UPPSALA
UNIVERSITET

The Importance of Bacterial Replichore Balance

Ender Efe Cerit

Master Degree Project in Infection Biology, 30 credits. Spring 2021
Department: Medical Biochemistry and Microbiology
Supervisor: Diarmaid Hughes

List of Contents

ABSTRACT

In most bacterial pathogens, the genome is comprised within a single circular chromosome which is typically organized by the origin-to-terminus axis that divides the chromosome into equally-sized arms of replication (replichores). This similarity in length is presumed to be required for the synchronization of the two replication forks to meet at the terminus for efficient chromosome segregation. Transfer of genes between organisms, different from the route of parent to offspring, is called horizontal gene transfer (HGT). Acquiring foreign DNA through HGT is an important factor for the evolution of virulence in bacteria since it provides access to new features such as new toxins and antibiotic resistance genes. Chromosomes of many pathogenic bacteria such as *Salmonella* spp. carry such horizontally-transferred DNA fragments called pathogenicity islands. However, after such HGT events, the existing organization of chromosome can be disrupted and an imbalance between the two halves of the circular chromosome might occur. The predicted outcome of a replichore imbalance is the retardation of growth which in turn might result in the out-competition by other faster-growing bacteria in the environment. For that reason, we have investigated the association of the fitness cost and the replichore imbalance with isogenic strains with varying degrees of inter-replichore inversions. Our results showed that there is a correlation between the magnitude of replichore imbalance and fitness cost, for example 2.49-fold imbalance (one replichore 2.49-fold longer than the other) resulted in 11% reduction of fitness in comparison with balanced replichores. Therefore, our data suggest that the replichore imbalance could be utilized to predict the fitness cost of HGT events.

POPULAR SUMMARY

Bacteria that have their cake and eat it too: How do they stay "fit" after a heavy "gene" meal?

Transfer of genes between organisms, different from the route of parent to offspring, is called horizontal gene transfer (HGT). Obtaining new DNA through HGT is an important factor for the evolution of disease-causing bacteria since it enables bacteria to gain new features such as new toxins and antibiotic resistance genes. However, after acquisition of new genetic material, the existing organization of chromosome can be disrupted, and an imbalance between the two halves of the circular chromosome can occur. The balance is needed for the timing of two DNA replication reactions happening simultaneously at the opposite halves, so that reactions can finish at the same time. A change in this balance can slow down the bacterial growth which in turn result in the reduction of their "fitness", their capacity to compete with other bacteria in the environment for survival. Nevertheless, it is also known that chromosomes of many pathogenic bacteria such as *Salmonella* species carry such horizontally-transferred DNA fragments successfully.

This project aimed to investigate how much the growth of bacteria is affected by the imbalance in the chromosome structure by measuring and comparing the growth rates of bacteria with the imbalance and without the imbalance, and to observe if there is a correlation with the imbalance and the hindrance of growth. Our results show that there is indeed a strong association between chromosomal imbalance and the reduction in fitness. We think that this study will deepen our understanding of bacterial virulence evolution by giving us more details to the story of what happens after a bacterium inserts new genetic material into its chromosome, and more importantly, help us to predict the cost (calories) of obtaining new DNA (cake) into the bacterial chromosome (stomach).

INTRODUCTION

The genome of most pathogenic bacteria is composed of a single circular chromosome. Typical organization of this circular chromosome is an origin of replication at one pole and a terminus of replication at the opposite pole (Figure 1) (1). Replication starts at the origin and proceeds bi-directionally until the two replication forks conclude at the terminus. At that moment, the mother and the daughter chromosomes resolve. Co-occurrence of multiple replication cycles enable bacteria to grow faster than the actual time needed for the completion of whole chromosome replication. A remarkable feature of bacterial chromosome organization is that the distance in nucleotides from origin to terminus in each arm of replication (replichore) is approximately equal in length. It is presumed that this similarity in length is conserved to secure the synchronization of two replication forks for the proper segregation of chromosomes (2 - 4).

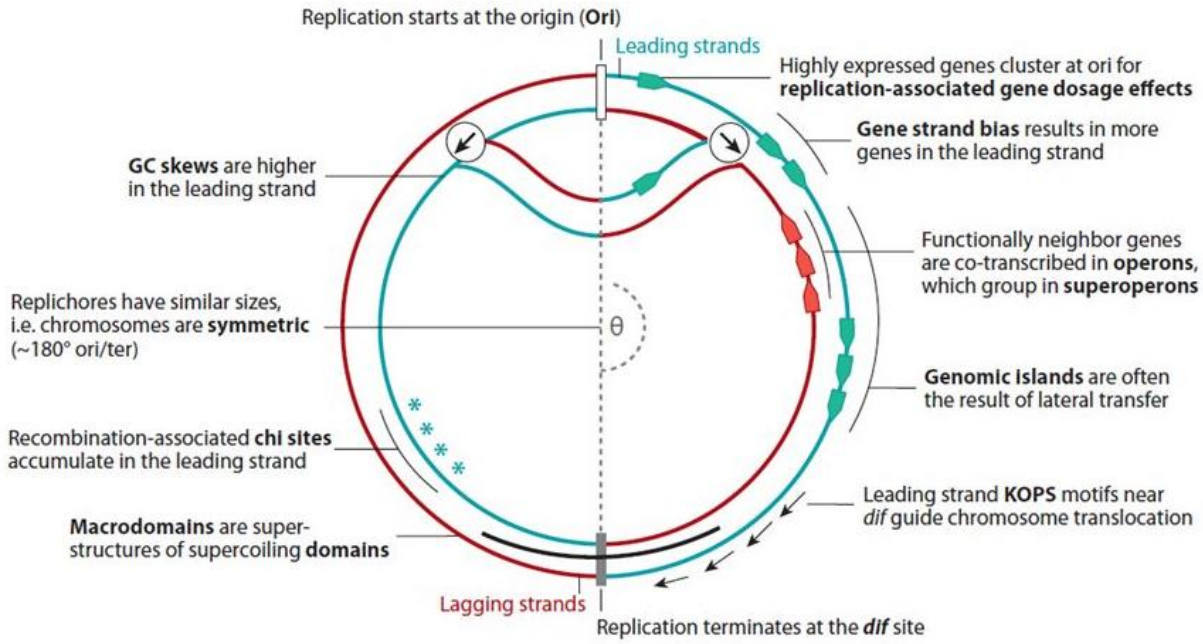

Figure 1. Organization of bacterial chromosome (1).

Chromosomal rearrangements can happen via recombination during replication and repair, and can disrupt existing chromosome organization. Homologous recombination between inverted repeats can result in the orientation change of the intervening region (5). Even though inversions take place frequently in laboratory settings, they are seldomly detected in natural isolates since there are mechanistic and functional limitations regarding creation and maintenance of inversions, respectively (6, 7, 8). A previous study showed that imbalanced replichores originated from asymmetrical inversions beget a notable fitness cost, and the severity of fitness cost is correlated with the degree of replichore imbalance. Following experimental evolution steps revealed that chromosomal rearrangements in the form of re-inversions occur within the genomes with asymmetrical inversions in order to restore the replichore balance (6). A study on chromosomal inversions in *Escherichia coli* genome showed that replichore imbalance in the magnitude of 15-20% may result in detrimental effects on cell growth (9). Surprisingly, researchers found that *Lactobacillus lactis* genome can tolerate asymmetrical chromosomal inversions up to 15% of length difference between the replichores unless the location of inversion endpoints is not close to the origin and terminus regions (10).

Apart from these two essential regions, it is shown that operon concatenation is another important feature that limits chromosomal rearrangements, and co-linear order of operons is maintained due to a strong selective pressure. Researchers demonstrated that upon interruption of operon concatenation with multigene insertions, growth rate and cellular fitness are significant reduced (11). In another study that supports this argument, the researchers indicated that by extending the duplication time, asymmetrical chromosome inversions might have slowed down the growth rate, and led to the emergence of attenuated CDC 684 strain of *Bacillus anthracis* (12).

Nonetheless, one exception to aforementioned argument might be the clinical isolates of pathogenic bacteria which indicates the human immune system possibly creates a selective pressure in the favor of the chromosomal rearrangements within the genomes of these bacteria (13). *Pseudomonas aeruginosa* isolates acquired from cystic fibrosis patients demonstrate this feature to a great extent where large chromosomal inversions facilitate the adaptation and persistence of these *P. aeruginosa* isolates to lung tissue with cystic fibrosis (14). This phenomenon can be explained by SNAP (selection during niche adaptation) hypothesis which asserts that upon chromosomal duplication under non-lethal selective pressure, duplicated genes that improve fitness can be preserved while the others can be lost or inactivated until the rearranged genome reaches a new equilibrium and the new gene order is fixated (15).

Two independent studies on two Gram-positive bacteria, *Staphylococcus aureus* and *L. lactis*, demonstrated that even if the asymmetrical chromosomal inversions are reversible, they give rise to phenotype switching a.k.a. phase variation in bacterial populations, and increase the possibility of survival against rapidly changing environmental conditions such as exposure to inhibitory concentrations of antibiotics (16, 17). Findings of a similar study on *Bacteroides fragilis*, a Gram-negative bacterium with commensal and opportunistic pathogen tendencies, supported aforementioned statements and stated that the bacteria utilize DNA inversions in order to diversify their physiologies and make them compatible with their particular ecological niches (18).

Even though chromosomal inversions come with a fitness cost most of the time, and therefore reversible inversions are more favored among bacterial populations, gene duplications due to chromosomal inversion also play a role in short- and long-term bacterial evolution as shown by a study about outer membrane protein gene duplications in *Helicobacter pylori* genome (19).

Apart from homologous recombination, the other cause of imbalanced replichores is the integration of a large DNA fragment acquired via horizontal gene transfer (HGT) that would shift the size of one replichore drastically. Therefore, even though HGT provides opportunities to bacteria by giving access to new environmental niches, it also burdens the bacteria with a fitness cost via disruption of the existing chromosome organization (20). According to a study, after such HGT event, the bacteria can compensate for the fitness reduction with amplifications of these newly acquired genes to adjust the expression to normal levels (21).

Evolution of bacterial virulence is closely related with the novel DNA acquisition via HGT. Incorporation of this novel DNA into the chromosome must cause a replichore imbalance. However, it is a well-known fact that the bacteria can harbor multiple pathogenicity islands which can be up to 100 kb in length (22). Quantification of fitness cost after such a chromosome rearrangement event will deepen the knowledge regarding the virulence evolution in bacteria.

In our study, we have chosen chromosomal inversion as our preferred model of chromosomal rearrangement since it changes the genome organization without changing the genetic content. Therefore, we utilized a set of *Salmonella* Typhimurium strains with two inversely-oriented selectable cassettes positioned at opposite replichores and at different location combinations to obtain chromosomal inversions that cause replichore imbalance. With this approach, we aimed to (i) calculate homologous recombination and inversion rates between these positions, (ii) investigate whether there is a fitness cost associated with the replichore imbalance, and (iii) investigate whether the fitness cost correlates with the degree of imbalance.

MATERIALS & METHODS

Bacterial strains

All the bacterial strains utilized in this study are derived from *Salmonella enterica* serovar Typhimurium, strain LT2 (23). Genotypes of the strains are given in Supplementary Table 1.

Bacterial growth conditions

Bacteria were grown in lysogeny broth (LB) (10% tryptone, 5% yeast extract (Oxoid, Basingstoke, England), 10% NaCl (Merck, Darmstadt, Germany)), or in M9 minimal liquid medium (per liter: 6g $Na_2HPO_4.2H_2O$, 3g KH_2PO_4 (Sigma-Aldrich, Sweden), 0.5g NaCl, 1g NH_4Cl with 0.1 mL of 1M $CaCl_2$, 1 mL of 1M $MgSO_4$ (Merck, Darmstadt, Germany) and 10 mL of 20% D-glucose (GIBCO, Invitrogen, Lithuania)) added after autoclaving. Solid medium was LA (LB with 1.5% agar (Oxoid, Basingstoke, England)). All cultures were incubated at 37^{o}C, and liquid cultures were agitated at 200 rpm for aeration. Concentration of kanamycin (Sigma-Aldrich, Sweden) utilized in the experiments is 50 mg/L.

Polymerase chain reaction

Polymerase chain reaction (PCR) was performed with Fermentas PCR Master Mix (Thermo Fisher Scientific, USA) in S1000^{TM} Thermal Cycler (Bio-Rad, USA) with following conditions: 95^{o}C for 5'; 30 cycles of 95^{o}C for 30", 56^{o}C for 30", 72^{o}C for 2'; 72^{o}C for 5'. All oligonucleotides were purchased from Sigma-Aldrich, and their sequences are given in Supplementary Table 2.

Measurement of recombination rates

Independent cultures were initiated by inoculating a single colony into 2 mL LB, and the inocula were incubated at 37^{o}C and 200 rpm for 18 hours until the cell density reaches approximately 2 x 10^9 cfu/mL. After serial dilution with 0.9% NaCl solution, approximately 2 x 10^5 cfu were plated onto LA plates with kanamycin, and incubated at 37^{o}C for 18 hours. Recombination rates were calculated with the following formula: $\mu = -(1/N) \ln(P_0)$, where μ is the recombination rate, N is the number of viable cells and P_0 is the proportion of cultures with no scored recombinants.

Estimation of relative fitness

Doubling times (T_d) of cultures were measured as the rate of increase in optical density at 600 nm (OD_{600}) using a Bioscreen C device (Oy Growth Curves Ab Ltd, Finland). For the initiation of the experiment, 1 µL of each independent overnight culture was inoculated into 1 mL of fresh medium. Then, 300 µL of the inoculum were transferred into the corresponding well of a Honeycomb microtiter plate (Oy Growth Curves Ab Ltd, Finland), and the plate was

incubated at 37°C for 18 hours with continuous shaking. Optical density of each well was measured with an interval of 5 minutes, and the doubling time of each strain was calculated with the following formula: T_d = ln2/slope, based on 50-minutes intervals of linear region of the growth curve during the exponential phase. For the calculation of relative fitness, the independent doubling times of a particular strain were compared with the average of the doubling time of the corresponding reference strain.

Light microscopy

Visualization of cell morphology was performed with Nikon Eclipse E100 light microscope (Nikon Europe B.V., Amsterdam, Netherlands), and images were photographed with Lumenera Infinity Lite B camera (OEM Optical, Danville, Canada) and Infinity Capture firmware version 247.06. For sample preparation, a single colony was resuspended in 1 mL of PBS solution (per liter: 8 g NaCl, 200 mg KCl (Merck, Darmstadt, Germany), 1.44 g Na_2HPO_4, 245 mg KH_2PO_4 (Sigma-Aldrich, Sweden)) and 15 µL of the bacterial suspension were placed on microscopy slide which then covered with a coverslip. Immersion oil (Motic, Reagecon Diagnostics Ltd., Ireland) was used for 100x magnification.

RESULTS

Rates of interreplichore recombination and inversion vary across the chromosome

Prior to this study, a set of *Salmonella* Typhimurium strains were constructed with Lambda Red recombineering system to measure the rates of recombinational repair between different locations on the chromosome by the introduction of two selectable gene cassettes. Each of these cassettes contains an inactive version kanamycin resistance gene, one copy with a non-sense mutation at residue 3 and the other one at residue 138. Expected outcome of the recombination between them is the accumulation of non-sense mutations in one copy and the restoration of the function of the other one. Apart from that, each cassette also contains a separate antibiotic resistance gene (chloramphenicol and ampicillin) to select the clones with recombination cassettes during the strain construction, and then to select the inversion strains with PCR primers specific to these genes upon selection for inversion (Figure 2) (24).

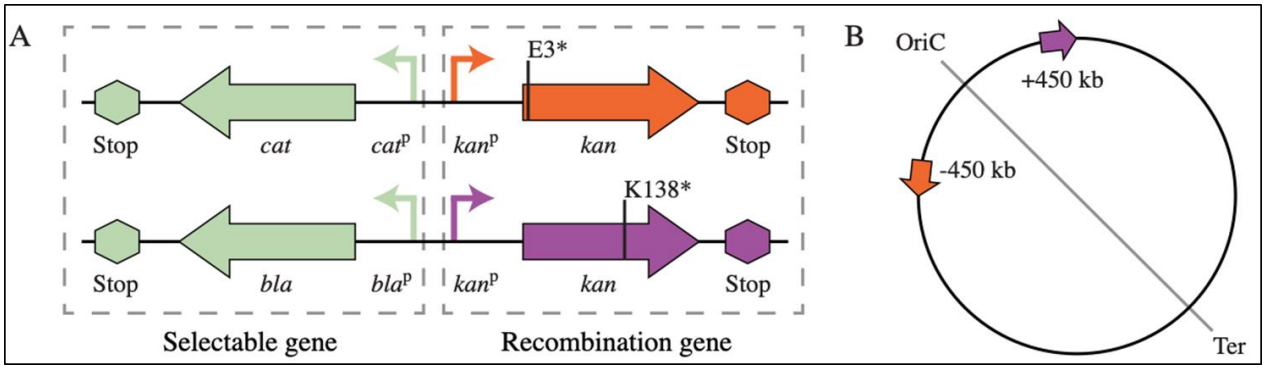

Figure 2. Experimental setup. (A) Design of selectable recombination cassettes. *kan*, *cat* and *bla* genes provide resistance against kanamycin, chloramphenicol and ampicillin, respectively. Each *kan* gene has one stop codon (E3* or K138*) that renders the gene non-functional. (B) Orientation and positioning of selectable recombination cassettes on the chromosome (24).

Two chromosomal positions proximal to *tuf* genes were selected as the initial insertion sites of these cassettes since the homologous recombination rates and the recombinational repair mechanisms among these sites were studied extensively by our group. Both *tuf* genes (*tufA* and *tufB*) encode EF-Tu (elongation factor thermo unstable), and while *tufA* is positioned on the left replichore 480 kb distant from the origin of replication (*oriC*), *tufB* is positioned on the right replichore 280 kb distant from *oriC* (25). Therefore, both recombination cassettes were inserted 450 kb distant from *oriC* (-450/+450), and the kanamycin resistance genes within the cassettes were positioned in an inverse orientation to create inversion between these sequences. Following that, by keeping the location of one of these initial cassettes fixed, the other cassette was moved to a variety of different locations on the chromosome, with varying distances from the origin of replication. The positions of variating cassettes were chosen randomly for the sake of sampling with initial 150 kb intervals but then the intervals were increased to obtain drastic inversions. With this approach, 9 cassette strains were created and designated according to the pairing of the cassette locations as follows: -450/+150, -450/+300, -450/+450, -450/+600, -450/+1500, -150/+450, -300/+450, -600/+450, -1500/+450 (Figure 3, Supplementary Table 1).

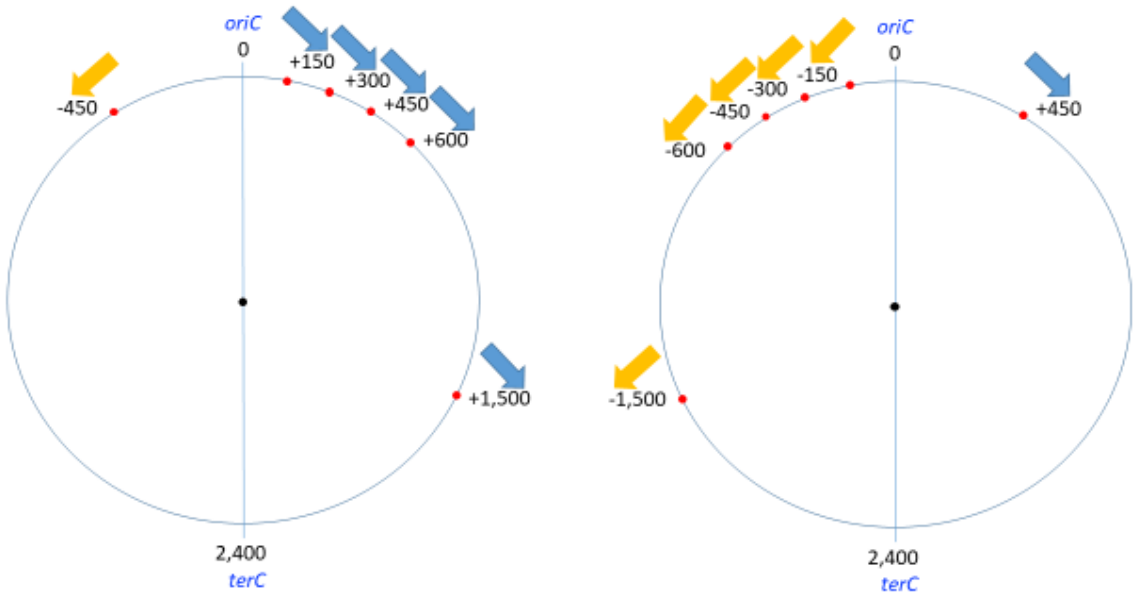

Figure 3. Positioning of the recombination cassettes on the chromosome. Arrows indicate the location of each recombination cassette on each replichore. By keeping the location of one recombination cassette fixed at +/-450, the opposite replichore is traversed with the other recombination cassette, resulting in the creation of 9 cassette strains.

After selection for recombination (kanamycin-resistant phenotype), the presence of an inversion was screened among the recombinant clones with PCR and primers specific for the novel chromosomal junction(s) expected to be formed after an inversion. While the reverse primer is specific for chromosomal position of the cassette, the forward primer is specific for the antibiotic gene alternating between ampicillin and chloramphenicol resistance genes according to the inversion outcome (Figure 4). Then, rates of recombination and inversion (a sub-set of the recombination rate) were calculated with the formula given in Materials and Methods. Expected outcomes of each inversion and the ratio of replichore lengths as a degree

of imbalance were shown in Figure 5. Numbers of plated cultures, rates of recombination and rates of inversion for each strain are given in Supplementary Table 3.

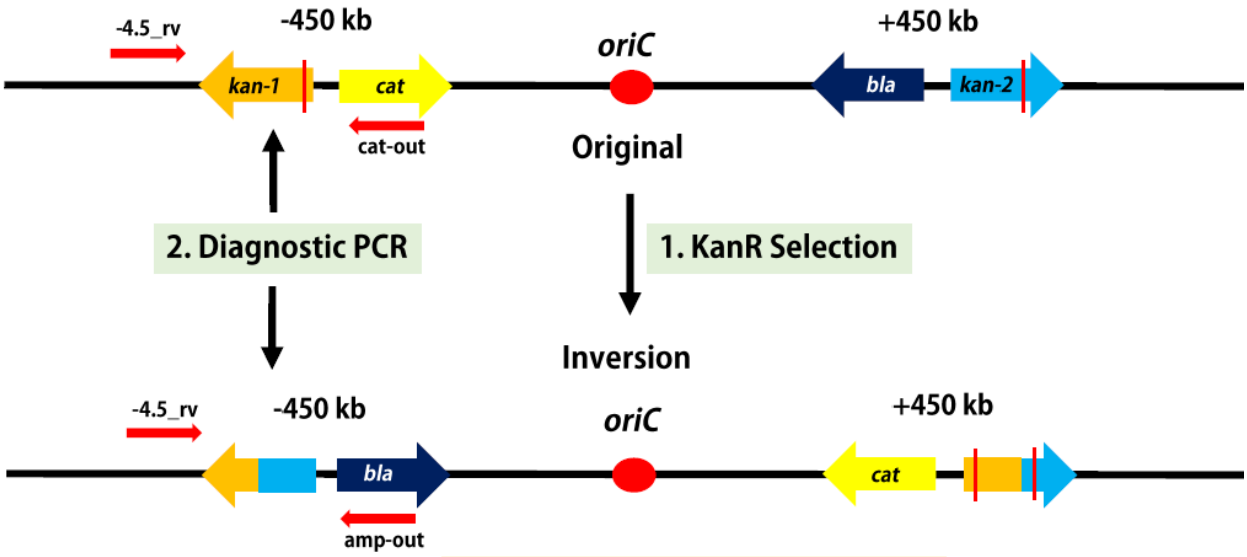

Figure 4. Detailed schematic diagram of the selection experiment for inversion between recombination cassettes equidistant from *oriC*. Horizontal red rods represent stop codons in *kan* genes. Small red arrows represent primers and their directions.

In our previous studies, the rates of recombination and the estimation of fitness cost of replichore imbalance were only studied with -450 fixed strains. And even though the inversion strains were selected and obtained from -450 fixed strains, the rate of inversion was measured only for -450/+450 strain (6, 24). Therefore, in this study, we aimed to measure the rates of both recombination and inversion for both -450 and +450 fixed strains, and also select and obtain inversion strains from +450 fixed strains as well as from -450 fixed strains to utilize them in the following fitness cost estimation experiments. For this study, the initial plan was to plate 100 cultures for each strain. Due to time constraints, 53 cultures were plated for 4 strains (-450/+150, -450/+300, -450/+600, -450/+1500) while the goal of 100 cultures was achieved for the remaining 5 strains (-450/+450, -150/+450, -300/+450, -600/+450, -1500/+450). Moreover, in the selection and validation steps, the great majority of kanamycin resistant recombinant colonies were found upon PCR screening to not to have an inversion. Screening every single selected kanamycin-resistant colony with PCR was an arduous work. For that reason, instead of screening individual colonies, groups of 10 colonies were mixed, their DNA was extracted by boiling lysis, and PCR was performed on the mixture. This method could detect a single inversion strain among 10 non-inversion strains (Figure 6). When an inversion was detected in the mixture, a second round of PCR was made on the individual clones to identify the inversion strain. The implementation of this procedure saved time, resources and sped up screening of every selected kanamycin-resistant colony.

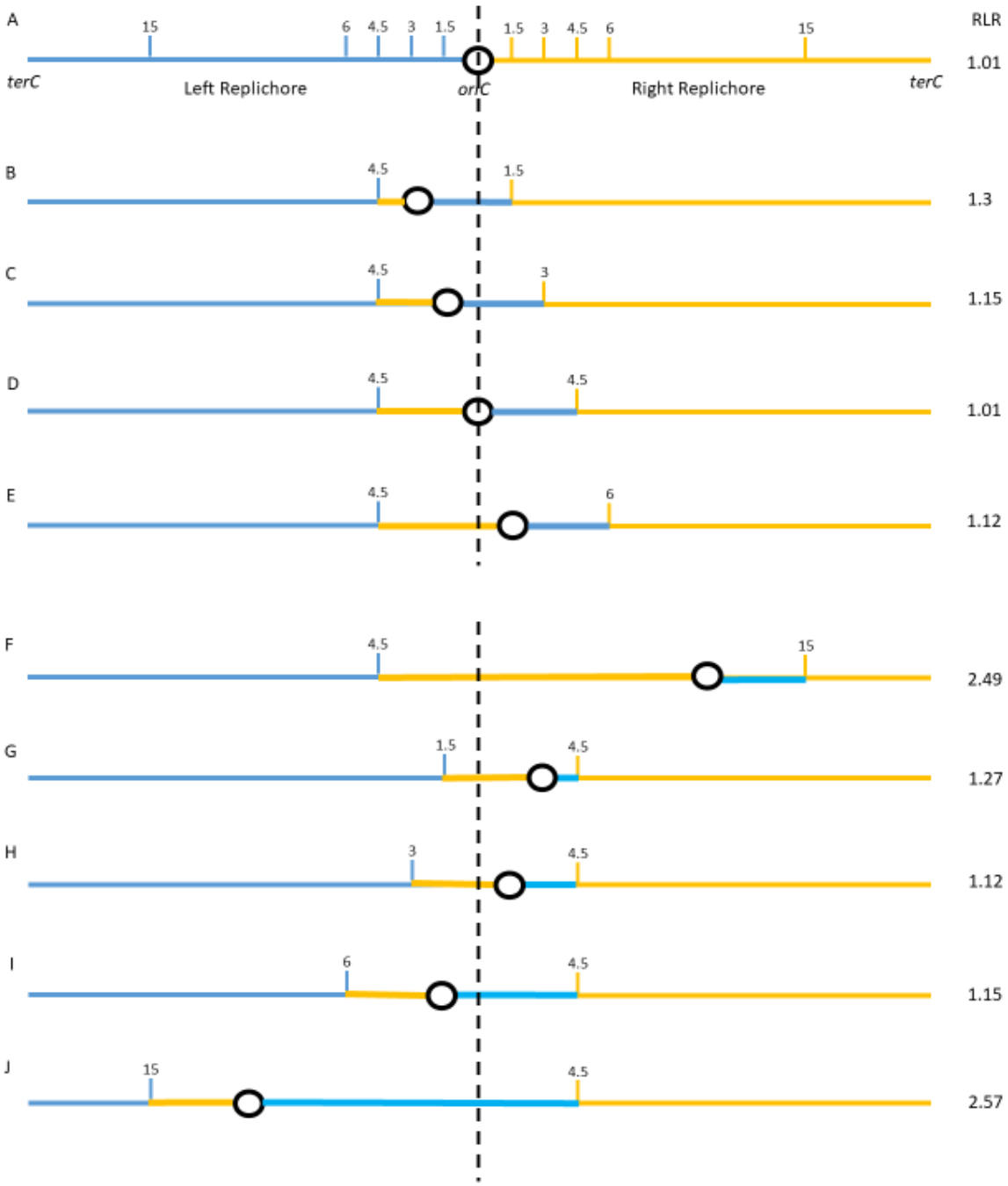

Figure 5. Selected inversions. A) Linear layout of the wild-type *Salmonella* Typhimurium chromosome with position indicators of recombination cassette insertion sites. The numbering refers to distance from *oriC* in terms of kilobase (*e.g.* 4.5 is 450 kb). B-J) Expected chromosome structure of inversion outcomes between 9 cassette pairs. Replichore length ratio (RLR) describes the imbalance between two replichores via comparison of the long replichore to the short one.

Pool Name	WT Colony #	INV Colony #
A	1	5
B	1	10
C	5	1
D	10	1
E	1	1

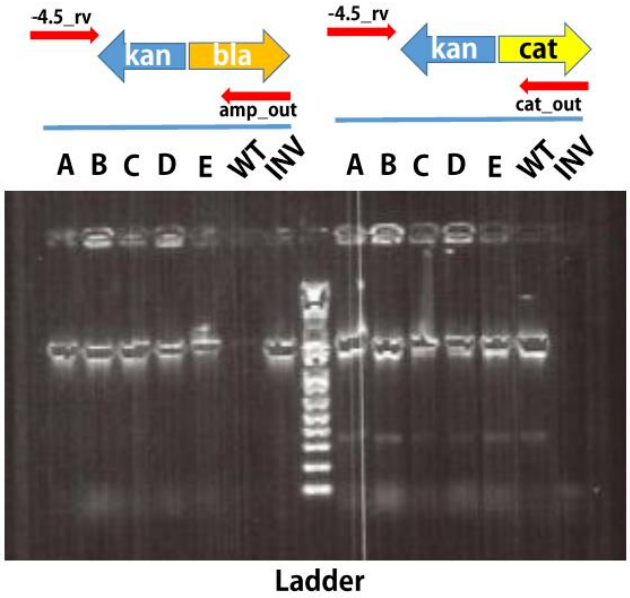

Figure 6. Proportion of Parental (Wild-type) / Inversion clones against band intensity. Primer pair -4.5_rv and amp_out is used for detection of inversion junction, while primer pair -4.5_rv and cat_out is used for detection of parental (wild-type) junction. According to the image, DNA of 1 inversion clone can be detected among DNA of 10 parental clones. On top of the gel image, WT and INV represent positive controls for parental and inversion strains, respectively.

The results showed that the correlation between the frequencies of recombination and the distance between the cassette positions varies according to the position of fixed cassettes, in other words whether the fixed cassette 450 kb distant from *oriC* is on the right or on the left replichore. While there is no reliable correlation (R^2 = 0.28) between the frequencies of recombination and the distance between the cassette positions when the fixed position is on the left replichore, there is a correlation (R^2 = 0.65) when the fixed position is on the right replichore: the rate of recombination decreases with the increase of the distance between the cassettes (Figure 7).

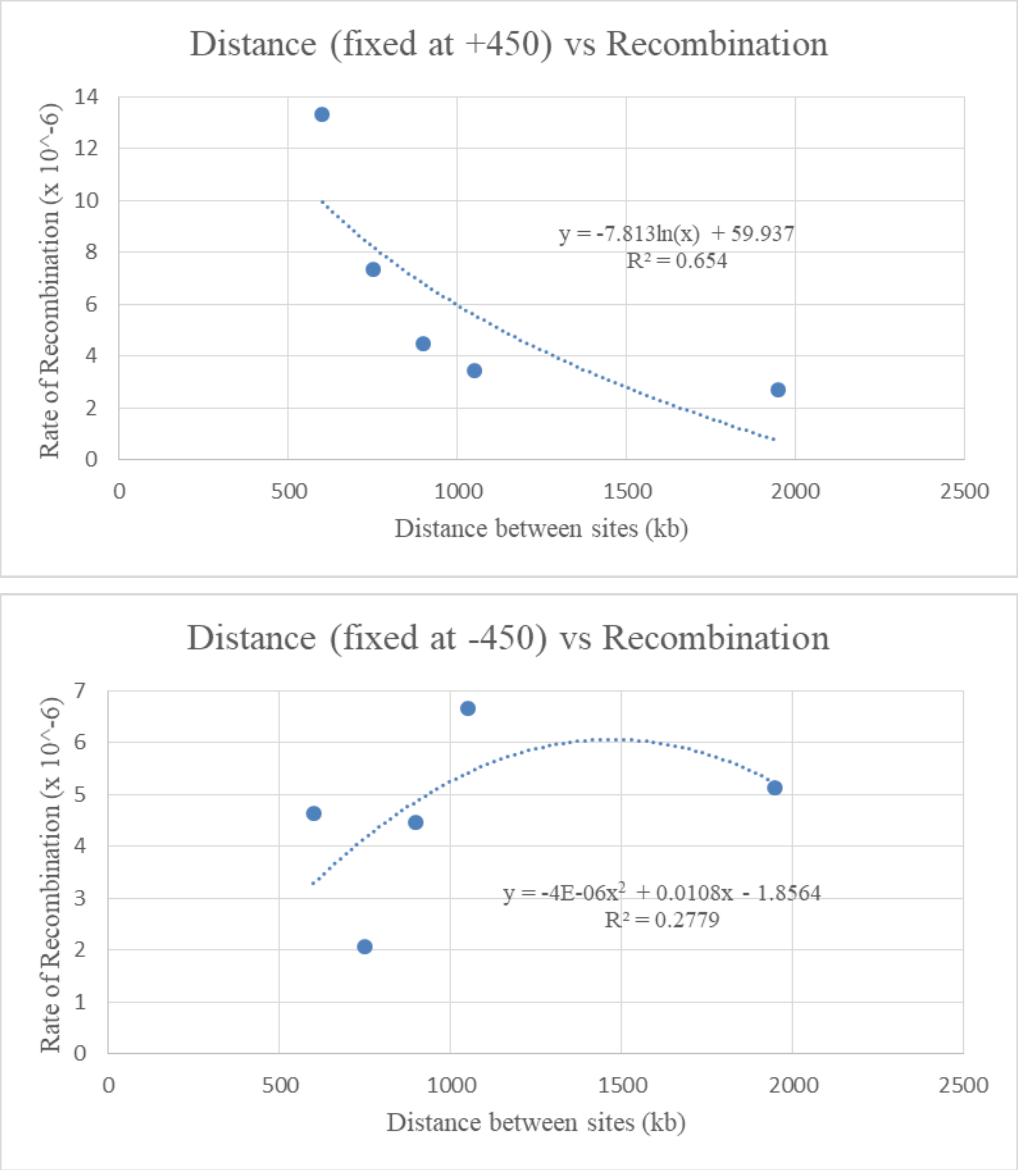

Figure 7. Comparative analysis of recombination rate as a function of distance between cassette positions. Graphic of the top shows the correlation between the recombination rate and the distance between the positions when the fixed position is +450, while graphic on the bottom shows the correlation the two factors when the fixed position is -450.

Interestingly, while inversions were detected with frequencies ranging from 0.1 to 0.53 x 10^{-6} per cell per generation among recombinants carrying a cassette 450 kb distant from *oriC* on the left replichore (-450), no inversions were detected among recombinants with a cassette 450 kb distant from *oriC* on the right replichore (+450). Therefore, these recombination cassettes could be utilized to estimate the individual accessibility of each position for inter-replichore recombination. However, with the exceptions of inversion between -450/+450 and an anomaly observed after inversion selection with cassettes at -150/+450, the inversions expected to originate from cassette location fixed at 450 kb distant from *oriC* on the right-hand replichore (+450) could not be obtained at aforementioned frequencies.

The recombinant anomaly was observed after selection involving the cassette locations at -150/+450. Several of the selected kanamycin-resistant recombinants gave positive results for the presence, within a single clone, of both non-inversion and inversion genotypes as

determined via diagnostic PCR. Even after several rounds of serial dilution, colonies containing only the inversion genotype could not be isolated (Figure 8). Based upon a previous study where the existence of filamentous cells was documented after inversions between the Ori and the Right macrodomains took place (9), a preliminary microscopic analysis was performed on the recombinant anomalous clones and the parental strain. Results of the preliminary microscopic analysis were in agreement with the previous study. Accordingly, a sample taken from a single colony of an anomalous recombinant clone contained both cells with regular morphology and filamentous cells while in contrast the parental strain solely contained cells with a regular morphology (Figure 9).

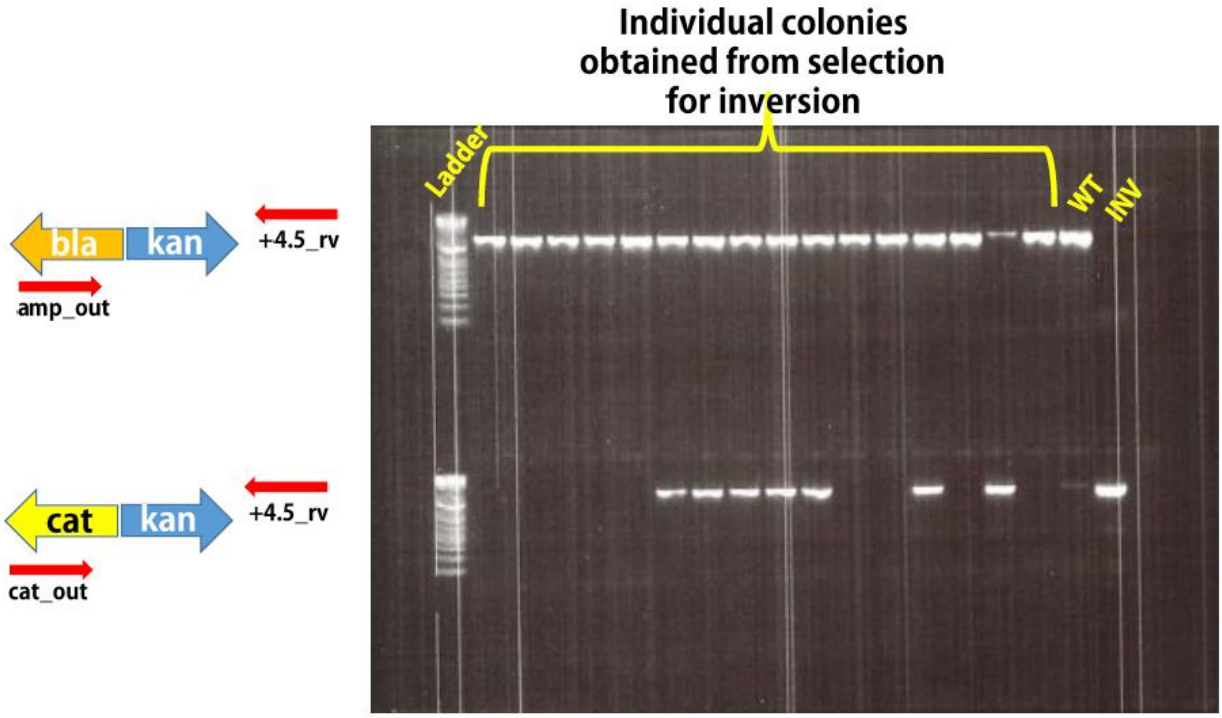

Figure 8. Presence of both non-inversion and inversion genotypes within a single clone. Primer pair +4.5_rv and cat_out is used for detection of inversion junction, while primer pair +4.5_rv and amp_out is used for detection of parental (wild-type) junction. On top of gel image, WT and INV represent positive controls for parental and inversion strains, respectively.

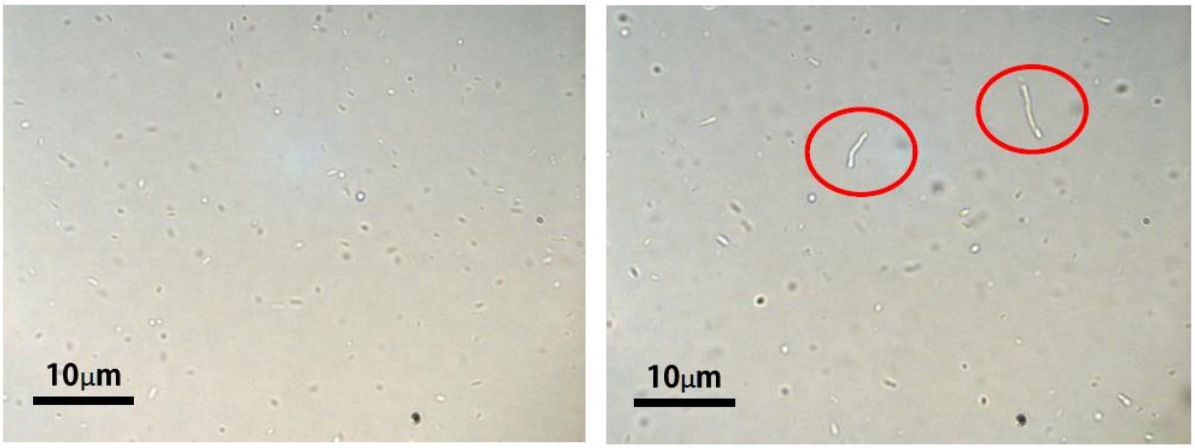

Figure 9. Morphological analysis of recombinants with light microscope. Image on the left shows the normal cell morphology of parental strain TH10742, while the image on the right shows the filamentous cells (encircled) among cells with normal morphology within a single clone of anomalous recombinant. Magnification: 1000X

Fitness cost of an inversion is positively correlated with the degree of replichore imbalance

In our previous study, fitness costs of replichore imbalance were estimated only for -450 fixed strains with head-to-head competition experiments against the wild-type strain (6). In head-to-head competition experiments, since both strains are incubated together within a single culture, even a slight growth retardation of one strain at the beginning could result in the domination of the other strain within the culture due to exponential nature of bacterial growth. In addition, since the wild-type strain does not harbor any of the recombination cassettes, comparing wild-type strain to the inversion strains with imbalanced chromosomes on top of harboring those cassettes might not be fair. Therefore, in this study, we have compared the individual growth rates of inversion strains and their parental cassette strains. However, since we could not obtain inversion strains from +450 fixed strains, apart from the balanced inversion (+450/-450), we decided to focus the study on strains with -450 fixed, and we expanded our data set by including two additional inversion strains (-450/+1200 and -450/+1650) selected and obtained in the previous study (24). This set of 7 inversion strains was used in the measurements of relative fitness.

Growth competition measurements showed that all inversions were associated with a fitness cost and caused a reduction in growth rates. The magnitude of the effect was up to 11% reduction per generation in rich media (LB), and up to 22% reduction per generation in minimal media (M9). There was a greater fitness cost associated with replichore imbalance during growth in minimal media relative to growth in rich media, but the difference between them was not substantially severe (Figure 10).

During growth in rich media, there is a positive correlation between the degrees of replichore imbalance and the associated fitness cost ($R^2 = 0.62$). There was no fitness cost associated with the inversion where the replichore balance remained unchanged. Moreover, there was a trend among inversions that indicated small replichore imbalances led to smaller reductions in growth (up to 2%), while large replichore imbalances led to larger reductions in growth (up to 11%).

In minimal media, the level of correlation between degrees of replichore imbalance and fitness cost had increased ($R^2 = 0.72$). Same as the growth in LB, there was no fitness cost to the inversion where the replichore balance was unchanged. Moreover, a similar trend with the growth in LB was observed at growth in M9 as well (up to 1% growth reduction in smaller imbalances, up to 22% growth reduction in larger imbalances). Two outliers to this trend (with significant standard deviations) were the strains with inversion between cassettes at -450/+600 and -450/+1200. Therefore, these two outliers were excluded from the correlation estimations.

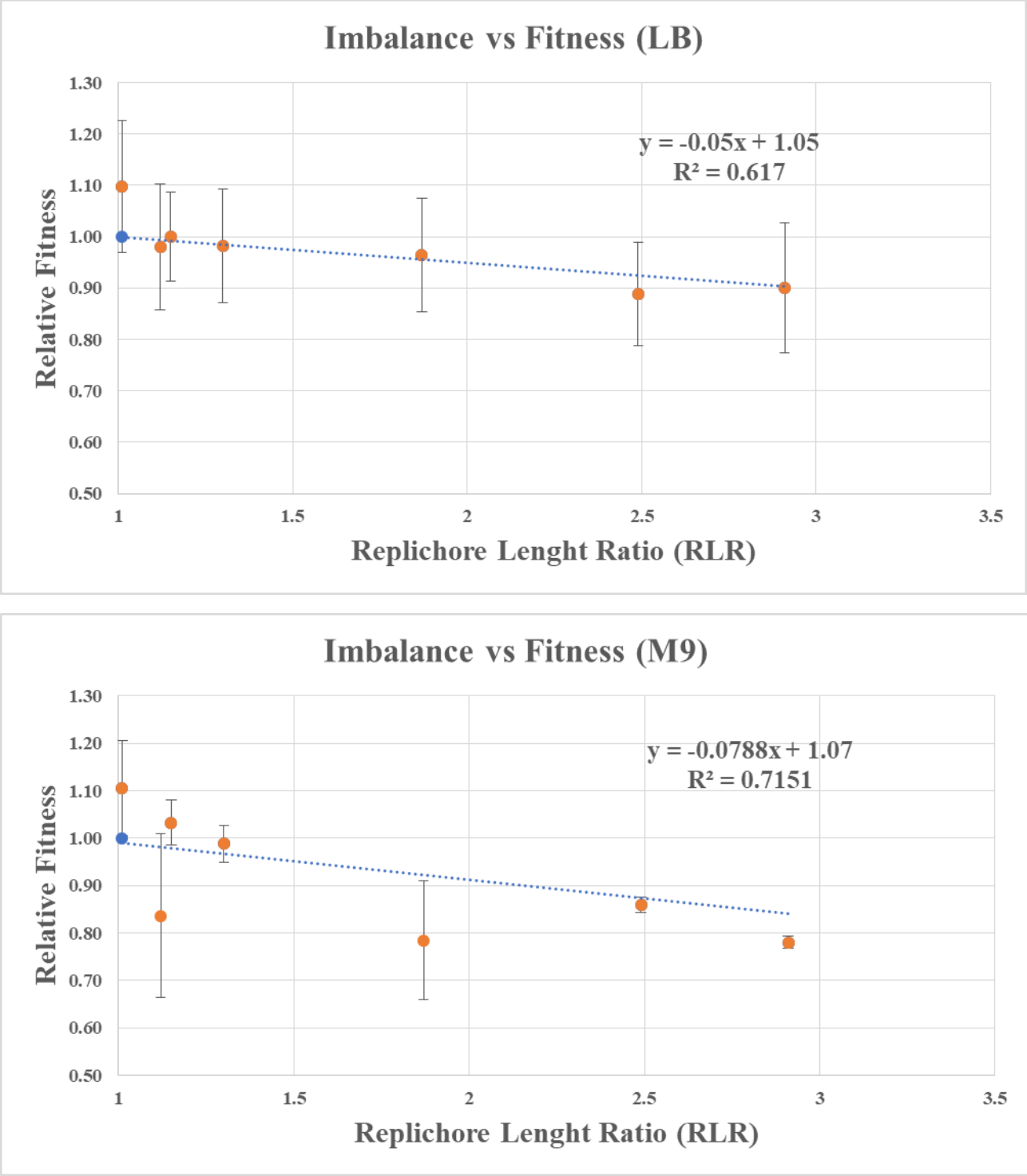

Figure 10. Comparative analysis of degree of replichore imbalance as a function of relative fitness between growth in rich (LB) and minimal (M9) media. Graphics on the top and on the bottom describe the negative correlation between the replichore length ratio and the relative fitness derived from growth rates in LB and M9 media, respectively. Blue point on the graphs represents the wild-type strain (RLR: 1.01, Relative Fitness: 1).

DISCUSSION

Homologous recombination (HR) is a type of recombination where the genetic information exchange takes place between two sequences with adequate similarity. The function of HR in horizontal gene transfer (HGT) has been studied extensively, since, in addition to site-specific recombination, it plays an important role in the integration of newly-acquired foreign genetic material into the host genome (26). In addition to that, HR can happen between similar sequences within the same chromosome, leading to co-evolution of genes via gene homogenization, as in the case of *tuf* genes which encode elongation factor thermo-unstable, or chromosomal rearrangements like inversions between inverted repeats which create variation among the genetics of bacterial population (5, 25). Therefore, the role of HR upon the evolution of bacterial genomes is undeniably important.

Even though there is a possibility for HR to occur between any two regions with sufficient sequence similarity, recent studies on the fluidity of *Enterobacteriaceae* genome organization showed that not every chromosomal region has the same chance of access for HR. By utilizing inversely-oriented *att* sites stationed on 20 different position combinations, a study on the plasticity of *Salmonella* Typhimurium chromosome structure showed that the rate of site-specific recombination between a chromosome and an extrachromosomal genetic entity like plasmid is generally higher than the rate of intrachromosomal recombination. In addition, same study also demonstrated that specific combination of sites is one of the causal factors for recombination while neither the distance between sites nor the individual accessibility of sites is individually relevant for recombination to occur. Therefore, the study indicated the presence of barriers which limit the plasticity of chromosome (27).

With a similar approach, Boccard and his colleagues defined the macrodomain (MD) structure of *E. coli* chromosome, which divides the chromosome into 4 macrodomains (Ori, Left, Right, Ter) and two non-structural (NS) regions that flank the Ori MD. They reported that NS regions are not determined by the sequences of the loci in which they reside, but by the localization of *oriC*. In other words, if the position of *oriC* changes due to an inversion, the loci that flank the new Ori MD reshuffle into new NS regions, regardless of their previous occupancies. Other interesting feature they reported is the recombinational accessibility of MDs and NS regions via their collision probability: while the interactions between NS regions and their adjacent MDs and the interactions within a MD are possible, the interactions between two different MDs are restricted and not favored (28 – 30).

In our first experiment, we measured the HR rates between two inversely-oriented identical sequences located at 5 different positions on each replichore, with 9 different combinations in total (-1500/+450, -600/+450, -300/+450, -150/+450, -450/+450, -450/+150, -450/+300, -450/+600, -450/+1500). Our results showed that the rates of recombination vary by a factor of 6-fold ($13.3 – 2.07 \times 10^{-6}$ per cell per generation, average: 5.53×10^{-6} per cell per generation) among the pairwise combinations across all the chromosomal positioning. Moreover, there is no reliable correlation between the distance of identical sequences from each other and the rate of recombination when the fixed position is -450, while a correlation is present when the fixed position is +450 with the decline of recombination rate as the distance between cassettes increases. This might be due to a statistical issue, since 53 cultures were plated for the strains with -450 fixed position whereas the goal of 100 plated cultures were achieved with strains

with +450 fixed position. Nonetheless, this discrepancy could also originate from the recombinational diverseness of two replichores as it is observed with their rates of inversion.

In the case of inversion rates, a significant divergence is observed between the inversions originating from positions fixed at -450 kb and positions fixed at +450. Apart from -450/+450 inversion and a recombinant anomaly originated from -150/+450, no viable inversion could be obtained from positions fixed at +450, while the average rate of inversion is 0.2×10^{-6} per cell per generation in positions fixed at -450 (Supplementary Table 3). Hence, +450 position might not be a competent zone for inversion. A previous study proved that intrareplichore inversions intermingling Ori and Right MDs have a detrimental effect on cell physiology and morphology. With a combination of fluorescence and differential interference contrast microscopic imaging, they depictured the presence of filamentous cells with multiple segregated nucleoids as a result of the septum formation inhibition via detrimental inhibitions (9). At this point, it is best to mention that formation of division septum depends on the complex interactions between the chromosome, septum-associated cell division proteins and the lipid membranes (31), which evidently interrupted by altered topology of chromosome due to certain types of inversions. Filamentous cells mixed with normal-sized cells are also observed in our preliminary light microscope analysis of the recombinant anomaly originated from -150/+450 which supports the results of the previous study on *E. coli* (9). The detection of both inversion and parental junctions in PCR screening could be interpreted as the presence of inverted and non-inverted chromosomes within a single filamentous cell. Nevertheless, more detailed studies are required for the detection of segregated nucleoids within the filamentous cells, and the estimation of filamentous cell percentage within the bacterial culture.

Even though macrodomain organization is not studied on the *Salmonella* genome, they both share a common ancestor and have similar-sized genomes, therefore their genomic maps are accepted as superimposable (32). Since our findings are in agreement with the results of the previous studies mentioned above, it is possible to suggest that the regional alterations in chromosome topology and structure could determine the probability of HR by influencing the individual recombinational access of those regions.

In our second experiment, we wanted to measure growth rate as a function of degree of imbalance and investigate to which degree the imbalance can predict fitness. Since we could not get inversions from -1500/+450, -600/+450, -300/+450 and -150/+450 positions, we have included two more inversions that were created previously, namely -450/+1200 and -450/+1650, to expand our test group. In addition, we have conducted our experiments in both rich and minimal media to determine the effect of fast-growth versus slow-growth on fitness cost resulting from replichore imbalance. We hypothesized that during fast-growth, there will be a greater fitness cost to the replichore imbalance since the bacteria with replichore balance would multiply without a problem while the bacteria with imbalanced replichores would stagger during replication, and the gap between them would eventually increase exponentially. However, during slow-growth, since replication is slow even for the bacteria with replichore balance as well as the bacteria with imbalanced replichores, there would be a lesser fitness cost in comparison.

According to our results, there is a correlation between the replichore imbalance and the fitness cost in terms of growth competition. As we predicted, there is no fitness cost to the -450/+450 inversion, where the replichore balance is retained. For the rest of the inversions, there is a

trend indicating lesser imbalances to have lesser fitness costs while greater imbalances to have greater fitness costs. Nonetheless, our findings indicate that, for the bacteria with greater replichore imbalance, the measured fitness cost is lower than predicted fitness cost. Even though we do not know the explanation yet, we could speculate that ultimately there is a physiological or metabolic barrier to the fitness cost originating from chromosomal rearrangements. It is best to keep in mind that, even in the case of a fitness cost of 1% per cell per generation, this implies that the acquired genes will need to confer a net benefit equal or greater than this cost on the recipient to balance this cost. Although a cost of 1% might seem very small and irrelevant, the evolution has selected a strong codon usage bias in bacteria like *Salmonella* spp. and *E. coli*, where the fitness advantage is approximately 10^{-4} per codon per generation (33). This shows that very small differences in relative fitness can be selected by the evolution.

Apart from that, there are two outliers to the trend mentioned in the paragraph above: inversions -450/+600 and -450/+1200 in minimal medium. In rich medium, inversion -450/+1200 has greater fitness than predicted. In previous head-to-head competition experiments conducted before this study, a similar result was obtained (data not shown). While additional genetic alterations in this strain could be the reason for the lesser fitness cost, a whole genome sequence analysis is required to pinpoint the reason. Alternatively, the experiment could be repeated with a newly-selected inversion strain. As for the minimal medium, both outliers have substantial standard deviations in their doubling times, therefore a technical error could be accounted for the deviation from the trend.

When we compared the effect of fast-growth versus slow-growth on the fitness cost resulting from inversions, we detected no significant difference for the inversions that cause lesser replichore imbalance, while we observed the opposite of the predicted results for the inversions that cause greater replichore imbalance, namely -450/+1500 and -450/+1650. Apparently, the growth in minimal medium has more significant effect on imbalanced replichores than the growth in rich medium. This phenomenon could be explained by the gene dosage during replication (34). While the rich medium contains all the vitamins, growth factors, complex nitrogen and carbon sources necessary for growth as ready to be used, the minimal medium contains only a single monosaccharide as carbon source, and also mineral nitrogen and phosphor sources and mineral salts. Thereby, the bacteria have to synthesize all the amino acids and vitamins required for growth themselves. Moreover, in bacteria, transcription and translation take place concomitantly during DNA replication, and the genes adjacent to *oriC* have approximately 3-fold more copies than the ones closer to *terC*. Since DNA replication is slow in minimal medium, thereby we can extrapolate that the gene expression is also slow. In the inversions where the replichore imbalance is greater, the genes that are adjacent to *oriC* are not adjacent anymore in the new organization. If the important metabolic genes or the genes of important transcription factors are among the ones that fall afar from *oriC*, they would have lesser copies and lesser expression in the cell than their original state, which might explain the greater fitness cost in minimal medium. However, in rich medium, this situation does not create a significant impact since the nutrition is already present and ready to be utilized.

With all the findings considered, it is possible to create a statistical modelling to estimate the fitness cost of any HGT event (*e.g.* conjugation) involving the insertion of a genetic fragment with a precise size (*e.g.* 200 kb) according to its insertion position on either replichore. Without

a doubt, this hypothetical modelling would assist us to predict future happenstances of antibiotic resistance development via HGT.

To conclude, with our system, we could successfully measure the fitness cost as a degree of replichore imbalance, and we also suggested that the rate of recombination could be determined by the combination of the recombinational probabilities of two individual positions. Follow-up studies could (I) test the predictions by inserting noncoding DNA of different lengths and composition into either replichore and measuring the associated fitness, (II) examine why it was apparently more difficult to select inversions in strains with +450 as the fixed cassette location, relative to strains with -450 as the fixed cassette location.

REFERENCES

1) Rocha, E. P. C. (2008). The organization of the bacterial genome. *Annual Review of Genetics, 42*(1), 211-233. https://doi.org/10.1146/annurev.genet.42.110807.091653

2) Cooper, S., & Helmstetter, C. E. (1968). Chromosome replication and the division cycle of *Escherichia coli*. *Journal of Molecular Biology, 31*(3), 519-540. doi:10.1016/0022-2836(68)90425-7

3) Eisen, J. A., Heidelberg, J. F., White, O., & Salzberg, S. L. (2000). Evidence for symmetric chromosomal inversions around the replication origin in bacteria. *Genome Biology, 1*(6), RESEARCH0011-RESEARCH0011.

4) Song, J., Ware, A., & Liu, S. (2003). Wavelet to predict bacterial ori and ter: A tendency towards a physical balance. *BMC Genomics, 4*(1), 17-17. doi:10.1186/1471-2164-4-17

5) Segall, A. M., & Roth, J. R. (1989). Recombination between homologies in direct and inverse orientation in the chromosome of *Salmonella*: Intervals which are nonpermissive for inversion formation. *Genetics (Austin), 122*(4), 737-747. https://doi.org/10.1093/genetics/122.4.737

6) Garmendia, E., Bueno-Galera, C., Huseby, D. L., & Hughes, D. (n.d.). The Selective Advantage of Replichore Balance in *Salmonella* Typhimurium. Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-332484

7) Miesel, L., Segall, A., & Roth, J. R. (1994). Construction of chromosomal rearrangements in *Salmonella* by transduction: Inversions of non-permissive segments are not lethal. *Genetics (Austin), 137*(4), 919-932. https://doi.org/10.1093/genetics/137.4.919

8) Rebollo, J. E., François, V., & Louarn, J. M. (1988). Detection and possible role of two large nondivisible zones on the *Escherichia coli* chromosome. *Proceedings of the National Academy of Sciences - PNAS, 85*(24), 9391-9395. https://doi.org/10.1073/pnas.85.24.9391

9) Esnault, E., Valens, M., Espéli, O., & Boccard, F. (2007). Chromosome structuring limits genome plasticity in *Escherichia coli*. *PLoS Genetics, 3*(12), e226-e226. doi:10.1371/journal.pgen.0030226

10) Campo, N., Dias, M. J., Daveran-Mingot, M. -L., Ritzenthaler, P., & Le Bourgeois, P. (2004). Chromosomal constraints in Gram-positive bacteria revealed by artificial inversions. *Molecular Microbiology, 51*(2), 511-522. doi:10.1046/j.1365-2958.2003.03847.x

11) Brandis, G., Cao, S., & Hughes, D. (2019). Operon concatenation is an ancient feature that restricts the potential to rearrange bacterial chromosomes. *Molecular Biology and Evolution, 36*(9), 1990-2000. doi:10.1093/molbev/msz129

12) Okinaka, R. T., Price, E. P., Wolken, S. R., Gruendike, J. M., Chung, W. K., Pearson, T., Xie, G., Munk, C., Hill, K. K., Challacombe, J., Ivins, B. E., Schupp, J. M., Beckstrom-Sternberg, S. M., Friedlander, A., & Keim, P. (2011). An attenuated strain of *Bacillus anthracis* (CDC 684) has a large chromosomal inversion and altered growth kinetics. *BMC Genomics, 12*(1), 477-477. doi:10.1186/1471-2164-12-477

13) Hughes, D. (2000). Evaluating genome dynamics: The constraints on rearrangements within bacterial genomes. *GenomeBiology.Com*, *1*(6), reviews0006-reviews0006. doi:10.1186/gb-2000-1-6-reviews0006

14) Kresse, A. U., Dinesh, S. D., Larbig, K., & Römling, U. (2003). Impact of large chromosomal inversions on the adaptation and evolution of *Pseudomonas aeruginosa* chronically colonizing cystic fibrosis lungs. *Molecular Microbiology, 47*(1), 145-158. doi:10.1046/j.1365-2958.2003.03261.x

15) Brandis, G., & Hughes, D. (2020). The SNAP hypothesis: Chromosomal rearrangements could emerge from positive selection during niche adaptation. *PLoS Genetics, 16*(3), e1008615-e1008615. doi:10.1371/journal.pgen.1008615

16) Cui, L., Neoh, H., Iwamoto, A., & Hiramatsu, K. (2012). Coordinated phenotype switching with large-scale chromosome flip-flop inversion observed in bacteria. *Proceedings of the National Academy of Sciences - PNAS, 109*(25), E1647-E1656. doi:10.1073/pnas.1204307109

17) Kojic, M., Jovcic, B., Miljkovic, M., Novovic, K., Begovic, J., & Studholme, D. J. (2020). Large-scale chromosome flip-flop reversible inversion mediates phenotypic switching of expression of antibiotic resistance in lactococci. *Microbiological Research, 241*, 126583-126583. doi:10.1016/j.micres.2020.126583

18) Cerdeño-Tárraga, A. M., Patrick, S., Crossman, L., Blakely, G., Abratt, V., Lennard, N., . . . Parkhill, J. (2005). Extensive DNA Inversions in the *B. fragilis* Genome Control Variable Gene Expression. *Science*, *307*(5714), 1463-1465. Retrieved December 15, 2020, from http://www.jstor.org/stable/3841726

19) Furuta, Y., Kawai, M., Yahara, K., Takahashi, N., Handa, N., Tsuru, T., Oshima, K., Yoshida, M., Azuma, T., Hattori, M., Uchiyama, I., Kobayashi, I., & Sung, P. (2011). Birth and death of genes linked to chromosomal inversion. *Proceedings of the National Academy of Sciences - PNAS, 108*(4), 1501-1506. doi:10.1073/pnas.1012579108

20) Baltrus, D. A. (2013). Exploring the costs of horizontal gene transfer. *Trends in Ecology & Evolution (Amsterdam), 28*(8), 489-495. doi:10.1016/j.tree.2013.04.002

21) Lind, P. A., Tobin, C., Berg, O. G., Kurland, C. G., & Andersson, D. I. (2010). Compensatory gene amplification restores fitness after inter-species gene replacements. *Molecular Microbiology, 75*(5), 1078-1089. doi:10.1111/j.1365-2958.2009.07030.x

22) Messerer, M., Fischer, W., & Schubert, S. (2017). Investigation of horizontal gene transfer of pathogenicity islands in *Escherichia coli* using next-generation sequencing. *PloS One*, *12*(7), e0179880-e0179880. https://doi.org/10.1371/journal.pone.0179880

23) McClelland, M., Sanderson, K., Spieth, J. et al. (2001). Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature*, *413*, 852–856. https://doi-org.ezproxy.its.uu.se/10.1038/35101614

24) Garmendia, E., Brandis, G., Guy, L., Cao, S., Hughes, D. (2021). Chromosomal location determines the rate of intrachromosomal homologous recombination in *Salmonella*. *mBio 12*:e01151-21. http://doi.org/10.1128/mBio.01151-21

25) Abdulkarim, F., & Hughes, D. (1996). Homologous recombination between the *tuf* genes of *Salmonella* Typhimurium. *Journal of Molecular Biology, 260*(4), 506-522. https://doi.org/10.1006/jmbi.1996.0418

26) Brochet, M., Rusniok, C., Couvé, E., Dramsi, S., Poyart, C., Trieu-Cuot, P., Kunst, F., & Glaser, P. (2008). Shaping a bacterial genome by large chromosomal replacements, the evolutionary history of *Streptococcus agalactiae*. *Proceedings of the National Academy of Sciences - PNAS, 105*(41), 15961-15966. https://doi.org/10.1073/pnas.0803654105

27) Garcia-Russell, N., Harmon, T. G., Le, T. Q., Amaladas, N. H., Mathewson, R. D., & Segall, A. M. (2004). Unequal access of chromosomal regions to each other in *Salmonella*: Probing chromosome structure with phage λ integrase-mediated long-range rearrangements. *Molecular Microbiology, 52*(2), 329-344. https://doi.org/10.1111/j.1365-2958.2004.03976.x

28) Duigou, S., & Boccard, F. (2017). Long range chromosome organization in *Escherichia coli*: The position of the replication origin defines the non-structured regions and the right and left macrodomains. *PLoS Genetics, 13*(5), e1006758-e1006758. https://doi.org/10.1371/journal.pgen.1006758

29) Valens, M., Penaud, S., Rossignol, M., Cornet, F., & Boccard, F. (2004). Macrodomain organization of the *Escherichia coli* chromosome. *The EMBO Journal, 23*(21), 4330-4341. https://doi.org/10.1038/sj.emboj.7600434

30) Espeli, O., Mercier, R., & Boccard, F. (2008). DNA dynamics vary according to macrodomain topography in the *E. coli* chromosome. *Molecular Microbiology, 68*(6), 1418-1427. https://doi.org/10.1111/j.1365-2958.2008.06239.x

31) Angel Robles-Ramos, M., Margolin, W., Sobrinos-Sanguino, M., Alfonso, C., Rivas, G., Monterroso, B., & Zorrilla, S. (2020). The nucleoid occlusion protein SlmA binds to lipid membranes. *Mbio, 11*(5). https://doi.org/10.1128/mBio.02094-20

32) Lawrence, J. G. (2005). The dynamic bacterial genome. In: Higgins, N.P., editor. The bacterial chromosome. Washington (D.C.): ASM Press. pp. 19–37.

33) Brandis, G., & Hughes, D. (2016). The selective advantage of synonymous codon usage bias in *Salmonella*. *PLoS Genetics, 12*(3), e1005926-e1005926. https://doi.org/10.1371/journal.pgen.1005926

34) Schmid, M. B., & Roth, J. R. (1987). Gene location affects expression level in *Salmonella* Typhimurium. *Journal of Bacteriology, 169*(6), 2872-2875. https://doi.org/10.1128/jb.169.6.2872-2875.1987

ACKNOWLEDGEMENTS

Foremost, I would like to offer my endless and sincere gratitude to my supervisor **Professor Diarmaid Hughes** for giving me a chance to study with him. It is not possible to summarize his support throughout my study in just one paragraph. His guidance, motivation and trust encouraged me to use my imagination in my experiments and shift my point of view towards problems in case of need.

I would like to thank **Gerrit Brandis** for the discussions we had and also for his valuable contribution to my data analysis.

I cannot thank **Cao Sha** enough for her positive attitude and her advices during my orientation to the experiments. I am also grateful to **Talia Berruga Fernández** for her tutorials in sequence analysis and her help in the laboratory.

Last but not least, I express my sincere thanks to my family. My mother **Mehtap Cerit** and my father **Şükrü Cerit** were always there when I needed them; they never got tired of supporting me in every aspect of life without questioning or judging me. Ultimately, my "elder" brother, **Erdem Cerit** who always supports me with his knowledge and experiences. I admire his intelligence, motivation and quick-witted nature, and I am hoping for the day I might possess half of those features.

SUPPLEMENTARY DATA

Supplementary Table 1: Strain list

Strain Name	Genotype
TH4527	LT2
TH10603	LT2 bla-kan(K138Stop)<>+1.5 , cam-kan(E3Stop)<>-4.5
TH10604	LT2 bla-kan(K138Stop)<>+3 , cam-kan(E3Stop)<>-4.5
TH10605	LT2 bla-kan(K138Stop)<>+4.5 , cam-kan(E3Stop)<>-4.5
TH10606	LT2 bla-kan(K138Stop)<>+6 , cam-kan(E3Stop)<>-4.5
TH10607	LT2 bla-kan(K138Stop)<>+12 , cam-kan(E3Stop)<>-4.5
TH10608	LT2 bla-kan(K138Stop)<>+16.5, cam-kan(E3Stop)<>-4.5
TH10645	LT2 bla-kan(K138Stop)<>+15, cam-kan(E3Stop)<>-4.5
TH10650	LT2 $INV2 (kan-kan) bla-kan(K138Stop-> K138wt)<>+4.5 , cam-kan(E3Stop)<>-4.5
TH10677	LT2 $INV3 (kan-kan) bla-kan(K138Stop??)<>+1.5 , cam-kan(E3Stop??)<>-4.5
TH10691	LT2 $INV5 (kan-kan) bla-kan(K138Stop??)<>+3 , cam-kan(E3Stop??)<>-4.5
TH10694	LT2 cam-kan(E3Stop)<>-3 , bla-kan(K138Stop)<>+4.5
TH10702	LT2 cam-kan(E3Stop)<>-15 , bla-kan(K138Stop)<>+4.5
TH10713	LT2 cam-kan(E3Stop)<>-6 , bla-kan(K138Stop)<>+4,5
TH10721	LT2 $INV6 (kan-kan) bla-kan(K138Stop??)<>+16,5 , cam-kan(E3Stop??)<>-4,5
TH10722	LT2 $INV7 (kan-kan) bla-kan(K138Stop??)<>+15 , cam-kan(E3Stop??)<>-4,5
TH10742	LT2 cam-kan(E3Stop)<>-1.5 , bla-kan(K138Stop)<>+4,5
TH10760	LT2 $INV8 (kan-kan) bla-kan(K138Stop??)<>+6 , cam-kan(E3Stop??)<>-4,5
TH10761	LT2 $INV9 (kan-kan) bla-kan(K138Stop??)<>+12 , cam-kan(E3Stop??)<>-4,5

Supplementary Table 2: Sequences of oligonucleotides used for diagnostic PCR

Name	Sequence (3’ – 5’)
-4.5_rv	CAGCAGCTCTCCAACTAT
+4.5_rv	GGAGACTAAGCGAATGAA
cat-out	GGCGGGCAAGAATGTGAATAAAGG
amp-out	GGTGAGCAAAAACAGGAAG

Supplementary Table 3: Numbers of plated cultures, rates of recombination and rates of inversion for each strain

Strain Name	TH10603		TH10604		TH10605		TH10606		TH10645	
Junction	-450/+150		-450/+300		-450/+450		-450/+600		-450/+1500	
# Plated Cultures	53		53		100		53		53	
Recombinant Type	KanR	INV	KanR	INV	KanR	INV	KanR	INV	KanR	INV
Frequency ($x10^{-6}$)	4.63	0.1	2.07	0.1	4.46	0.53	6.66	0.19	5.13	0.1

Strain Name	TH10742		TH10694		TH10713		TH10702	
Junction	-150/+450		-300/+450		-600/+450		-1500/+450	
# Plated Cultures	100		100		100		100	
Recombinant Type	KanR	INV	KanR	INV	KanR	INV	KanR	INV
Frequency ($x10^{-6}$)	13.3	<0.05	7.35	<0.05	3.47	<0.05	2.72	<0.05