Molecular Microbiology (2018) 108(6), 697–710 ■
doi:10.1111/mmi.13959
First published online 25 April 2018

Co-evolution with recombination affects the stability of mobile genetic element insertions within gene families of *Salmonella*

Gerrit Brandis,† **Sha Cao**† **and Diarmaid Hughes***
Department of Medical Biochemistry and Microbiology, Box 582 Biomedical Center, Uppsala University, Uppsala, Sweden.

Summary

Bacteria can have multiple copies of a gene at separate locations on the same chromosome. Some of these gene families, including *tuf* (translation elongation factor EF-Tu) and *rrl* (ribosomal RNA), encode functions critically important for bacterial fitness. Genes within these families are known to evolve in concert using homologous recombination to transfer genetic information from one gene to another. This mechanism can counteract the detrimental effects of nucleotide sequence divergence over time. Whether such mechanisms can also protect against the potentially lethal effects of mobile genetic element insertion is not well understood. To address this we constructed two different length insertion cassettes to mimic mobile genetic elements and inserted these into various positions of the *tuf* and *rrl* genes. We measured rates of recombinational repair that removed the inserted cassette and studied the underlying mechanism. Our results indicate that homologous recombination can protect the *tuf* and *rrl* genes from inactivation by mobile genetic elements, but for insertions within shorter gene sequences the efficiency of repair is very low. Intriguingly, we found that physical distance separating genes on the chromosome directly affects the rate of recombinational repair suggesting that relative location will influence the ability of homologous recombination to maintain homogeneity.

Accepted 26 March, 2018. *For correspondence. E-mail diarmaid.hughes@imbim.uu.se; Tel. +46-18-4714507; Fax +46-18-4712000.
†These authors contributed equally to this work.

Introduction

Gene families with multiple copies of the same gene separately located on the chromosome are a frequent feature in prokaryotic and eukaryotic genomes. One common gene family in bacteria are the genes for translation elongation factor EF-Tu, *tufA* and *tufB*. The duplication of the *tuf* gene, that led to the creation of this gene family, most likely occurred early in the evolution of the eubacterial taxon (Sela *et al.*, 1989; Lathe and Bork, 2001). Despite the ancient origin of the duplication, the *tuf* genes of *Salmonella enterica* serovar Typhimurium strain LT2 differ at only 13 of 1185 nucleotides (Abdulkarim and Hughes, 1996). This high degree of nucleotide identity indicates that the *tuf* genes evolve in concert. It has been shown that homologous recombination between the genes leads to the exchange of genetic information, which facilitates this co-evolution (Abdulkarim and Hughes, 1996; Paulsson *et al.*, 2017) and a similar mechanism is expected to operate in other gene families. Co-evolution by recombination can limit the divergence of sequences in gene homologs (Abdulkarim and Hughes, 1996) or of functional domains within protein paralogs (Kruithof *et al.*, 2007).

Mobile genetic elements (MGE) play an important role in genome evolution in both eukaryotes (Feschotte and Pritham, 2007; Lisch, 2013) and prokaryotes (Chao *et al.*, 1983; Gaffé *et al.*, 2011). Insertion of a MGE into the chromosome can inactivate genes (Hayakawa *et al.*, 2001; Kumar *et al.*, 2014), with global effects if transcriptional regulators are affected (Jadoun and Sela, 2000; Gaffé *et al.*, 2011) or activate expression of adjacent genes by read-through transcription (Boyen *et al.*, 1978). MGEs also play an important role in the development of multidrug-resistance plasmids (Sandegren *et al.*, 2012) and clinical antibiotic resistance development (Komp Lindgren *et al.*, 2003; Sandegren and Andersson, 2009). Many MGEs use site-specific recombination for insertion into the chromosome at a highly preferred integration site (Hoess and Landy, 1978; Reiter *et al.*, 1989; Whittle *et al.*, 1999). This mechanism of integration makes these MGEs vulnerable to genetic

© 2018 The Authors. *Molecular Microbiology* Published by John Wiley & Sons Ltd.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

alterations within the integration site. One way to overcome this vulnerability is to use a highly conserved integration site where genetic alterations are associated with repercussions to the bacteria. A downside to this strategy is that genes that fit this description are likely to be essential and integration within them would reduce the viability of the bacterial cell. A possibility to circumvent this dilemma is site-specific insertion into conserved genes that exist in multiple copies on the bacterial chromosome. There are three main gene families in the *Salmonella* genome that fit this description, namely (i) the *tuf* genes (*tufA* and *tufB*) that encode elongation factor EF-Tu (Jaskunas *et al*., 1975; Hughes, 1986), (ii) up to eight copies of ribosomal RNA genes, *rrs* (16S rRNA), *rrl* (23S rRNA) and *rrf* (5S rRNA) (McClelland *et al*., 2001) and (iii) various tRNA gene families with up to six tRNA genes that encode the same anticodon (McClelland *et al*., 2001). Each of these gene families presents a potential target site for integration that is present in all (or most) bacteria, is highly conserved, and where inactivation of an individual gene within the gene family would not be a lethal event. Not surprisingly, tRNA genes frequently serve as integration sites for mobile genetic elements (Reiter *et al*., 1989; Hacker and Kaper, 2000) and enable such elements to integrate into widely divergent bacteria (Reinhold-Hurek and Shub, 1992). Other gene families, such as the *tuf* genes and ribosomal RNA genes do not seem to be frequently targeted by MGEs. This apparent lack of targeting is surprising due to the excellent characteristics for integration sites that these gene families display.

In this study, we set out to investigate why some gene families are not primary targets for integration of MGEs while others are. We constructed a selectable 3.3 kb long insertion cassette and used this to mimic integration of a MGE. We inserted this cassette into various positions within the *tuf* and *rrl* genes and studied the rates and the mechanism of recombinational repair that removes the inserted cassette from the chromosome. Our results indicate that homologous recombination protects the *tuf* and *rrl* genes from long-term inactivation by mobile genetic elements while tRNA genes probably lack this protection due to their smaller size. Additionally, we found evidence that the linear distance between genes on the chromosome affects the rate of recombinational repair.

Results

Recombination protects the tuf genes from inactivation by a large insertion

We constructed a 3.3 kb insertion cassette to mimic the size of a mobile genetic element. The cassette is a few hundred base pairs larger than large IS elements (around 2.5 kb) and could also represent the size of a small transposon consisting of a single gene flanked by two IS elements. The cassette consists of three genes: (i) a chloramphenicol resistance gene (*cat*) as a selectable marker to facilitate insertion of the cassette into the genome; (ii) a *sacB* gene to select for isolates that lost the cassette by recombination; and (iii) a *yfp* gene to screen for false positive colonies in the selection of recombinants (Fig. 1A). We inserted this cassette at seven locations in the *tufA* and *tufB* genes (Fig. 1B) and measured the rate of recombinational repair for each of the 14 insertion sites. One hundred independent cultures of each of the 14 different strains were grown overnight at 37°C in LB media before aliquots containing approximately 4×10^8 cells were spread on salt-free LA medium containing 5% sucrose to select putative recombinants that had lost the inserted cassette (Experimental Procedures). This experiment is setup as a fluctuation assay and the recombination rate was calculated using the P_0 method (Luria and Delbrück, 1943). A potential source of false positives in this selection will be provided by mutants in which the *sacB* gene is mutationally inactivated rather than removed, allowing growth of a colony on the selective medium. Such false positives were removed from the analysis by screening for simultaneous loss of YFP (nonfluorescent colonies under UV-light) and the remaining colonies were then assayed for the loss of the cassette from the *tuf* gene by PCR and subsequent sequencing. Successful recombination was observed for 11 of the 14 insertion sites (Supporting Information Table S1) and for each of the seven distinct locations in at least one of the *tuf* genes

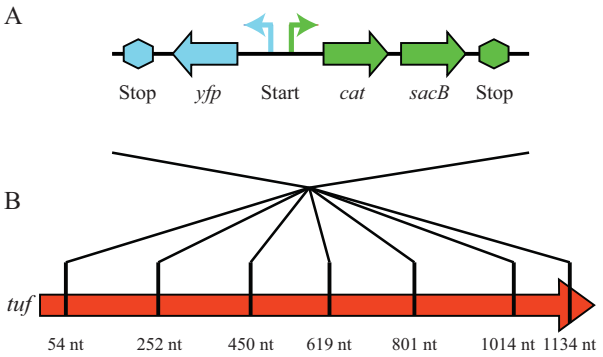

Fig. 1. Schematic overview of the insertion cassette and its insertion sites in the *tuf* genes.
A. The insertion cassette contains a *yfp* gene (blue) and a *cat-sacB* operon (green) that are transcribed in opposite directions. Each end of the insert contains a transcriptional terminator to inhibit transcription out of the cassette.
B. The cassette was inserted into seven distinct positions throughout the length of each *tuf* gene (red arrow representing *tuf*). The nucleotide position of each insertion, relative to the start of the *tuf* coding sequence, is indicated below the gene.

Table 1. Rates of recombination in different genetic backgrounds, selecting for the removal of the cassette inserted after nucleotide 619.

Strain[a]	Recombination rate (per cell per generation)	
	tufA(619 nt)<>*cat-sacB-yfp*	*tufB*(619 nt)<>*cat-sacB-yfp*
Wild type	1.81×10^{-10}	5.05×10^{-11}
Δ*recA*	$< 2.51 \times 10^{-11}$	$< 2.51 \times 10^{-11}$
Δ*recB*	$< 2.51 \times 10^{-11}$	$< 2.51 \times 10^{-11}$
Δ*recO*	1.81×10^{-10}	7.61×10^{-11}
Δ*mutS*	5.27×10^{-8}	4.66×10^{-8}
Δ*mutH*	1.28×10^{-8}	1.00×10^{-7}
tufB::*tufA* [b]	2.55×10^{-8}	n. d.

a. The mutations affecting recombinational repair present in each strain.
b. The *tufB* coding sequence was replaced with the *tufA* coding sequence.

(Supporting Information Table S1). The average recombination rate was 1.8×10^{-10} per cell per generation, but varied by an order of magnitude (2.5×10^{-11} per cell per generation to 6.9×10^{-10} per cell per generation). This variation is possibly dependent on the location of the insert, with the highest rates occurring between nucleotides 600-1,000, and very low rates close to each end of the *tuf* genes (Supporting Information Table S1). The average recombination rate is 10-fold to 100-fold lower than the rate previously measured for the repair of single point mutations within the *tuf* genes (Abdulkarim and Hughes, 1996). These data show that recombination between the *tuf* genes can repair, not only point mutations within the *tuf* genes but can also remove relatively large cassette insertions to restore gene function.

Analysis of the mechanism of recombination

We next analysed the underlying mechanism of recombination for the removal of our insertion cassette. It has previously been shown that recombination that repairs single point mutations within the *tuf* genes is dependent on RecA and RecBCD function but independent of RecFOR (Abdulkarim and Hughes, 1996). We deleted the *recA*, *recB* and *recO* genes in strains carrying the cassette inserted at position 619 nt in the *tufA* or *tufB* gene respectively. No recombination was detected for either of the *tuf* genes when *recA* or *recB* was deleted (Table 1). This shows that recombination to remove the inserted cassette is RecA- and RecB-dependent (below our detection limit of 2.5×10^{-11} per cell per generation). In contrast, inactivation of RecO showed no significant effect on the recombination rate (Table 1). These results indicate that the observed recombination originates from a double-stranded break repair event (Fig. 2). A double-stranded break would be generated within or close to the insert followed by a RecBCD-dependent bidirectional degradation of the insert. In this case, the existence of Chi sites close to the ends of the insert or within the *tuf* genes would be expected to have a major impact on the recombination event. Insertion sites that are flanked by a pair of appropriately oriented Chi sites would be expected to display a significantly increased recombination rate, which could explain the variation in the recombination rates observed for the various insertion sites (Supporting Information Table S1) (Smith *et al.*, 1981). A sequence analysis showed that the insertion cassette, as well as both *tuf* genes are devoid of Chi sites. It has previously been shown that Chi-like sites that differ in only one nucleotide from Chi can have some residual Chi-activity (Cheng and Smith, 1984) but the *tuf* genes contain only one such Chi-like site and only five of the seven inserts are located on the correct side of that Chi-like sequence to affect recombination (Supporting Information Fig. S1B). The insertion cassette does contain two Chi-like sites but only one is located close to the end of the insertion cassette and that one is in the wrong orientation to affect RecBCD-dependent DNA degradation that would be needed to remove the insert (Supporting Information Fig. S1A). It is, therefore, unlikely that these Chi-like sites contribute significantly to the recombinational repair rate (Fig. 2C). Previous studies have proposed RecB-dependent but Chi-independent recombination mechanisms. One possibility is that transcription complexes act as an obstacle to RecBCD progression on the DNA and thereby induce RecA loading (Fig. 2D) (Sivaramakrishnan *et al.*, 2017). Another possibility is the existence of a RecBC subpopulation that lacks RecD and can promote recombination without the need for Chi (Fig. 2E) (Churchill *et al.*, 1999).

A prediction that can be made for Chi-independent recombination events is that the rate of recombination should diminish as the insertion site approaches the end of homology. Due to the lack of Chi-sites within the homology region RecB-dependent degradation will eventually remove all homology and thereby render homologous recombination with another gene within the gene family impossible. Insertion sites closer to the end of

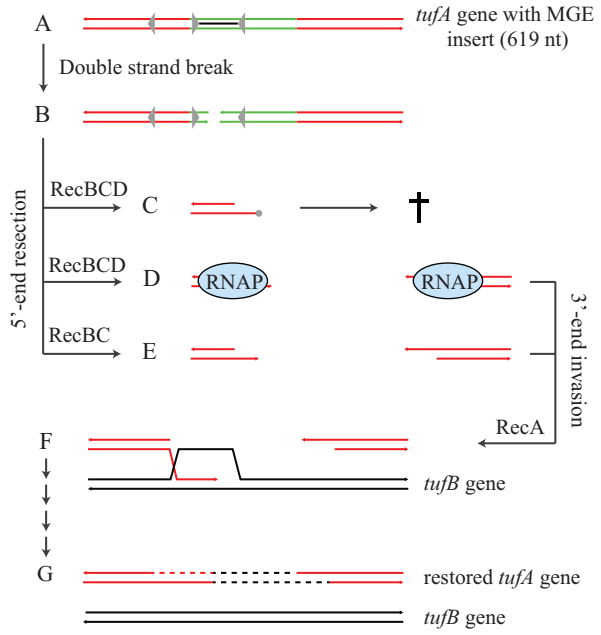

Fig. 2. Model for the recombinational removal of a mobile genetic element (green) inserted in the *tufA* gene (red) at position 619 nt. Grey arrows in the genes show Chi-like sequences that differ in only one nucleotide from Chi (5′-GCTGGTGG-3′) and indicate their direction of activity on RecBCD.
A, B. A double-stranded DNA break within or close to the MGE initiates the recombination event. The black line within the MGE sequence indicates the region where a double-stranded DNA break is predicted to be the most productive (no interfering Chi-like sequences within the insertion cassette).
C–E. 5′-end resection of the DNA leads to the removal of the MGE sequence by RecBCD in a (C) Chi-dependent manner; or (D) in a Chi-independent manner involving stalling of RecBCD by RNA polymerase complexes (Sivaramakrishnan *et al.*, 2017); or (E) Chi-independent recombination by RecBC complexes lacking RecD (Churchill *et al.*, 1999). A Chi-dependent recombination mechanism (C) by itself cannot lead to a successful removal of the MGE due to the lack of Chi-sites downstream the insertion site. Either of the Chi-independent recombination mechanisms (D, E), or a combination of the three potential mechanisms is proposed to lead to RecA-mediated 3′-end invasion of the homologous *tufB* sequence (black) (F). Polymerase extension from the invading 3′-end leads to the re-synthesis of the degraded DNA (shown as dashed lines) without the MGE sequence and the formation of a functional *tufA* gene (F).

homology reduce the time frame in which recombination between genes of a gene family can occur thus leading to a reduced recombination rate. The recombination rates of the various insertion sites showed some dependency on the location of the cassette insert, with the highest rates occurring between nucleotides 600–1,000 and very low rates close to each end of the *tuf* genes (Supporting Information Table S1). This skew in recombination rates towards the second half of the gene is somewhat unexpected. The insertion site in the centre of the gene has the greatest length of *tuf* sequence on both sites of the insertion and is, therefore, expected to result in the highest rate of recombinational repair. A possible explanation for this observation might be sequence differences between the *tuf* genes. The coding sequences of *tufA* and *tufB* in LT2 differ at thirteen nucleotide positions (Supporting Information Fig. S2) and almost all (10 out of 13) of these nucleotide differences are located in the first 600 bp of the gene, which might limit recombination in that region (Abdulkarim and Hughes, 1996). Mismatches in the homology are recognized by MutS and the endonuclease MutH is responsible for the specificity of strand recognition and cleavage. The removal of MutS or MutH function was shown to increase the recombination rate to repair single nucleotide point mutations by 100-fold to 1,000-fold. We deleted *mutS* and *mutH* in the two strains that carry the cassette insertion at location 619 nt in the *tufA* and *tufB* genes and measured recombination rates to test the influence of MutS and MutH activity on repair of the insertion. The results (Table 1) show that recombination rates increase 100-fold to 1,000-fold in strains with an inactive mismatch repair system. These data indicate that more than 99% of all recombination events are aborted due to the presence of nucleotide mismatches between the *tuf* genes. To further test this hypothesis, we constructed a strain in which the *tufB* coding sequence was replaced by the *tufA* sequence, thus removing all mismatches between the genes. We inserted the insertion cassette in the *tufA* gene at position 619 nt and measured the rate of recombinational repair. The recombination rate increased 100-fold and was indistinguishable from the rate measured in strains with *mutS* or *mutH* deletions (Table 1). These data show that imperfect homologies in the *tuf* gene sequences have a significant impact on the rate of recombination to remove an inserted cassette. To confirm the prediction that the recombination rate should diminish as the insertion site approaches the end of homology, we deleted the *mutS* gene in the seven strains that carry the insertion cassette in the different positions of the *tufA* gene. The recombination rates in these Δ*mutS* strains should be independent from nucleotide differences and, therefore, show the influence of location on recombination. Recombination rates were measured for these Δ*mutS* strains and the results show that the rates are correlated to the insertion site (Fig. 3). As expected, the highest recombination rate was observed for the insert closest to the centre of the gene and the rates decreased symmetrically when the cassette is moved closer to either end of the *tuf* gene, as expected for a two-ended recombination event (Fig. 3). The only exception is the insert at position 801 nt which displays a twofold higher recombination rate than expected (Fig. 3, Supporting Information Table S1). Growth measurements show that this discrepancy is most likely explained by differences in fitness. The insertion at

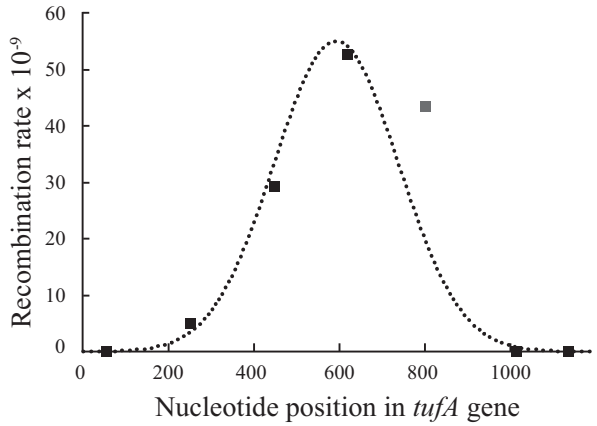

Fig. 3. Rate of recombinational repair as a function of insertion location in the *tufA* gene.
All rates were measured in mismatch repair-deficient strains. The insert at position 801 nt (grey marker) displays a significantly larger decrease of cellular fitness. The dashed line indicates a normal distribution of recombination rates with the maximum at the centre of the *tuf* gene.

position 801 nt causes a 43% reduction of fitness while all other six insertion sites display a reduction of only about 25% (Supporting Information Table S1). Recombinational repair of the insert at position 801 nt will, therefore, lead to an almost twofold higher improvement of cellular fitness compared to the other positions and lead to a higher apparent recombination rate. These data show that recombination rates are a function of location within the *tuf* gene and that insertions that maximize the homology lengths on both sites of the insert display the highest rates of recombination.

We conclude that the recombinational removal of the insert is initiated by a double-stranded DNA break within or close to the insert (Fig. 2B), followed by a RecB-dependent degradation of the inserted MGE DNA in a Chi-independent manner (Fig. 2C–E) and finally the break is repaired using RecA-dependent strand-invasion of a homologous DNA region (Fig. 2F and G).

Recombinational repair can lead to sequence transfer between homologous DNA segments

As a final test to show that recombination occurs between the two *tuf* genes, rather than a precise deletion of the inserted cassette, we analysed the sequences of the recombinant *tuf* genes after removal of the insert. During recombination that repairs single nucleotide point mutations DNA sequence information can be transferred from one *tuf* gene into the other (Abdulkarim and Hughes, 1996; Paulsson *et al*., 2017). The coding sequences of *tufA* and *tufB* in *Salmonella* LT2 differ at thirteen nucleotide positions (Supporting Information Fig. S2). Alterations at these nucleotide positions facilitate the determination of the region of sequence information that has been transferred during recombination. It has previously been shown that recombination that repairs single point mutations can: (i) transfer sequence information that contains up to 12 of the 13 nucleotide differences (all but the nt difference in the start codon) (Abdulkarim and Hughes, 1996) and (ii) that the probability of sequence transfer reduces as a function of the distance from the mutation that is selected for repair (Abdulkarim and Hughes, 1996; Paulsson *et al*., 2017). If the repair of an inserted genetic element were done by recombination between the *tuf* genes, as for repair of single nucleotide point mutations, then it would be expected that similar signatures in the patterns of sequence transferred should be observed. We analysed the sequences of all recombinants that were isolated and sequenced in this study. The results of that analysis showed that the pattern of sequence transfer associated with the repair of the inserted cassette is similar to that previously reported for the repair of single nucleotide point mutations. We observed sequence transfers from one *tuf* gene to the other, and these include 10 of the 13 nucleotide differences, with the exception of the two earliest ones and the last one. It is worthwhile noting that the length of sequence homology separating these three nucleotide differences from the beginning or the end of the *tuf* coding sequence, respectively, is less then 14 nucleotides (Supporting Information Fig. S2). This short length of homology makes successful sequence transfer very unlikely, as previously noted for repair of single nucleotide point mutations (Abdulkarim and Hughes, 1996). As expected the probability that any particular sequence (the 13 nts used as markers for recombination) was transferred, during the recombination event that removed an inserted cassette, was a function of the distance separating the marker nucleotide and the selected insertion element (Fig. 4). The pattern of sequence transfers observed in this study (Fig. 4) mimics the sequence transfers observed during recombination events that repair nucleotide point mutations in the *tuf* genes (Paulsson *et al*., 2017). These data confirm that the insertion cassettes were removed by recombination between the *tuf* genes rather than by precise deletions of the inserted DNA sequence.

Insertions within the tuf genes are unstable

Insertions in either of the *tuf* genes cause a growth fitness cost of around 25% due to inactivation of the targeted gene. This high fitness cost makes it very unlikely that an isolate with an insertion will be able to compete within a population. Nevertheless, there are three

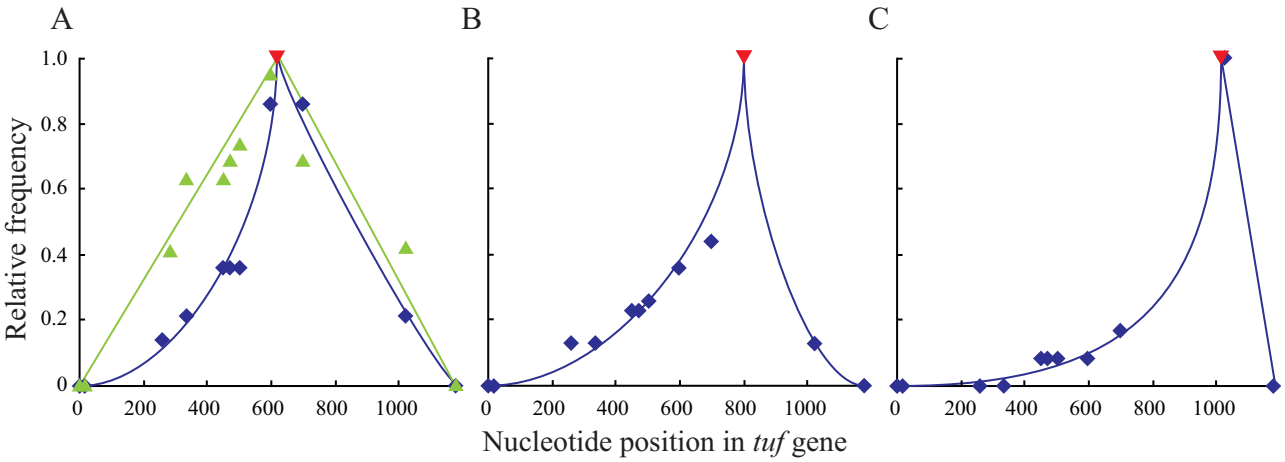

Fig. 4. Relative frequency of converted positions after recombinational repair of the insertion cassette in position (A) 619 nt, (B) 801 nt and (C) 1014 nt, as indicated by the red marker.
Blue diamond's indicates frequencies in mismatch repair-proficient strains and green triangles indicate frequencies of conversion in mismatch repair-deficient isolates. Solid lines indicate trend.

possible scenarios in which a stable population with an insertion can form, namely (i) the population undergoes a single-cell bottleneck so that the isolate with the insertion is the only cell present, (ii) the insert hitchhikes in an isolate with a secondary unrelated mutation that is beneficial and (iii) the insert carries a gene that gives a selective advantage (e.g., an antibiotic resistance gene). Each of these scenarios will lead to a population in which the insert is present in all cells. We next asked how stable the insert would be within such a population. To do this we modelled the evolution of a population with the insert in *tufA* at position 619 nt using experimentally determined parameters (growth rate and recombinational repair rate). The model is based on a *Salmonella* population with the insertion in *tufA* that is serially passaged to grow a total of 100 generations. We allowed growth to various total population sizes (4×10^7 cells to 4×10^{10} cells) and different bottleneck sizes during each passage (400 cells per passage to 4×10^9 cells per passage) as described in Experimental Procedures. In total, 20 distinct combinations were modelled and for each model 1,000 independent experiments were run. In each of these experiments recombinants that repair the insert in the *tuf* gene were allowed to appear at random and according to the experimentally determined rates of recombinational repair. After 100 generations of evolution, we determined the proportion of cells within the population that had lost the insert in the *tuf* gene (Supporting Information Table S2). The results showed that a stable subpopulation of isolates with a repaired *tuf* gene was present after 100 generations of growth for bottleneck sizes as low as 5×10^6 cells per passage (Fig. 5, Supporting Information Table S2). This subpopulation will rapidly outcompete the population with the insertion due to the large fitness benefit that comes with the recombinational repair of the *tuf* gene. We next asked how this evolution would look like in isolates that had fewer mismatches between the *tuf* genes. To mimic this evolution we used the rates of recombinational repair determined in the isolate that had the mismatches between the *tuf* genes removed and modelled 100 generations of evolution according to the previously mentioned parameters. As expected,

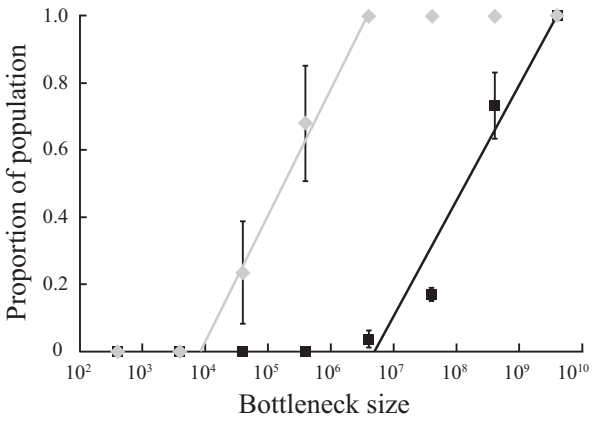

Fig. 5. Evolution model of a population with the insertion cassette in *tufA* at position 619 nt.
One hundred generation of growth were modelled with various total population and bottleneck sizes (Experimental Procedures) and the proportion of cells within the population that have repaired the insertion cassette by recombination was calculated. Markers indicate the average proportion of recombinant cells (± standard deviation) for each bottleneck size in wild-type *Salmonella* (black) and a *Salmonella* isolate that has no mismatches between the *tuf* genes (grey). Solid lines show the trend line that was used to calculate the minimal bottleneck size that yields a stable population of recombinant cells. All values are shown in Supporting Information Table S2.

recombinants appeared more frequently and a stable subpopulation was formed with only 8,500 cells per passage (Fig. 5, Supporting Information Table S2). These modelling results, based on experimentally determined parameters, indicate that insertions within the *tuf* genes are unstable. Even when an insert is successfully spread within a population, for example due to a transient selective advantage, recombinational repair together with a significant selective advantage will quickly lead to the removal of the insertion from the population.

Recombination leads to the removal of transposon-sized inserts

The insertion cassette that we used to measure the rates of recombinational repairs has a size of 3.3 kb and is in the size range of IS elements and small transposons. We decided to test whether an insertion cassette that is even larger (in the size range of larger transposons and small bacteriophages) could also be removed by recombination between the *tuf* genes. To test this we used a strain that had a transposition-defective transposon, MudJ, originally called MudI (Castilho *et al*., 1984) inserted in the *tufB* gene at position 606 nt. We inserted our small insertion cassette inside the sequence of the MudJ transposon so that the total size of the insert was about 15 kb. We deleted the *mutS* gene in this strain to increase the chances to observe recombination and measured the rate of recombinational repair. The results showed that the rate of recombinational repair was 7.6×10^{-9} per cell per generation, which is only a sixfold reduction compared to the strain with the 3.3 kb insertion in a comparable genetic background (Supporting Information Table S1). This shows that recombination between the *tuf* genes can lead to the removal of insertions of at least 15 kb length. This size range includes IS elements, transposons and small bacteriophages. It furthermore demonstrates that the recombinational repair is not limited to the artificial small cassette used in this study but was successful in the removal of the MudJ transposon.

Recombinational repair of insertions in ribosomal RNA genes

Another major class of gene families that are present in all bacteria are ribosomal RNA genes. We decided to test the rate of recombinational repair of our insertion cassette in the *rrl* genes that encode the 23S ribosomal RNA, to include a second gene family in our study. There are two features that distinguish the *rrl* gene family from the *tuf* genes; (i) the *rrl* genes are approximately 3 kb long, which is 2.5-fold larger than the *tuf* genes and (ii) there are seven copies of the *rrl* gene in the *Salmonella* chromosome, compared to two *tuf* genes (McClelland *et al*., 2001). We inserted the 3.3 kb insertion cassette in the middle of each of the seven *rrl* genes and measured the respective rates of recombinational repair (Supporting Information Table S1). Removal of the insert was observed for all seven *rrl* genes and the average recombination rate was 2.8×10^{-6} per cell per generation (1.7×10^{-7} per cell per generation to 5.6×10^{-6} per cell per generation), which is 15,000-fold more frequent than the average rate of recombinational repair when inserted in the *tuf* genes.

The rate of recombinational repair is a function of the distance to other homologs

There is a 30-fold variation in the rates of recombinational repair in the different *rrl* genes. This variation could be explained by differences in fitness costs caused by the insertions. Insertions that cause a higher fitness cost are expected to show a higher apparent rate of recombinational repair. We measured the fitness costs of the seven insertions and found that all insertions lead to a similar reduction in fitness of about $12\% \pm 2\%$. It is, therefore, unlikely that differences in relative fitness effects can account for the variation in recombination rates. The presence of Chi sequences within the homology region might be a possible explanation for this variation but the various *rrn* operons do not contain any Chi sequences and all appropriately oriented Chi-like sequences are present in all seven *rrn* operons (Supporting Information Fig. S1C and D). Another possibility is that imperfect homologies between the various *rrl* genes lead to the observed variation. We made a phylogenetic analysis of the *rrl* coding sequences to obtain a measure of sequence diversity (Fig. 6A). If imperfect homologies were responsible for reduced recombination rates, then it is most likely that individual *rrl* genes would recombine most efficiently with their most similar homolog. The level of imperfect homology towards the most similar homolog should, therefore, be the rate-limiting factor for recombinational repair. We used the phylogenetic distance (branch length of the phylogenetic tree) between any given *rrl* gene and its closest homolog as a measure of imperfect homology. Our results show that the rates of recombinational repair are indeed correlated with the phylogenetic distance to its closest homolog ($R^2 = 0.83$) (Fig. 6A). These data indicate that *rrl* genes that diverge from the other genes within the family will recombine less frequently. This would lead to a reduced level of co-evolution that would lead to further sequence divergence. This negative

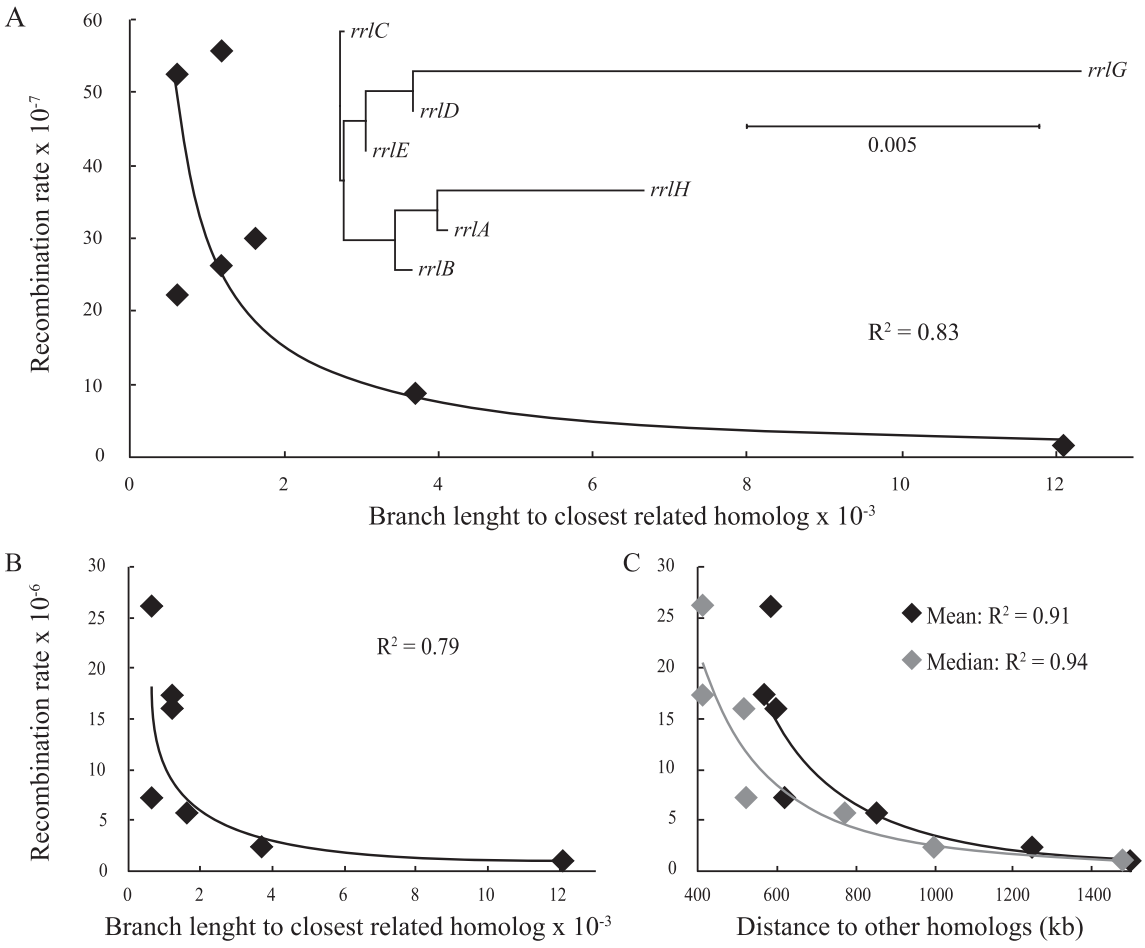

Fig. 6. Recombinational repair rates of the insertion cassette in the *rrl* genes as a function of phylogenetic and physical distance. A phylogenetic analysis of the *rrl* genes was performed (A) to measure the level of imperfect homologies between the different homologs. Recombination rates as a function of the branch length to the closest homolog was plotted for mismatch repair-proficient (A) and mismatch repair-deficient cells (B). (C) Recombinational repair rates of mismatch repair-deficient cells as a function of the mean (black) and median (grey) distance to the other *rrl* homologs on the chromosome.

feedback loop might be an explanation for why the *rrlG* and *rrlH* genes of *Salmonella* have evolved to be more different from the other *rrl* genes (Fig 6A). We decided to test this hypothesis by deleting the *mutS* gene in the seven isolates with insertion in the *rrl* genes. If imperfect homologies were the reason for the differences in the rates of recombinational repair, then deleting MutS should remove this variation. Our results show that deleting MutS leads to a twofold to sevenfold increase in rates of recombinational repair. Surprisingly, the variation in recombination rates was not reduced in the isolates without *mutS* gene (Supporting Information Table S1). Furthermore, recombinational repair rates in the Δ*mutS* strains still correlated with the phylogenetic distance to its closest homolog ($R^2 = 0.79$) (Fig. 6B). This indicates that the imperfect homologies between the *rrl* genes are not the cause of the reduced recombination rates. It is more likely that the causality is the reverse, such that reduced rates of recombinational repair lead to more diversification in the *rrl* genes. A second possible explanation for the differences in recombinational repair rates is the linear distance on the chromosome between the *rrl* genes and their homologs. To test for this we compared the recombinational repair rates of the *rrl* genes, in strains with deletion of MutS to reduce the impact of imperfect homologies, with the mean as well as the median distance to their six respective homologs. We found that the recombinational repair rates correlated very well with both the mean ($R^2 = 0.91$) and median ($R^2 = 0.94$) linear distance on the chromosome such that recombination rates decreased with increasing distance to the homologs (Fig. 6C). These results indicate that the chromosomal position of genes within a gene family could have a direct impact on their evolutionary divergence. Genes that are physically more distant to other genes within

the family (such as *rrlG* and *rrlH*, relative to *rrlA, B, C, D and E*) display a reduced rate of recombination, which leads to a more relaxed co-evolution. These genes are, therefore, more likely to accumulate genetic changes and diverge at the nucleotide level from the other genes within the family.

Recombination is a function of distance to the nearest end of the homologous sequence

Recombination rates for the removal of the insertion cassette in the *tuf* genes were a function of location within the gene (Fig. 3). The highest rate was observed for the cassette insertion closest to the middle of the gene and rates decreased symmetrically when the inserted cassette was moved towards either end of the *tuf* gene (Fig. 3). This pattern is consistent with the idea that the distance of the insert to the nearest end of the homologous sequence is the main factor determining the rate of recombination. The absence of appropriately oriented Chi or Chi-like sequences in the *rrn* operons indicates that the recombinational removal of the insert from the *rrl* genes is also achieved in a Chi-independent recombination mechanism (Supporting Information Fig. S3). The recombinational repair rate for the *rrl* genes should, therefore, display a similar dependence on distance of the insert to the nearest end of the homologous sequence. We compared the rates of recombination for insertions in the *tuf* and the *rrl* genes to the distance to the respective nearest ends of the homologous sequence. For this analysis, we used the values from Δ*mutS* strains to reduce the impact of mismatch repair at SNPs within the homologous sequence on the rates of recombinational repair (Fig. 7). The results show that the rate of recombinational repair increases exponentially as a function of increasing length of sequence identity ($R^2 = 0.89$). We decided to extrapolate from these results a potential rate of recombinational repair in tRNA genes. This gene family is, unlike the *tuf* and *rrl* genes a frequent insertion target of mobile genetic elements (Reiter *et al.*, 1989). Although in many cases the site of insertion is at the 3′end of a tRNA gene (Hacker and Kaper, 2000), mobile elements have also been shown to insert within tRNA genes, adjacent to the anticodon sequence (Reinhold-Hurek and Shub, 1992). tRNA genes are approximately 80 bp long, which means that the length of perfect sequence identity to the nearest site of mismatch will not exceed 40 bp. Under these circumstances we calculated that the rate of recombinational repair would be 5×10^{-12} per cell per generation. This number is an estimate of the upper limit of the rate of recombinational repair of an insertion into a tRNA gene but indicates that insertions in tRNA genes are very unlikely to be removed by recombination. This conclusion is consistent with previous measurements showing that the lower limit for detectable RecA-mediated homologous recombination in *E. coli* requires about 20 to 40 bps of perfect homology (Watt *et al.*, 1985; Shen and Huang, 1986). Taken together, these data indicate that the distance to the nearest end of the homologous sequence is rate limiting for RecA-mediated recombination and that recombinational repair of insertions in tRNA genes will occur at a very low rate, which might be one explanation (in addition to relative differences in fitness cost) for why tRNA genes are targeted for insertion by mobile genetic elements.

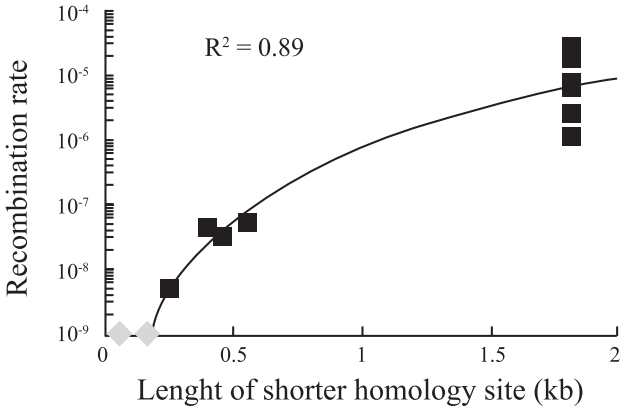

Fig. 7. Recombinational repair rates as a function of the length of sequence identity to the nearest end of homologous sequence. Values are rates of recombinational repair of the insertion cassette in the *tufA* and *rrl* genes in mismatch repair-deficient cells. Isolates that displayed no recombination are indicated in grey. Solid line indicates trend that was used to estimate recombinational repair rates for tRNA genes.

Discussion

In this study, we have shown that gene inactivation by MGE insertion is reversible in two very important bacterial multigene families, *tuf* and *rrl*. Removal of the insertion was not limited by the insertion size but became progressively less efficient as the length of sequence identity from the site of insertion to the end of the region of homology becomes shorter. Furthermore, we found evidence that the linear distance separating genes on the chromosome affects the rate of recombinational repair.

The underlying mechanism of the recombinational removal of MGEs is RecA and RecB dependent but independent of RecO function, which implies that it involves the repair of double stranded DNA breaks by homologous recombination (Fig 2, Supporting Information Fig. S3). One of the most important functions of this homologous recombination pathway may be the repair of collapsed replication forks (Asai *et al.*, 1994;

© 2018 The Authors. *Molecular Microbiology* Published by John Wiley & Sons Ltd.,

Kuzminov, 1995). Our data suggest that homologous recombination might also be crucially important in the maintenance of dispersed gene families, to counteract the affects of mutational damage associated with Muller's ratchet (Muller, 1964). Muller's ratchet, the accumulation of mutations over time leading to long-term declines in relative fitness, has been shown to operate on bacterial populations (Muller, 1964; Andersson and Hughes, 1996). Genes within gene families should be relatively more susceptible to the effects of Muller's ratchet since the fitness costs associated with inactivating mutations in individual genes within the family would have a lower effect on bacterial fitness than they would have on equivalent single-copy genes. Without a recombinational repair mechanism individual members of a gene family would be susceptible to inactivation by mutations or insertions of MGEs over time leading eventually to only a single copy being active. Since recombination between genes within gene families is not unidirectional it is equally likely that mutations/insertions are further spread within the family, rather than removed. Our assay is setup to measure loss of the insert and is not able to detect its spread. However, for the insertions in the *tuf* genes a spread into the second copy of *tuf* would be a lethal event and, therefore, not result in a viable cell. Since *Salmonella* contains seven ribosomal RNA operons it would be possible that an MGE could be moved into additional copies without it being lethal to the cell. An insertion within a single *rrl* gene causes a 12% fitness cost (Supporting Information Table S1) so every further copy of the *rrl* gene that is silenced by an MGE would result in additional substantial loss of cellular fitness. These cells would rapidly been outcompeted by cells within the population that have more functional copies of *rrl*, unless this fitness cost is counterbalanced by an selective advantage conferred by the MGE. These expectations would be opposite for a mutation or insert that would confer a selective advantage. In this case recombination coupled with purifying selection would lead to the spread of the mutation or insertion within the gene family, rather than its loss. The existence of multiple copies of a gene increases the total probability of beneficial mutations/insertions within that gene and recombination can facilitate spreading to the other copies within the family. Gene families are, therefore, expected to adapt faster to selective conditions than single-copy genes, at least for dominant mutations. Recombinational repair, as described in this study, may, therefore, play an important part in maintaining the integrity and increasing the adaptive speed of gene families over evolutionary timescales.

From the point of view of the MGE, a major challenge is how to break the association between insertion into a chromosome and reducing bacterial fitness. Our data show that the inactivation of even a single member of a gene family can have significant fitness costs (Supporting Information Table S1) and purifying selection would be expected to remove these isolates from the population. There are different ways in which MGEs could potentially reduce the probability of loss due to purifying selection, namely (i) chance (e.g., a small population bottleneck), (ii) a counter-balancing selective benefit supplied by the MGE (e.g., drug resistance) and (iii) removal of the fitness cost. There is a class of MGEs that have evolved a way to prevent the inactivation of the gene they are inserted into, thus removing the fitness cost associated with their integration. Self-splicing introns are capable of cutting themselves post-transcriptionally out of the RNA leading to the production of a mature functional RNA (Cech, 1990). In bacteria, self-splicing introns are mainly found in tRNA genes (Paquin *et al.*, 1999) but there are examples of introns found in bacterial 23S rRNA *rrl* genes (Everett *et al.*, 1999; Edgell *et al.*, 2000) and 16S rRNA *rrs* genes (Salman *et al.*, 2012). Thus, *Simkania negevensis*, a bacterium in the *Chlamydiales* order, contains a group I intron within the *rrl* gene (Everett *et al.*, 1999). The intron in this case is not a target for recombinational repair since *S. negevensis* has only one ribosomal RNA operon (Kahane *et al.*, 1999). Another example of self-splicing group I introns within ribosomal RNA genes is in the ciliated protozoa *Tetrahymena* (Kruger *et al.*, 1982). Some species of *Tetrahymena* contain an intron within their 26S rRNA gene. Interestingly, this intron is present in all copies of the 26S rRNA gene, thus inhibiting recombinational removal of the intron. These examples show that ribosomal RNA genes can indeed be targets for MGE insertion. The fact that the phenomenon is observed in organisms that have only one rRNA operon (Kahane *et al.*, 1999) or where the insertion occurs in all rRNA operons (Kruger *et al.*, 1982) is consistent with there being a selective force against insertions in only some genes within the gene family. Our data indicate that recombinational removal of the insert from the genome, associated with increased relative fitness, could be this selective force. Even though gene families are promising target sites for MGEs it might be that the combination of purifying selection and recombinational removal pose too big a challenge. In this context, the frequent targeting of tRNA genes by MGEs may be a result of the much lower fitness costs associated with gene inactivation, and the much lower rates of recombinational repair due to their small size (Fig. 7).

The biological rationale for having multiple copies *tuf* and *rrn* is at least partly explained by the need to produce very high concentrations of their products (EF-Tu and rRNA) to support protein synthesis and fast growth. This reasoning is consistent with the chromosomal locations of these genes, being situated relatively close to the origin of

replication to benefit from the effects of gene dosage associated with overlapping rounds of chromosome replication (Cooper and Helmstetter, 1968; Helmstetter and Cooper, 1968), and being transcribed in the same direction as replication, which reduces the probability of potentially disruptive clashes between DNA and RNA polymerases involved in replication and transcription respectively (French, 1992; Rocha and Danchin, 2003; Mirkin and Mirkin, 2005; De Septenville *et al.*, 2012; Lang *et al.*, 2017). Nevertheless, regardless of their critical importance for bacterial fitness, the individual members of gene families can be expected to accumulate mutations independently of each other, and over time, subject to the frequency of recombinational repair, this would lead their sequences to diverge. In the case of the *tuf* and *rrn* genes, given the functional constraints on these genes such divergence might be expected to significantly reduce bacterial fitness. Our data indicate that there might be an extra advantage, in addition to gene dosage and polymerase collision avoidance, to concentrating the members of a gene family within the same chromosomal region. The *rrl* genes that are closer to each other on the chromosome show a higher rate of recombinational repair than those that are more distant, thus, reducing the rate of accumulation of mutations and insertions that would potentially reduce bacterial fitness. This might be a selective force favouring the location of members of gene families in close proximity to each other. The obverse of this argument is that the relative location of genes on the chromosome could also affect the evolution of new gene functions. Genes that are physically more distant to other members within the family (such as *rrlG* and *rrlH*, relative to *rrlA, B, C, D* and *E*) display a reduced rate of recombination and are, therefore, more likely to accumulate genetic changes and diverge at the nucleotide level from the other genes within the family. Although this is unlikely to be favoured for *rrn* operons (because of their functional association with fast growth rate) it could play a role in the divergent evolution of gene families less constrained by strong selection. This divergence can lead to the evolution of new genes and functions (Ohno, 1970; Bergthorsson *et al.*, 2007; Nasvall *et al.*, 2012). Accordingly, greater distance between genes in a family on the chromosome could favour the ability to successfully evolve separate functional genes in the face of recombinational repair.

Experimental procedures

Bacterial strains and strain constructions

All strains are derivatives of *Salmonella enterica* serovar Typhimurium, strain LT2 (McClelland *et al.*, 2001). The *cat-sacB-yfp* cassette was inserted into *tufA*, *tufB* and *rrl* genes using dsDNA recombineering (Yu *et al.*, 2000). All recombineering primers are shown in Supporting Information Table S3. Genetic markers were moved between strains by P22 HT phage-mediated transduction (Schmieger and Backhaus, 1973).

Bacterial growth conditions

Bacteria were grown in LB medium (10% tryptone, 5% yeast extract (Oxoid, Basingstoke, England) and 10% NaCl, (Merck, Darmstadt, Germany)) or on LA plates (LB with 1.5% agar, Oxoid). Sucrose counter-selection was performed on salt-free LA plates with 5% sucrose (Sigma Aldrich, Steinheim, Germany). When indicated, the media was supplemented with tetracycline (15 mg l^{-1}) or chloramphenicol (25 mg l^{-1}; Sigma Aldrich). Growth rates were measured using a Bioscreen C machine (Oy Growth Curves Ab Ltd). Overnight cultures were diluted 2,000-fold in LB and 300 μl of the diluted culture were incubated at 37°C with continuous shaking in honeycomb microtiter plates, with optical density readings at 5 min intervals and the doubling time was calculated from the increase in optical density at 600 nm.

Measuring recombination rates

Cultures were inoculated in 2 ml LB, which were grown with shaking at 37°C to a final cell density of approximately 4×10^9 cfu ml^{-1}. An appropriate number of cells (Supporting Information Table S4) was plated on sucrose plates to select for cells that had lost the inserted *cat-sacB-yfp* cassette (Gay *et al.*, 1985). After overnight growth colonies were assessed under UV light for fluorescence, and non-fluorescent colonies (lack of fluorescence indicates loss of *yfp* activity and, therefore, possible loss of the entire cassette) and nonfluorescent colonies were restreaked on sucrose plates for purification. Removal of the cassette was confirmed by PCR and sequencing.

Recombination rates were calculated according to the formula

$$\mu = -(1/N)\ln(P_0) \quad (1)$$

where μ is the recombination rate, N is the number of viable cells plated and P_0 is the proportion of cultures where no recombination was scored. The total number of independent cultures for each strain is shown in Supporting Information Table S4.

Ninety five percent C. I. were estimated for each P_0 value according to the formula

$$P_0 \pm 1.96 \cdot \sqrt{\frac{P_0(1-P_0)}{N}} \quad (2)$$

and used to calculate estimated 95% C. I. for each recombination rate according to formula (1).

PCR and DNA sequencing

Polymerase chain reactions (PCR) for recombineering were done using Phusion High Fidelity PCR Master Mix (New

© 2018 The Authors. *Molecular Microbiology* Published by John Wiley & Sons Ltd.,

England Biolabs, Ipswich, USA). For screening the presence or absence of cassette inserts, and for preparing DNA for sequencing, the Taq polymerase based PCR Master Mix from Thermo Scientific (Waltham, USA) was used. Reactions were run on a S1000 Thermal Cycler from Bio-Rad (Hercules, USA). Oligos for PCR were all purchased from Sigma Aldrich. Denaturing temperature for PCR reactions was either 98°C or 95°C, for Phusion or Taq polymerase respectively. The primer annealing temperature was calculated as 5°C below the primer melting temperature and the elongation time set as one minute per kb of product length. PCR products were prepared with the QIAquick PCR Purification Kit (Qiagen). Local sequencing was carried out by Macrogen Incorporated, Amsterdam, The Netherlands. The software CLC Main Workbench 7.7.2 from Qiagen was used for primer design, sequence analysis and sequence comparisons.

Phylogenetic analysis

Phylogenetic analysis of the *rrl* genes was performed using software CLC Main Workbench 7.7.2 from Qiagen.

Evolution model

The evolution model was designed to model 100 generations of growth of a population with an insertion in the *tufA* gene (ins) and to model the appearance of a wild-type population (wt) that arises by a recombinational event that removes the insertion in the *tufA* gene. Fitness cost and rates of recombinational repair of the insertion cassette were determined experimentally (Supporting Information Table S1) and used for the evolution model.

The bacterial growth rate is a monotonically increasing function of the concentration of a limiting resource, R (mg l^{-1}) (Monod, 1949)

$$\psi_i(R) = V_i\left(\frac{R}{R+k}\right) \tag{3}$$

where V_i is the relative fitness of the *i*th strain of bacteria and k is the concentration of the resource at which V_i is at half its maximum value. With these definitions the change in densities of bacterial populations and the concentration of resources are given by the following two coupled differential equations:

$$\frac{dR}{dt} = -\sum_{i=1}^{2} n_i * \psi_i(R) * e_n \tag{4}$$

$$\frac{dn_i}{dt} = n_i * \psi_i(R) \tag{5}$$

where n_i is the density of strain i (cfu ml^{-1}) and e is the conversion efficiency parameter (μg $cell^{-1}$). The standard parameters $R_{t=0} = 100$ mg l^{-1}, $k = 1$ mg l^{-1}, $e_1 = 3 \times 10^{-6}$ μg $cell^{-1}$, $e_2 = 3 \times 10^{-7}$ μg $cell^{-1}$, $e_3 = 3 \times 10^{-8}$ μg $cell^{-1}$ and $e_4 = 3 \times 10^{-9}$ μg $cell^{-1}$ result in a growth cycle that leads to a final density of approximately 4×10^7 cfu ml^{-1} for e_1, 4×10^8 cfu ml^{-1} for e_2, 4×10^9 cfu ml^{-1} for e_3 and 4×10^{10} cfu ml^{-1} for e_4. After every cycle the culture is diluted with the dilution factor d into fresh media and grown to full density. Serial passaging was repeated until a total growth of 100 generations. A Monte Carlo procedure was used to determine the appearance of wild-type cells by recombinational repair of the insertion. The probability $p_{wt}(t)$ that the wild-type strain is generated by a recombinational event at time point t is

$$p_{wt}(t) = g_{ins} * \mu_{ins} \tag{6}$$

where g_{ins} is the number of generations of growth of the strain with the insertion at time point t, and μ_{ins} is the rate of recombinational repair that changes the strain with insertion into a wild-type strain. A random number $\times$ $(0 < \times < 1)$ is generated. A single cell of strain wt generated at time point t if $\times < p_{wt}(t)$. The simulation was programmed in Berkeley Madonna (Version 9.0.100) and run with varying bottlenecks and total population sizes. All results are averages of 1,000 independent simulations.

Acknowledgements

This work was supported by grants to DH from the Swedish Science Research Council (Vetenskapsrådet, grant numbers 2016-04449, 2017-03953) and the Carl Trygger Foundation, (grant numbers CTS16:194, CTS17:204). The authors have no conflict of interest.

Author contributions

G.B., S.C. and D.H. designed the study. G.B. and S.C. performed the experiments and analysed the data. G.B., S.C. and D.H. wrote the manuscript.

References

Abdulkarim, F., and Hughes, D. (1996) Homologous recombination between the tuf genes of *Salmonella typhimurium*. *J Mol Biol* **260**: 506–522.

Andersson, D.I., and Hughes, D. (1996) Muller's ratchet decreases fitness of a DNA-based microbe. *Proc Natl Acad Sci USA* **93**: 906–907.

Asai, T., Bates, D.B., and Kogoma, T. (1994) DNA replication triggered by double-stranded breaks in *E. coli*: dependence on homologous recombination functions. *Cell* **78**: 1051–1061.

Bergthorsson, U., Andersson, D.I., and Roth, J.R. (2007) Ohno's dilemma: evolution of new genes under continuous selection. *Proc Natl Acad Sci USA* **104**: 17004–17009.

Boyen, A., Charlier, D., Crabeel, M., and Cunin, R. (1978) Studies on the control region of the bipolar argECBH operon of *Escherichia coli*. *Mol Gen Genet* **161**: 185–196.

Castilho, B.A., Olfson, P., and Casadaban, M.J. (1984) Plasmid insertion mutagenesis and lac gene fusion with

mini-mu bacteriophage transposons. *J Bacteriol* **158**: 488–495.
Cech, T.R. (1990) Self-splicing of group I introns. *Annu Rev Biochem* **59**: 543–568.
Chao, L., Vargas, C., Spear, B.B., and Cox, E.C. (1983) Transposable elements as mutator genes in evolution. *Nature* **303**: 633–635.
Cheng, K.C., and Smith, G.R. (1984) Recombinational hotspot activity of Chi-like sequences. *J Mol Biol* **180**: 371–377.
Churchill, J.J., Anderson, D.G., and Kowalczykowski, S.C. (1999) The RecBC enzyme loads RecA protein onto ssDNA asymmetrically and independently of chi, resulting in constitutive recombination activation. *Genes Dev* **13**: 901–911.
Cooper, S., and Helmstetter, C.E. (1968) Chromosome replication and the division cycle of *Escherichia coli* B/r. *J Mol Biol* **31**: 519–540.
De Septenville, A.L., Duigou, S., Boubakri, H., and Michel, B. (2012) Replication fork reversal after replication-transcription collision. *PLoS Genet* **8**: e1002622.
Edgell, D.R., Belfort, M., and Shub, D.A. (2000) Barriers to intron promiscuity in bacteria. *J Bacteriol* **182**: 5281–5289.
Everett, K.D.E., Kahane, S., Bush, R.M., and Friedman, M.G. (1999) An Unspliced Group I Intron in 23S rRNA Links Chlamydiales, Chloroplasts, and Mitochondria. *J Bacteriol* **181**: 4734–4740.
Feschotte, C., and Pritham, E.J. (2007) DNA transposons and the evolution of eukaryotic genomes. *Annu Rev Genet* **41**: 331–368.
French, S. (1992) Consequences of replication fork movement through transcription units in vivo. *Science* **258**: 1362–1365.
Gaffé, J., McKenzie, C., Maharjan, R.P., Coursange, E., Ferenci, T., and Schneider, D. (2011) Insertion sequence-driven evolution of *Escherichia coli* in chemostats. *J Mol Evol* **72**: 398–412.
Gay, P., Le Coq, D., Steinmetz, M., Berkelman, T., and Kado, C.I. (1985) Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. *J Bacteriol* **164**: 918–921.
Hacker, J., and Kaper, J.B. (2000) Pathogenicity islands and the evolution of microbes. *Annu Rev Microbiol* **54**: 641–679.
Hayakawa, T., Satta, Y., Gagneux, P., Varki, A., and Takahata, N. (2001) Alu-mediated inactivation of the human CMP- N-acetylneuraminic acid hydroxylase gene. *Proc Natl Acad Sci USA* **98**: 11399–11404.
Helmstetter, C.E., and Cooper, S. (1968) DNA synthesis during the division cycle of rapidly growing *Escherichia coli* B/r. *J Mol Biol* **31**: 507–518.
Hoess, R.H., and Landy, A. (1978) Structure of the lambda att sites generated by int-dependent deletions. *Proc Natl Acad Sci USA* **75**: 5437–5441.
Hughes, D. (1986) The isolation and mapping of EF-Tu mutations in *Salmonella typhimurium*. *Mol Gen Genet* **202**: 108–111.
Jadoun, J., and Sela, S. (2000) Mutation in csrR global regulator reduces *Streptococcus pyogenes* internalization. *Microb Pathog* **29**: 311–317.
Jaskunas, S.R., Lindahl, L., Nomura, M., and Burgess, R.R. (1975) Identification of two copies of the gene for the elongation factor EF-Tu in *E. coli*. *Nature* **257**: 458–462.
Kahane, S., Everett, K.D.E., Kimmel, N., and Friedman, M.G. (1999) Simkania negevensis strain ZT: growth, antigenic and genome characteristics. *Int J Sys Bacteriol* **49**: 815–820.
Komp Lindgren, P., Karlsson, A., and Hughes, D. (2003) Mutation rate and evolution of fluoroquinolone resistance in *Escherichia coli* isolates from patients with urinary tract infections. *Antimicrob Agents Chemother* **47**: 3222–3232.
Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E., and Cech, T.R. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell* **31**: 147–157.
Kruithof, E.K.O., Satta, N., Liu, J.W., Dunoyer-Geindre, S., and Fish, R.J. (2007) Gene conversion limits divergence of mammalian TLR1 and TLR6. *BMC Evol Biol* **7**: 148.
Kumar, R., Grover, S., Kaushik, J.K., and Batish, V.K. (2014) IS30-related transposon mediated insertional inactivation of bile salt hydrolase (bsh1) gene of *Lactobacillus plantarum* strain Lp20. *Microbiol Res* **169**: 553–560.
Kuzminov, A. (1995) Collapse and repair of replication forks in *Escherichia coli*. *Mol Microbiol* **16**: 373–384.
Lang, K.S., Hall, A.N., Merrikh, C.N., Ragheb, M., Tabakh, H., Pollock, A.J., *et al*. (2017) Replication-transcription conflicts generate R-loops that orchestrate bacterial stress survival and pathogenesis. *Cell* **170**: 787–799.e18.
Lathe, W.C., and Bork, P. (2001) Evolution of tuf genes: ancient duplication, differential loss and gene conversion. *FEBS Lett* **502**: 113–116.
Lisch, D. (2013) How important are transposons for plant evolution? *Nat Rev Genet* **14**: 49–61.
Luria, S.E., and Delbrück, M. (1943) Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* **28**: 491–511.
McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W., Latreille, P., Courtney, L., *et al*. (2001) Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature* **413**: 852–856.
Mirkin, E.V., and Mirkin, S.M. (2005) Mechanisms of transcription-replication collisions in bacteria. *Mol Cell Biol* **25**: 888–895.
Monod, J. (1949) The growth of bacterial cultures. *Annu Rev Microbiol* **3**: 371–394.
Muller, H.J. (1964) The relation of recombination to mutational advance. *Mut Res* **1**: 2–9.
Nasvall, J., Sun, L., Roth, J.R., and Andersson, D.I. (2012) Real-time evolution of new genes by innovation, amplification, and divergence. *Science* **338**: 384–387.
Ohno, S. (1970) *Evolution by Gene Duplication*. New York: Springer Science & Business Media.
Paquin, B., Heinfling, A., and Shub, D.A. (1999) Sporadic distribution of tRNA(Arg)CCU introns among alpha-purple bacteria: evidence for horizontal transmission and transposition of a group I intron. *J Bacteriol* **181**: 1049–1053.
Paulsson, J., Karoui, El, M., Lindell, M., and Hughes, D. (2017) The processive kinetics of gene conversion in bacteria. *Mol Microbiol* **104**: 752–760.

© 2018 The Authors. *Molecular Microbiology* Published by John Wiley & Sons Ltd.,

Reinhold-Hurek, B., and Shub, D.A. (1992) Self-splicing introns in tRNA genes of widely divergent bacteria. *Nature* **357**: 173–176.
Reiter, W.D., Palm, P., and Yeats, S. (1989) Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. *Nucleic Acids Res* **17**: 1907–1914.
Rocha, E.P.C., and Danchin, A. (2003) Gene essentiality determines chromosome organisation in bacteria. *Nucleic Acids Res* **31**: 6570–6577.
Salman, V., Amann, R., Shub, D.A., and Schulz-Vogt, H.N. (2012) Multiple self-splicing introns in the 16S rRNA genes of giant sulfur bacteria. *Proc Natl Acad Sci USA* **109**: 4203–4208.
Sandegren, L., and Andersson, D.I. (2009) Bacterial gene amplification: implications for the evolution of antibiotic resistance. *Nat Rev Microbiol* **7**: 578–588.
Sandegren, L., Linkevicius, M., Lytsy, B., Melhus, A., and Andersson, D.I. (2012) Transfer of an *Escherichia coli* ST131 multiresistance cassette has created a *Klebsiella pneumoniae*-specific plasmid associated with a major nosocomial outbreak. *J Antimicrob Chemother* **67**: 74–83.
Schmieger, H., and Backhaus, H. (1973) The origin of DNA in transducing particles in P22-mutants with increased transduction-frequencies (HT-mutants). *Mol Gen Genet* **120**: 181–190.
Sela, S., Yogev, D., Razin, S., and Bercovier, H. (1989) Duplication of the tuf gene: a new insight into the phylogeny of eubacteria. *J Bacteriol* **171**: 581–584.
Shen, P., and Huang, H.V. (1986) Homologous recombination in *Escherichia coli*: dependence on substrate length and homology. *Genetics* **112**: 441–457.
Sivaramakrishnan, P., Sepúlveda, L.A., Halliday, J.A., Liu, J., Núñez, M.A.B., Golding, I., Rosenberg, S.M., and Herman, C. (2017) The transcription fidelity factor GreA impedes DNA break repair. *Nature* **550**: 214–218.
Smith, G.R., Kunes, S.M., Schultz, D.W., Taylor, A., and Triman, K.L. (1981) Structure of chi hotspots of generalized recombination. *Cell* **24**: 429–436.
Watt, V.M., Ingles, C.J., Urdea, M.S., and Rutter, W.J. (1985) Homology requirements for recombination in *Escherichia coli*. *Proc Natl Acad Sci USA* **82**: 4768–4772.
Whittle, G., Bloomfield, G.A., Katz, M.E., and Cheetham, B.F. (1999) The site-specific integration of genetic elements may modulate thermostable protease production, a virulence factor in *Dichelobacter nodosus*, the causative agent of ovine footrot. *Microbiology* **145**: 2845–2855.
Yu, D., Ellis, H.M., Lee, E.-C., Jenkins, N.A., Copeland, N.G., and Court, D.L. (2000) An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc Natl Acad Sci USA* **97**: 5978–5983.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web-site.