



UPPSALA
UNIVERSITET

*Digital Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Medicine 1392*

A Unified Multitude

*Experimental Studies of Bacterial Chromosome
Organization*

EVA GARMENDIA



ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2017

ISSN 1651-6206
ISBN 978-91-513-0140-2
urn:nbn:se:uu:diva-332471

Dissertation presented at Uppsala University to be publicly examined in room B42, Uppsala Biomedical Centre (BMC), Husargatan 3, Uppsala, Friday, 15 December 2017 at 13:15 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English. Faculty examiner: Distinguished Professor John R. Roth (University of California, Davis).

Abstract

Garmendia, E. 2017. A Unified Multitude. Experimental Studies of Bacterial Chromosome Organization. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine* 1392. 66 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-0140-2.

Bacteria are many, old and varied; different bacterial species have been evolving for millions of years and show many disparate life-styles and types of metabolism. Nevertheless, some of the characteristics regarding how bacteria organize their chromosomes are relatively conserved, suggesting that they might be both ancient and important, and that selective pressures inhibit their modification. This thesis aims to study some of these characteristics experimentally, assessing how changes affect bacterial growth, and how, after changing conserved features, bacteria might evolve.

First, we experimentally tested what are the constraints on the horizontal transfer of a gene highly important for bacterial growth. Second, we investigated the significance of the location and orientation of a highly expressed and essential operon; and we experimentally evolved strains with suboptimal locations and orientations to assess how bacteria could adapt to these changes. Thirdly, we sought to understand the accessibility of different regions of the bacterial chromosome to engage in homologous recombination. And lastly, we constructed bacterial strains with chromosomal inversions to assess what effect the inversions had on growth rate, and how bacteria carrying costly inversions could evolve to reduce these costs.

The results provide evidence for different selective forces acting to conserve these chromosome organizational traits. Accordingly, we found that evolutionary distance, functional conservation, suboptimal expression and impaired network connectivity of a gene can affect the successful transfer of genes between bacterial species. We determined that relative location of an essential and highly expressed operon is critical for supporting fast growth rate, and that its location seems to be more important than its orientation. We also found that both the location, and relative orientation of separated duplicate sequences can affect recombination rates between these sequences in different regions of the chromosome. Finally, the data suggest that the importance of having the two arms of a circular bacterial chromosome approximately equal in size is a strong selective force acting against certain type of chromosomal inversions.

Keywords: bacterial evolution, chromosome organization and structure, chromosomal inversions, EF-Tu, horizontal gene transfer

Eva Garmendia, Department of Medical Biochemistry and Microbiology, Box 582, Uppsala University, SE-75123 Uppsala, Sweden.

© Eva Garmendia 2017

ISSN 1651-6206

ISBN 978-91-513-0140-2

urn:nbn:se:uu:diva-332471 (<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-332471>)

*In memory of my grandmother,
mi abuela, mi tata.*

*Y a mis padres, por hacer de mi
la persona curiosa que soy.*

Portrait in the back-cover
by Laura Francesconi

The role of the infinitely small in nature is infinitely great.

- Louis Pasteur

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

- I Kacar, B.[☉], **Garmendia, E.**[☉], Tuncbag, N., Andersson, D.I. and Hughes, D. Functional constraints on replacing a gene with its ancient and modern homologs. 2017. *mBio*. 8, e01276-17.
- II **Garmendia, E.** and Hughes, D. Phenotypic and genotypic responses to relocating a highly expressed bacterial operon. 2017. *Manuscript*.
- III **Garmendia, E.**[☉], Brandis, G.[☉] and Hughes, D. Chromosomal location determines the rate of homologous recombination. 2017. *Manuscript*.
- IV **Garmendia, E.**, Bueno-Galera, C., Huseby, D.L. and Hughes, D. The selective advantage of replicore balance in *Salmonella* Typhimurium. 2017. *Manuscript*.

Other articles co-authored but not included in this thesis:

Hammarlöf, D.L.[☉], Bergman, J.M.[☉], **Garmendia, E.** & Hughes, D. Turnover of mRNAs is one of the essential functions of RNase E. 2015. *Molecular Microbiology*. 98(1), 34-45.

Marton, S., Berzal-Herranz, B., **Garmendia, E.**, Cueto, F. J. & Berzal-Herranz, A. 2011. Anti-HCV RNA aptamers targeting the genomic cis-acting replication element. *Pharmaceuticals*. 5(1), 49-60.

[☉] authors contributed equally to the work.

Reprints were made with permission from the respective publishers.

Examining Committee

Faculty Examiner (Opponent)

Distinguished Professor **John R. Roth**

Department of Microbiology and Molecular Genetics

University of California, Davis, USA

Research areas: mutations under selection, recombination, chromosome rearrangements, gene amplifications, bacterial physiology.

Evaluating Committee

Professor **Lars Hederstedt**

Department of Biology.

Lund University, Lund, Sweden.

Research areas: molecular biology and genetics of Gram-positives.

Senior Lecturer **Margareta Krabbe**

Biology Education Centre

Uppsala University, Uppsala, Sweden

Research areas: molecular microbiology, gene regulation, biotechnology.

Professor **Stefan Åström**

Department of Molecular Biosciences

The Wenner-Gren Institute, Stockholm University, Stockholm, Sweden.

Research areas: recombination, mating-type switching, DNA repair.

Chairperson

Professor **Staffan Svärd**

Department of Molecular Cell Biology

Uppsala University, Uppsala, Sweden

Research areas: molecular biology, parasitology, genetics.

Contents

| | |
|--|----|
| Preface..... | 11 |
| Background..... | 13 |
| <i>Escherichia coli</i> and <i>Salmonella enterica</i> serovar Typhimurium as model organisms | 13 |
| Bacterial fitness..... | 14 |
| Experimental evolution | 14 |
| Horizontal gene transfer (HGT) | 16 |
| Mechanisms | 17 |
| Barriers | 18 |
| Role in evolution | 18 |
| Elongation Factor Tu (EF-Tu) and bacterial growth..... | 19 |
| The protein and its genes | 19 |
| EF-Tu and bacterial growth..... | 20 |
| Organization of the bacterial genome and chromosome..... | 21 |
| Functional units: the operons..... | 22 |
| Location bias and gene dosage effect | 23 |
| Gene orientation bias | 24 |
| Chromosome structure..... | 25 |
| Chromosomal rearrangements..... | 27 |
| Homologous recombination | 28 |
| Mechanism..... | 28 |
| Effects..... | 29 |
| Present Investigations | 31 |
| Aims of this thesis | 31 |
| Paper I: Constraints on HGT..... | 32 |
| Paper II: Location & orientation bias of a highly expressed operon...35 | |
| Paper III: Location/orientation and homologous recombination | 38 |
| Paper IV: Unbalanced inversions..... | 42 |
| Concluding Remarks..... | 45 |
| Future Perspectives | 47 |
| Svensk sammanfattning | 48 |
| Resumen en Español..... | 49 |
| Words of gratitude | 52 |
| References..... | 57 |

Abbreviations

| | |
|-------------|---|
| bp | Base pair |
| bya | Billion years ago |
| CID | Chromosomal interaction domains |
| DNA | Deoxyribonucleic acid |
| DSB | Double-stranded break |
| dsDNA | Double-stranded DNA |
| EF-Tu | Elongation factor Tu |
| Fis | Factor for inversion stimulation |
| HGT | Horizontal gene transfer |
| H-NS | Heat-stable nucleoid-structuring protein |
| HR | Homologous recombination |
| HU | Histone-like DNA binding protein |
| IHF | Integration host factor |
| Kb | Kilo base |
| Mb | Mega base |
| mRNA | Messenger RNA |
| NAPs | Nucleoid-associated proteins |
| <i>oriC</i> | Origin of replication |
| RNA | Ribonucleic acid |
| rRNA | Ribosomal RNA |
| ssDNA | Single-stranded DNA |
| Ter | Terminus of replication |
| tRNA | Transfer RNA |
| SMC | Structural maintenance of chromosomes complex |
| SNP | Single-nucleotide polymorphism |

Background

Escherichia coli and *Salmonella enterica* serovar Typhimurium as model organisms

Theodore Escherich (1857-1911) discovered what he called *Bacterium coli commune* in 1885, during the golden age of bacteriology. During his work on the bacterial flora of the gastrointestinal tract he described many strains of this bacteria, each one showing a different morphology and behavior. The ease of culturing it, together with its short generation time, were key factors in the continuance of Escherich's work. The now well-known name of *Escherichia coli* (from now on *E. coli*) to refer to *Bacterium coli commune* was not proposed until 1918, although the early nomenclature is not so clear. From the field of bacteriology, *E. coli* soon started to be the key microorganism in studies in biochemistry, and today is the most popular organism in studies in molecular biology and genetics¹. The strain of *E. coli* used in this work is known as *Escherichia coli* K12 MG1655². In this thesis, only **Paper I** uses *E. coli* as model organism.

Most of the strains of *Salmonella enterica* serovar Typhimurium (from now on *S. Typhimurium*) used in genetic analysis derive from the strain LT2. *S. Typhimurium* LT2 was originally isolated and studied in the 1940s³, and since then it has been used widely to study bacterial physiology and genetics. In this strain, the *rpoS* gene is defective (start codon changed from ATG to TTG), making it non-virulent⁴. The whole genome sequence of this strain was published in 2001⁵. Here, **Paper II, III and IV** use *S. Typhimurium* as model organism.

Both of the bacterial strains used in this work are good as model organisms in the study of molecular genetics for the following reasons:

- They are easy to grow on the lab using in both rich and defined media.
- They have a short generation time.
- They are a-virulent, making them safe to work with under normal laboratory conditions.
- It is possible to make genetic modifications in them by different procedures (e.g. transduction with prophages, P1 for *E. coli* and P22 for *S. Typhimurium*, and by lambda-red recombineering).
- They are single-celled haploid organisms, making the selection of preferred genotypes easy during manipulation.

Bacterial fitness

In general terms, fitness is the capacity of an organism to survive and reproduce in a given environment⁶. Depending on the particular organism, the specific factors that influence its survival and reproduction may differ, but for all organisms fitness is a key factor in evolution by natural selection. Proposed by Darwin in 1859, evolution by natural selection works on the variation between individuals in populations, where relative fitness can be inherited from generation to generation⁷. In this situation, the fittest individuals will over time accumulate in the population. For fitness to be inheritable it has to be genetic-based and consequently natural selection is one of the driving forces that, through selection for fitness, shapes and evolves genotypes.

When studying evolution, the term “relative fitness” of a genotype is often used, rather than absolute fitness. Relative fitness is the absolute fitness of a genotype in a particular environment normalized to that of a reference genotype⁸. This definition makes possible a meaningful comparison of genotypes and investigations into how genotypic changes affect fitness.

Bacterial fitness can be measured in the lab in many different ways. One of the most frequently used ways is to measure how long a bacterial cell takes to complete a generation (so-called doubling time) during exponential growth. This method has its limitations because the fitness of a bacterium is not only determined by how fast it grows in exponential phase but rather by its total growth performance throughout the whole life cycle. To address this limitation, pair-wise competitions can be made, where different (usually isogenic) strains are mixed in culture and let grow together (so-called head-to-head competition). After a certain number of generations, the ratio of the competitors is measured, compared to the initial ratio, and a selection coefficient can be calculated, reporting the relative fitness of one strain to the other. In this thesis, bacterial fitness is assessed in these two ways depending on the project and the particular goals. In **Paper I** and **Paper II**, bacterial fitness is estimated by means of doubling time because the effects on the mutants are expected to be maximal during exponential growth phase. On the other hand, in **Paper IV** bacterial fitness is assessed in head-to-head competitions because the chromosomal rearrangements made could potentially affect any part of the lifecycle of the bacterium, and also because the competition assays can discriminate smaller changes in relative fitness than can be achieved with growth rate assays.

Experimental evolution

For many years, evolutionary biology was limited to studying the gross phenotypes of living organisms and fossils as a way to understand the processes of evolution. More recently, the development of molecular biology

opened up possibilities to compare organisms at the more fundamental level of protein and nucleic acid sequences, but the traditional view, that evolution was something that happened in the past and was difficult to reconstruct, persisted. To better understand how evolution works, there was a need for methods that could allow us to observe the actual process as it happened and to make experiments designed to test different hypotheses about evolutionary processes. With this purpose in mind, the field of experimental evolution using model organisms, was created. Early work mainly focused on *Drosophila* and bacteria^{9,10}. Initially, technical issues limited progress and only in the last two decades has it become routine to use microorganisms to study evolution.

In experimental evolution, a population of organisms is presented with a problem (that affects their fitness) and given time (in terms of generations) to solve that problem (*Fig. 2*). Microorganisms are of great use to study evolution due to their short generation times, large population sizes, and the possibility to store frozen samples for later analysis. Also, experimental studies in bacteria can be performed under different conditions that promote evolution and adaptation of different traits. For example, work with bacterial colonies on agar can be used to create extreme population bottlenecks that are useful for studying genetic drift and mutation accumulation, whereas using liquid batch culture is very useful for studying adaptation within populations¹¹.

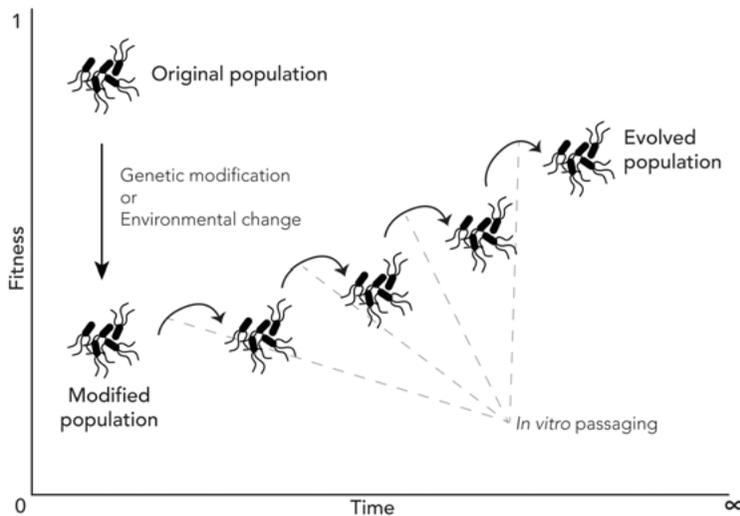


Figure 2. Experimental evolution. Schematic representation of an in vitro evolution experiment.

Several important evolutionary concepts have been tested in microbial experimental evolution studies. It has been shown that populations can adapt quickly to new environments but also that specific adaptations to one environment can be detrimental in another environment^{12,13}. Moreover, it has been shown how important population sizes are for evolution, from allowing

the spread of *a priori* detrimental mutator phenotypes to making potentially beneficial mutations invisible to selection¹⁴. A very important result from experimental evolution studies is that they have shown that parallel molecular evolution frequently occurs in independent replicate populations¹⁵, which helps in the understanding of eventual changes found in evolved populations, as parallel changes are presumably beneficial in particular given situations.

In this thesis, both **Paper II** and **Paper IV** present results from experimental evolution. For both projects, experimental evolution was performed in rich liquid media with a bottleneck of 10^6 cells per passage and for a total of 500 generations.

Horizontal gene transfer (HGT)

Horizontal gene transfer (also referred to as lateral gene transfer) is defined as the movement of genetic material between individuals that is not done through the vertical parent-offspring line. This type of transfer was first seen and identified in the late 1920s through the transfer of virulence factors in *Pneumococcus* infected mice¹⁶, and later described in the 1940s in the studies of recombination in bacteria¹⁷. But it was not until recently, with the explosion of genome sequence data from the three domains of life, and the development of tools to recognize HGT, that we could see instances of HGT between different species within domains, and also inter-domain HGT (bacteria and archaea; bacteria and eukarya and archaea and eukarya). Although the majority of the studied instances of HGT are between microorganisms, there are also instances of transfer between microorganisms and multicellular organisms¹⁸.

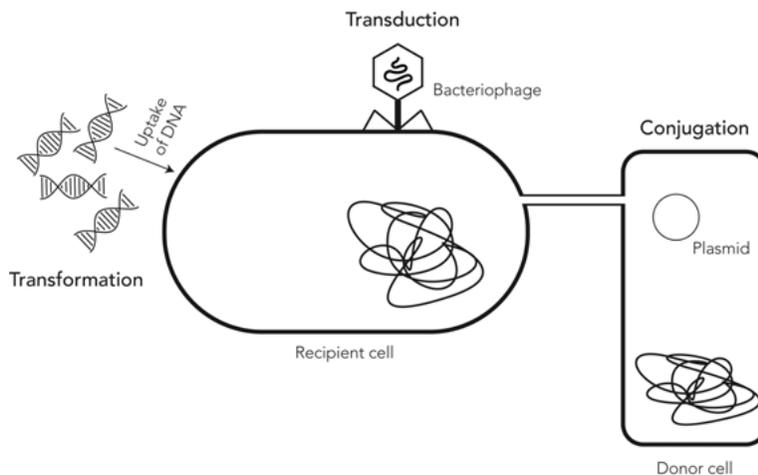


Figure 3. Main mechanisms of HGT. Representation of a cell receiving DNA via the three main ways for the horizontal transfer. Note: objects not to scale.

Mechanisms

The three most studied mechanisms of HGT between prokaryotic organisms are transformation, transduction and conjugation (*Fig. 3*).

Transformation is the uptake of free DNA present in the surroundings of the cell. For this to happen, the bacterium needs to be in what is called a state of competence, and such a state is generally coupled with specific environmental conditions and happens for a short amount of time, limiting the possibilities for the uptake to successfully occur. The particular environment will also play a role in determining the fate of the DNA and whether it is degraded or recombined into a new genome¹⁹.

The transfer of bacterial DNA between bacterial cells by means of bacteriophages (prokaryotic viruses) is called transduction. With a low frequency, phages incorporate part of the bacterial host DNA in their capsids during phage formation and excision (generalized transduction). Generalized transduction is of particular interest in HGT because phages can act as carriers of bacterial genetic material that can potentially be integrated into a new genome. The species range of phage infection potentially limits the transfer of DNA via transduction.

Conjugation is referred to as the transfer of DNA through a cell-to-cell bridge containing a pore. This process is mainly associated with transference of plasmids, although some plasmids can also be acquired through transformation. Conjugative transfer can occur through different mechanisms, but the most complex systems are encoded by large plasmids in Gram-negative bacteria and use a type-IV secretion system to form a pilus that mediates the cell-to-cell contact, and a pore through which a copy of the plasmid is transferred²⁰. The species or strain specificity of the conjugation event depends on the plasmid, and it is generally related with the outer membrane proteins and other surface components on the recipient. In many cases, a bacterium containing a plasmid cannot be the recipient of a second plasmid of the same type. The range of species through which a plasmid can spread varies depending on the plasmid, but some, such as the plasmids of the Inc-P group, are very promiscuous and can transfer between different species and bacterial groups (Gram-positive/Gram-negative)²¹.

More recently, new mechanisms of gene transfer have been described. Examples include gene transfer agents (GTAs), which are delivery systems encoded in chromosomes and under host regulatory control and which resemble transducing prophages, or the process of cell fusions, as it has been observed between archaeal groups²².

Barriers

Each process of HGT is restricted by particular barriers, such as the range of host species for conjugation or transduction, the rate of uptake and amount of DNA available for transformation, or the physical proximity of cells limiting the probability of conjugation. Additionally, for transformation, transduction and conjugation (unless the conjugative element is able to self-replicate in the host), an additional barrier to HGT is the successful integration of the DNA into the recipient genome. If host mechanisms such as restriction endonucleases or potential CRISPR systems do not degrade the incoming DNA, integration can happen through homologous recombination (HR) (see *Homologous recombination* section below). For HR to be effective the sequences involved should be at least 70-75% identical at the nucleotide level²³. However, a sequence divergence barrier can be overcome by integration through illegitimate recombination, albeit at a lower frequency²⁴.

If a piece of DNA overcomes all the barriers mentioned above, its fate within the host will ultimately depend upon its effect on the host's fitness. Transference of whole operons that work as functional units can provide recipients with the tools necessary to occupy new niches and therefore give an advantage to the host. This is the case, for example, with HGT of pathogenicity islands or of plasmids encoding genes that confer resistance to antibiotics. At present, the consensus is that the majority of the successful transfers are nearly neutral to the recipient and the selection acting upon them is weak^{25,26}. Nevertheless, the potential for a transferred DNA segment to be costly is big since it can disrupt conserved genomic and chromosomal features (see *Organization of the bacterial genome and chromosome* section below), have a cytotoxic effect, sequester resources, and/or disrupt established cellular networks²⁷.

According to the so-called complexity hypothesis, fixation of a gene after HGT is more probable for operational genes (genes whose function does not require many interactions) than for informational genes (genes involved in core functions such as replication, transcription and translation, with many cellular partners)²⁸. In **Paper I**, I aim to understand what are the barriers to the transfer of the latter group by using an ancient, conserved, and highly connected informational gene such as *tuf* (see section *Elongation Factor Tu (EF-Tu) and bacterial growth* below).

Role in evolution

The classical tree-like visualization of evolution of species is challenged by the presence of HGT between different branches. The binary, dichotomous schemes first proposed by Lamarck and Darwin^{7,29} represents the evolution of species from one common root (also known as the Last Universal Common Ancestor -LUCA-) and as a bifurcating process, due to the vertical

transmission of genetic material from parent to offspring and the ongoing processes of speciation. The presence of HGT became evident when evolutionary analyses of different genes and proteins resulted in incongruent phylogenies, splitting groups that otherwise (through morphology, physiology and other markers) would be cohesive^{30,31}.

The assessment and identification of HGT events is still a challenge in modern biology. One way to identify possible HGT events is by identifying different patterns of the codon usage bias and base composition in different parts of a genome, but over time acquired DNA, if preserved, is expected to undergo the same mutational processes as the host and over time these signals will eventually be lost, making it difficult to recognize ancient HGT events³². The identification of HGT by incongruent phylogenetic gene trees is also faulted with since a spotty gene distribution could also be associated with duplication events followed by deletions, and the determination of what is a proper “reference tree” remains always a hot topic²². Recent analyses have highlighted a biased gene transference between closely related organisms, for both phylogenetically related species and species sharing a community, which can potentially strengthen prior monophyletic classifications^{33,34}.

Despite the difficulties in recognizing and assessing transfer events, it is nevertheless evident that HGT plays an important role in the evolution of species both now at present and in the past^{18,35} and some authors propose that it should be included in modern paradigms of evolution, together with other mechanisms, to better understand the real shape of the tree, or web, of life.

Elongation Factor Tu (EF-Tu) and bacterial growth

The protein and its genes

The elongation factor Tu (EF-Tu) is one of the key players in the synthesis of proteins, bringing the aminoacylated-tRNA (aa-tRNA) to the ribosome as part of the ternary complex during the elongation step in the translation process³⁶. EF-Tu is formed of 394 amino acids and they are structured into three globular domains, one α/β domain where GTP/GDP binds (domain 1), and two β domains (domains 2 and 3)³⁷ (Fig. 4). All three domains are important for the binding of the aa-tRNA but it is domain 2 that is specifically involved in the binding of the ternary complex to the ribosome³⁸. Under normal growth situations, EF-Tu is the most abundant cytoplasmic protein in *Escherichia coli* and it accounts for up to 9% of all soluble proteins in the cell in *S. Typhimurium*^{39,40}.

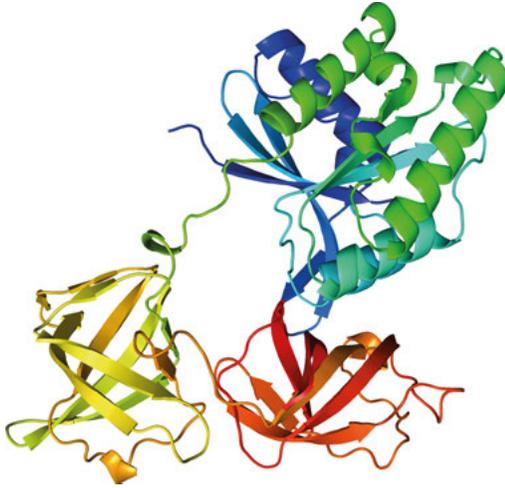


Figure 4. 3-Dimensional structure of EF-Tu in its inactive, GDP-bound conformation. Blue/green: domain 1, Red: domain 2, Yellow: domain 3. Figure created with UCSF Chimera package (PDB entry 1EFC).

In *S. Typhimurium*, as well as in other proteobacteria, EF-Tu is encoded by two separate genes *tufA* and *tufB*, which are almost identical in sequence^{41,42}. *tufA* and *tufB* are located in different arms of the chromosome, approximately equidistant from the origin of replication. *tufA* is the last gene in the operon containing the genes for two ribosomal proteins (*rpsL* and *rpsG*) and the gene for the elongation factor G (*fusA*)⁴¹. *tufB* is located as the last gene in an operon together with four tRNA genes (*thrU*, *tyrU*, *glyT* and *thrT*), two of which are essential tRNAs⁴³. It has been shown that *tufA* and *tufB* maintain their sequence similarity in *S. Typhimurium* by means of gene conversion, showing the importance of recombination in co-evolution of the genes and in repairing mutations to maintain bacterial fitness⁴⁴⁻⁴⁶. One proposal is that an ancient duplication followed by differential loss has caused the spotty distribution of the *tuf* duplication among bacterial species, but it has also been proposed that an horizontal gene transfer event is the cause of this duplication in *Enterococci*^{47,48}.

Here, I've used the EF-Tu system in the first two papers. For **paper I**, I used the *tuf* gene, as it is an ancient and conserved gene, to explore the limits of its transferability. For **paper II**, I used the whole *tufB* operon as a model to understand the biases on organization of highly expressed operons and the physiological consequences of its displacement.

EF-Tu and bacterial growth

EF-Tu is the most abundant cytoplasmic protein in the cell and there is a strong selective pressure to maintain its two genes intact because having functional EF-Tu is crucial for bacterial growth⁴⁹. Consequently, it has been shown that growth rate and translation rate are correlated with EF-Tu concentration in *S. Typhimurium*⁴⁰. In general, around two thirds of EF-Tu in the cell comes from

tufA and one third from *tufB* and, although the EF-Tu protein is essential to the cell, each of the *tuf* genes can be individually deleted without killing the cell^{50,51}. In the situation of a *tufB* knock-out mutation, the bacterial growth is, as expected, two thirds of the wild-type level, whereas for a *tufA* knock-out the expected one third of the wild-type growth is increased to two thirds, because there is a doubling of the amount of EF-Tu that comes from the *tufB* locus^{40,50}. This situation suggests that a specific process, dependent on EF-Tu concentration, regulates the expression of *tufB*⁵²⁻⁵⁶. The specific mechanism of autoregulation of EF-Tu expression from the *tufB* operon has recently been shown to work via transcriptional speed and post-transcriptional regulation⁵⁷.

Organization of the bacterial genome and chromosome

Bacterial genomes are variable in gene content and size all across the eubacterial kingdom, yet some organizational features are highly conserved (Fig. 5).

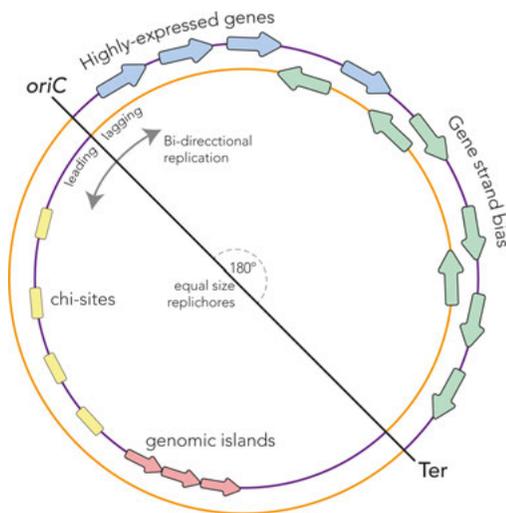


Figure 5. Schematic representation of conserved organizational features in bacterial genomes. Highly expressed genes are shown in blue, non-highly expressed genes are shown in green or red. Chi-sites are shown in yellow. Note: not to scale.

In general, bacterial chromosomes are circular and have symmetric replichores, with the replication starting at a unique origin (*oriC*), progressing bi-directionally, and terminating at a unique terminus (*Ter*) region. Replication occurs continuously on the leading strand (origin to terminus direction) but discontinuously on the lagging strand (terminus to origin direction). There is usually a gene strand bias resulting in more genes being transcribed from the leading strand, and frequently also a positional bias with highly expressed genes being located closer to the origin of replication and thus benefiting from a gene dosage effect because multiple replication cycles may be concurrent in the same cell (see specific sections below). Recombination-associated *chi* sites, GC skews and segregation-related motifs

are also typically more frequently encountered on the leading strand. Additionally, genomic islands are created as the result of horizontal gene transfer^{58,59}. Bacterial genes are often organized in operons, where multiple genes are co-transcribed from the same promoter. Frequently, but not always, operons carry functionally related genes. Some operons are grouped into superoperons or uber-operons, areas where functionally related genes coded in different operons are located in proximity to each other on the chromosome (for example, operons involved in flagellar structure and chemotaxis)⁶⁰. Finally, because the length of a bacterial genome is much greater than the length or breadth of a bacterial cell, it is tightly packed into structural domains at different scales (see section on *chromosome structure* below). That these organizational features are commonly found, even in distantly related species of bacteria, whereas their gene content and nucleotide sequence identity can be very different, is evidence that there are constraints and selections promoting genome-level organizational conservation.

Functional units: the operons

More than half a century ago, it was noted that some genes had the same order in the chromosome as their products were active in metabolic pathways, that the expression of those genes was often controlled by a common regulatory region, and that the genes were co-transcribed in a unit that was named the operon⁶¹⁻⁶³.

From the evolutionary point of view, many scientists have asked the question of why there are operons and more interestingly, how they have formed and evolved to their current status. One line of thought is that genes are linked together by selection for regulatory purposes^{64,65}. Here, regions with one regulatory sequence are under stronger selection and can develop complex regulatory strategies to meet the necessities of the cell. Additionally, in bacteria transcription and translation are coupled and co-transcribed genes will result in co-translated proteins, with potential benefits when products are to work together in metabolic pathways or protein complexes. Another argument for the regulatory model is that co-transcribed genes will undergo mRNA degradation at the same time and this might help the regulation of gene expression. Although this model might explain the existence of operons it does not explain how individual genes came to be organized into operons.

To find an answer to this latter question, another model for the existence and maintenance of operons was proposed by Lawrence and Roth, namely that operons are selfish entities that have formed due to a selection at the gene level making operons fitter for horizontal gene transfer⁶⁶. This way, the transferred information is integrated into the new cell and because genes in operons code for products involved in a specific function, it can increase the fitness of the host by providing a complete functional unit, for example allowing growth in a new niche. This model also presents some inconsistencies, for instance, it

predicts that clustering will happen preferentially for non-essential genes and will be underrepresented for essential genes, nevertheless we find many examples of essential genes in operons and non-essential genes in single-transcript units. An inconsistency pointed out by the authors is that gene clustering can differ among species and some operons seem to contain genes that to date are not related. Mechanistically, operons can evolve by insertions into transcriptional units, by deletions of neighboring genes, or by global rearrangements creating new junctions between genes⁶⁷.

Although these main proposed models for the formation and existence of operons seem to have some shortcomings, one can think of situations where it is not one way or the other, and that different selections and processes that bring together both models could have contributed to the actual organization of genes in operons.

Location bias and gene dosage effect

To be able to grow fast, bacteria can start new rounds of replication before the previous one finishes. Due to this, genes that are encoded in regions close to the origin of replication are more abundant in the cell than genes located closer to the terminus⁶⁸. For example, if a bacterium starts 2 rounds of replication before it divides, a single gene coded near the origin will be 4 times more abundant than a gene coded near the terminus (Fig. 6). As a result of this amplification, genes located near the *ori* can be expressed more highly under fast growth situations than the same gene if it were near the terminus^{69–71}.

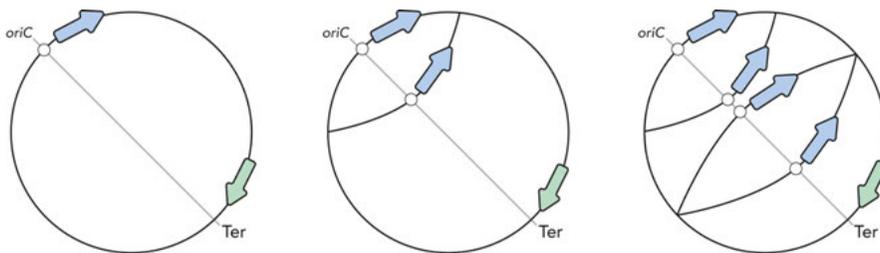


Figure 6. Gene-dosage hypothesis. Representation of genetic configuration of a cell undergoing overlapping replication rounds. The relative increase in copy number of a gene near the *oriC* is 2^n , being n the number of replication rounds. Note: single black line represents dsDNA.

This situation created by the replication process can be exploited in natural adaptive evolution where genes whose products are highly expressed and needed under fast-growth conditions are under selection to be located closer to the origin of replication. In accordance with this, genes coding for products involved in transcription and translation tend to map near the origin of replication in fast-growing bacteria^{72,73}. Recent work on *Vibrio cholera* has also shown that this advantage of gene-dosage can be seen at fast and slow

grow conditions, suggesting the importance of gene location throughout the whole life cycle and not only during conditions that are optimal for fast growth⁷⁴.

Replication-associated gene dosage effects may also be phenotypically important in situations where the progression of replication forks is slowed down. Thus, in the case of induction of genetic competence under stress conditions in *Streptococcus pneumoniae*, the relative locations of the genes involved pushes the dosage of early genes over a critical threshold and activates the competent state⁷⁵. Another recently described example of the effect of gene location in a global physiological process controlled by gene expression levels is in the sporulation of *Bacillus subtilis*. Here, the relative expression of two genes is involved in regulating sporulation-induction in a way that is dependent on their position on the chromosome relative to the origin of replication and how that position affects their relative gene dosage levels^{76,77}.

In this thesis, the importance and effect of location of a highly expressed operon whose product is directly related to fast growth rates is investigated in **Paper II**.

Gene orientation bias

Another situation created by the idiosyncrasy of the replication process is the existence of a leading strand of DNA (synthesized continuously) and a lagging strand of DNA (synthesized semi-discontinuously). The difference in their synthesis leads to differences between the leading and lagging strands, such as a difference in their relative base compositions, a difference known as GC skew⁷⁸.

In actively growing and dividing bacterial cells, chromosome replication and transcription take place at the same time, and both processes use DNA as their template. Due to the differences in speed between the DNA and RNA polymerases (the latter being around 10-fold slower), collisions between the different polymerases are expected to occur⁷⁹. These collisions are co-oriented if the transcribed gene is encoded on the leading strand, but will be head-on collisions if the gene is encoded on the lagging strand. A bias in favor of having genes encoded on the leading strand was originally thought to be the result of the selection against these head-on collisions⁸⁰. Initially, it was proposed that highly expressed genes (e.g. ribosomal RNA and ribosomal protein genes) would be selected to be on the leading strand because they are more likely to be highly transcribed during fast growth when the replication forks are also proceeding⁸⁰. Later, more comprehensive and extensive bioinformatics analysis found that the essentiality of genes, their functional category and their chromosomal configuration also play a significant role in this gene orientation bias⁸¹⁻⁸⁴. In *S. Typhimurium*, 95% of all genes involved

in transcription and translation, which are highly expressed, are coded in the leading strand⁸⁵.

But, why would head-on collisions be selected against? On one side, head-on collisions have been shown to affect DNA replication progression more than co-directional collisions^{86,87}. Experimental work has also shown that replication intermediates accumulate if head-on collisions cannot be resolved properly⁸⁸, and that total disruption of a replication fork can affect genome integrity and activate DNA damage responses⁸⁹. Therefore, high level transcription from the lagging strand could potentially lead to impairment of the ability to complete replication and to increased cell death. Additionally, it has been proposed that there is increased mutagenesis of genes transcribed from the lagging strand⁹⁰, a situation that could also drive the orientation bias observed for essential genes. Increased mutagenesis has also been proposed to be a selective driver of the maintenance of certain genes in the lagging strand, as these genes would benefit from increased mutation rates^{91,92}. The production of truncated transcripts, reducing gene expression, has also been hypothesized to play a role in the selection against head-on collisions, in particular for essential genes⁹³ and for long, highly-transcribed operons⁸³.

In **Paper II** of this thesis, the relative effect in growth of the orientation of a highly expressed operon is examined.

Chromosome structure

Were it to be fully stretched out, the average bacterial chromosome would be about a thousand times longer than the cell that contains it. Nevertheless, the whole chromosome is tightly packed and organized in a way that allows the proper execution, in space and time, of all the cellular processes that have DNA at their core: replication, transcription, repair, recombination, and segregation. This organization occurs at different levels (*Fig. 7*). At the smallest level, nucleoid associated proteins (NAPs) compact the DNA and help regulate gene expression⁹⁴. Then, the action of the replication and transcription machinery, together with the topoisomerases, creates patterns of supercoiling domains that can expand anywhere between a few Kb to several hundreds of Kb⁹⁵. At a higher scale, there are larger regions called macrodomains, which allow the compartmentalization of different chromosomal properties⁹⁶. All these organizational levels end up compacting the chromosomal DNA into a discrete structure that occupies part of the bacterial cell, called the nucleoid.

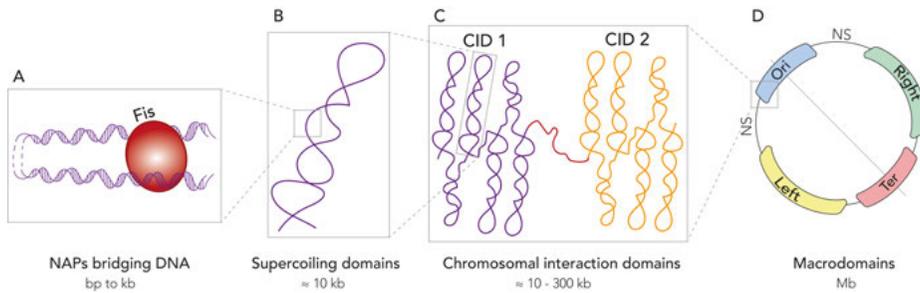


Figure 7. Levels of chromosome structure. A) Compaction by NAPs. B) Plectonemes formed by supercoiled DNA. C) Chromosome interaction domains separated by a highly expressed region (red). D) Macrodomain organization of the bacterial chromosome. In each case the average size is indicated at the bottom.

The bacterial chromosome is organized and compacted at a first level by the action of the NAPs (Fig. 7A). These proteins bind relatively non-specifically to DNA and can wrap, bend and bridge chromosomal regions⁹⁴. One of these NAPs is H-NS, a small protein that has been shown to bind to AT-rich or curved DNA and can bring distant loci together⁹⁷. H-NS binding seems to coincide with supercoiling-sensitive promoters⁹⁸ and it can regulate gene expression by blocking transcription-related binding sites⁹⁹. HU is another NAP that helps the organization of the chromosome by wrapping DNA around itself and inducing sharp bends. Work with HU mutants suggest a general role in chromosome compaction, and it possibly also influences supercoiling¹⁰⁰. IHF and Fis are two additional NAPs that can sharply bend and compact DNA and have been shown to play important roles in gene expression by connecting RNA polymerases with distant regulatory factors¹⁰⁰. Apart from those small NAPs, the nucleoid is also influenced by the action of larger protein complexes, such as SMC (or their homolog MukBEF in *E. coli*), which help to compact the DNA and maintain supercoiling, and are also involved in the positioning of the origin of replication¹⁰⁰. SMC has been recently show to juxtapose the chromosome arms in *B. subtilis* in a dynamic way throughout the cell-cycle¹⁰¹.

At a higher level of organization, loops of chromosomal DNA are supercoiled, which in itself are coiled up, forming what is known as a plectonemes, supercoil domains or topological domains¹⁰² (Fig. 7B). Work based on responses to DSB and site-specific recombination assays have estimated an average length of 10kb for these supercoil domains^{95,103}. The boundaries and dynamics of plectonemes vary a lot *in vivo*, depending partly on bound proteins and partly on gene expression, since these, directly or indirectly, modify DNA supercoiling and its diffusion.

Gene expression has recently been shown to be of special importance in chromosome organization¹⁰⁴. Long and highly expressed genes seem to act as boundaries of chromosomal interaction domains (CIDs) by physically separating DNA supercoils due to DNA unwinding, independently of active

translation¹⁰⁵. These CIDs seem to be a higher order of organization than supercoil domains (*Fig. 7C*), and are more stable across the cell-cycle¹⁰⁴, although the boundaries significantly change depending on the state of the cell, reflecting a different set of highly expressed genes being active¹⁰⁵.

At the highest level of organization, there are macrodomains, which can span through several Mb in length (*Fig. 7D*). These regions can be seen occupying the same space in the cell and sequences within macrodomains have been shown to recombine with higher frequency than across domains⁹⁶. The *E. coli* chromosome is organized in four stable macrodomains: Ori, Ter, Left and Right, with two unstructured DNA regions at both sides of the Ori. The macrodomains play an important role in the dynamic segregation of the newly replicated chromosome¹⁰⁶. Specific processes of compaction and organization can vary between macrodomains, as is illustrated by the MatP-*matS* complexes in the Ter macrodomain seen in *E. coli*¹⁰⁷ and the ParB-*parS*-SMC system at the Ori in *B. subtilis*¹⁰⁸.

This physical higher-order multi-level organization of the bacterial chromosome is, undoubtedly, tightly linked to the functional processes of the cell. New *in vivo* techniques are helping us elucidate the overall dynamics of the chromosome, how function and structure are related, and how structure, in turn, has evolved¹⁰⁹. In **Paper III**, I investigate the relative accessibility of different chromosomal regions to each other, which can drive chromosomal and gene organization if there is a selection for repair and increased interaction between regions.

Chromosomal rearrangements

Although the chromosome and genome organization seem to be conserved and finely tuned, recombination during DNA replication and repair can lead to rearrangements that can potentially disrupt or change the above-mentioned organizational traits (see *Homologous recombination - Effects* section below). The overall conservation of the chromosome map of *Salmonella* and *E. coli*, although they separated more than 100 million years ago, suggests that long-range rearrangements might be detrimental on a long-term scale¹¹⁰. Recombination between directly oriented repeats in the chromosome can lead to duplications of the intervening region, or deletions of the section, with the additional creation of novel joints in either of those two cases. If a segment is deleted, the DNA fragment can potentially integrate elsewhere, creating a DNA translocation¹¹¹⁻¹¹³. When recombination occurs between inverted repeats, it can lead to the inversion of the whole intervening region¹¹⁴⁻¹¹⁶. Although recombination between any two inverted repeats can lead to an inversion, a good body of research has shown that there are certain limitations to both the creation (mechanistic problems) and the maintenance (selection problems) of certain type of inversions¹¹⁷⁻¹²⁰. Genome-wide studies of related species frequently find that the relative positions of sequences conserved

between the species generate an X-shaped pattern in scatterplots, suggesting that inversions, when they occur, are usually symmetrical around the origin of replication¹²¹.

In this thesis, I investigate the effects on fitness of creating asymmetrical chromosomal inversions, and the process of their compensatory evolution, in **Paper IV**.

Homologous recombination

Homologous recombination (HR) can be defined as the process of informational exchange between two sequences with enough shared nucleotide identity, catalyzed by a dedicated group of enzymes¹²². In bacteria, this process can occur between two chromosomal DNA sequences, between the chromosome and an extrachromosomal DNA element (phage, plasmid or transformed linear DNA) and between two extrachromosomal elements. Depending on the topology of the substrates, the HR events can be classified into circular-by-circular (e.g. plasmid integration into the chromosome), circular-by-linear (e.g. integration of HGT linear piece of DNA into the chromosome) and linear-by-linear (e.g. phage crosses). Traditionally, HR has been detected by genetic approaches such as screens and selections^{123,124}, or physical methods that measure changes in density, size or configuration of the involved molecules¹²⁵⁻¹²⁷.

In this thesis, the accessibility of different regions of the *S. Typhimurium* chromosome to undergo HR is studied in **Paper III**.

Mechanism

The body of literature dedicated to identify the different main and alternative enzymes and pathways involved in HR in bacteria is a beautiful example of the power of genetics, driven by the fact that recombination-deficient mutants are sensitive to DNA damage. Brute-force screenings and selections of single, double and triple mutants of genes involved in HR has divided the process into three main stages: (i) pre-synapsis, where the substrate is recognized; (ii) synapsis, where the homologous sequence recognition and strand exchange happen; and (iii) post-synapsis where the formed junctions are resolved¹²². In bacteria the first step is performed by the RecBCD complex (alternatively RecFOR), the central homologous recombinase RecA performs the second step, and the third step occurs through the actions of RuvABC and, alternatively, RecG¹²⁸ (*Fig. 8*).

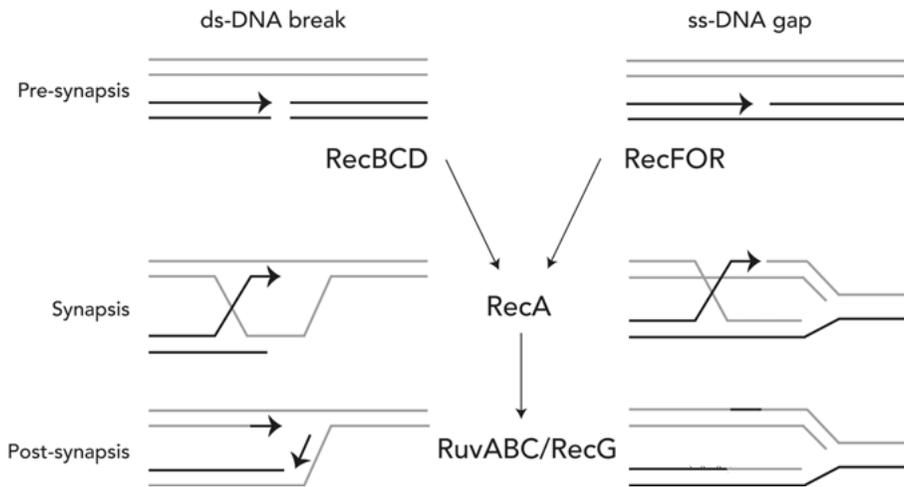


Figure 8. Pathways of HR in bacteria. Steps of repair of dsDNA breaks and ssDNA gaps is shown, together with the enzymes and complexes responsible for each step. Black and grey lines represent DNA strands from sister chromosomes.

When a double-stranded break (DSB) happens in the chromosome, the enzyme complex RecBCD quickly degrades the DNA until it reaches a recognition site known as chi (χ)¹²⁹. At this site, the enzyme complex continues degrading but only the 5' strand, creating a ssDNA overhang at the 3' end where the RecA enzyme binds and filaments. After filamentation, RecA binds to dsDNA of the sister chromosome and finds a homologous region. When homology is found, RecA mediates the replication restart by recruiting enzymes that polymerize from the 3' DNA end. When the degraded DNA is repaired, replication can restart. The junction formed as consequence of the strand exchange (known as a Holliday junction¹³⁰) will be resolved by the enzymes RuvABC (or alternatively RecG)¹³¹.

In the case of a single-stranded gap in the chromosome, the ssDNA is quickly protected by single strand binding proteins (SSB). The complex RecFOR removes the SSB and allows RecA to come in and polymerize on the single strand. RecA then searches for a homologous dsDNA sequence and promotes repair of the gap¹²⁸.

Effects

The observation that HR-deficient mutants were particularly sensitive to DNA damage hinted towards the main function that HR has evolved for: DNA repair and maintenance of chromosome integrity. If the bacterial chromosome gets any lesion that breaks it (double-stranded breaks -DSB- or single-stranded gaps), the DNA replication machinery is unable to continue, eventually

leading to cell death. To avoid this, HR is activated in the cell and assures that chromosome breaks get repaired before DNA replication can continue. Nevertheless, the effects of HR in the cell are not only restricted to those effect of *silent* nature, since basic recombinatorial repair should go unnoticed after DNA replication restarts, because the sequence is not changed. HR has also profound and very important consequences of an evolutionary nature that, over time, shape genomes and influence the evolution of bacterial species.

One of these effects is the evolution of gene families through homogenization of sequences, such as the rRNA operons, the genes coding for EF-Tu, or the genes involved in nitrogen-fixation^{44,132,133}. This type of evolution, through homogenization of sequences by HR, is known as gene conversion, and is responsible for counteracting the divergence of genes within gene families over time¹³⁴. Apart from gene families, other types of repeat sequences in the chromosome, such as *Rhs* elements, insertion sequences, and small extragenic repeats, can also undergo HR-mediated gene conversion and avoid sequence divergence^{135,136}.

Another consequence of HR between homologous repeats in the chromosome is the creation of chromosomal rearrangements (see *Chromosomal rearrangements* section above). These rearrangements can be seen in the form duplications or deletions when the HR happens between direct repeats, or inversions when HR occurs between inverted repeats^{114,137,138}. The creation of chromosomal rearrangements is an important source of genetic variation upon which evolution can act. For example, duplications, and subsequent amplifications, have been shown to be a source of genetic innovation¹³⁹, and can also act as a compensatory mechanism after transfer of unfit sequences¹⁴⁰. Selection of deletions during reductive evolution has also been observed in the lab¹⁴¹, and genome-wide inversions seem to have played an important role in the evolution of the pathogen *Salmonella enterica* serovar Typhi¹⁴².

Last, but not least important, HR is partially responsible for the evolution of bacterial species through HGT (see *Horizontal gene transfer* section above). All the three main mechanisms of HGT (conjugation, transformation and transduction) rely on HR in one way or another. For conjugation, integration of whole plasmids into bacterial chromosomes happens through HR, as exemplified by the creation of Hfr strains¹⁴³. Successful insertion of transformed or transduced sequences into the chromosome relies mostly on HR, although illegitimate recombination can also happen¹⁴⁴.

Present Investigations

Aims of this thesis

This thesis, as the title implies, has several aims which are intended to deepen our understanding of the bacterial chromosome organization. These aims can be detailed as follows:

1. HGT seems to not be homogenously successful for all gene classes. Can we identify the main reasons that could affect the transfer of a highly connected, highly expressed and essential gene into a new chromosome? Is there any relationship between evolutionary distance of the transferred sequence with the new host and the potential constraints?
2. Given that there is a widely-extended bias in the organization of highly expressed and essential genes, how do changes in these organizational traits affect the bacterial cell? How can bacterial populations adapt to those changes?
3. Homologous recombination has the power to profoundly change the organization and dynamics of the chromosome, we thus ask, how does the location and relative orientation of separated duplicate sequences affect the rates of homologous recombination between them?
4. There is typically a 180° axis between the replication origin and terminus in circular bacterial chromosomes. There seems to be a selection against the fixation of inversions that disturb this axis. Does the degree of imbalance predict the magnitude of the effect? How can unbalanced chromosomes evolve to restore fitness?

In this chapter, I present the main methods and findings of the studies I have worked on for these past years and discuss their conclusions and implications. However, highly detailed writings of those studies can be found in the attached papers.

Paper I: Constraints on HGT

The so-called complexity hypothesis proposes that the two main factors constraining the successful transfer of genes are the function and network interactivity of the gene's product^{28,145}. According to this, genes that carry out functions in complex and connected systems like central informational tasks (replication, transcription and translation) are potentially less successfully transferrable to a new genome than other type of genes. In this work, we use the bacterial gene *tuf*, coding for the elongation factor Tu (EF-Tu), as a model to study the constraints on transfer of a highly-connected and essential gene involved in information activities.

To study if there is a correlation between organism fitness and the evolutionary distance between the endogenous gene and the substituted variants, and the possible constraints on transfer of these variants, we replaced the native *tufA* gene of *E. coli* with 16 homologs coming from a wide span of extant taxa (*Yersinia enterocolitica*, *Vibrio cholerae*, *Pseudomonas aeruginosa*, *Legionella pneumophila*, *Bartonella henselae*, *Streptococcus pyogenes*, *Bacillus subtilis*, *Thermus thremophilus*, *Mycobacterium smegmatis* and *Thermotoga maritima*) and from six ancestral resurrected sequences representing *tuf* genes from nodes extending deep into the bacterial phylogenetic tree (0.7 to 3.6 bya). We successfully replaced *tufA* with the homologs in all cases, leaving the native homolog copy *tufB* in place. Analysis of fitness by means of exponential growth rate showed that, for some substitutions, the replacement of *tufA* was equivalent to a clean knock-out of the gene, suggesting that the inserted *tuf* might not be active. As shown before for *tufA*-deleted strains in *S. Typhimurium*⁵⁷, we found that some of the replacements led to selection of genomes with amplifications of the region containing *tufB*, further reinforcing the idea that inserted *tuf* genes might not be able to support viability by themselves. We then attempted the removal of the *tufB* copy in all the constructed strains, and found that only the genes from extant and ancestral homologs within the Gammaproteobacterial family were able to support growth as the only source of EF-Tu in the cell (*Fig. 9*). This shows a relationship between evolutionary distance and successful integration.

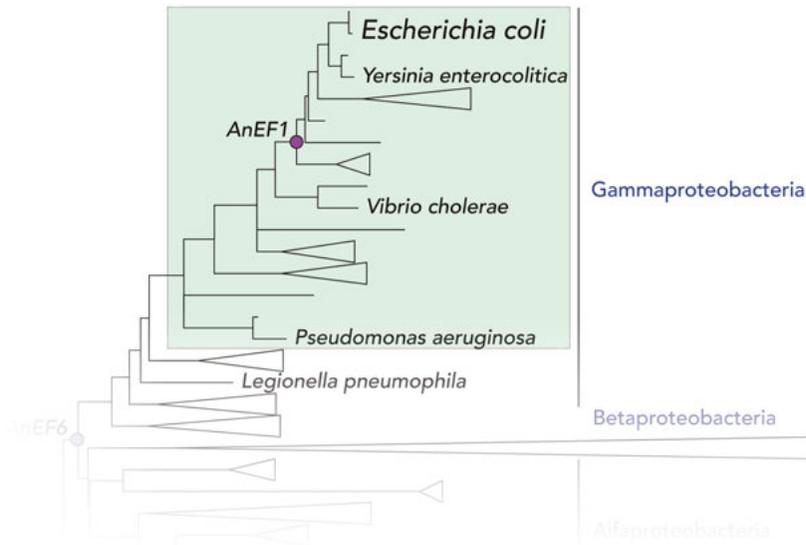


Figure 9. Interchangeability region for *EF-Tu*. The green area shows the homologs that yield a viable *E. coli* when their *EF-Tu* genes were transferred.

Although these four homologs were able to provide a viable organism, the effects of these replacements could be seen in both rich and poor media. We saw that the parameters of the growth cycle of the bacteria (lag-phase time, growth rate and maximum final density) were affected to a certain degree, showing an overall effect on global fitness. But, why do these replacements reduce fitness? One possibility is that some recently-acquired genes are not expressed enough (concentration problem), or it could be that they do not function properly (activity problem). For *tuf*, distinguishing between these two possibilities is not an easy task, given that the proper function of EF-Tu in the translation system is needed to produce a correct amount of itself. Nevertheless, we were able to find a correlation between the amount of EF-Tu relative to total protein and fitness, and between protein-synthesis step-time and fitness, showing that in this case both options seem to be playing a role. Additionally, we constructed strains carrying the same transferred gene as an additional copy under two constitutive promoters with different expression levels and showed that an increased expression of these homologs was associated with increased fitness. Taken together, these data suggest that a major reason for the fitness defects observed could be a sub-optimal expression of the gene's product, although we did not rule out the possibility of a contribution from defects in enzyme specific activity and/or interaction with other proteins, since overexpression increased fitness but not to wild-type levels for the most distant homologs. Furthermore, the results from the strain carrying the only viable ancestral gene showed that, even with a high protein

abundance, the reduction in fitness was large and unrelated to expression levels, pointing towards a problem in specific activity.

When looking at the amino acid sequences of all the homologs we worked with, including the native *E. coli* EF-Tu, we could see that higher sequence similarity (93-84%) was associated with viability, but that the differences between viable and non-viable EF-Tu sequences were small (84% versus 82%, respectively). Previous work with *S. Typhimurium*'s EF-Tu showed that individual single amino-acid substitutions in critical residues of the protein are enough to abolish its function^{38,146}, so here it is reasonable to hypothesize that for those inviable homologs close in sequence similarity to viable ones, one or more critical residues might be affected. EF-Tu's function is essential to the cell and to work properly it must interact with GTP, EF-Ts, tRNAs and the ribosome. Potentially, a change affecting any of these interactions might yield an inviable organism. In addition, EF-Tu undergoes major conformational changes throughout the elongation cycle, so variations in residues facilitating these changes might also affect EF-Tu's function. In accordance with this, we found changes in residues potentially affecting GTP binding/hydrolysis and interactions with EF-Ts and/or proteins L7/L12 of the ribosome for the inviable sequences. At least for the homologs coming from extant organisms, the EF-Tu must be able to interact with and hydrolyze GTP to support protein elongation in their native system, so it is more plausible than that the specific interactions with other cellular partners such as EF-Ts or the ribosome are what is limiting their functionality in the new environment. Co-evolution of EF-Tu with its partners might have fine-tuned these interactions, and this could create a barrier to the successful transfer beyond a certain point. This is in agreement with the complexity hypothesis, and members of complex and widely extended interaction networks could show a resistance to transfer due to their co-evolution and finely-tuned interactions.

In summary, we have demonstrated the limited transferability of the bacterial elongation factor EF-Tu between species and from reconstructed ancestral nodes. Analysis of amino-acid sequences showed that possible effects to the interactions with other members of the translation system might be limiting the successful integration. We also showed that when integration does happen and the genes are able to support growth, cellular fitness is affected mainly because of a suboptimal expression, but also partly because of a decreased activity.

Paper II: Location & orientation bias of a highly expressed operon

In bacterial genomes, the location and orientation of genes is often biased. Highly expressed genes involved in transcription and translation are often located near the origin of replication⁷². This bias is thought to be a consequence of selection for maximum growth, since the products of those genes are required in larger amounts during fast growth, and an increase in relative copy number near the origin can be achieved as a result of overlapping cycles of bacterial chromosome replication occur (gene dosage effect). Additionally, essential genes and highly expressed genes in operons are preferentially coded on the leading strand^{83,93}. This orientation bias is explained by the selection against head-on collisions between DNA and RNA polymerases, which are expected to be detrimental⁸⁰. With the increase in the amount of genome data available we know now that these biases are highly prevalent in the bacterial world, but no one has experimentally tested the significance of these biases for any essential and/or highly expressed gene whose product is directly responsible for fast growth.

To achieve this, we moved the highly expressed and essential *tufB* operon to different locations, in each orientation, within the bacterial chromosome of *S. Typhimurium*. This operon encodes the genes for four tRNAs, two of them essential, and one of the duplicate genes coding for the elongation factor EF-Tu (*tufB*). In *Salmonella* and related species, the two genes coding for EF-Tu (*tufA* and *tufB*) are encoded in the leading strand, relatively near to, on opposite sides of, and approximately equidistant from, the origin of replication. These organizational features make *tuf* a good model to study the effects of changing location and orientation on bacterial growth. To be sure that we studied only the effects of moving the operon without interference due to expression from nearby regions, we created a custom-made transcriptionally-insulated operon by including a terminator in the front of the operon. This operon was then inserted at 5 different locations in the chromosome in the native orientation and, for a subset of the locations, in the opposite orientation. As a last step, we deleted the remaining copy of *tuf* (*tufA*), thus constructing a set of strains relying on expression of *tuf* from non-canonical locations and/or orientations.

How did these modifications in location and orientation of an essential and highly expressed gene affect bacterial fitness? Measurements of exponential growth rate showed a significant effect on fitness as a function of location, with a reduction in fitness of 1.3% for every additional 100Kb from *oriC* (Fig. 10). This linear relationship between distance of the operon to the origin of replication and the magnitude of the associated fitness cost suggests that growth rate is being primarily limited by the concentration of EF-Tu. This fitness reduction (1.3%/100Kb) is in good agreement with the theoretical reduction in relative copy number of a given position in a fast-growing

bacterial population with an average of 4 *oriC* copies per every *ter* (3%/100Kb). In contrast, the inversion of the *tufB* operon, so that it was transcribed from the lagging strand, opposite to the direction of replication, did not further reduce fitness (within the standard deviation of the measurements).

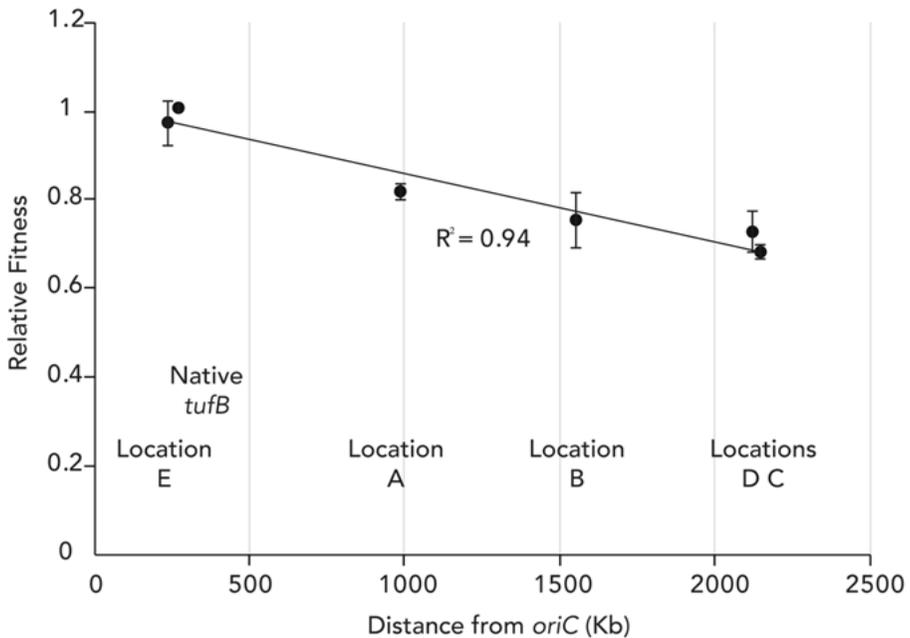


Figure 10. Relative fitness of strains as a function of *tufB* distance from *oriC*. Linear regression shows a correlation between fitness and distance.

Thus far, our hypothesis was that longer distances from the origin resulted in lower expression of EF-Tu and, because EF-Tu concentration and growth rate are directly related⁴⁰, we could see an effect on fitness. To explore this hypothesis further, we experimentally evolved a subset of the mutants to obtain insight into mechanisms by which sub-optimal operon locations and orientations could be compensated for. If the main reason for the fitness' effects is indeed the reduced expression of EF-Tu, we expected to find changes that would directly increase expression. Following an experimental evolution of 500 generations in rich media all evolved populations showed an increased fitness and clones from these populations presented changes that potentially increased the expression of EF-Tu from the *tufB* gene. We found that all clones had either acquired an amplification of a region containing the *tufB* operon, or had acquired one of several different point mutations (SNP) in the region immediately upstream of, or in the beginning of the *tufB* coding sequence. Amplifications are expected to increase the level of EF-Tu by providing more gene copies for transcription and translation. Four out of ten

of the selected SNPs have previously been shown to increase the expression of *tufB* up to 60% by disturbing the mechanisms of *tufB* autoregulation⁵⁷. There was no significant difference between the spectra of mutations selected in the two different orientations. Taken together, these results indicate that the major part of the fitness cost associated with the displacement of the *tufB* operon can be ameliorated by increasing the expression of EF-Tu from the *tufB* gene.

These experiments do not address how or why the *tuf* genes have come to be at their specific current locations and orientation, near the origin of replication. Our experimental evolution results suggest that increased expression to support increased growth rate can be readily achieved by local adaptations, regardless of the location of the *tufB* operon, and that the orientation of the operon does not seem to be critical for growth. However, these conclusions come with several significant limitations. Firstly, relative fitness after evolution improved, and approached wild-type fitness, but did not equal wild-type fitness. This might reflect the crudeness of the measurements, or the short duration of the evolution, but it definitely does not rule out the possibility that, on an evolutionary timescale, the original location might have a significant fitness advantage. Secondly, the evolution experiments were made with strains carrying only a single *tuf* gene, leaving open the possibilities that, either on physiological or evolutionary timescales, the specific location of both *tuf* genes might have a selective advantage. Thirdly, local changes in copy number (amplifications) may not be easy to fix in natural populations since they generally include surrounding regions of the genome with additional genes, for which an increase in copy number may have pleiotropic and/or detrimental effects. Additionally, the mechanisms of duplication/amplification creation, mediated by direct repeats, make the amplification unstable, and segregation it is expected to happen at a high frequency in the absence of selection¹⁴⁷. Finally, selecting SNPs in the regulatory region of *tufB* that constitutively *turn up* expression will affect the intrinsic regulatory mechanism and may be detrimental in situations where fine-tuning *tufB* expression is needed. Accordingly, it is plausible that the local changes selected in these evolution experiments are good “short-term” solutions in this particular case, but in the grander scheme of the evolution of bacterial species these may simply not be the good enough, and rearrangements or transfers which position this kind of genes near the origin of replication may be preferentially selected.

All of the limitations and arguments presented above apply to both gene location and orientation. However, an additional feature, namely polymerase collisions, is relevant to consider with respect to the lack of detrimental effects on growth rate associated with *tufB* operon inversion at each of the tested locations. Caveats in this case are that the previously observed negative affects of inversion concerned very highly expressed and relatively long *rrn* genes, and were only observed in the absence of a fully functioning recombination

repair pathway¹⁴⁸. In the study by De Septenville, deleting the *rrn* operon from an inverted region of over 100 kb was sufficient to abolish the detrimental phenotype, showing that it was specifically inverted *rrn* and not other inverted genes that were responsible. Accordingly, the lack of detrimental effect seen when the *tufB* operon was placed in inverse orientation could mean that detrimental collisions occur less frequently, possibly due to it being transcribed at a lower frequency than *rrn*, or due to the shorter length of its transcript relative to that of *rrn*, or even due to it being assayed in strains with a functional recombination repair system. It has also been shown that genes encoded in the lagging strand have a higher mutation rate⁹⁰ and on evolutionary timescales this could be an additional selective pressure against inverse orientation for conserved genes, such as *tuf* and *rrn*.

The two *tuf* genes are known to co-evolve by gene conversion through homologous recombination⁴⁴, a process by which the *tuf* sequences are homogenized and prevented from diverging. In this case, if one of the *tuf* genes acquire a null mutation, this change can be reverted by copying the sequence information for the other gene. Since any deleterious mutation in either of the *tuf* genes will result in a decrease in fitness, the bacteria with a repaired gene will outcompete the unrepaired ones. We can hypothesize that the current location and orientation of the *tuf* genes reflects a selection for a chromosomal configuration that allows recombinatorial repair between them at a higher frequency. This hypothesis is indirectly investigated in **Paper III**, where we developed a system to measure homologous recombination rates between duplicate genes at different locations on the chromosome. In that work, we also investigated if the orientation of the sequences involved has any effect on the rate of homologous recombination.

In conclusion, we have shown that there is an approximately linear correlation between the location of a highly essential expressed operon, coding for a gene whose product is required at high concentrations to achieve maximum growth, and fitness. This correlation is in agreement with the gene dosage hypothesis and implies that the selection for faster growth rates can potentially drive the location of genes.

Paper III: Location/orientation and homologous recombination

Homologous recombination (HR) is an important mechanism directly involved in the repair, organization and evolution of bacterial chromosomes. HR between similar sequences within a chromosome can result in co-evolution by gene homogenization⁴⁴, but also it can lead to chromosomal rearrangements such as duplications and inversions^{114,138}. In addition, HR is involved in the repair of chromosomal lesions¹⁴⁹, and horizontally-transferred

genes can integrate in new genomes by means of HR¹⁵⁰. It has been previously shown that structural domains that affect the physical proximity of different parts of the chromosome can potentially affect the rates of intrachromosomal recombination^{96,104}, but it is still unclear how the position and orientation of identical sequences in a chromosome affects the HR between them.

Here, we have constructed an experimental system based on two mutationally-inactive kanamycin-resistant genes (*kan*) that allows us to measure HR rates between any two desired regions in the chromosome of *S. Typhimurium*. Recombinational repair of the engineered stop-codons in these sequences was shown to be fully dependent on RecA activity, largely dependent on RecB activity (90%), and fully independent of RecF activity, which indicates that the mechanism generating an active *kan* gene involves homologous recombination to repair double-strand breaks in DNA.

We inserted the two cassettes at many different chromosomal positions and in each orientation, creating a total of 76 different pairwise combinations that allowed us to explore the effects of location and orientation on the rate of intrachromosomal HR (Fig. 11). We measured an average rate for all pairs tested of approximately 2.6 and 4.6×10^{-6} per cell per generation, depending on the relative genetic orientation, with the individual values ranging over at least an order of magnitude, from 0.3×10^{-6} up to 16.1×10^{-6} per cell per generation.

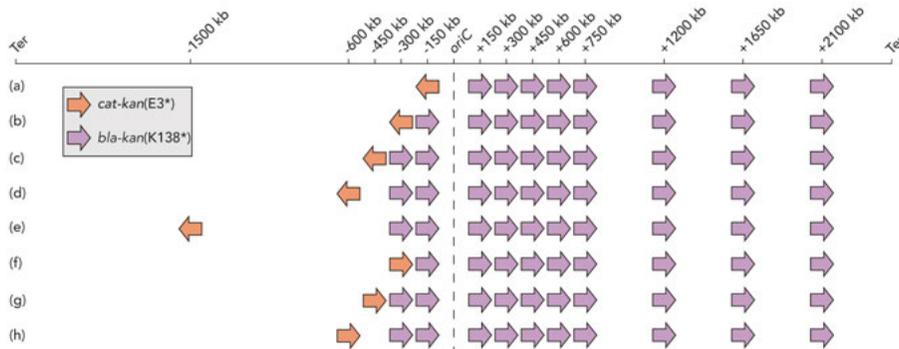


Figure 11. Experimental setup. Eight sets of strains (a total of 76 strains) were constructed to measure recombinational repair rates between *kan* cassettes placed at various locations around of the bacterial chromosome. The fixed position for each set of strains is shown in orange, with the variable positions shown in purple.

We can hypothesize several ways in which location might influence rates of HR between pairs of *kan* cassettes. One is that the location of one or both cassettes sequences relatively close to the origin of replication might increase HR rates. The hypothesis is that increased copy number for genes close to the origin, associated with overlapping rounds of replication, would increase the number of cassettes available for recombination and this increases the rate of HR. A second factor could be the location of the cassette sequences relative

to each other. We can imagine that closely linked sequences might be able to find each other more easily than more distantly located cassettes, and consequently, genetic distance might correlate with the rate of HR. Finally, as seen for *Bacillus subtilis*¹⁰¹, there is the possibility that cassette sequences located diametrically opposite each other, in different arms of the circular chromosome, could find each other more easily in space during the replication cycle and if that were the case it could correlate with the rate of HR. In fact, our data did not provide compelling support for any of the three hypotheses outlined above.

If none of the three hypotheses outlined above is supported by our results, how then can we explain the differences observed in HR rates? We re-analyzed the data asking whether each cassette location might individually influence the rate of HR, independently of the location of its partner sequence. For each cassette location, we calculated an ‘individual accessibility’ value and found that these could provide a good quantitative prediction of the measured rates of HR for pairs of locations (Fig 12A). This analysis also showed that some locations in the chromosome are up to 4-fold more accessible for HR than other locations. In particular, we found that three locations with the highest accessibility for recombination are three consecutive positions spanning 300 kb of the chromosome while the two locations with the lowest accessibility are two consecutive positions 150 kb apart of each other. These results are compatible with the idea that potentially large domains within the chromosome can be more or less accessible for recombination than other areas, but it also highlights that HR can effectively occur between any two given regions of the chromosome.

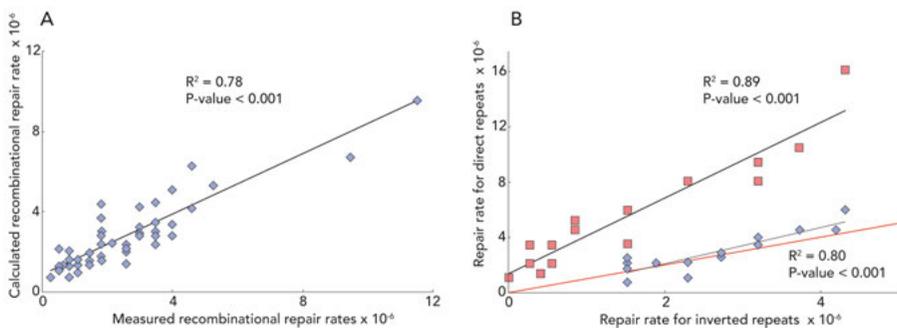


Figure 12. Measured HR rates as a function of A) calculated rates based on the determined accessibility values of the individual locations; and B) rates of the correspondent pairs as direct repeats. Blue diamonds indicate cassette pairs where gene orientation does not affect the HR rate and red squares indicate gene pairs where HR between direct repeats is higher. The red line indicates a one to one ratio for the HR rates of inverted versus direct repeats.

We then asked whether *kan* cassette orientation influenced HR rates. Our working hypothesis was that the relative orientation of two distantly separated *kan* cassettes would not have any influence on the mechanics of HR between them, but it could affect the rate of formation of chromosomal duplications and inversions involving the intervening chromosomal region. The experimental data showed that the average rate of HR measured for *kan* cassettes placed as direct repeats relative to each other is approximately 2-fold higher than the average rate measured for the same set of cassettes placed as inverted repeats. This difference suggests that relative orientation of cassette sequences can affect the recombinatorial repair rate between them. Furthermore, we could divide the pairs of *kan* cassettes into two classes, depending on whether or not their relative orientation affected the rate of HR (Fig. 12B). The first class shows no significant difference in recombinatorial repair rates between inverted and direct repeats at the same locations. This indicates that gene orientation does not affect the rate of recombinatorial repair of the *kan* genes at these locations. The second class shows gene pairs where there is an approximately 3-fold higher recombinatorial repair rate when the *kan* genes are in the direct repeat orientation compared to when they are in the inverted repeat orientation. This difference in rates of HR as a function of relative cassette orientation could be explained in two different ways. One possibility is that an increase in the frequency of tandem duplications, associated with HR between cassettes in direct orientation, could also be associated with an increased probability of forming an active KanR gene (the phenotype being screened for as a proxy of successful HR). A second possibility is that, despite the large distances between recombining cassettes, for mechanistic reasons HR is more efficient when it involves partners in direct orientation rather than in inverse orientation.

Looking into the two groups, we could see that an increase in HR for direct repeats was more likely to affect HR for pairs located in the same arm of the chromosome, than if they were in different chromosomal arms, possibly indicating a bias against the formation of larger duplications and/or duplications that include the origin of replication, and therefore supporting the hypothesis of duplication formation as the reason for the increased HR rate. We tested several KanR mutants from each group by both PCR and phenotypical analysis and we were not able to find any evidence of duplications. This rules out duplications as the primary source for the increased recombinatorial repair rates and indicates that, for certain combinations of chromosomal locations, insertions in direct repeat orientation can recombine more efficiently than in inverted repeat orientation, hinting at the possibility that the mechanics of recombination is sensitive to chromosomal orientation, even when the partners are separated by hundreds of kilobases.

In conclusion, we have found that that each region of the chromosome influences the rate of recombination independently of the location of the homologous copy, with some regions more accessible than others. Also, we have shown that homologous recombination can happen at a global scale in the chromosome. And finally, we demonstrated that the relative orientation of the duplicated copies involved in the recombination event can affect the recombination rates for some specific locations.

Paper IV: Unbalanced inversions

Chromosomal inversions are rearrangements in which a region of the chromosome has changed orientation. These rearrangements can arise as a result of homologous recombination between inverted repeat sequences¹¹⁵, and although under laboratory conditions inversions occur at least as frequently per genome/generation as nucleotide substitutions, they are rarely seen in natural bacterial populations, including in *S. Typhimurium* and in *E. coli*. This bias against the presence of chromosomal inversions suggests that there is a strong selection in many bacterial species to conserve the existing chromosomal organization.

The bacterial chromosome is often circular, with DNA replication starting at a single point (origin of replication, or *oriC*), and proceeding bidirectionally until the replication forks reach the terminus region (Ter), where the segregation of the newly synthesized sister chromosomes occurs. This situation creates two replichores, one on each side of *oriC*, that are approximately equal in length¹⁵¹. Genome analyses in several different species have shown that, when present, inversions are most often symmetrical around the origin of replication, thus preserving replichore length¹²¹. In addition, experimental work in *E. coli* has demonstrated that inversions creating a replichore imbalance of 15-20% can have a negative effect on growth¹⁵². These data suggest that replichore balance might be a factor constraining the fixation of inversions in bacteria, but we do not yet know if there is a systematic relationship between the degree of replichore imbalance and the magnitude of the fitness effects, or whether, and by which mechanisms, costly unbalanced inversions could be compensated for.

In this work, we used the experimental system presented in **Paper III** to create a series of inversions in the *S. Typhimurium* chromosome, each of them of different lengths and traversing *oriC*. Construction of inversions was done by: i) inserting two mutationally-inactivated *kan* genes at different positions in the chromosome, ii) selecting KanR colonies on plates containing kanamycin (some of these KanR mutants are expected to have the intervening region inverted); and iii) screening by PCR to detect the junctions associated with the predicted inversions (*Fig. 13*).

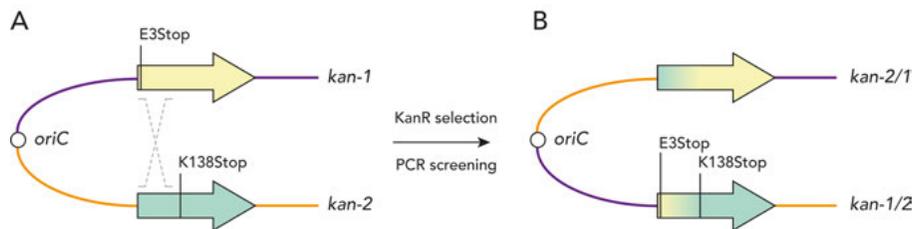


Figure 13. Schematic representation of inversion selection. A) Chromosome carrying two mutationally-inactivated *kan* genes (*kan-1* and *kan-2*). B) Homologous recombination between the *kan* genes resulting in one kanamycin-resistant gene (*kan-2/1*), one double-mutated gene (*kan-1/2*), and the inversion of the chromosomal region between them.

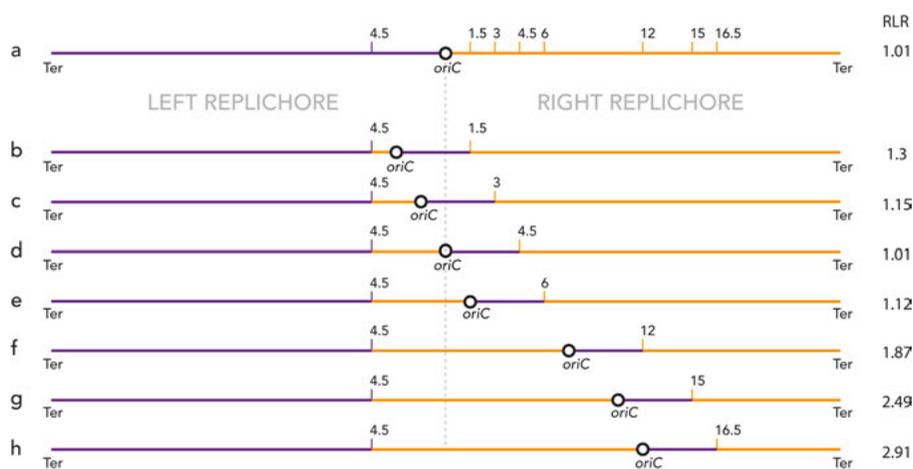


Figure 14. Constructed inversions. a) Chromosome structure of wild-type *Salmonella* indicating the location of insertions of the *kan* genes to generate inversions. The numbering refers to the kbp away from the *oriC* (4.5 is 450kbp). b-h) Chromosome structure of the strains carrying the seven constructed inversions. Replichore length ratio (RLR) represents how much longer one replicore is relative to the other.

The selected inversions resulted in replicore imbalances of different degrees, allowing us to study the fitness effects of these inversions in relation to imbalance, and subsequently, to study the evolution and mechanisms of compensation to reduce the associated costs (Fig 14).

The data showed that all the inversions that resulted in unbalanced replicore lengths resulted in a significant fitness cost (measured in head-to-head growth competition experiments), while the one inversion that did not alter replicore balance was competitively neutral. There was a good correlation between the magnitude of the fitness cost and the degree of

replichore imbalance, supporting the hypothesis that replichore imbalance is a major reason for the fitness cost associated with chromosomal inversions.

After the initial fitness assessment, we then were interested in understanding how bacteria carrying these unbalanced and costly inversions could evolve. If, as suggested by previous work and our own results, the main cost of these inversions comes from the chromosome imbalance we expected to see that changes to chromosome organization that restored the replichore balance would be selected. On the other hand, if the costs of inversion are actually associated with other problems, for example the effects of changing the location or orientation of a specific gene, the expectation was that the evolutionary answers might be more varied. Experimental evolution of several lineages of three strains carrying costly unbalanced inversions demonstrated that the fitness cost associated with the inversions was compensated in the majority of cases by rearrangements that restored the degree of replichore balance to wild-type or near-wild-type levels. Apart from re-inversions mediated by recombination between the original *kan* genes, a subset of cases presented re-inversions involving recombination between additional repetitive sequences present in the chromosome, such as the *rrn* operons and the *oadAB/dcoAB* operons.

In nature, there are at least two types of genetic event that could result in a strain with an unbalanced chromosome. One is the integration into the chromosome of a large fragment of DNA, acquired by HGT, that could significantly increase the size of one of the replichores. Another is equivalent to the experimental set-up tested here, where HR between inversely oriented repetitive sequences present in different replichores, and at different distances from *oriC*, results in the inversion of the intervening region, leading to different replichore sizes. Previous work showed that the frequent rearrangements seen in *Salmonella* Typhi, which were previously thought to reflect selection for replichore balance, are actually more likely a consequence of a weak selection associated with the host-specific lifestyle, rather than a selection for balance^{153,154}. Here we have shown that inversions resulting in a change of replichore balance in *Salmonella* Typhimurium have a detrimental effect, the magnitude of which is directly correlated with the degree of imbalance, and that compensatory evolution selects re-inversions that restore the balance between the two replichores.

In conclusion, our work suggests that replichore balance can act as major constraint to the maintenance of unbalanced inversions in bacterial populations.

Concluding Remarks

The work in the four studies forming this thesis touches on several concepts related directly or indirectly to the organization of the bacterial chromosome (Fig. 15), and provide insights that may explain the biases and conservations we see regarding some of these organizational traits.

First, we have shown that the amount of protein expressed, the functional conservation of protein activity and the network connectivity act to constrain the successful transfer of an essential, highly-conserved, bacterial gene. We observed a correlation between evolutionary distance, protein similarity and transferability, limiting successful and functional integration of genes to those from evolutionarily close species. This suggests that the co-evolution of cellular partners might play a major role as a constraint limiting HGT.

Second, we found that changing the location of an essential and highly expressed operon, whose function is required to support fast growth rates, affects bacterial growth in a systematic way. The native location of the operon is close to the origin of replication and we showed that re-locating it further away from the origin reduced bacterial growth rate. We suggest that the increased gene dosage resulting from location close to the origin of replication may drive the selection of the native operon location. Although many highly expressed and conserved genes are encoded in the leading strand, we found that changing the orientation of this operon did not impose measurable costs. We therefore argue that for this operon, location, more than orientation, is under strong selection to support fast growth.

Homologous recombination can cause rearrangements in chromosome organization potentially affecting physiology and evolution. Our third study showed that any region of the chromosome can potentially access and undergo HR with any other given region and that each region of the chromosome influences the rate of recombination independently of the location of the homologous copy. Our results suggest and agree with the presence of higher-order structural domains in the *Salmonella* chromosome that influence the rate of recombination. Additionally, we found that the relative orientation of the involved sequences can affect HR rates.

Finally, our study on the effects of large unbalanced inversions in the chromosome showed that there is a direct relationship between the degree of replicore imbalance and the associated fitness cost, and that maintenance of chromosome replicore balance can be a major factor acting against the fixation of chromosomal inversions.

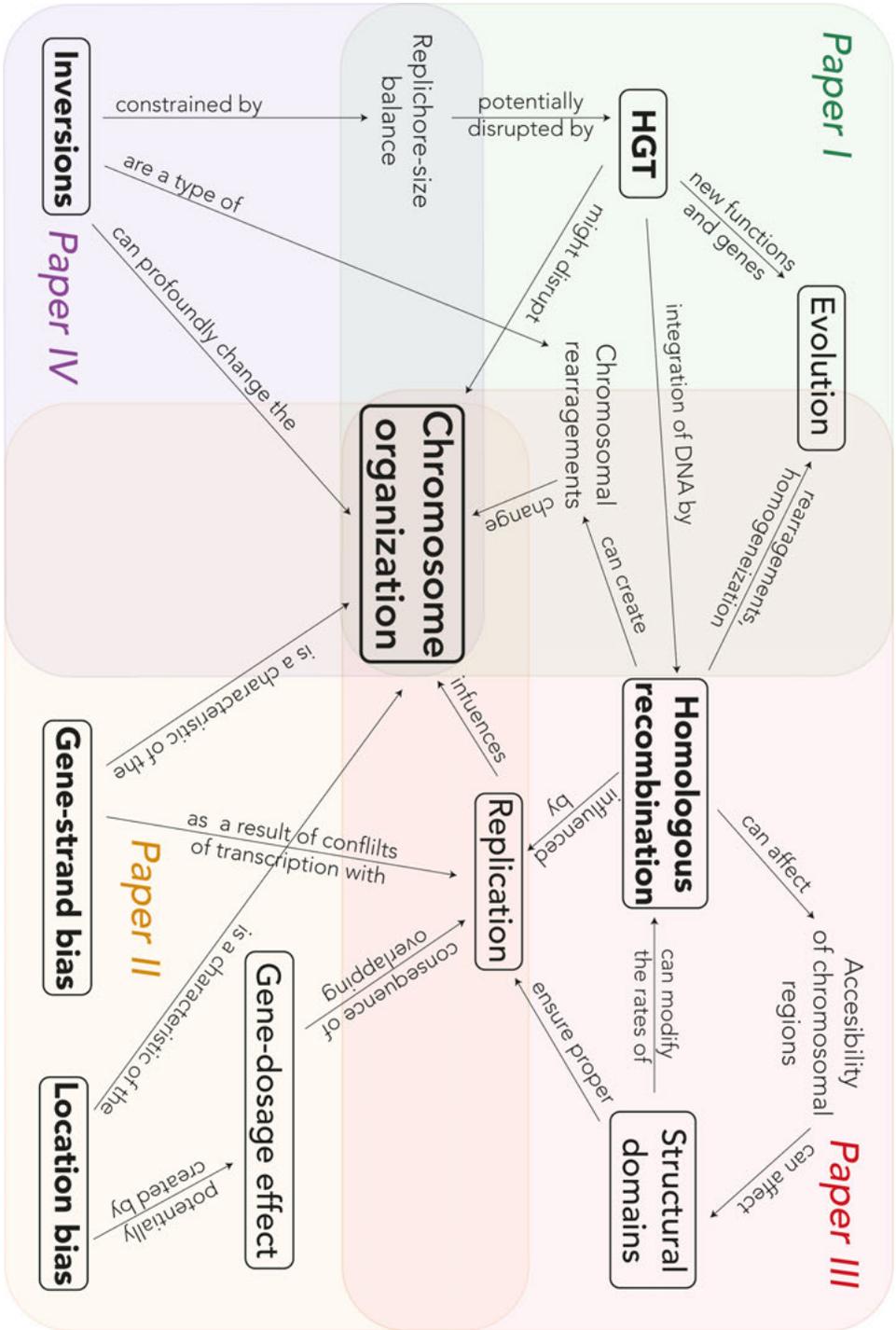


Figure 13. Conceptual map of the works presented in this thesis.

Future Perspectives

As expected, the work of this thesis can serve as starting point for further projects and extended studies that could address questions remaining. In particular, I am very interested in the following topics:

Given that the more distant replacements we made in **Paper I** imposed a cost to the cell, how can that cost be compensated? Our work suggests that both low amount and impaired activity of the protein are at fault, so experimental evolution of those strains might select both changes that increase expression and enhance activity. Either type of change would be of interest. First, to date we have not found a mechanism of upregulation for the operon involved in the replacements, *tufA*, and we don't know if increased expression of *tuf* from that operon is achievable. Second, although the sequence comparison between the viable and non-viable replacements points toward an effect on the interactions between EF-Tu and its cellular partners, we do not know which particular interaction(s) are important. Experimental evolution that selects changes improving these interactions could provide clues to the nature of the important interactions and partners. Also, if the cost is due to impairment on a specific activity like, for example, hydrolysis of GTP, we could expect to see changes affecting that activity.

In **Paper II** we demonstrated the effect of changing the position of EF-Tu in the chromosome but did not detect any effect associated with orientation, despite observational studies showing a strong bias in orientation. This would be of interest to test further, to ask if there is a selective advantage for EF-Tu to be encoded by a gene in the leading strand, and how that advantage could have effect the evolution of these sequences.

The results of **Paper III** show that the relative orientation of repetitive sequences apparently affects the rate of homologous recombination between some regions of the chromosome but not others. It would be interesting to study this further and understand the mechanisms underlying this phenomenon.

Lastly, in **Paper IV** we see that the greater the unbalance of the chromosome's replichores, the stronger the reduction in bacterial growth fitness. But why is that so? Further research could focus on understanding how the degree of imbalance imposes a cost to the cell.

Svensk sammanfattning

Bakterier är många, gamla och varierande; olika bakteriearter har evolverat under miljontals år och påvisar många olika livsstilar och typer av metabolism. Trots detta så har vissa egenskaper gällande hur bakterier organiserar sin kromosom visat sig vara oförändrade, vilket antyder att de är uråldriga och viktiga, och att selektivt tryck kan förhindra modifikationer av dessa. Denna avhandling ämnar studera några av dessa egenskaper experimentellt, för att utvärdera hur förändringar av dessa grunddrag påverkar bakteriens tillväxt, och hur bakterier skulle kunna evolvera efter dessa förändringar.

Först undersökte vi experimentellt vad det fanns för begränsningar för horisontell genöverföring av en essentiell gen för bakterietillväxt. Sedan undersökte vi betydelsen av läge och riktning hos ett starkt uttryckt och essentiellt operon; vi utförde experimentell evolution av stammar med operonet i suboptimala positioner och riktningar, för att utvärdera hur bakterier anpassade sig till dessa kostsamma förändringar. Vidare så ville vi förstå hur potentiellt tillgängliga olika regioner av kromosomen är till varandra för att kunna genomgå homolog rekombination. Till sist, så konstruerade vi bakteriestammar med kromosomala inversioner för att utvärdera vilka effekter dessa kunde ha på tillväxt, och hur bakterier med kostsamma inversioner skulle kunna evolvera för att minska dessa kostnader.

Resultaten gav bevis för att olika selektiva faktorer kan spela in för att bevara egenskaper hos den kromosomala organisationen. Vi kom fram till att evolutionär distans, funktionell likhet, suboptimalt uttryck och försämrade nätverkskontakt av en gen kan påverka överföringen av gener mellan bakteriearter. Vi drog slutsatsen att positionen för en essentiell och högt uttryckt gen är kritisk för snabb tillväxt, och att dess position verkar vara viktigare än dess riktning. Dessutom visade vi att både position och den relativa riktningen hos separerade, duplicerade sekvenser kan påverka rekombinationshastigheten mellan dessa sekvenser i olika regioner av kromosomen. Slutligen så pekar resultaten på vikten av att behålla de två replikationsarmarna hos en cirkulär bakteriell kromosom ungefär lika i storlek som en stark selektiv faktor, som motarbetar vissa typer av kromosomala inversioner.

Resumen en Español

Una multitud unificada: estudios experimentales de la organización del cromosoma bacteriano

El mundo bacteriano es grande, antiguo y muy variado. Las diferentes especies bacterianas, que llevan evolucionando millones de años, tienen distintos estilos de vida, así como metabolismos diferentes. Sin embargo, se ha observado que ciertas características de sus genomas no siguen tal variedad. Por ejemplo, sabemos que la mayoría de los genes altamente expresados se encuentran cerca del origen de replicación y están codificados en la misma dirección en la que progresa la replicación del genoma, también que la mayoría de los genomas bacterianos son circulares y tienen brazos de replicación de longitud similar, que la transferencia horizontal de genes no parece ser uniforme y finalmente que la estructuración física del cromosoma en la célula puede potencialmente interferir en la accesibilidad de unas regiones a otras. Todo esto nos hace plantearnos algunas preguntas que esta tesis pretende responder como son: ¿Por qué son estas características importantes? ¿Podemos estudiar experimentalmente su significado? ¿Qué pasa si cambiamos esas características en el laboratorio? ¿Pueden las bacterias evolucionar y sobrepasar los efectos de esos cambios?

Para ello hemos estudiado en el laboratorio qué ocurre cuando modificamos las características mencionadas analizando sus efectos.

Adicionalmente, en dos de los cuatro proyectos hemos realizado experimentos de evolución en el laboratorio para entender cómo las poblaciones bacterianas pueden resolver los problemas que surgen cuando se realizan cambios en el genoma.

En particular, los distintos objetivos de esta tesis se pueden definir en los siguientes puntos:

1. Si la transferencia horizontal de genes no es uniforme, y tiende a ser más exitosa para cierto tipo de genes, ¿se pueden identificar las razones principales que afectan la transferencia satisfactoria de un gen cuyo producto está conectado con muchos componentes celulares y es esencial para la supervivencia? ¿Existe algún tipo de relación entre la distancia evolutiva de los genes transferidos y el reemplazado y el efecto de cada transferencia a nivel celular?

2. Dado que hay una preferencia en la localización y la orientación de genes altamente expresados y esenciales, queremos saber cómo afectan a la célula los cambios en estas características y nos preguntamos cómo pueden las poblaciones bacterianas evolucionar y adaptarse a esos cambios.
3. La recombinación homóloga entre distintas partes del genoma tiene el poder de cambiar extensivamente la organización genómica y dinámica cromosómica. Debido a esto nos preguntamos cómo afecta a la frecuencia de recombinación la localización y orientación de dos secuencias duplicadas en el genoma.
4. El cromosoma bacteriano presenta típicamente un eje de 180° entre el origen de replicación y la zona de terminación. La hipótesis actual es que hay una selección en contra de reorganizaciones genómicas, como las inversiones, que perturban ese equilibrio. ¿Podemos predecir el nivel de los efectos según el grado de perturbación que las inversiones desequilibradas imponen? ¿Cómo evolucionan los cromosomas desequilibrados para restablecer el crecimiento?

En primer lugar, hemos demostrado que la cantidad de proteína expresada, la conservación funcional de la proteína y la red de conectividad en la célula, actúan en conjunto limitando la transferencia con éxito de un gen cuyo producto es esencial para el crecimiento bacteriano. Además, hemos constatado que, para determinar el éxito de la transferencia, existe una correlación entre la distancia evolutiva de las secuencias transferidas y la secuencia reemplazada. Esto hace que sólo se pueda llevar a cabo una transferencia con éxito entre especies más cercanas evolutivamente. En conjunto, estos resultados sugieren que la co-evolución de las interacciones celulares puede tener un papel importante en la limitación de la transferencia horizontal.

En el segundo estudio, hemos mostrado que el cambio de localización de una región genómica esencial y altamente expresada, afecta al crecimiento bacteriano de forma sistemática. Esta región con la que hemos trabajado se encuentra en su forma natural cerca del origen de replicación, pero cuando la movemos a posiciones más alejadas, se reduce el crecimiento bacteriano. Cuando la célula bacteriana crece rápido, las zonas cercanas al origen se encuentran temporalmente en un estado amplificado permitiendo así máxima expresión. Los resultados obtenidos, sugieren que esta situación puede impulsar la selección de la localización de ciertos genes cerca del origen de replicación. Por otro lado, los experimentos realizados no han mostrado ningún efecto en el mismo sentido, cuando lo que se produce es un cambio de orientación de esa misma región, con lo que podemos concluir que es la posición, más que la orientación, la seleccionada para obtener un crecimiento rápido.

En relación a la recombinación homóloga, nuestro tercer estudio demuestra que cualquier región del genoma de *Salmonella* Typhimurium puede potencialmente acceder a otra región del cromosoma y llevar a cabo dicha recombinación. La frecuencia de esta recombinación entre diferentes regiones del genoma solo cambia, como mucho, en un orden de magnitud dependiendo de las regiones involucradas. Los resultados sugieren y coinciden con la presencia de estructuras cromosómicas que influyen en la accesibilidad del genoma.

Finalmente, el cuarto trabajo indica que existe una relación directa entre el grado de perturbación en el equilibrio de los brazos de replicación y el efecto del crecimiento bacteriano. Esto sugiere que posiblemente el mantenimiento de ese equilibrio sea un factor principal en contra de la fijación de inversiones desequilibradas en el cromosoma.

Los resultados de estos cuatro estudios nos proporcionan nuevos conocimientos que nos ayudan a entender el porqué de las diferentes preferencias y rasgos conservados en la estructura cromosómica bacteriana.

Words of gratitude

"Listen to me, Morty. I know that new situations can be intimidating. You lookin' around and it's all scary and different, but y'know... meeting them head-on, charging into 'em like a bull—that's how we grow as people."

Rick Sanchez

There are many to whom I am forever indebted for being part of the road that has led me to today. It feels like it was yesterday when I was a little girl (not that I've grown that much since...) that dreamed of being a scientist and working towards answering the "hows" and "whys" of the world we live in. Well, I guess today I can, *with a reasonable degree of certainty*, say that the dream came true, regardless of where my life path continues hereafter.

Let me start by saying that if you feel you are one of the persons who was part of this journey, you most surely were; and for that I thank you greatly. Nonetheless, I take what probably is going to be my first and last doctoral thesis (although who knows, the world has seen crazier people...) to get in indelible print some particular words of gratitude.

First of all, I want to thank my main supervisor **Diarmaid**; I've always been so impressed with your generosity and altruistic nature, both professionally and personally. Scientifically, you have been the best *boss* I could have asked for: you are knowledgeable and always there; you are creative yet cautious with letting us express our own interests; and you deeply care for the people around you. In particular, I must mention how thankful I am that you trusted me to be your right hand for two consecutive summers as teaching assistant at the *Advanced Bacterial Genetics* course in CSHL. The intense weeks we spent there are some of the best memories I have from my time as a Ph.D. candidate, and will remain forever in my heart. Likewise, I really appreciate that you understood my love and interest in science communication and outreach, never complaining about the numerous hours that I have devoted to it throughout my time with you. I could really continue for a longer stretch, for I admire you so, but I will simply leave it here: *Go raibh maith agat*.

Also, thanks to my current co-supervisor **Linus**! You have always been there for any questions I have had, and you have saved me quite a number of times from endless hours on a stuck train in the middle of dark (*and full of terrors*) nights.

A special mention is needed to thank my collaborators in the studies that form this thesis, **Betül Kaçar**, **Nurcan Tuncbag**, **Dan Andersson**, **Gerrit Brandis**, **Douglas Huseby**, and **Conchi Bueno**; you have helped me to get better at something that is at the core of science: unite minds and efforts in scientific endeavors and create works where we all put the best *of* each other, at the same time that we learn *from* each other. Also thanks to the various people that have proofread parts of this thesis, and have helped me feel less crazy and more secure about my writing: **Ana** (a.k.a. *mum*), **Dani**, **Doug**, **Gerrit**, **Hugo**, **Diarmaid** and **Linus**. And thanks to **Linnéa** for the translation of the Swedish summary (yeah, told you, my Swedish is not *that* good!). Additionally, I want to personally thank you, **Klas**, **Lionel** & **Sanna** for your enthusiasm when you were asked to be a special part of this thesis; I am sorry that things had to take a sudden turn.

Thanks to past and present members of my own group, for forming this wonderful family we have. **Jessica**, you were my first contact with practical bacterial genetics so I owe you so much! Thanks for putting in a good word for me when I applied for the master project back in 2010, and thank you for all the help you have provided over the years. I've missed you a lot since you've left, specially those characteristic *lab notes* of yours, and the jokes we used to crack day in day out. I wish you the best on your professional life as well as in you newly minted married life, and other personal ventures. **Franzi**, I am so happy you got that position you wanted. I also miss you a lot, although I stole your desk after your departure, *that's life*, as they say. Thank you for all the moments shared. **Lisa**, I wish you the best in this new leg of your life after the Ph.D. **Gerrit**, second mention, thanks for being an awesome office/lab colleague and for your immense willingness to help anyone around you, in spite of the hard times you have been through. I've learnt a great deal from you and, as you said in your own acknowledgements, this time around our scientific baby is here! **Doug**, we are, DH-lab-wise, the same age and I feel like we have grown together, so to speak. Can you be the brother I never had?! I can thank you for a lot of things, but specially I should mention the out-of-office times watching GoT at your place with **Bethany**, and the interesting conversations we've held over countless lunch times. Ah! And remember I will always be your *analog* friend. **Linnéa**, I am so happy you did your M.Sc. thesis with us and that Diarmaid didn't let you go. You are one of the most sweet and heartfelt persons I know. Thank you for being my go-to roomie at conferences and meetings, and for not getting annoyed with me when I can't stop talking early in the morning (and late at night) on our train rides. **Cao Sha**, I love your sense of humor and your hard-work. It has been impressive to see how much better you've become at speaking English, and to see grow your beautiful family. Other people have joined the group recently but I wish you, of course, the all the best: **Onur**, **Katrin**, **Anna**. **Rachel**, I am so happy that, after sharing lab during our M.Sc. thesis, you are back! Others came for a somewhat shorter time and left, but nevertheless they left a mark:

Rahel, we spent many really nice times in and out of the lab, thank you for all your wise words. **Kenessa**, thank you for your warmth and closeness, and for sharing your own struggles and being so open to talk; and of course, thank you for choosing me to be your M.Sc. thesis opponent, a role I had never had before. **Conchi**, what an adventure we had, right? You were the first and only student I've supervised. I am not sure yet how well I took on the job, but I hope our time together has helped you learn and grow. I think we made a pretty good damn team! Can't wait to see you again at my defense. Related to the DH family as well, thanks to **Disa & Marie** for the fun times at Diarmaid's house; and **Marie**, our HM god, thanks for being so kind to me from the very beginning, I still own the mugs you gave me when I moved to Sweden with literally nothing.

I've had the luck to work in a very social corridor, where one really needs to put quite some amount of effort to choose labwork vs. Swedish fika times (which in many occasions turn into scientific discussion as such, so no time wasted, don't you think?). Thank you, **DA, JN, LG, & LS families' members**, present and past! If any of you is particularly interested in knowing what I am grateful to you for, please come talk to me (*don't be shy!*); I will invite you to a coffee and tell you all the reasons why you are special to me, because you are!

The department I've worked at is big and very varied in scientific nature. In it, I had the opportunity to help out other Ph.D. students by being chair of the association and successfully coordinating some good achievements next to **Julia, Marlen, Andreas, Jessica, Erik, Franzi, Emma, Ravi, Iris &** others. Thanks to the new IPhA members for wanting to continue the tasks and activities. Special thanks to the administrative staff who have always been so thoughtful; and to the IMBIM board for letting me be the Ph.D. students' representative at their meetings, which were fully in Swedish, when my Swedish skills were quite poor; I've gotten much better thanks to them!

Also, I'd like to take the opportunity to thank the people that have made the Association of Spanish Scientists in Sweden (**ACES**) possible. I spent a lot of my non-lab time in this grand project of us and I am so happy we took it to where it is today. Meeting all of you has been a pleasure. There is a lot more to be done and hard work ahead, but I am sure it is going to continue being a successful association. In addition, I also devoted part of my time outside the lab (an even at some point inside!) to photography. The time being part of, and running, the photo group at Kalmar Nation brought me a lot of joy. In particular, thank you **Ana Rosa, Carlos, Mahsa & Phil**, for the times we shared there. Keep spreading the love for photography, please! Finally, in this section about my *volunteer* work, I want to thank **Erik, Anna & Co.**, for the times "preaching" science all around SciFest, TomTits and Nobel Museum.

One person I want to specially mention is **Marlen**! You instantaneously became my favorite colleague, neighbor and friend within the first week!

Although we are no longer neighbors nor colleagues, we shall forever remain friends, and I miss you every day we are apart (seriously, we might not talk that often but I think of you all the time). Countless are the instances I can thank you for, and you know them! I am super excited about you coming back to Sweden after your Seattle adventure, I am waiting for you.

Other two special persons have supported me in spite of the long distances, and have remained my *besties* after sharing five long years of university courses. **Lau & Ali**, I still remember how we were called out for speaking too much during a botany class on our first year (*señoritas Derrac, García and Garmendia, una palabra al acabar la clase, por favor*); I think this anecdote tells a lot about our relationship. I love how we balance each other and how much we've always cared for one another. I will always be there for the both of you, now and forever. Lau, good luck on your life as a postdoc, and Ali, you will do just great in your Ph.D., do not ever doubt yourself! And to our other half of the Top6, **Chris, Antonio, Mario**, can we once and for all meet at a single point in time and space? I miss you a lot!

From the beginning of my academic life journey, I shall mention some people whose actions and work has made me who I am today. My primary-school teacher **Ana**, for stating that I will never be good at English (or any other language for that matter); I am sure she didn't mean any harm, but I grew up to prove her wrong, and look at me today - I can now have a decent conversation in English, and almost so in Swedish as well! To my high-school biology teacher **Ana Mellado**, you were such an inspiration and you saw the potential in me and steered me towards the right choice: biology! I have never doubted for a second ditching medical school and becoming a biologist; thank you for your hard work. **Juan Pedro & Pepi**, you were my first scientific family and I can't thank you enough for seeing something in me that could be of use in your laboratory. I miss the grasshoppers and I miss you. Likewise, from my time at the evolutionary genetics lab in Granada I want to say thanks to **Paquillo, Maria, Tati, Eli, Bea**, you will always be in my hearts; I learned so much next to you! And to **Moha** and **Jesús**, we shared many (maybe even *too many*) hours late in the days and over weekends sharing office and meal/coffee times; your experiences and teachings have also contributed to who I am today. From my time at the HCV lab in Granada I want to thank **Alfredo**, for the heart-felt recommendation letter you wrote for my Ph.D. position; and **Cristina**, for teaching me an insane amount of techniques in two months. To my scholarship-holder mate in that experience, **Fran**, I wish you the best in and out of the lab!

As I mentioned above, I spent the craziest but some of the best weeks of my Ph.D. at Cold Spring Harbor Laboratory being the teaching assistant for the *E. coli* & *Salmonella* part at the *ABG* course. The nirvana-like experience that it was is quite difficult to describe, but I'll say that I doubt I will ever feel as connected to science as I did back there. Special mention to my assistant fellows, **David** (Daviiiiiii), **Sara** and **Nick**, I am amazed at what we pulled

off! **Fitnat & Beth**, thank you for an awesome time at the Delbruck building. There, I worked hard, late and very close to the students, and some of them really left a mark: **Tiffany**, our sincere conversations that summer have stayed with me ever since, you are gold; **Kathryn** you were the baby of that year and I am so glad we got to see each other when you moved to Copenhagen; **Vijay**, our string theory physicist turned into lab-rat, very impressive, very impressive! **Eli**, your laid-back personality and dark humor really lightened your year's course; thanks for the day-trip to NYC and taking us to that ramen place (never had any better until I went to Japan!).

Other friends and acquaintances have kept me sane with out-of-work times. Special mention to **Fred**, my oldest Uppsala friend and someone I met on my first night out trying to dance salsa during my ERASMUS year. You were back then in the *middle* of your own PhD and had always a lot to share to help me. You also introduced me to the wonderful dance of kizomba, and took me *that* day to Stockholm; this is something I will be always in debt for! We talked about it many times, life is not easy, but knowing ourselves first can allow us to help others the best. I also shared a lot of fun times with the international gang formed by **Martoffer, Emill, Jeino** and **Jenni**; and with a Russian speaking gang: **Kate, Asker, Tanja, Victor**, thank you for being so open and adopting me much as if I would understand what *cyka* means; and for being the best models for my photography work! **Paul**, you are kinda part of the Russian speaking gang, but I specially thank you for your generosity, and your (ehmm) *horrible* dancing skills, haha! To others that also shared many dance times with me: **Paola, Junior, Olga, Guilhe, Kelvin, John**, thank you!

I will be wrapping up this now, since I consider you all had enough (because I know that EVERYONE who touches this thesis is just only going to read this chapter). So, final words to my family, who deeply know me and accept me how I am; who wish the best for me and help me be better day after day; whom I so profoundly love! En especial, a mis padres **Ana & Samuel**; gracias por todo el apoyo y la ayuda durante mis años académicos, os debo la curiosidad que me ha llevado hasta aquí y la perseverancia de querer hacer un buen trabajo; y a mis abuelos **Domingo & Mariluz**, por todo el amor incondicional que me habéis dado. And **Morito**, for being the best dog in the whole world! Last but not less important, my deepest words of gratitude to my *bebito*; I am SO glad our paths crossed in that park that late summer evening and that the dancefloor later brought us together. This is meant to be, and we will make of it the best story of our lives. Thank you for all the encouragement and help, specially over the last tough period; and for your logical calm mind when mine is overflowed with emotions. The world is waiting for us, I love you so!

Eva Garmendia
October 2017

References

1. Friedmann, H. C. Escherich and *Escherichia*. *Adv. Appl. Microbiol.* **60**, 133–96 (2006).
2. Blattner, F. R. *et al.* The Complete Genome Sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1462 (1997).
3. Lilleengen, K. Typing of *Salmonella dublin* and *Salmonella enteritidis* by means of bacteriophage. *Acta Pathol. Microbiol. Scand.* **27**, 625–640 (1948).
4. Swords, W. E. & Cannon, B. M. Avirulence of LT2 Strains of *Salmonella typhimurium* Results from a Defective *rpoS* Gene. *Infect. Immun.* **65**, 2451–2453 (1997).
5. McClelland, M. *et al.* Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature* **413**, 852–6 (2001).
6. Barker, J. S. F. *Adaptation and Fitness in Animal Populations*. (Springer, 2009).
7. Darwin, C. *On the origins of species by means of natural selection*. (1859).
8. Orr, H. A. Fitness and its role in evolutionary genetics. *Nat. Rev. Genet.* **10**, 531–539 (2009).
9. Atwood, K. C., Schneider, L. K. & Ryan, F. J. Periodic selection in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **37**, 146–155 (1951).
10. Ryan, F. J. Evolution observed. *Scientific American* 78–83 (1953).
11. Dykhuizen, D. Experimental Studies Of Natural Selection In Bacteria. *Annu. Rev. Ecol. Syst.* **21**, 373–398 (1990).
12. Chao, L., Levin, B. R. & Stewart, F. M. A Complex Community in a Simple Habitat: An Experimental Study with Bacteria and Phage. *Ecology* **58**, 369–378 (1977).
13. Lenski, R. E., Rose, M. R., Simpson, S. C. & Tadler, S. C. Long-Term Experimental Evolution in *Escherichia coli*. I. Adaptation and Divergence During. *Am. Nat.* **138**, 1315–1341 (1991).
14. Elena, S. F. & Lenski, R. E. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat. Rev. Genet.* **4**, 457–69 (2003).
15. Wichman, H. A., Scott, L. A., Yarber, C. D. & Bull, J. J. Experimental evolution recapitulates natural evolution. *Philos. Trans. R. Soc. London* **355**, 1677–1684 (2000).

16. Griffith, F. The Significance of Pneumococcal Types. *J. Hyg. (Lond)*. **27**, 113–159 (1928).
17. Tatum, E. L. & Lederberg, J. Gene recombination in the bacterium *Escherichia coli*. *J. Bacteriol.* **53**, 673–684 (1947).
18. Boto, L. Horizontal gene transfer in evolution: facts and challenges. *Proc. R. Soc. B Biol. Sci.* **277**, 819–827 (2010).
19. Thomas, C. M. & Nielsen, K. M. Mechanism of, and barriers to, horizontal gene transfer between bacteria. *Nat. Rev. Microbiol.* **3**, 711–721 (2005).
20. Christie, P. J., Whitaker, N. & González-Rivera, C. Mechanism and structure of the bacterial type IV secretion systems. *Biochim. Biophys. Acta* **1843**, 1578–91 (2014).
21. Popowska, M. & Krawczyk-Balska, A. Broad-host-range IncP-1 plasmids and their resistance potential. *Front. Microbiol.* **4**, 44 (2013).
22. Soucy, S. M., Huang, J. & Gogarten, J. P. Horizontal gene transfer: building the web of life. *Nat. Rev. Genet.* **16**, 472–482 (2015).
23. Townsend, J. P., Nielsen, K. M., Fishert, D. S. & Hartl, D. L. Horizontal acquisition of divergent chromosomal DNA in bacteria: Effects of mutator phenotypes. *Genetics* **164**, 13–21 (2003).
24. Ehrlich, S. D. *et al.* Mechanisms of illegitimate recombination. *Gene* **135**, 161–166 (1993).
25. Gogarten, J. P. & Townsend, J. P. Horizontal gene transfer, genome innovation and evolution. *Nat. Rev. Microbiol.* **3**, 679–687 (2005).
26. Knoppel, A., Lind, P. A., Lustig, U., Nasvall, J. & Andersson, D. I. Minor fitness costs in an experimental model of horizontal gene transfer in bacteria. *Mol. Biol. Evol.* **31**, 1220–1227 (2014).
27. Baltrus, D. A. Exploring the costs of horizontal gene transfer. *Trends Ecol. Evol.* **28**, 489–95 (2013).
28. Jain, R., Rivera, M. C. & Lake, J. A. Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 3801–3806 (1999).
29. Lamarck, J. B. P. A. de M. de. *Philosophie zoologique*. (Nouvelle édition. Paris : J. B. Baillière; Londres : Même Maison; Bruxelles : Au Dépôt de la Librairie Médical Française, 1830., 1830).
30. Hilario, E. & Gogarten, J. P. Horizontal transfer of ATPase genes — the tree of life becomes a net of life. *Biosystems* **31**, 111–119 (1993).
31. Golding, G. B. & Gupta, R. S. Protein-based phylogenies support a chimeric origin for the eukaryotic genome. *Mol. Biol. Evol.* **12**, 1–6 (1995).
32. Langille, M. G. I., Hsiao, W. W. L. & Brinkman, F. S. L. Detecting genomic islands using bioinformatics approaches. *Nat. Rev. Microbiol.* **8**, 373–382 (2010).
33. Andam, C. P. & Gogarten, J. P. Biased gene transfer in microbial evolution. *Nat. Rev. Microbiol.* **9**, 543–555 (2011).

34. Polz, M. F., Alm, E. J. & Hanage, W. P. Horizontal gene transfer and the evolution of bacterial and archaeal population structure. *Trends Genet.* **29**, 170–175 (2013).
35. Brown, J. R. Ancient horizontal gene transfer. *Nat. Rev. Genet.* **4**, 121–132 (2003).
36. Lucas-Lenard, J. & Lipmann, F. Protein Biosynthesis. *Annu. Rev. Biochem.* **40**, 409–448 (1971).
37. Kjeldgaard, M. & Nyborg, J. Refined structure of elongation factor EF-Tu from *Escherichia coli*. *J. Mol. Biol.* **223**, 721–742 (1992).
38. Tubulekas, I. & Hughes, D. A single amino acid substitution in elongation factor Tu disrupts interaction between the ternary complex and the ribosome. *J. Bacteriol.* **175**, 240–250 (1993).
39. Pedersen, S., Bloch, P. L., Reeh, S., Neidhardt, F. C. & Farrell, O. Patterns of Protein Synthesis in *E. coli*: a Catalog of the Amount of 140 Individual Proteins at Different Growth Rates. *Cell* **14**, 179–190 (1978).
40. Tubulekas, I. & Hughes, D. Growth and translation elongation rate are sensitive to the concentration of EF-Tu. *Mol. Microbiol.* **8**, 761–770 (1993).
41. Jaskunas, S. R., Lindahl, L. & Nomura, M. Identification of two copies for the elongation factor EF-Tu in *E. coli*. *Nature* **257**, 458–462 (1975).
42. Hughes, D. The isolation and mapping of EF-Tu mutations in *Salmonella typhimurium*. *Mol. Gen. Genet.* **202**, 108–111 (1986).
43. An, G. & Friesen, J. D. The nucleotide sequence of *tufB* and four nearby tRNA structural genes of *Escherichia coli*. *Gene* **12**, 33–39 (1980).
44. Abdulkarim, F. & Hughes, D. Homologous Recombination between the *tuf* Genes of *Salmonella typhimurium*. *J. Mol. Biol.* **260**, 506–522 (1996).
45. Arwidsson, O. & Hughes, D. Evidence against reciprocal recombination as the basis for *tuf* gene conversion in *Salmonella enterica* serovar Typhimurium. *J. Mol. Biol.* **338**, 463–7 (2004).
46. Paulsson, J., El Karoui, M., Lindell, M. & Hughes, D. The processive kinetics of gene conversion in bacteria. *Mol. Microbiol.* **104**, 752–760 (2017).
47. Ke, D. *et al.* Evidence for horizontal gene transfer in evolution of elongation factor Tu in enterococci. *J. Bacteriol.* **182**, 6913–20 (2000).
48. Lathe, W. C. & Bork, P. Evolution of *tuf* genes: ancient duplication, differential loss and gene conversion. *FEBS Lett.* **502**, 113–6 (2001).
49. Kondrashov, F. A., Gurbich, T. A. & Vlasov, P. K. Selection for functional uniformity of *tuf* duplicates in gamma proteobacteria. *Trends Genet.* **23**, 215–218 (2007).
50. Hughes, D. Both genes for EF-Tu in *Salmonella typhimurium* are individually dispensable for growth. *J. Mol. Biol.* **215**, 41–51 (1990).
51. Zuurmond, A. M., Rundlöf, A. K. & Kraal, B. Either of the chromosomal *tuf* genes of *E. coli* K-12 can be deleted without loss of cell viability. *Mol. Gen. Genet.* **260**, 603–7 (1999).

52. Van der Meide, P. H., Vijgenboom, E., Dicke, M. & Bosch, L. Regulation of the expression of *tufA* and *tufB*, the two genes coding for the elongation factor EF-Tu in *Escherichia coli*. *FEBS Lett.* **139**, 325–330 (1982).
53. Van der Meide, P. H., Vijgenboom, E., Talens, A. & Bosch, L. The role of EF-Tu in the expression of *tufA* and *tufB* genes. *Eur. J. Biochem.* **130**, 397–407 (1983).
54. Van Delft, J. H. *et al.* Control of the tRNA-*tufB* operon in *Escherichia coli*. 1. rRNA gene dosage effects and growth-rate-dependent regulation. *Eur. J. Biochem.* **175**, 355–362 (1988).
55. Van Delft, J. H., Talens, A., De Jong, P. J., Schmidt, D. S. & Bosch, L. Control of the tRNA-*tufB* operon in *Escherichia coli*. 2. Mechanisms of the feedback inhibition of *tufB* expression studied in vivo and in vitro. *Eur. J. Biochem.* **175**, 363–374 (1988).
56. Van Delft, J. H. & Bosch, L. Control of the tRNA-*tufB* operon in *Escherichia coli*. 3. Feedback inhibition of *tufB* expression by an EF-Tu with a deletion in the guanine-nucleotide-binding domain. *Eur. J. Biochem.* **175**, 375–378 (1988).
57. Brandis, G., Bergman, J. M. & Hughes, D. Autoregulation of the *tufB* operon in *Salmonella*. *Mol. Microbiol.* **100**, 1004–1016 (2016).
58. Rocha, E. P. C. The organization of the bacterial genome. *Annu. Rev. Genet.* **42**, 211–33 (2008).
59. Touchon, M. & Rocha, E. P. C. Coevolution of the organization and structure of prokaryotic genomes. *Cold Spring Harb. Perspect. Biol.* **8**, 1–18 (2016).
60. Lathe, W. C., Snel, B. & Bork, P. Gene context conservation of a higher order than operons. *Trends Biochem. Sci.* **25**, 474–479 (2000).
61. Demerec, M. & Hartman, P. E. Complex Loci in Microorganisms. *Annu. Rev. Microbiol.* **13**, 377–406 (1959).
62. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356 (1961).
63. Jacob, F., Perrin, D., Sánchez, C. & Monod, J. L'opéron: Groupe de gènes à expression coordonnée par un opérateur [C. R. Acad. Sci. Paris 250 (1960) 1727-1729]. *Comptes Rendus - Biol.* **328**, 514–520 (2005).
64. de Daruvar, A., Collado-Vides, J. & Valencia, A. Analysis of the Cellular Functions of *Escherichia coli* Operons and Their Conservation in *Bacillus subtilis*. *J. Mol. Evol.* **55**, 211–221 (2002).
65. Price, M. N., Huang, K. H., Arkin, A. P. & Alm, E. J. Operon formation is driven by co-regulation and not by horizontal gene transfer. *Genome Res.* **15**, 809–819 (2005).
66. Lawrence, J. G. & Roth, J. R. Selfish operons: Horizontal transfer may drive the evolution of gene clusters. *Genetics* **143**, 1843–1860 (1996).
67. Price, M. N., Arkin, A. P. & Alm, E. J. The life-cycle of operons. *PLoS Genet.* **2**, 0859–0873 (2006).

68. Cooper, S. & Helmstetter, C. E. Chromosome replication and the division cycle of *Escherichia coli* B/r. *J. Mol. Biol.* **31**, 519–40 (1968).
69. Schmid, M. B. & Roth, J. R. Gene location affects expression level in *Salmonella typhimurium*. *J. Bacteriol.* **169**, 2872–5 (1987).
70. Sousa, C., De Lorenzo, V. & Cebollat, A. Modulation of gene expression through chromosomal positioning in *Escherichia coli*. *Microbiology* **143**, 2071–2078 (1997).
71. Block, D. H. S., Hussein, R., Liang, L. W. & Lim, H. N. Regulatory consequences of gene translocation in bacteria. *Nucleic Acids Res.* **40**, 8979–92 (2012).
72. Couturier, E. & Rocha, E. P. C. Replication-associated gene dosage effects shape the genomes of fast-growing bacteria but only for transcription and translation genes. *Mol. Microbiol.* **59**, 1506–18 (2006).
73. Sobetzko, P., Travers, A. & Muskhelishvili, G. Gene order and chromosome dynamics coordinate spatiotemporal gene expression during the bacterial growth cycle. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E42–50 (2012).
74. Soler-Bistue, A., Timmermans, M. & Mazel, D. The proximity of ribosomal protein genes to *oriC* enhances *Vibrio cholerae* fitness in the absence of multifork replication. *MBio* **8**, 1–13 (2017).
75. Slager, J., Kjos, M., Attaiech, L. & Veening, J. W. Antibiotic-induced replication stress triggers bacterial competence by increasing gene dosage near the origin. *Cell* **157**, 395–406 (2014).
76. Narula, J. *et al.* Chromosomal Arrangement of Phosphorelay Genes Couples Sporulation and DNA Replication. *Cell* **162**, 328–337 (2015).
77. Lazazzera, B. A. & Hughes, D. Genetics: Location affects sporulation. *Nature* **525**, 7567 (2015).
78. Lobry, J. R. Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol. Biol. Evol.* **13**, 660–665 (1996).
79. Bremer, H. & Dennis, P. P. in *Escherichia coli and Salmonella: cellular and molecular biology* **2**, 1527–1542 (1987).
80. Brewer, B. J. When Polymerases Collide : Replication and the Transcriptional Organization of the *E. coli* Chromosome. *Cell* **53**, 679–686 (1988).
81. Rocha, E. P. C. & Danchin, A. Essentiality, not expressiveness, drives gene-strand bias in bacteria. *Nat. Genet.* **34**, 377–8 (2003).
82. Omont, N. & Képès, F. Transcription/replication collisions cause bacterial transcription units to be longer on the leading strand of replication. *Bioinformatics* **20**, 2719–2725 (2004).
83. Price, M. N., Alm, E. J. & Arkin, A. P. Interruptions in gene expression drive highly expressed operons to the leading strand of DNA replication. *Nucleic Acids Res.* **33**, 3224–34 (2005).
84. Lin, Y., Gao, F. & Zhang, C.T. Functionality of essential genes drives gene strand-bias in bacterial genomes. *Biochem. Biophys. Res. Commun.* **396**, 472–6 (2010).

85. Brandis, G. Biased Evolution. Doctoral Thesis. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 1198* (2016).
86. French, S. Consequences of Replication Fork Movement Through Transcript. *Science*. **258**, 1362–1365 (1992).
87. Mirkin, E. V & Mirkin, S. M. Mechanisms of transcription-replication collisions in bacteria. *Mol. Cell. Biol.* **25**, 888–95 (2005).
88. Boubakri, H., Langlois De Septenville, A., Viguera, E. & Michel, B. The helicases DinG, Rep and UvrD cooperate to promote replication across transcription units in vivo. *EMBO J.* **29**, 145–157 (2009).
89. Srivatsan, A., Tehrani, A., MacAlpine, D. M. & Wang, J. D. Co-orientation of replication and transcription preserves genome integrity. *PLoS Genet.* **6**, e1000810 (2010).
90. Paul, S., Million-Weaver, S., Chattopadhyay, S., Sokurenko, E. & Merrikh, H. Accelerated gene evolution through replication-transcription conflicts. *Nature* **495**, 512–5 (2013).
91. Merrikh, H. Spatial and Temporal Control of Evolution through Replication – Transcription Conflicts. *Trends Microbiol.* **25**, 515–521 (2017).
92. Lang, K. S. *et al.* Replication-Transcription Conflicts Generate R-Loops that Orchestrate Bacterial Stress Survival and Pathogenesis. *Cell* **170**, 787–799.e18 (2017).
93. Rocha, E. P. C. & Danchin, A. Gene essentiality determines chromosome organisation in bacteria. *Nucleic Acids Res.* **31**, 6570–6577 (2003).
94. Dillon, S. C. & Dorman, C. J. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat. Rev. Microbiol.* **8**, 185–195 (2010).
95. Postow, L., Hardy, C. D., Arsuaga, J. & Cozzarelli, N. R. Topological domain structure of the *Escherichia coli* chromosome. *Genes Dev.* **18**, 1766–1779 (2004).
96. Valens, M., Penaud, S., Rossignol, M., Cornet, F. & Boccard, F. Macrodome organization of the *Escherichia coli* chromosome. *EMBO J.* **23**, 4330–4341 (2004).
97. Song, D. & Loparo, J. J. Building bridges within the bacterial chromosome. *Trends Genet.* **31**, 164–173 (2015).
98. Higgins, N. P., Yang, X., Fu, Q. & Roth, J. R. Surveying a supercoil domain by using the gamma delta resolution system in *Salmonella typhimurium*. *J. Bacteriol.* **178**, 2825–35 (1996).
99. Singh, S. S. *et al.* Widespread suppression of intragenic transcription initiation by H-NS. *Genes Dev.* **28**, 214–219 (2014).
100. Badrinarayanan, A., Le, T. B. K. & Laub, M. T. Bacterial Chromosome Organization and Segregation. *Annu. Rev. Cell Dev. Biol.* **31**, 171–99 (2015).

101. Wang, X., Brandão, H., Le, T. B. K., Laub, M. T. & Rudner, D. *Bacillus subtilis* SMC complexes juxtapose chromosome arms as they travel from origin to terminus. *Science* **355**, 524–527 (2017).
102. Kavenoff, R. & Bowen, B. C. Electron microscopy of membrane-free folded chromosomes from *Escherichia coli*. *Chromosoma* **59**, 89–101 (1976).
103. Stein, R. A., Deng, S. & Higgins, N. P. Measuring chromosome dynamics on different time scales using resolvases with varying half-lives. *Mol. Microbiol.* **56**, 1049–1061 (2005).
104. Le, T. B. K., Imakaev, M. V., Mirny, L. A. & Laub, M. T. High-Resolution Mapping of the Spatial Organization of a Bacterial Chromosome. *Science* **342**, 731–734 (2013).
105. Le, T. B. K. & Laub, M. T. Transcription rate and transcript length drive formation of chromosomal interaction domain boundaries. *EMBO J.* **35**, 1582–1595 (2016).
106. Espeli, O., Mercier, R. & Boccard, F. DNA dynamics vary according to macrodomain topography in the *E. coli* chromosome. *Mol. Microbiol.* **68**, 1418–1427 (2008).
107. Dupaigne, P. *et al.* Molecular Basis for a Protein-Mediated DNA-Bridging Mechanism that Functions in Condensation of the *E. coli* Chromosome. *Mol. Cell* **48**, 560–571 (2012).
108. Sullivan, N. L., Marquis, K. A. & Rudner, D. Z. Recruitment of SMC by ParB-*parS* Organizes the Origin Region and Promotes Efficient Chromosome Segregation. *Cell* **137**, 697–707 (2009).
109. Lagomarsino, M. C., Espéli, O. & Junier, I. From structure to function of bacterial chromosomes: Evolutionary perspectives and ideas for new experiments. *FEBS Lett.* **589**, 2996–3004 (2015).
110. Ochman, H. & Groisman, E. A. in *Molecular Ecology and Evolution: Approaches and Applications* (eds. Schierwater, B., Streit, B., Wagner, G. P. & DeSalle, R.) 479–493 (Birkhäuser Basel, 1994).
111. Anderson, R. P. & Roth, J. R. Tandem Genetic Duplications in Phage and Bacteria. *Annu. Rev. Microbiol.* **31**, 473–505 (1977).
112. Birch, A., Hausler, A., Ruttener, C. & Hutter, R. Chromosomal deletion and rearrangement in *Streptomyces glaucescens*. *J. Bacteriol.* **173**, 3531–3538 (1991).
113. Hughes, D. in *Organization of the Prokaryotic Genome* (ed. Charlebois, R. L.) 109–128 (American Society for Microbiology Publishers, 1999).
114. Hill, C. W. & Harnish, B. W. Inversions between ribosomal RNA genes of *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **78**, 7069–72 (1981).
115. Segall, A. M. & Roth, J. R. Recombination between homologies in direct and inverse orientation in the chromosome of *Salmonella*: intervals which are nonpermissive for inversion formation. *Genetics* **122**, 737–747 (1989).
116. Hughes, D. Co-evolution of the *tuf* genes links gene conversion with the generation of chromosomal inversions. *J. Mol. Biol.* **297**, 355–64 (2000).

117. Rebollo, J.-E., Ois, V. F. & Louart, J.-M. Detection and possible role of two large nondivisible zones on the *Escherichia coli* chromosome. *Biochemistry* **85**, 9391–9395 (1988).
118. Segall, A., Mahan, M. J. & Roth, J. R. Rearrangement of the bacterial chromosome: forbidden inversions. *Science* **241**, 1314–1318 (1988).
119. Hill, C. W. & Gray, J. A. Effects of chromosomal inversion on cell fitness in *Escherichia coli* K-12. *Genetics* **119**, 771–778 (1988).
120. Miesel, L., Segall, A. & Roth, J. R. Construction of chromosomal rearrangements in *Salmonella* by transduction: Inversions of non-permissive segments are not lethal. *Genetics* **137**, 919–932 (1994).
121. Eisen, J. a, Heidelberg, J. F., White, O. & Salzberg, S. L. Evidence for symmetric chromosomal inversions around the replication origin in bacteria. *Genome Biol.* **1**, research0011.1 (2000).
122. Kuzminov, A. Homologous Recombination—Experimental Systems, Analysis, and Significance. *EcoSal Plus* **4**, (2011).
123. Konrad, E. B. Method for the isolation of *Escherichia coli* mutants with enhanced recombination between chromosomal duplications. *J. Bacteriol.* **130**, 167–172 (1977).
124. Galitski, T. & Roth, J. R. Pathways for Homologous Recombination Between Chromosomal Direct Repeats in *Salmonella typhimurium*. *Genetics* **146**, 751–767 (1997).
125. Meselson, M. & Weigle, J. J. Chromosome Breakage Accompanying Genetic Recombination in Bacteriophage. *Proc. Natl. Acad. Sci. U. S. A.* **47**, 857–868 (1961).
126. Kolodner, R. Genetic recombination of bacterial plasmid DNA: electron microscopic analysis of in vitro intramolecular recombination. *Proc. Natl. Acad. Sci. U. S. A.* **77**, 4847–51 (1980).
127. Zaman, M. M. & Boles, T. C. Chi-dependent formation of linear plasmid DNA in exonuclease-deficient *recBCD*⁺ strains of *Escherichia coli*. *J. Bacteriol.* **176**, 5093–5100 (1994).
128. Michel, B. & Leach, D. Homologous Recombination—Enzymes and Pathways. *EcoSal Plus* **5**, (2012).
129. Anderson, D. G. & Kowalczykowski, S. C. The translocating RecBCD enzyme stimulates recombination by directing RecA protein onto ssDNA in a chi-regulated manner. *Cell* **90**, 77–86 (1997).
130. Holliday, R. A mechanism for gene conversion in fungi. *Genet. Res.* **5**, 282 (1964).
131. West, S. C. Processing of Recombination Intermediates By the RuvABC Proteins. *Annu. Rev. Genet.* **31**, 213–244 (1997).
132. Liao, D. Gene conversion drives within genic sequences: concerted evolution of ribosomal RNA genes in bacteria and archaea. *J. Mol. Evol.* **51**, 305–317 (2000).

133. Santoyo, G., Martínez-Salazar Jaime M, Rodríguez, C. & Romero, D. Gene Conversion Tracts Associated with Crossovers in *Rhizobium etli*. *J. Bacteriol.* **187**, 4116–4126 (2005).
134. Santoyo, G. & Romero, D. Gene conversion and concerted evolution in bacterial genomes. *FEMS Microbiol. Rev.* **29**, 169–83 (2005).
135. Sadosky, A., Davidson, A., Lin, R.-J., and Hill, C. *rhs* gene family of *Escherichia coli* K-12. *J. Bacteriol.* **171**, 636–642 (1989).
136. Higgins, C. F., McLaren, R. S. & Newbury, S. F. Repetitive extragenic palindromic (REP) sequences, mRNA stability and gene expression: evolution by gene conversion? - a review. *Gene* **72**, 3–14 (1988).
137. Chumley, F. G. & Roth, J. R. Rearrangement of the bacterial chromosome using Tn10 as a region of homology. *Genetics* **94**, 1–14 (1980).
138. Anderson, P. & Roth, J. R. Spontaneous tandem genetic duplications in *Salmonella typhimurium* arise by unequal recombination between rRNA (*rrn*) cistrons. *Proc. Natl. Acad. Sci. U. S. A.* **78**, 3113–3117 (1981).
139. Nasvall, J., Sun, L., Roth, J. R. & Andersson, D. I. Real-Time Evolution of New Genes by Innovation, Amplification, and Divergence. *Science* **338**, 384–387 (2012).
140. Lind, P. a., Tobin, C., Berg, O. G., Kurland, C. G. & Andersson, D. I. Compensatory gene amplification restores fitness after inter-species gene replacements. *Mol. Microbiol.* **75**, 1078–1089 (2010).
141. Koskiniemi, S., Sun, S., Berg, O. G. & Andersson, D. I. Selection-driven gene loss in bacteria. *PLoS Genet.* **8**, 1–7 (2012).
142. Liu, S. L., Sanderson, K. E. & Anderson, K. E. E. S. Highly plastic chromosomal organization in *Salmonella typhi*. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 10303–10308 (1996).
143. Cullum, J. Chromosome Transfer and Hfr Formation by F in *rec+* and *recA* Strains of *Escherichia coli* K12. **365**, 358–365 (1979).
144. Bailly, X., Olivieri, I., Brunel, B., Cleyet-Marel, J. C. & Béna, G. Horizontal gene transfer and homologous recombination drive the evolution of the nitrogen-fixing symbionts of *Medicago* species. *J. Bacteriol.* **189**, 5223–5236 (2007).
145. Cohen, O., Gophna, U. & Pupko, T. The complexity hypothesis revisited: Connectivity Rather Than function constitutes a barrier to horizontal gene transfer. *Mol. Biol. Evol.* **28**, 1481–1489 (2011).
146. Abdulkarim, F., Tuohy, T. M. F., Buckingham, R. H. & Hughes, D. Missense substitutions lethal to essential functions of EF-Tu. *Biochimie* **73**, 1457–1464 (1991).
147. Reams, A. B. & Roth, J. R. Mechanisms of gene duplication and amplification. *Cold Spring Harb Perspect Biol* **7**, a016592 (2015).
148. De Septenville, A. L., Duigou, S., Boubakri, H. & Michel, B. Replication Fork Reversal after Replication–Transcription Collision. *PLoS Genet.* **8**, e1002622 (2012).

149. Kuzminov, A. Recombinational repair of DNA damage in *Escherichia coli* and bacteriophage lambda. *Microbiol. Mol. Biol. Rev.* **63**, 751–813 (1999).
150. Brochet, M. *et al.* Shaping a bacterial genome by large chromosomal replacements, the evolutionary history of *Streptococcus agalactiae*. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 15961–15966 (2008).
151. Song, J., Ware, A. & Liu, S. Wavelet to predict bacterial ori and ter: a tendency towards a physical balance. *BMC Genomics* **4**, 17 (2003).
152. Esnault, E., Valens, M., Espéli, O. & Boccard, F. Chromosome Structuring Limits Genome Plasticity in *Escherichia coli*. *PLoS Genet.* **3**, e226 (2007).
153. Matthews, T. D., Edwards, R. & Maloy, S. Chromosomal rearrangements formed by *rrn* recombination do not improve replicore balance in host-specific *Salmonella enterica* serovars. *PLoS One* **5**, e13503 (2010).
154. Liu, G. R. *et al.* Genome plasticity and ori-ter rebalancing in *Salmonella typhi*. *Mol. Biol. Evol.* **23**, 365–371 (2006).

Acta Universitatis Upsaliensis

*Digital Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Medicine 1392*

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-332471



ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2017