± 0.002	This work
BG1633	+ Δ*recA*	<0.001	16

3 [a]All *B. subtilis rec*-deficient strains are isogenic with BG1359. The genotype of the BG1359
4 strain is *trpCE metA*5 *amyE1 ytsJ*1 *rsbV*37 *xre*1 *xkd*A1 *att*^{SPß} *att*^{ICEBs1} Δ*rok*. This strain lacks
5 restriction-modification, CRISPR-Cas systems, different prophages and MGEs that might reduce
6 the transformation rate.
7

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.13.090589; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 **FIGURES**

2 **Figure 1**



3

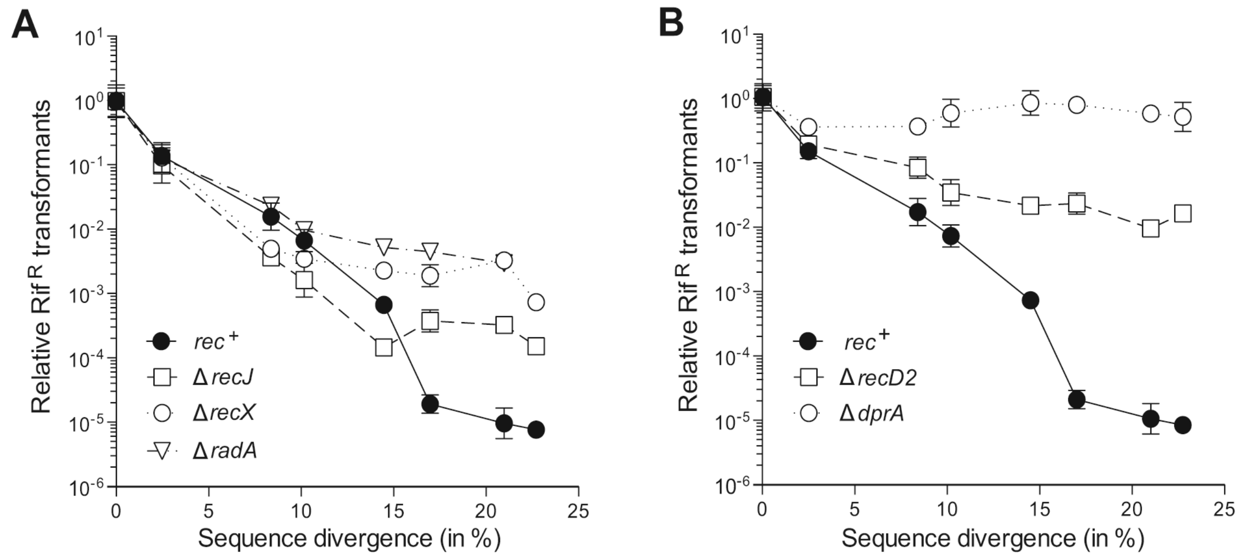
4 **Figure 1**. CT frequencies as a function of SD in different *rec* mutants. Donor DNA was a *rpoB*482 DNA
5 conferring Rif^R, derived from *B. subtilis* 168 (0.04% SD, homologous DNA), *B. subtilis* W23 (2.47%),
6 *B. atrophaeus* 1942 (8.35%), *B. amyloliquefaciens* DSM7 (10.12%), *B. licheniformis* DSM13 (14.52%),
7 *B. gobiensis* FJAT-4402 (17%), *B. thuringiensis* MC28 (20.83%) and *B. smithii* DSM4216 (22.74% SD).
8 The *rpoB*482 DNA (0.1 µg DNA/ml) from these different *Bacillus* species with the selectable Rif^R
9 mutation was used to transform BG1359 (*rec*$^+$, ●) competent cells and its isogenic derivatives. The
10 values are plotted dividing the number of transformants/CFUs obtained in each condition by the number
11 of transformants/CFUs obtained when the *rec*$^+$ cells are transformed with *Bsu*168 *rpoB*482 DNA. In (A):
12 BG1611 (*recF*15, ☐), BG1641 (Δ*recO*,○) and BG1633 (Δ*recA*, ▽); in (B): BG1653 (Δ*recU*, ☐),
13 BG1485 (Δ*ruvAB*, ○) and BG1631 (Δ*addAB*, ▽); in (C): BG1813 (Δ*recJ*, ☐), BG1397 (Δ*recX*, ○) and
14 BG1647 (Δ*radA*, ▽); in (D): BG1549 (Δ*recD*2, ☐) and BG1811 (Δ*dprA*, ○). All data points are mean
15 ± standard error of the mean (SEM) derived from 3 to 5 independent experiments

16

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.13.090589; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 **Figure 2**

2



3

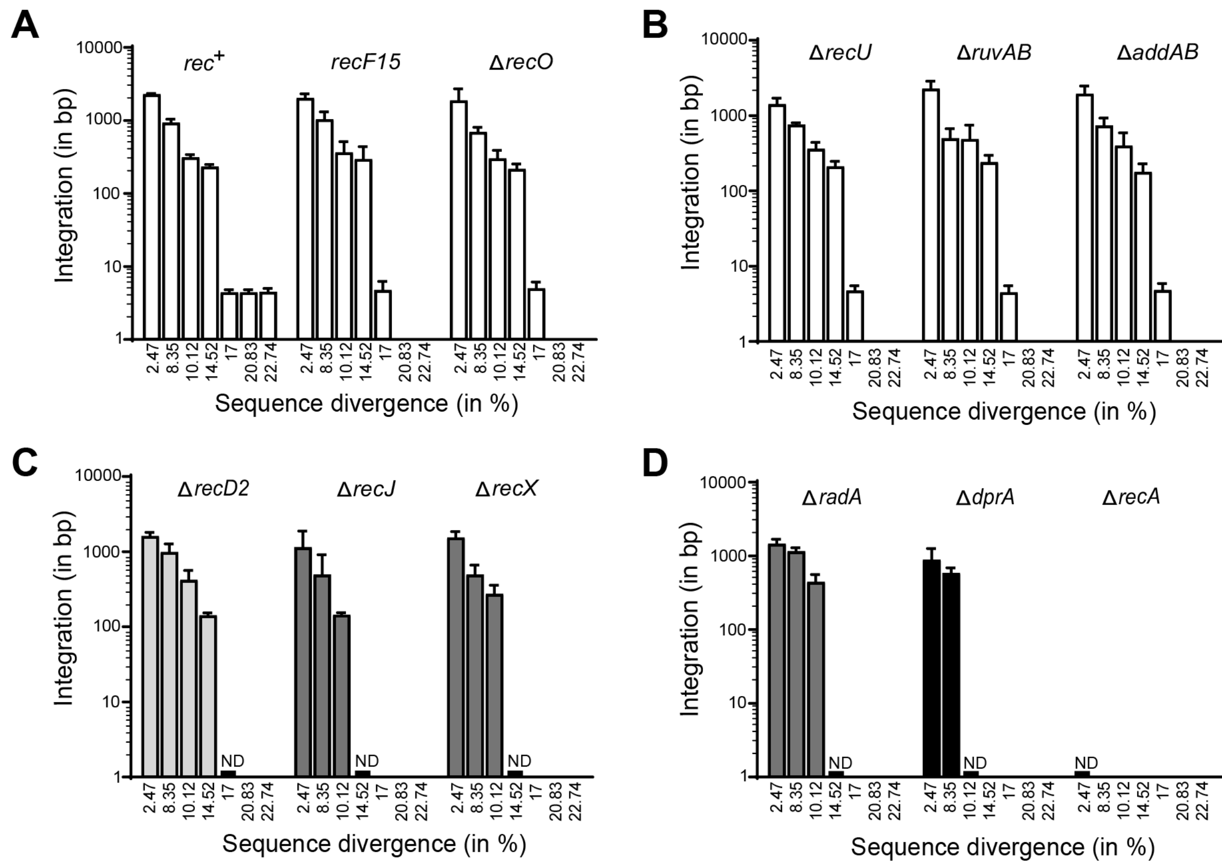
4

5 **Figure 2**. Normalised CT frequencies. The CT frequencies were normalized to give a value of 1 to the
6 transformation frequency of the indicated *rec*-deficient strain when transformed with homogamic DNA,
7 and heterogamic CTs are plotted relative to these values. In (A): Δ*recJ*, Δ*recX* and Δ*radA* (A); in (B):
8 Δ*recD*2 and Δ*dprA* Rif^R cells. For comparison, the values obtained for the *rec*$^+$ strain are also plotted.

9

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.13.090589; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 **Figure 3**



2

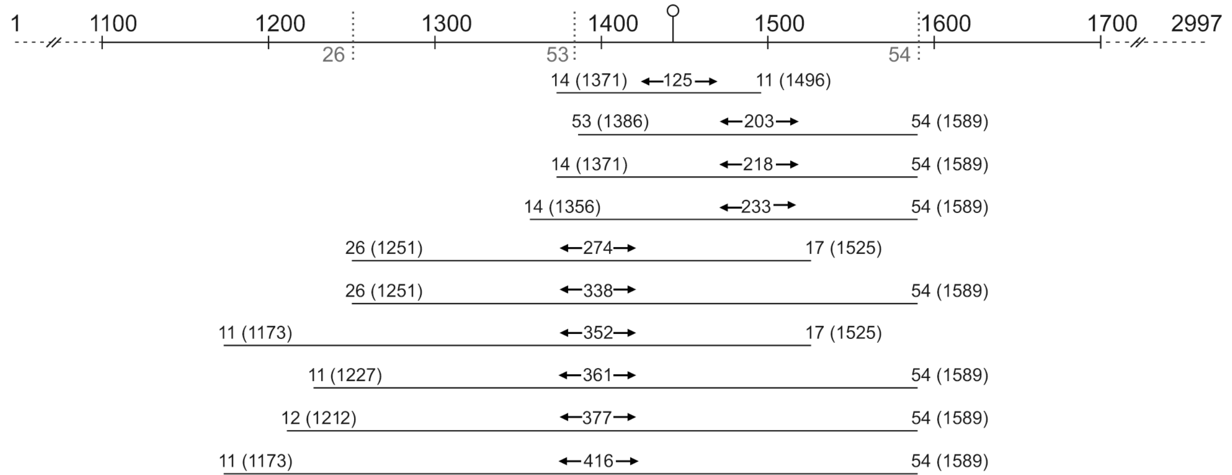
3

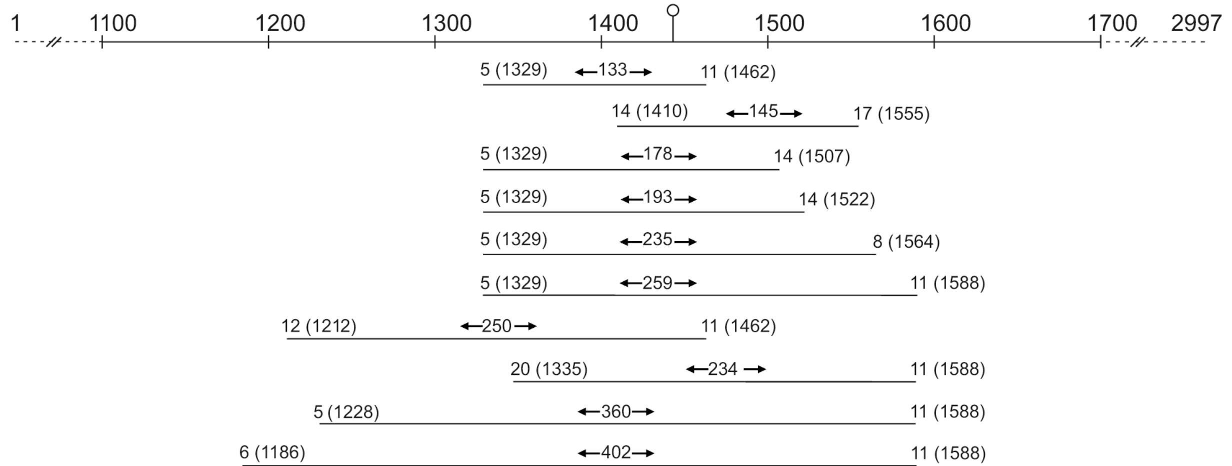
4 **Figure 3**. Determination of integration length of donor DNA with increased SD in different *rec* mutants.
5 The mean length of integration was determined by sequencing the 2997-bp *rpoB*482 DNA from different
6 transformants obtained in *rec*$^+$, *recF*15 and Δ*recO* (A); Δ*recU,* Δ*ruvAB* and Δ*addAB* (B); Δ*recD*2, Δ*recJ*
7 and Δ*recX* (C) and Δ*radA*, Δ*dprA* and Δ*recA* (D). Integration endpoints were defined as the average
8 between the last single nucleotide polymorphism present and the next SNP absent in the sequence of
9 trasformants (integration endpoint). The empty bars highlight the values obtained with strains that
10 undergo both homology-directed and homology-facilitated CT, and filled bars denote strains impaired in
11 homology-directed and blocked in homology-facilitated CT. ND, not detected.

12

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.13.090589; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 **Figure 4**





2

3

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

14

15